2010-07-21

Acquisition of Otoacoustic Emissions Using Swept-Tone Techniques

Christopher Lee Bennett
University of Miami, bennett@miami.edu

UNIVERSITY OF MIAMI

ACQUISITION OF OTOACOUSTIC EMISSIONS USING SWEPT-TONE TECHNIQUES

By

Christopher Lee Bennett

A DISSERTATION

Submitted to the Faculty
of the University of Miami
in partial fulfillment of the requirements for
the degree of Doctor of Philosophy

Coral Gables, Florida

August 2010

UNIVERSITY OF MIAMI

A dissertation submitted in partial fulfillment of
the requirements for the degree of
Doctor of Philosophy

ACQUISITION OF OTOACOUSTIC EMISSIONS USING SWEPT-TONE TECHNIQUES

Christopher Lee Bennett

Approved:

_____	_____
Özcan Özdamar, Ph.D.	Terri A. Scandura, Ph.D.
Professor of Biomedical Engineering	Dean of the Graduate School
_____	_____
Jorge Bohórquez, Ph.D.	Colby Leider, Ph.D.
Professor of Biomedical Engineering	Professor of Music Engineering
_____	_____
Fred F. Telischi, M.D.	Rafael E. Delgado, Ph.D.
Professor of Otolaryngology	Intelligent Hearing Systems, Inc.

BENNETT, CHRISTOPHER (Ph.D., Biomedical Engineering)
Acquisition of Otoacoustic Emissions Using Swept-Tone Techniques (August 2010)

Abstract of a dissertation at the University of Miami.

Dissertation supervised by Professor Özcan Özdamar.
No. of pages in text. (135)

Otoacoustic emissions (OAEs) have been under investigation since their discovery 30 years ago (Kemp, 1978). Otoacoustic emissions are quiet sounds generated within the cochlea that can be detected with a sensitive microphone placed within the ear canal. They are used clinically as a hearing screening tool but have the potential for diagnostic and monitoring purposes.

For this dissertation, high-resolution instrumentation was developed for improving the acquisition of OAEs. It was shown that a high bit-depth device is required in order to simultaneously characterize the ear canal and the cochlear responses. This led to a reduction in the stimulus artifact that revealed early latency, high-frequency otoacoustic emissions.

Next, a swept-tone technique originally developed for use in acoustical systems was formally developed for use in the human ear. The swept-tone technique allows for the simultaneous acquisition of a system's impulse response and its distortion components. The swept-tone was first used in this study to characterize the ear canal transfer properties. From that transfer function, a compensation routine was developed which equalized the magnitude and phase distortions of the ear canal. As a result, an improved acoustical click could be presented to the ear, which allowed for further reduction of the stimulus artifact, revealing early latency emissions. Spectral flatness and effective

duration measurements of the compensated click showed an improvement over traditional click stimuli. Furthermore, wavelet analysis and time-frequency latency computations showed that higher frequency otoacoustic emissions were recoverable when using a compensated click stimulus.

The swept-tone technique was then utilized for the direct acquisition of otoacoustic emissions. The swept-tone response was compressed to an impulse response and compared to a standard click response. It was found that several similarities exist between the two response types. The divergences, primarily in the low-frequencies, have implications in the generation mechanisms involved in a click-evoked otoacoustic emission. The swept-tone response provided some clinical benefits, namely in an improved signal-to-noise ratio, and in the removal of obstructive synchronized spontaneous OAEs when compared to a standard click response. Current methods are restricted by noise contamination, and the use of a swept-tone technique can reduce the acquisition time by up to a factor of four, compared to standard click methods. These implications and future potential studies are discussed.

DEDICATION

To Alicia Jackson, my loving and beautiful wife who challenges and inspires me...

ACKNOWLEDGMENT

My time in the Neurosensory Lab was made successful due to the people that I met who helped me along the way. First and foremost, I would like to thank the contribution of my mentor, Dr. Özdamar, for his guidance not only in my Ph.D. project, but in helping me in all aspects of my academic development. He helped me to formulate my research questions, but provided me ample independence in answering those questions, for which I am very grateful.

Next, I would like to thank my lab mates, Jonathon, Alessandro, Nuri, Erdem, and Scott, with whom thoughts, beach barbeques, sports (most notably baseball), drinks, (most notably beer), and philosophies were shared. Without their friendship, it is certain that my time here would have been an unproductive misery. I am also thankful to Ryan Twilley and Dan Harris for their help in the development of the hardware portion of my Ph.D. project. Without their expert advice in embedded systems, my entire project surely would have taken months if not years longer to complete, so thank you guys.

Finally, I would like to acknowledge the contribution of my committee members. First, to Colby for his advice in audio, life, and all things cool. Next to Dr. Bohórquez for teaching me by example what it means to be an ethical researcher and to always produce work with the utmost integrity. Also I am thankful to Rafael, who brought me to the IHS family during my first year in the doctoral program, and allowed me to be a part of a top hearing research corporation.

TABLE OF CONTENTS

LIST OF ABBREVIATIONS

BM	basilar membrane
BPF	band-pass filter
CLAD	continuous loop averaging deconvolution
CRF	coherent reflection filtering
cTEOAE	click transient evoked otoacoustic emission
dB	decibel
DNLR	derived nonlinear response
DPOAE	distortion product otoacoustic emission
FFT	fast Fourier transform
HL	hearing level
HPF	high-pass filter
IR	impulse response
kHz	kilohertz
LPF	low-pass filter
LTI	linear time-invariant
MLS	maximum length sequence
mPa	millipascal
MPEG	motion pictures expert group
MR	meatal response
ms	millisecond
NIHL	noise induced hearing loss
OAE	otoacoustic emission
pe	peak-equivalent
pp	peak-to-peak
pSPL	peak sound pressure level
SFM	spectral flatness measure
SFOAE	stimulus frequency otoacoustic emission
SNR	signal to noise ratio
SPL	sound pressure level
sTEOAE	swept-tone transient evoked otoacoustic emission
STFT	short time Fourier transform
TEOAE	transient evoked otoacoustic emission
WSS	wide sense stationary
UART	universal asynchronous receive and transmit
WT	wavelet transform

Chapter 1 Introduction

The human ear, and in particular the inner ear, is a sensory transduction organ that converts compression and rarefaction pressure waves into electrical signals that can be transmitted to the brain via the auditory nerve. If thought of as an electro-acoustic system, the ear accepts mechanical energy and produces electrical energy. Furthermore, in order to generate signals that are within a sensitive operation range, the inner ear also has an amplification system, known as the cochlear amplifier. When the cochlear amplifier is turned up (e.g., in a quiet environment), an unstable feedback can occur, resulting in the ear emitting sounds as it relates to the cochlear amplifier. In fact, these sounds are known as otoacoustic emissions (OAEs) and can be recorded with a small, sensitive microphone placed within the auditory canal. OAEs are currently used as a screening mechanism in hearing diagnostics, and they are particularly useful in non-cooperative subjects, for example newborns, who cannot participate in a standard subjective hearing examination. OAEs have other potential roles, including monitoring of cochlear functioning during the delivery of ototoxic therapies, e.g., cisplatin or tobramycin (Stavroulaki, *et al.*, 2001; Dreisbach, *et al.*, 2006), monitoring in other high noise-exposure environments, quick objective audiogram determination, or even biometric registration since OAEs are unique personal identifiers (Swabey *et al.,* 2009). OAEs have several current and potential uses that merit their further exploration and understanding.

OAEs are typically separated into different groups based on the stimulus that is used to evoke them. If no stimulus is used (in other words, if the OAEs arise without external

input), then these are classified as spontaneous otoacoustic emissions (SOAEs). SOAEs can be observed in a quiet environment as high-pitch, sort duration tone. OAEs can also be evoked by presenting a stimulus to the ear, and through post-processing techniques, an OAE response can be recovered. Typically the stimulus is a single-frequency tone, two simultaneous single-frequency tones, or a click. A response to a single frequency tone is called a stimulus-frequency OAE (SFOAE), and the OAE response occurs at the same frequency as the stimulus. A response to two simultaneous tones is called a distortion product OAE (DPOAE), and the OAE frequency occurs at several frequency locations that are equal to one of the intermodulation distortions between the two stimulus frequencies. Finally, if a click stimulus is used, then the resulting response is considered to be a click-evoked (or transient-evoked) OAE (TEOAE). For the remainder of this manuscript, TEOAEs will be generally explored, with some reference to SFOAEs in Chapter 4.

The purpose of this dissertation is to examine three contributions to current TEOAE acquisition technologies and methodologies. The first step was to improve acquisition through enhanced instrumentation. Section 2.1 explains a custom OAE acquisition device that was developed with particular specifications with regard to quantization of bit-depth and sampling rate. It will be shown that enhancing the instrumentation can improve standard TEOAE recordings and allow for alternate methodologies that require large dynamic range capability. The second step was to improve on existing click methodologies. Standard methods utilize an electric click that undergoes phase and magnitude distortions due acoustic properties of the speaker, ear canal, and microphone, resulting in an acoustic stimulus that no longer resembles a click. Chapter 3 examines a

compensation routine for the equalization of these magnitude and phase distortions in order to develop an idealized acoustical click inside the ear canal. With this compensated click, improved TEOAE responses can be recovered for enhanced diagnostics, especially in the early latency, high-frequency regions. Finally, in Chapter 4, a novel method for the recovery of TEOAE responses is explored, which utilizes a swept-tone stimulus followed by a deconvolution post-processing method to obtain the TEOAE. One of the primary contributions to the field is in the application of a swept-tone stimulus and deconvolution process to the acquisition of OAEs. The swept-tone method allows for determination of a nonlinear system's impulse response and furthermore allows for characterization of higher-order distortion components. The swept-tone method can be used to characterize the auditory canal, and it can also be used to acquire OAEs. A rigorous development of the swept-tone method and its signal-to-noise ratio properties is presented in Section 2.2.

1.1 BACKGROUND

1.1.1 AUDITORY CONDUCTIVE PATHWAYS

The peripheral auditory conductive pathways are shown in Figure 1. Sound arrives at the ear either via direct path, or as a reflection off of the pinna. The pinna acts as a 10dB passive amplifier by reflecting sound into the auditory canal. As well, the pinna has ridges which reflect high-frequencies at such an angle that high-frequency gain and attenuation notches are created in the spectrum that are highly dependent on the angle of incidence. This is known as pinna flange, and is used in sound source localization. The auditory canal acts as an amplifier around 15 dB and as a resonator of speech-related frequency content, as it funnels sound directly to the tympanic membrane (Durrant and

Lovrinic, 1995). A first-order approximation of the auditory canal can be represented by a single open-ended Helmholtz resonator. In a typical adult ear, the auditory canal measures about 27mm long, which produces a resonant frequency around 3.2 kHz.

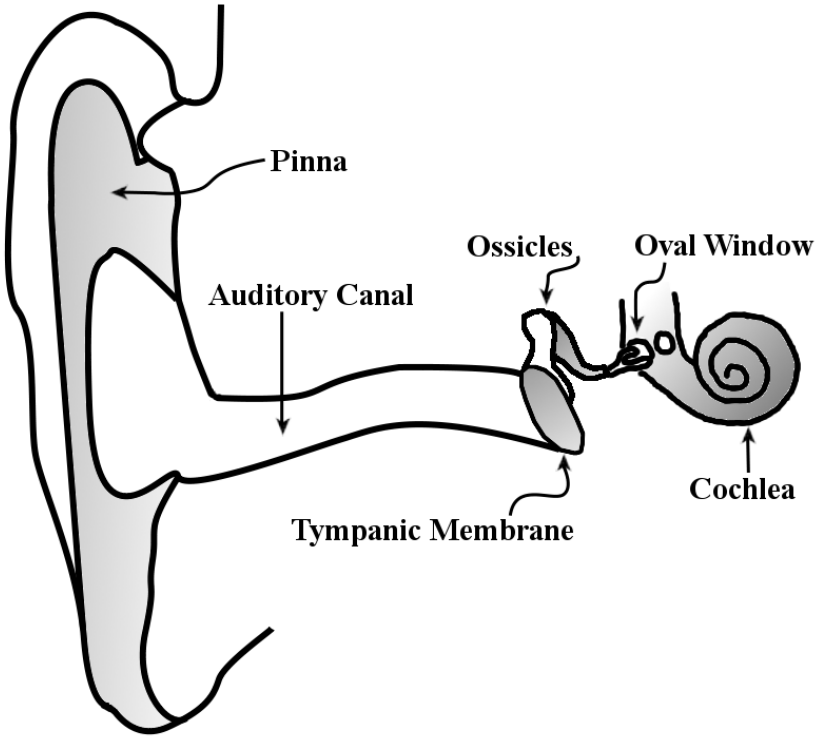

Figure 1. The pinna and auditory canal provide amplification in the speech-related frequency regions. The ossicles provide amplification and impedance matching to couple the air-filled auditory canal to the fluid-filled cochlea. Sensory transduction occurs within the cochlea.

The tympanic membrane is stretched across the end of the auditory canal and vibrates due to sound pressure level differences. It is slightly canonical in shape, as it is attached to the ossicles on one side. The ossicles are comprised of three distinct bones: the malleus, the incus, and the stapes. The malleus is rigidly attached to the tympanic membrane and

forms a diarthrodial (mobile) joint with the incus. The incus makes a flexible connection to the stapes which is attached to the oval window by an annular ligament. The ossicles act as an impedance matcher between the air-filled auditory canal and the fluid-filled cochlea. Furthermore, the ossicles amplify the sound by about 20 dB due ratio of the surface area of the malleus to the surface area of the stapes (Rosowski, 1996). The oval window is where sound energy passes into the cochlea as a forward-travelling wave.

The cochlea can be thought of as a coiled tube that is divided in the middle by two membranes to form three separate sections, as shown in Figure 2. The membranes and cochlea are embedded in the bony otic capsule. Each of the two membranes inside the cochlea is soft and pliable. The three chambers in the cochlea are the scala vestibuli, the scala media, and the scala tympani. Reissner's membrane separates the scala vestibuli from the scala media, and the basilar membrane (BM) separates the scala tympani from the scala media. In the scala media, and affixed to the BM, is the organ of Corti, which contains the sensory hair cells. At the apex of the cochlea, the scala vestibuli and the scala tympani are adjoined through a tiny hole called the helicotrema, so that the upper and lower chambers are contiguous. At the base of the cochlea, the scala tympani and the middle ear are separated by the flexible round window, which is situated adjacent to the oval window. Both of these windows are semi-compliant. Vibrations on the oval window cause movement of the perilymph, the fluid inside the scalas tympani and vestibuli. The vibrations in the perilymph cause fluctuations along the BM. The BM swells from the base to the apex in width by a factor of 5, and is stiffer at the base than the apex by a factor of 100. In part due to these reasons, the length of the BM can be considered frequency-mapped with high-frequency resonance at the basal end and low-

frequency resonance at the apex. A certain place along the BM will tend to resonate in response to a particular frequency; this is said to be that characteristic frequency's tonotopic place.

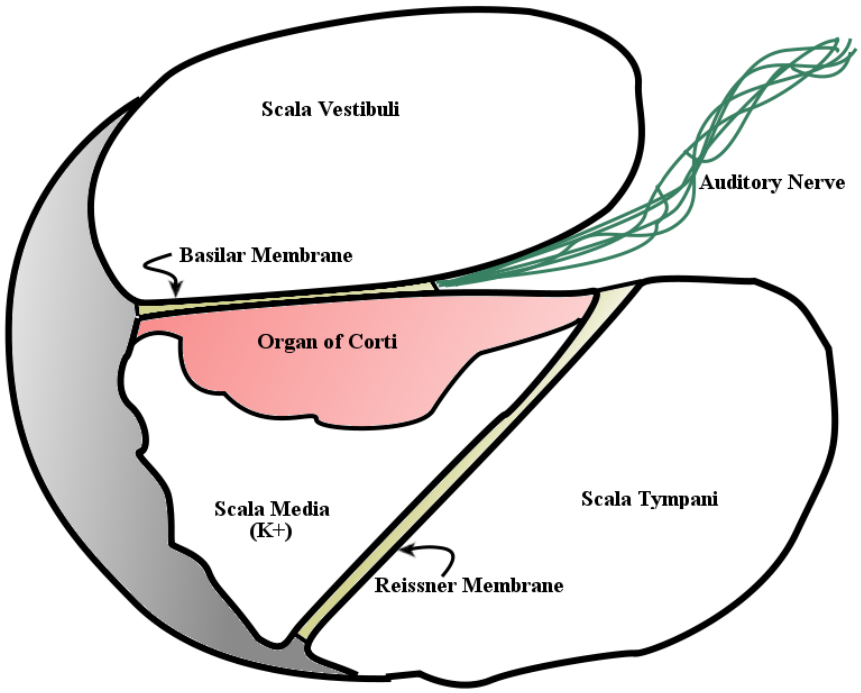

Figure 2. The cochlea is comprised of three chambers, the middle containing the organ of Corti. Ionized fluid in the scala media de- and hyper-polarizes the neurosensory cells which stimulate afferent conduction to the VIIIth cranial nerve.

The fluid in the scala media, the endolymph, has in ionic concentration of high K^+ and low Na^+, whereas the perilymph has a concentration of low K^+ and high Na^+. The difference in concentrations and the permeability of Reissner's membrane generates a standing potential that is approximately 80 mV higher in the endolymph than in the perilymph. In the scala media, the basilar membrane is lined with the organ of Corti.

This sensory organ transforms the mechanical energy along the BM into neural messages via the hair cells. Each hair cell is covered with stereocilia, or tiny hair-like structures extending from the dorsal end, as shown in Figure 3. These hair cells are situated in between the BM and the tectorial membrane forming a tight connection. So movement of the BM causes the hair cells, which are connected together by a tip-link, to bend in tandem. Movement in one direction causes depolarization, and movement in the other direction causes hyperpolarization. The spiral ganglion, which when bundled together form a portion of the eighth cranial nerve, makes synaptic connections to these hair cells for sending afferent information to the central nervous system.

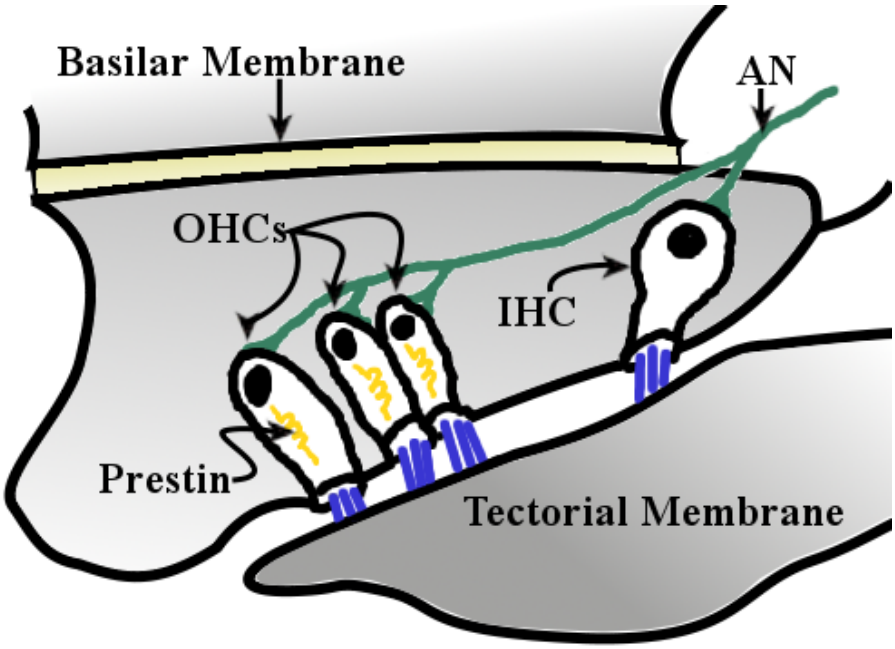

Figure 3. The hair cells are ventrally supported by the Basilar membrane and dorsally attached via stereocilia to the tectorial membrane. The stereocilia mechanically gate the cellular membrane and are connected together with tip links. Extra motility is provided by the outer hair cells through a motor protein known as prestin.

1.1.2 OTOACOUSTIC EMISSIONS

1.1.2.1 Hair cells

There are two distinct types of hair cells, inner hair cells (IHCs) and outer hair cells (OHCs). These hair cells are grouped together in rows, each row containing one IHC and about three OHCs. The IHCs send 95% of the afferent information, but the OHCs provide amplification and fine frequency tuning and are modulated by efferent neural connections. The OHCs contain a special motor protein known as prestin that allow the cells to grow and shrink in size rapidly. This motion has a pronounced effect on the wave traveling down the BM. The OHCs amplify extremely low-intensity sounds to a range that can be transduced by the ICHs, extending the hearing range of the ear down to a pressure change as small as 20 μPa. OHCs amplify the fluctuations of the BM, resulting in an enhanced neural response of the IHCs.

It is possible to detect the motility of hair cells completely noninvasively. When the OHCs induce motion along the basilar membrane, a disturbance in the perilymph is created that is associated with the active mechanisms of the cochlea. That is to say, the motion introduced along the BM by the OHCs is unassociated with the stimulus, as would be present in a passive environment. This creates a secondary, backwards travelling wave that reaches the oval window. The backwards propagation follows the same path as forward propagation (except, obviously, in reverse); the wave travels from the oval window, through the ossicles, to the tympanic membrane, and eventually to the outer auditory canal. However, instead of incurring amplification in the ossicles, as happens in the forward direction, the emitted wave is actually attenuated in its backwards propagation through the ossicles. The wave can be detected in the auditory canal with a

very sensitive microphone in low-noise environments. These types of waves are termed otoacoustic emissions (OAEs).

1.1.2.2 Types of OAEs

There are two basic types of OAEs: spontaneous (SOAE) and evoked (EOAE) emissions. SOAEs, though indicative of a working cochlea, only have a prevalence of 52% in females and 30% in males (Prieve *et al.*, 1996). They are usually narrow-band acoustic signals generated in the cochlea naturally, in the absence of a stimulus (Probst *et al.*, 1991). SOAEs are generally a pure tone in the 1-2 kHz frequency range lasting up to a few seconds. SOAEs are not practical to measure in the lab because they are neither prevalent, predictable, nor frequent enough in occurrence. There are many types of EOAEs, including two tone distortion-product (DPOAE), single tone stimulus frequency (SFOAE), chirp evoked (ChEOAE), speech evoked (SpEOAE), but this dissertation will focus on OAEs evoked by transient stimuli (TEOAE).

In a functioning cochlea OAEs occur when the OHCs, which normally are used to amplify environmental stimuli, become activated. The OHCs stimulated by a transient, e.g., a short-duration (~0.1 ms) click, will respond by driving the BM and subsequently the cochlear fluid, which in turn vibrates the oval window, the ossicles, and finally the tympanic membrane. A transient will stimulate the entire length of the BM, and so a frequency-dispersed response is emitted into the ear canal. High-frequency emissions will have an earlier latency due to the close proximity of the high-frequency OHCs to the oval window. It takes a round trip through the cochlea for the stimulus sound energy to reach the low frequency OHCs, and then travel back to the oval window, and so lower frequencies are emitted later. An example TEOAE is shown in Figure 4.

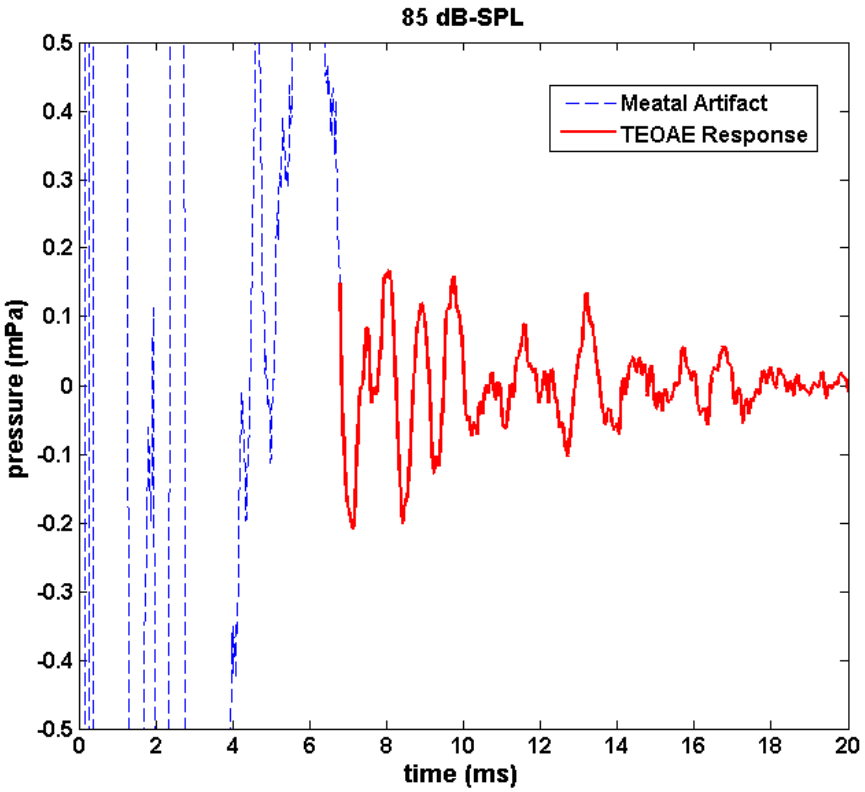

Figure 4. The TEOAE response (solid) is windowed @ 6 ms and band-pass filtered between 0.3 and 7.5 kHz. The meatal artifact (dashed) obstructs the early latency responses.

TEOAEs are a good indicator of hearing loss, as they are very prevalent (>99%) in human subjects. Commonly used transient stimuli are rectangular or Gaussian shaped pulses, single sinusoids, half sinusoids, or windowed tone bursts. The use of transients in evoking TEOAEs is desirable for two main reasons. First, it is short in duration, so the resulting TEOAE can be recorded with minimal interference from the stimulus. Second, a transient is broadband in its frequency response, so it will stimulate OHCs along the basilar membrane from base to apex.

However, these ideal generalizations are not readily observed in clinical settings. To the first point, the transient response in the auditory canal, known as the meatal response (MR), creates ringing that lasts typically around 5 ms. But early latency (high-frequency) TEOAEs arrive in the ear canal in less than 1 ms, so there is an overlap where the MR is an artifact obscuring these TEOAEs. As to the second point, the auditory canal by the same mechanism produces peaks and notches in the magnitude response, distorting an otherwise broadband signal. As a result, the length of the BM is supplied with unequal amounts of energies across the frequency spectrum.

1.1.2.3 OAE Generation Mechanisms

OAEs are thought to arise by two fundamental mechanisms within the cochlea. One of these sources of OAE results from reflections from cochlear irregularities. These sources are indiscriminate with respect to the stimulus, and are dependent solely on mechanical perturbations along the BM. These sources go by several names, but will be referred to as simply a reflection source. The second source type occurs primarily at high input levels and is related to stimulus-induced distortions along the BM. Unlike reflection sources, which are fixed in place within the cochlea, the distortion sources are dependent upon the frequency of the stimulus, and so are said to be wave-fixed.

The generation mechanisms of TEOAEs are often compared to those that generate stimulus-frequency OAEs (SFOAE) (Shera and Guinan Jr., 2003; Kalluri and Shera, 2007b). Under the SFOAE modality, a single frequency is used as the stimulus, and OAE time-delays can be acquired for multiple stimulation frequencies. To make the comparison with TEOAE responses, group delays must be determined. At low- to moderate-intensity levels, delays of frequency components of these two modalities are

quite similar. Delays of low to mid range frequency components (500 Hz to 4,000 Hz) under both modalities are approximately twice the group delay of the travelling wave reaching its tonotopic place, which accounts for both forward and reverse transmission time. These group delay characteristics are explained by the well-accepted coherent reflection filtering (CRF) theory (Shera and Zweig, 1993; Talmadge *et al.*, 1998; Shera and Guinan Jr., 1999). Under CRF theory, mechanical irregularities along the basilar membrane result in partial reflections, or back-scattering. These reflections exhibit group delay and frequency characteristics in accordance with its place along the BM. Since this source of OAE is reliant on the topology of the cochlea, it is also often referred to as a place-fixed source. There are considerable similarities between SFOAE and TEOAE responses with respect to the frequency-dependent group delay and frequency-dependent growth (input/output) functions.

Additional response components associated with high intensity stimuli can also be observed in OAE responses. This secondary source, while not completely understood, is sometimes called a wave-fixed source, as it tends to vary little with respect to time-frequency delay or tonotopic place. This component is difficult to study because it occurs within the first several milliseconds of the TEOAE response, which is typically contaminated by stimulus artifact in transient recordings. These wave-fixed distortion sources are characterized by a short-latency, slowly rotating (or in other words, short latency) TEOAE phasor component (Withnell *et al.*, 2008). The wave-fixed component is a nonlinear distortion originating at the tonotopic place, which generates higher-order harmonics that appear basal to the fundamental frequency's tonotopic place.

1.1.2.4 Clinical Usage of OAEs

The greatest contributor to noise-induced sensorineural hearing loss (NIHL) and age-related hearing loss (presbycusis) is damage to the OHCs. Without these cells, the cochlea loses frequency resolution and low-level amplitude sensitivity. The result is an overall attenuation by anywhere from 20 to 80 dB in hearing and also a loss of frequency discrimination (Bray, 1989).

A fundamental hearing test used in audiology clinics and required in most states on newborn infants is OAE screening. Early detection is especially important in newborns, because early intervention in hearing loss is imperative for proper speech-language and cognitive development. OAEs occur as a result of a properly functioning cochlea, so its absence can be an indicator of hearing loss or damage to the ear. The discovery of this phenomenon was made nearly three decades ago by David Kemp in a series of experimentations (Kemp, 1978). Measurements are repeatable due to the extremely stable responses found in OAE recordings over long durations of time. Since a significant portion of hearing loss involves age related, noise induced, ototoxic, or hereditary cochlear damage, OAE screenings potentially provide a quick means of detection. OAEs have become a popular tool for quickly screening subjects, and more recently has become a means of monitoring cochlear functioning in real-time when subjective testing is not possible (for example, if the subject is undergoing surgery or receiving therapeutic ototoxic agents).

1.2 Overview of Current Methods

1.2.1 Stimulus Reduction Methods

1.2.1.1 Derived Nonlinear Response Averaging

The most common method for reducing the MR artifact is a simple subtraction method that takes advantage of the compressive growth function of OAEs. At low intensities, OAE amplitude will grow linearly with stimulus intensity level. However, at moderate to high intensities, OAE amplitude will begin to increase at a slower rate than stimulus intensity. The so-called 'derived nonlinear residual' (DNLR) method uses a train of mixed intensity and mixed polarity stimuli in order to effectively subtract out linear components of the response (which are assumed to be stimulus-related) while only partially subtracting nonlinear components of the response (which are assumed to be physiologically related), as shown in Figure 5.

It has been shown that the response of TEOAEs begins to saturate with increasing stimulus intensity at 40 dB HL (Zwicker, 1979). The MR is approximately linear with respect to the stimulus intensity; however, the TEOAE response is nonlinear and grows compressively. So at high stimulus levels, an increase in intensity will comparably increase the amplitude of the MR, but only marginally increase the amplitude of the TEOAE. The growth of the TEOAE with respect to the input stimulus level can be given by a dB/dB slope. A slope of 1 dB/dB represents perfect linearity, and a slope of 0 dB/dB represents complete saturation (Lineton *et al.*, 2006). It was found that a slope of approximately 0.33 dB/dB exists in the nonlinear region (Kemp and Chum, 1980).

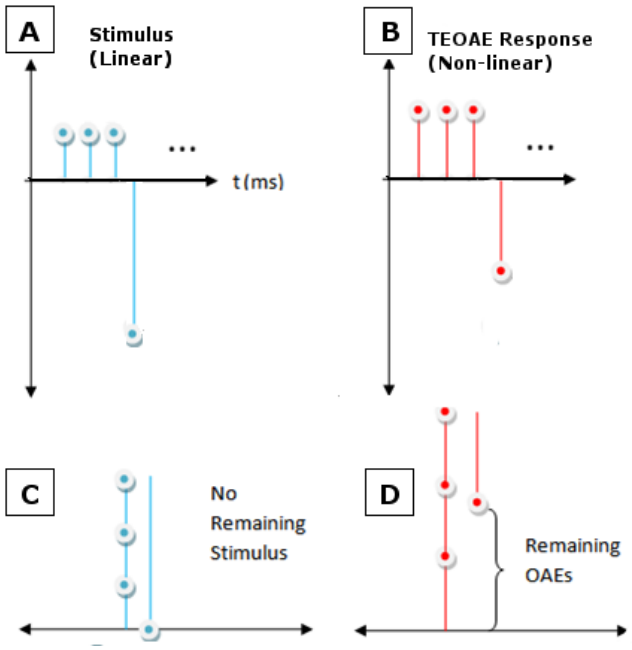

Figure 5. (A) The stimulus sequence contains 3 low-level clicks followed by a high-level click of opposite polarity. (B) However, the TEOAE saturates at high intensities, so the response to the -3s stimulus is not equal to the sum of responses to s. (C) The stimulus is linear, and so a summation of s+s+s-3s is identically zero. (D) But since OAE responses sum nonlinearly, after summation, some TEOAE information remains.

The DNLR recovery method is a technique to partially cancel the stimulus artifact from the recorded TEOAE signal. Under the DNLR modality, a train of four stimuli are presented with the first three at the nominal intensity and polarity and the fourth inverted and at an intensity of 3x (or approximately +9.5 dB). The four responses (containing both MR artifact and OAE response) are summed together. Let the stimuli be represented by s, the responses by r, the intensity by p, and the noise by η. The sum can be written by the expression

$$\sum_{i=0}^{3} s_i(p) + r_i(p) + \eta_i \quad \textbf{(1.1)}$$

By the definition of DNLR s_0, s_1, and s_2 are identical, and so comparably their responses should be identical, given that OAEs are indeed time-invariant (the case could be made against time-invariance, but that would generally be over much longer time interval, say weeks to months, as opposed to milliseconds). However, a distinction lies in the linearity, whereby $s(3p) = 3s(p)$ but $r(3p) \neq 3r(p)$. Finally, assuming the additive noise is white Gaussian, its amplitude will grow as a function of $1/\sqrt{K}$, where K is the number of epochs; in this case K=4, so the amplitude of the noise in 4 epochs will be twice the amplitude of a single epoch. Grouping terms and simplifying, the above expression can be re-written as

$$3s(p) - s(3p) + 3r(p) - r(3p) + 2\eta \quad \textbf{(1.2)}$$

Since s is LTI, its two terms will cancel. Since r is nonlinear, its terms will not cancel. In the best-case scenario, the response r will be in complete saturation, such that $r(3p) = 3r(p)$. Then the power of the signal can be represented as a ratio to the power of the noise

$$SNR_{DNLR} = \frac{2r(p)}{2\eta} \quad \textbf{(1.3)}$$

Compared to standard synchronous averaging, DNLR is effective at removing much of the stimulus-related artifact. However, this comes at the cost of a reduced SNR. For example, if the same response were acquired under linear modality, then the four responses would sum to $4r(p)$, thus doubling the power of the response. So it can be

generalized that under ideal conditions, the DNLR modality will necessarily reduce SNR by 6 dB.

1.2.1.2 Digital Subtraction

The well known physiological characteristics of OAEs offer two principle ways to eliminate or attenuate the MR in TEOAEs: masking and nonlinearities. Physiologic masking can suppress the OAEs by the introduction of another sound in the same (ipsilateral) or opposite (contralateral) ear, or both (binaural) ears (Wilson, 1980). If the OAEs are masked or suppressed by a simultaneous tone or noise, then the pure meatal response can be recorded (Arslan *et al.*, 2001; Tavartkiladze *et al.*, 1994). Generally, an ipsilateral reproducible pseudo-random noise masker is presented simultaneously with the stimulus, and a digital subtraction method is used to recover the MR free TEOAEs by post-processing. At low stimulus levels, such a technique revealed the existence of TEOAE components at latencies around 2.5-5.0 ms that were more linear than later components (Kruglov *et al.*, 1997). These components, however, were not found in all subjects and appeared to consist of lower frequencies than expected, probably due to MR contamination.

1.2.1.3 Stimulus Compensation

In this approach the aim is to reduce the MR by presenting a special transient to the ear that will generate a more impulse like waveform in the ear canal than the rectangular click. Chertoff and colleagues used the recorded MR to model the ear canal transfer function with finite impulse response (FIR) pulses, band-limited from 256 to 10,512 Hz (Chertoff and Chen, 1996; Chertoff and Guruprasad, 1997). In these studies, FIR pulses were used to guarantee stability in frequency-domain division and to have a band-limited

template. The computation procedure involves calculating the transfer function of the ear by dividing the recorded meatal response with the FIR pulses, then inverting and low-pass filtering the result.

However, the FIR pulse is not identical to a rectangular click that is used clinically. An FIR pulse necessarily contains ringing after the point of maximum energy. Since one of the motivations of modifying a TEOAE stimulus is to reduce ringing, computing with an FIR pulse is not desirable since it will add ringing. To correct for this, Chertoff and Chen (1996) modified the computed click spectrum to have a roll-off, much like a rectangular pulse. Acoustic clicks were generated with this method, and TEOAEs were obtained, as shown in Figure 6. However, Chertoff and Guruprasad (1997) found that spectrally flattened stimuli actually reduced the overall TEOAE level and its reproducibility. This was attributed to the minimization of the naturally occurring ear canal-related 2-3 kHz spectral peak, which in turn reduced the saturation effect of the TEOAEs, which were acquired in DNLR mode.

Other previous studies have utilized a compensated click for the purposes of extending the high-frequency stimulation. The earliest such method was developed for the purposes of reducing ringing in auditory evoked responses (Van Campen *et al.*, 1994). In that study, supra-aural earphones, instead of inserts, were calibrated through complex frequency-domain division to produce tone-bursts and low-pass filtered DC pulses with reduced ringing. In another study, a compensated click was used for the purposes of studying the effects of click stimulus intermodulation distortion in evoking TEOAEs (Yates and Withnell, 1999). In that study, a high-pass filtered compensated click was used in guinea pigs. The MR was derived from a pseudo-random noise input, but the

compensation technique was not elaborated upon. Furthermore, it is not known if the compensated click had reduced ringing, as the motivation was solely to deliver a spectrally flat stimulus.

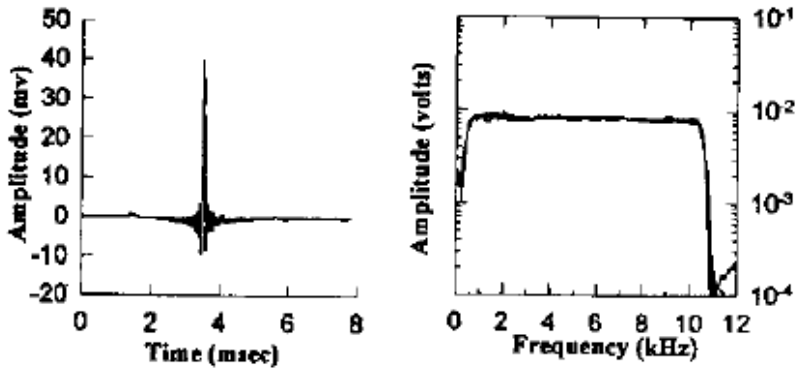

Figure 6. Chertoff and Chen use an FIR pulse, which has no zeroes in its frequency transformation, for click calibration. The result is a flat magnitude response, but with time-domain ringing. From (Chertoff and Chen, 1996).

1.2.2 NONLINEAR ANALYSIS

A Volterra series can be used to represent dynamic, nonlinear systems, including time memory. Full Volterra information can be analyzed using a maximum length sequence (MLS) method. Generally, the linear transient response is generated by cross-correlating the MLS response with the MLS stimulus sequence. But higher-ordered responses can also be obtained on TEOAEs, as well (Thornton, 1997). Thornton and colleagues were able to obtain reproducible 2nd and 3rd order TEOAE responses from 1 to 15 ms post-stimulus onset (Thornton *et al.*, 2001). They found that higher-order responses were much lower in amplitude than linear responses, and that the growth function was much less compressive than in 1st order responses. In a later study, they found that the higher-

order responses lasted, on average, for only about 12 ms (Slaven *et al.*, 2003). But similar to linear responses, the higher order responses followed the same morphology of high-to-low frequency dispersion. Furthermore, Thornton and colleagues theorized that high ordered responses may be a better indicator of high-frequency hearing loss than standard TEOAEs (de Boer and Thornton, 2006).

1.2.3 TIME-FREQUENCY ANALYSIS METHODS

The goals in representing TEOAEs in time-frequency plots are two-fold: (1) to identify the relative amplitudes of individual frequency components, and (2) to identify these frequency components' corresponding latencies. A short-time Fourier transform (STFT) is an obvious choice, and is readily advocated by some in the field; e.g., time-frequency analyses performed at the Boys Town National Research Hospital visualized data with the STFT (Goodman *et al.*, 2009), as did analyses performed at the Center of Bioacoustics at the University of Ferrara (Hatzopoulos *et al.*, 2000; Hatzopoulos *et al.*, 2003). The STFT is subject to a time-frequency window, such that improving resolution along one axis limits resolution along the other (eg., if good frequency resolution is used, then the time resolution suffers, and vice versa).

Another method is the Wavelet transform (WT), which uses a frequency-based variable window to give good time resolution at high-frequencies and good frequency resolution at low-frequencies (Daubechies, 1988; Wit *et al.*, 1994; Tognola *et al.*, 1997; Zheng *et al.*, 1999; Janušauskas *et al.*, 2001). Some caveats to this method include analysis windows with lengths that are fixed to frequency location, and various artifacts that may be introduced depending on the selection of mother wavelet. Minimum variance spectral

estimation (MVSE) builds upon principles of both the WT and the STFT (Zhang *et al.*, 2008). MVSE uses an adaptive filtering routine to minimize the out-of-band energy, and implements a frequency-dependently sized sliding window. The Bionic Wavelet transform (BWT) adds a variable to the formulation of the WT, while maintaining the same window center in the time and frequency domains, but altering the envelope of the mother function with the added parameter (Yao and Zhang, 2000).

The Wigner-Ville distribution (WVD) exhibits high time-frequency resolution, but there is a cross-term "leakage" between frequency components, causing energy from one band to distort the results in a neighboring band (Cheng, 1995). Finally, the matching pursuit (MP) algorithm employs a dictionary of "atoms" of various frequencies, latencies, amplitudes, and durations (Blinowska *et al.*, 1997; Jedrzejczak *et al.*, 2004; Notaro *et al.*, 2007). The algorithm then searches for a combination of these atoms which can be constructed in some fashion to liken a synthetic signal to the actual recording within some prescribed error. However, the computational complexity is immense, taking several minutes to perform the calculations for just one recording.

1.2.4 USE OF SWEPT-TONES IN OAE RECORDINGS

The use of frequency-swept sinusoids is not new in the acquisition of OAEs. The basilar membrane itself is frequency-dispersed, whereby a click impulse results in a swept-frequency response (known as click-evoked OAEs). In addition, models map the frequency layout of the cochlea, all of which base their frequency mapping on logarithmic models. These include the Mel scale (Stevens *et al.*, 1937), the Bark scale on which critical bands are based (Zwicker, 1961), and the Greenwood cochlear map

(Greenwood, 1961). It should be noted that in these cases, and probably in general, the function of frequency is actually exponential, which means they are linear in the logarithmic domain, hence the commonly used terminology of referring to them as logarithmic.

One use of swept-sinusoids in OAE acquisition is a short duration (4 to 10 ms) signal that uses a frequency mapping of the cochlea to derive a stimulus that would simultaneously reach all portions of the cochlea. This type of stimulus is known as a chirp, and was first developed by Dau and colleagues for acquiring ABRs, but were later modified and used in the collection of transient OAEs (Dau *et al.*, 2000; Fobel and Dau, 2004). These chirp-evoked OAEs (ChEAOE) were found to have improved SNR characteristics over traditional click stimuli (Medri and Özdamar, 2000).

Longer swept-tone stimuli are also used in the acquisition of SFOAEs and DPOAEs that are hundreds or thousands of ms in duration. Swept-tones are used often to acquire SFOAEs over a broad frequency range (Shera and Guinan Jr., 1999). Removal of SFOAE response from stimulus information is possible via several mechanisms, which are discussed in Section 1.2.5. Furthermore, the swept-tone method was adapted for use in acquiring DPOAEs (Long *et al.*, 2008). In DPOAE measurements, two swept-tones are needed, one for each of the f_1 and f_2 frequencies. The strongest DP frequency (which is at the cubic distortion place, or $2f_2$-f_1) is at a different third frequency. The DPOAE sweeping response is typically analyzed by an adaptive filter which tracks the expected response frequency (Talmadge *et al.*, 1999). An additional method was developed by Douglas Keefe to "de-chirp" the DPOAE swept-tone response to obtain a click-like response (Keefe, 1998; Keefe and Ling, 1998). The Keefe chirp is defined in the

frequency domain, beginning with a prescribed group delay that is a function of frequency, with some parameters controlling duration, and directionality. The constant-magnitude, logarithmic-group delay chirp is converted to the time-domain via inverse DFT. The de-chirped, or time-compressed signal is obtained by using an all-pass filter with an inverted phase response.

1.2.5 PREVIOUS SFOAE ANALYSIS METHODS

Acquisition of SFOAEs requires some clever methodology and processing to reveal, as the response is embedded within the microphone's direct recording of the stimulus itself. Three acquisition modalities are explored and analyzed in (Kalluri and Shera, 2007a). Very briefly, the three methods are compression, suppression, and spectral smoothing. The compression technique is similar to the commonly used derived nonlinear residue (DNLR) technique in which the stimulus is presented at two intensities. The difference in the responses to those intensities can be used to eliminate the linear portion and retain the nonlinear portion, which is thought to consist primarily of OAE. The two-tone suppression technique and the spectral smoothing technique both subtract the contamination of the stimulus by some form of stimulus estimation. Suppression estimates the stimulus by suppressing the SFOAE response upon introduction of a second, proximal frequency tone to obtain a recording of pure stimulus through vector subtraction. The spectral smoothing method estimates the stimulus by making a prediction of background spectral content (which arises presumably from the stimulus) by applying a low-pass smoothing filter to the magnitude response. The SFOAE response is obtained by subtracting the background estimation from the total response spectrum.

Recently a fourth method, digital heterodyne analysis, was proposed as a way to analyze swept-tone SFOAEs (Choi *et al.*, 2008). Heterodyne analysis involves baseband-shifting the original response waveform such that the stimulus frequency is set to 0 Hz. Then a Blackman low-pass filter (LPF) is applied to the shifted spectrum, the lobe width of which can be shortened or lengthened to obtain late or early latency responses, respectively.

In the subsequent section, a proposed alternative analysis of SFOAEs will be provided, in which TEOAE-like responses are derived from a swept-tone single frequency stimulus. The response can be considered as an SFOAE because, although the stimulus frequency is constantly rising, at any instantaneous moment there is only a single pure tone being presented to the subject. However, the swept-tone response can be presented as a response to an impulse through some post-processing techniques. This derived impulse response closely resembles a click-evoked OAE in phase and time-frequency characteristics. Through this sort of analysis, direct comparisons between SFOAE and TEOAE responses can be made.

1.3 STUDY OVERVIEW

Swept-tone analysis was first theorized for audio systems by Craven and Gerzon (1992) and Griesinger (1996) but formally developed and applied by Farina (2000, 2007) and independently by Muller and Massarani (2001). The swept-tone analysis method utilizes a swept frequency tone in conjunction with an inverse swept-tone in order to extract the impulse response of an acoustical system. Furthermore, the use of a logarithmic sinusoidal argument allows for the separation of nonlinear components from the linear

response, whereby the distortion itself can be characterized. Here, two uses of swept-tone analysis are developed for the purpose of improving acquisition of OAEs. Swept-tone analysis will first be used to characterize the transfer properties of the outer auditory canal in order to develop a compensatory filtering scheme. Swept-tone analysis will also be used for the acquisition of OAEs. The swept-tone response is hardly different from a swept-tone SFOAE response (except slightly shorter in duration), but once deconvolved and transformed into a click response, the OAE resembles that of a TEOAE. For this reason, the deconvolved swept-tone OAEs will be referred to as sTEOAEs (contrast that to standard click TEOAEs, which will be denoted as cTEOAEs).

1.3.1 SPECIFIC AIMS

The aims of this dissertation are to (i) develop a subject- and test- dependent equalization scheme for producing idealized (short duration, flat magnitude, linear phase) acoustic clicks within the auditory canal for the acquisition of TEOAEs; and (ii) acquire deconvolved swept-tone OAE responses (sTEOAEs) and methodically compare to standard click TEOAE responses (cTEOAEs).

Specific Aim #1 will be tested by using swept-tone analysis to obtain the transfer function of the ear canal. From this transfer function, a real-time magnitude and phase equalization processes as an inverse filter to the auditory canal will be developed. A compensated click will be produced with the inverse filter and OAEs using standard and compensated clicks will be compared with respect to meatal response duration and magnitude spectrum. It is hypothesized that a compensated click will have a flatter

magnitude spectrum and shorter effective duration, thus revealing early latency, high-frequency OAEs.

Specific Aim #2 will be tested by using swept-tine analysis to directly obtain an OAE swept-tone response. The response will be time-compressed and compared to standard click OAE responses. Comparisons will be made with respect to coherence/correlation, signal-to-noise ratio (SNR), amplitude characteristics, time-frequency properties, and input/output growth functions. It is hypothesized that swept-tone derived and click derived OAEs will be substantially similar in time-alignment. Furthermore, clinical applications will be discussed.

Chapter 2 Methods and Materials

2.1 INSTRUMENTATION

A custom biomedical device was built at the Neurosensory Lab for high-fidelity acquisition of OAEs. The device interfaces with an Etymōtic Research ER-10D OAE probe and to MATLAB via UART (universal asynchronous receiver/transmitter) communication and a custom software program. The device is capable of acquiring a large dynamic range at a high sampling rate using an Analog Devices ADSP-21369 EZ-Kit Lite. The device is shown in Figure 7.

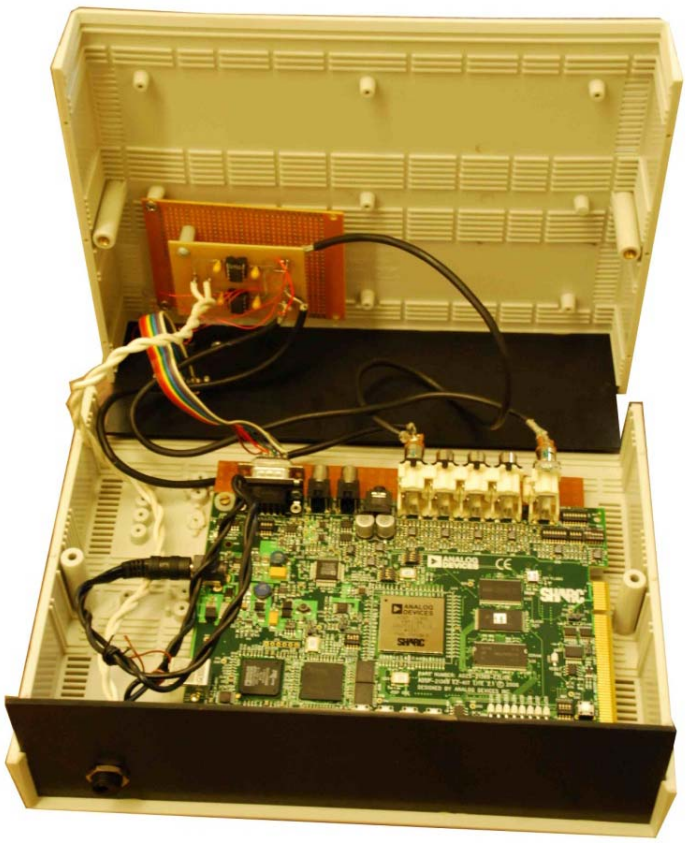

Figure 7. A high-resolution (24 bit, 48 kHz) OAE acquisition device was built. This board interfaces with custom software, allowing for adaptable functionality and real-time communications with a PC.

2.1.1 BIT DEPTH

In TEOAE acquisition, the peak response, R_{max}, is typically less than 1 mPa in amplitude. While the MR will vary, its peak amplitude, MR_{max}, can reach 500 mPa. The number of bits needed to cover this dynamic range is given by the $\log_2$ of the quotient of these numbers. The result must be rounded up (indicated by the ceiling function), since the number of bits are an element of the set of positive integers. In the case of a medium bit-depth system, the analog-to-digital converter (ADC) has 12- or 16-bit resolution, given by N_{system}. For these systems, it is not feasible to record the entire dynamic range of the MR and reserve a sufficient number of bits for the TEOAE, given by N_{TEOAE}. For example, if the entire amplitude-range of the MR is recorded via a 12-bit ADC (such as in the Otodynamics ILO88), then only 4 bits remain for the TEOAE response, corresponding to a -24 dB digital noise floor, which is entirely inappropriate for a biomedical signal (see Figure 8a). The equation governing the relationship between the number of bits utilized by the system and the size of the MR is given by

$$N_{system} = N_{TEOAE} + \left\lceil \log_2 \left(\frac{MR_{max}}{R_{max}} \right) \right\rceil \tag{2.1}$$

On the other hand, if all 12 bits are reserved for the TEOAE, then a more appropriate digital noise floor of -72 dB can be obtained; however, the stimulus will saturate the ADC (see Figure 8b). In order to recover both the MR and the TEOAE, a higher bit-depth will be needed. In fact, if 12 bits are reserved exclusively for the TEOAE, which could have a magnitude of approximately 1, then in order to also recover the meatal response, or stimulus artifact, then at least 21 bits will be required. If greater bit-depth

than 12 is desired for the TEOAE, then the system bit-depth must be comparably increased as well (eg., a 24-bit ADC is required if reserving 15 bits for N_{TEOAE}).

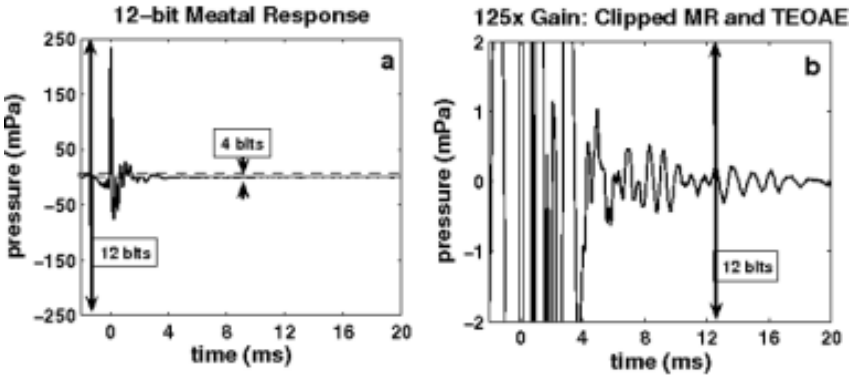

Figure 8. A typical MR and TEOAE from a subject (01). (a) The entire metal response recovered using a 12-bit system leaves only 4 bits for representation of the TEOAE. (b) 125x gain applied to the recovered response which gives a full 12 bits to the TEOAE, but saturates the MR.

2.1.2 ACQUISITION DEVICE

Figure 9 shows the block diagram of the designed system. The probe interfaced with an embedded system (Analog Devices ADSP-21369 Sharc EZ-Kit Lite) with an on-board ADC/DAC (AD1835: 24-bit, 192kHz) and DSP (ADSP21369). The output of the DAC was calibrated to the sound source specifications of the ER-10D probe, in which a 1 V_{RMS} signal is equivalent to 86 dB-SPL. Since the AD1835 DAC has a maximum peak output voltage of 1 V, then the maximum stimulus intensity level is 89 dB pSPL, or about 564 mPa. The ADC voltage levels were calibrated to pressure by taking into account the voltage range of the ADC, V_{pp}, and the sensitivity of the probe microphone, V_{sens}. The microphone sensitivity indicates how the microphone converts acoustic pressure into

voltage and is given in its datasheet as 50 $^{mV}/_{Pa}$. The digital-to-pressure scaling equation is given by Eqs. (2.2) and (2.3).

$$R_{mPa} = \frac{V_{pp}}{2^{N_{bits}} V_{sens}} \cdot R_{digital} \tag{2.2}$$

$$R_{mPa} = \frac{5.0V}{2^{24} \cdot 0.05 \frac{V}{Pa}} \cdot R_{digital} \tag{2.3}$$

So the equivalent pressure value in mPa, R_{mpa}, becomes a function of the input in fixed-point integer, $R_{digital}$; whereby, $R_{mPa} = 5.9605 \times 10^{-3} \times R_{digital}$.

Noise floor recordings were taken by measuring the RMS value of a zero input to the ADC, and using the equation above, to convert to pressure. The noise floor of the probe and amplifier is nominally -15 dB SPL, and the noise floor of the Sharc EZ-Kit was measured to be -14 dB SPL, for a combined system noise floor of about -8.5 dB SPL, or about 7.5 μPa. This noise floor will be reduced even further through the process of synchronous averaging.

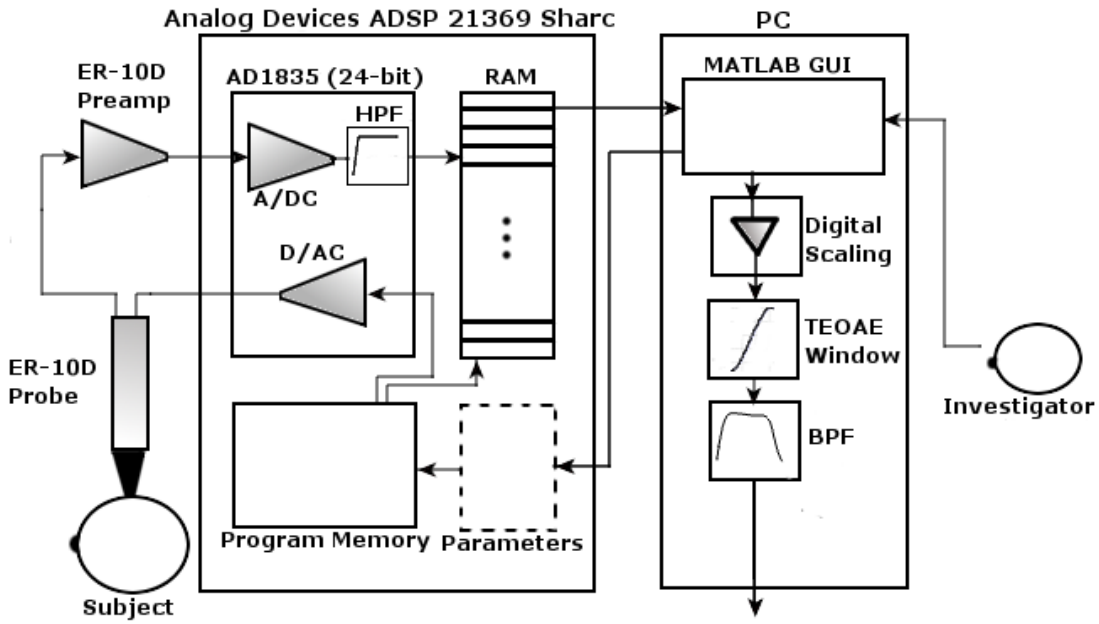

Figure 9. High resolution OAE system block diagram. The investigator chooses the stimulus properties from a GUI, and this information is sent to the AD Sharc EZ-Kit via UART communication. The main program acquires from the ADC, using the DAC as a trigger. After a DC-Blocking filter, the response is synchronously averaged in a data memory buffer. After recording, the signal is passed back to MATLAB for post-processing, which includes digital scaling, windowing, and zero phase-distortion BPF.

During subject testing, the investigator makes parametric adjustments to the stimulus type (such as pulse width, rate, number of epochs, etc.) in a MATLAB GUI. The GUI communicates to the Sharc EZ-Kit through via UART. A memory buffer is updated with the stimulus information and is loaded via serial port into the Sharc EZ-Kit in real time. The on-board program memory triggers the stimulus on the DAC and acquires the amplified recording on the ADC which passes through a hardware DC-blocking filter, built-in to the AD1835. The signal is synchronously averaged in a memory buffer, which

is passed back to MATLAB for post-processing. A block diagram of the system is shown in Figure 9, and a screen capture of the GUI is shown in Figure 10.

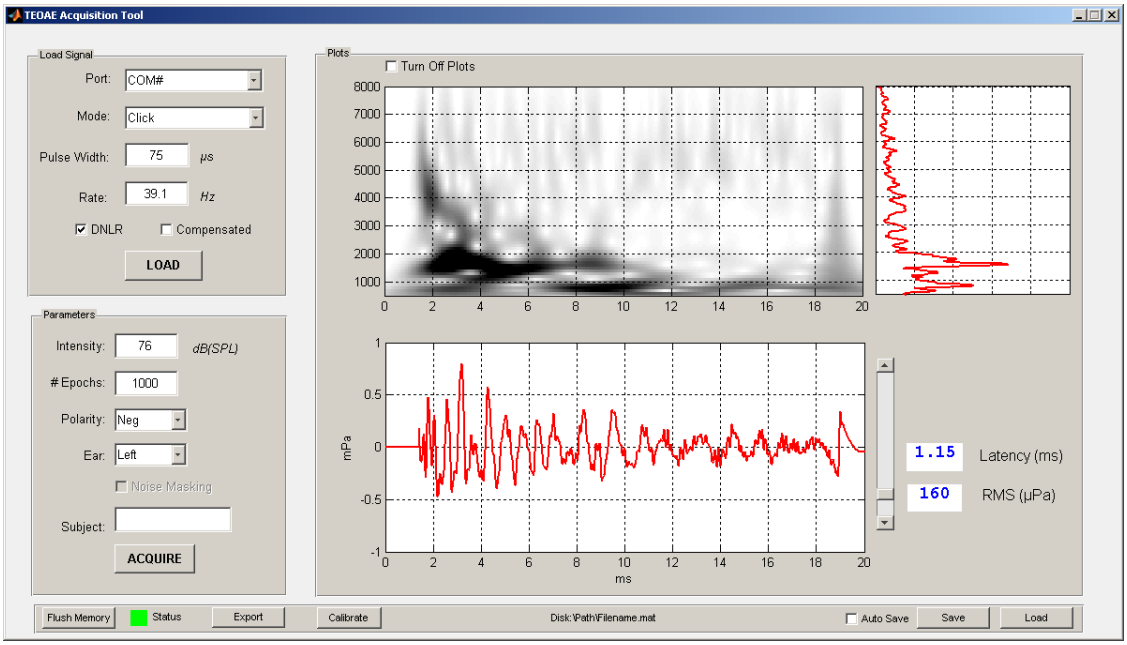

Figure 10. A MATLAB program was written to interface with the acquisition board. The user selects the serial port, the modality (click, swept-tone, etc.), the click pulse width, the repetition rate, intensity, number of epochs, polarity, and subject information. After acquisition, a time plot of the acquired OAE, as well as a time-frequency display and FFT is shown.

2.1.3 NOISE ESTIMATION

In the synchronous averaging procedure, rather than using a single memory buffer to store and average every epoch, two memory buffers are used to interleave the epochs. Therefore, one memory buffer averages the even-numbered epochs, and one averages the odd-numbered epochs. If the system is assumed to be static, that is, the response to the same input does not change from moment to moment then the only difference between

the two buffers should constitute the contribution of noise. So, when the even and odd buffers are averaged together, the total signal average can be obtained. But, when the difference of the even and odd buffers is calculated, an estimate of the noise can be obtained. The summation and difference averages represent the signal and noise estimates, respectively. This process is shown for click acquisition in Figure 11. The SNR can be obtained by obtaining the ratio of the RMS of the signal to the RMS of the noise, given in Eq. (2.4).

$$SNR_{click} = \frac{\sqrt{\frac{1}{N}\sum_{n=1}^{N}(r_e[n] + r_o[n])^2}}{\sqrt{\frac{1}{N}\sum_{n=1}^{N}(r_e[n] - r_o[n])^2}} \tag{2.4}$$

The total number of epochs for the click modality is given by *K*, so each memory buffer contains *K*/2 epochs each. The total length of each buffer for the click modality is given by *N*. The total duration of a trial is the product of the length of each epoch and the total number of epochs, scaled by the sampling rate.

2.2 SWEPT-TONE METHOD

A novel impulse response measurement technique was developed for the characterization of weakly nonlinear, approximately time-invariant systems (Farina, 2007). The method uses an exponentially-swept sinusoid, and was originally developed for the measurement of audio systems, such as loudspeakers and room acoustical environments. It was developed as an alternative to traditional impulse response, MLS, and time-stretched pulse (TSP) methods, which require strict linearity and time-invariance. The nature of the swept-tone technique allows for explicit characterization of the distortion of a system. Furthermore, the swept-tone method is much better at reducing noise than the other

impulse response methods. A generalized description of the swept-tone, its construction, post-processing methods, nonlinear analysis, and noise properties will be given in the following sections.

2.2.1 HAMMERSTEIN SYSTEM

A Hammerstein system involves a static (or memoryless), weakly nonlinear system followed by an LTI system. A weakly nonlinear system is one that can be represented by a converging Volterra series. Figure 12 shows a static nonlinear system, $d[n]$, being driven by in input, $s[n]$, and followed by an LTI system, $h[n]$, to produce the output, $r[n]$.

The overall Hammerstein system is represented by the Eq. (2.5)

$$r[n] = (d(s) * h)[n] \tag{2.5}$$

Standard linear convolution cannot fully describe a nonlinear system, so dynamic convolution must be used. Dynamic convolution system represents a full Volterra kernel. The output, $q[n]$, of a nonlinear system $d[n]$, given some input $s[n]$, can be analyzed using Volterra analysis, or dynamic convolution, given in Eq. (2.6). Since the Volterra analysis operates on a nonlinear and time variant system, the output depends on all system inputs.

$$q[n] = \sum_{i_1=0}^{L-1} d[i_1] \cdot s[n-i_1] + \sum_{i_1=0}^{L-1}\sum_{i_2=0}^{L-1} d[i_1,i_2] \cdot s[n-i_1] \cdot s[n-i_2] + \sum_{i_1=0}^{L-1}\sum_{i_2=0}^{L-1}\sum_{i_3=0}^{L-1} d[i_1,i_2,i_3] \cdot s[n-i_1] \cdot s[n-i_2] \cdot s[n-i_3] + \cdots \tag{2.6}$$

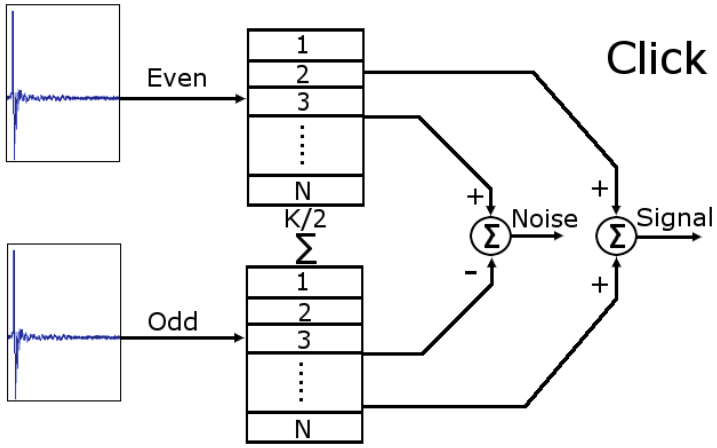

Figure 11. The click response is averaged in an interleaved manner to produce two memory buffers, even and odd. The sum average of the even and odd buffers produces a signal estimate, and the difference average produces a noise estimate. The ratio of average power of these signal and noise averages gives an SNR value.

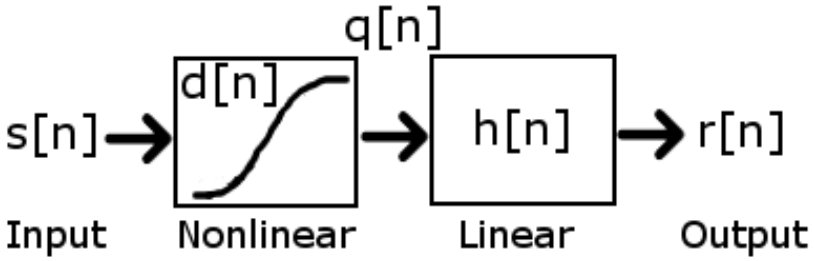

Figure 12. Swept-tone analysis is valid for Hammerstein models, as shown here. The nonlinear system, *d*[*n*], should be only time-invariant and weakly nonlinear, and followed by an LTI system.

The 1st-order (LTI) convolution can be recognized as the first term in this series, and successive orders are given by higher order terms, ad infinitum. This approached provides a complete representation of nonlinear behavior of a time-invariant system,

provided that enough terms of the expansion are used which capture the memory of the system. It can be seen that the dimension of each term (called a kernel), is related to the order of that kernel. So the first order term is only one-dimensional, and is plotted as a line. The second term is two-dimensional, and is plotted as a plane, and so on. Considering just the second order kernel, the diagonal of the kernel represents the instantaneous response of that kernel, and the lower triangle represents the output to previous inputs, and the upper triangle the output to future inputs (these are obviously zero for causal systems). Note that at higher orders the mathematics becomes extremely intensive, and computationally inefficient. But this model can be simplified if a restriction of memorylessness is placed on the system, whereby $d[i_1,i_2,...] = 0$ for all instances that $i_1 \neq i_2 \neq ...$ such that

$$q[n] = \sum_{i=0}^{L-1} d[i] \cdot s[n-i] + \sum_{i=0}^{L-1} d[i] \cdot s^2[n-i] + \sum_{i=0}^{L-1} d[i] \cdot s^3[n-i] + \cdots \quad \textbf{(2.7)}$$

The simplified Volterra model, given in Equation (2.7) considers only the diagonals of the full Volterra matrix, where each dimensional index is equal to the others. The diagonals contain information about the current state only, without regard to previous (or future) states. Swept-tone analysis analyzes such a time-invariant, memoryless, weakly nonlinear system. In the simplified Volterra representation, the computation load increases linearly, with each successive term requiring only one additional multiplication. If a system satisfies the restrictions of a Hammerstein model, then swept-tone analysis can be used to classify linear and nonlinear properties of that system.

2.2.2 Stimulus and Post-Processing Design

2.2.2.1 Swept-Tone Signal

Since the swept-tone stimulus represents a minimum-phase system, the amplitude of the swept-tone and its group delay are intimately intertwined in order to create a flat envelope stimulus. The group delay must be logarithmic such that each octave increase in stimulus frequency is an equal step in time. The group delay, $\tau_g(f)$, is given by the indefinite integral of the power spectrum, scaled by the duration, T, and the energy, as described in Eqs. (11, 12) by Muller and Massarani (2001).

$$\tau_g(f) = T\frac{\int A^2(f)\mathrm{d}f}{\int_{f_1}^{f_2} A^2(f)\mathrm{d}f} \tag{2.8}$$

And the magnitude as a function of frequency is chosen such that the amplitude envelope in the temporal domain will remain constant. As such, the magnitude function, $A(f)$ is designed to have a pink spectrum, with the magnitude attenuating at -10 dB/decade (Muller and Massarani, 2001). This amplitude function is defined as

$$A(f) = \sqrt{\frac{f_1}{f}} \tag{2.9}$$

So substituting back in, the group delay equation becomes

$$\tau_g(f) = \frac{T}{\int_{f_1}^{f_2}\frac{f_1}{f}\mathrm{d}f}\int\frac{f_1}{f}\mathrm{d}f \tag{2.10}$$

All of the constants can be pulled out of the integrals

$$\tau_g(f) = \frac{T}{f_1\int_{f_1}^{f_2}\frac{1}{f}\mathrm{d}f}f_1\int\frac{1}{f}\mathrm{d}f \tag{2.11}$$

Solving the integral

$$\tau_g(f) = \frac{T}{[\log f]_{f_1}^{f_2}} \log f + C \tag{2.12}$$

Evaluate and apply the a rule of subtraction of logarithms

$$\tau_g(f) = \frac{T}{\log \frac{f_2}{f_1}} \log f + C \tag{2.13}$$

The constant *C* can be determined by setting the initial condition of $\tau_g(f_1) = 0$

$$C = -\frac{T}{\log \frac{f_2}{f_1}} \log f_1 \tag{2.14}$$

Substituting in *C*, the group delay can be re-written as

$$\tau_g(f) = \frac{T}{\log \frac{f_2}{f_1}} \left(\log \frac{f}{f_1} \right) \tag{2.15}$$

The dividend in Eq. (2.15), which is a constant, can be thought of as a change of base of the log operand containing the frequency variable, *f*. And the variable, *T*, scales the group delay to the proper duration. The group delay is shown in Figure 13, with regard to its bounding frequencies, f_1 and f_2, and the duration of the stimulus, *T*. The phase, $\varphi(f)$, is defined as the negative integral of the group delay

$$\varphi(f) = -\frac{2\pi}{f_2 - f_1} \int \tau_g(f) \, \mathrm{d}f \tag{2.16}$$

The group delay expression can be re-written by making some substitutions

$$\tau_g(f) = a(\log f + b) \tag{2.17}$$

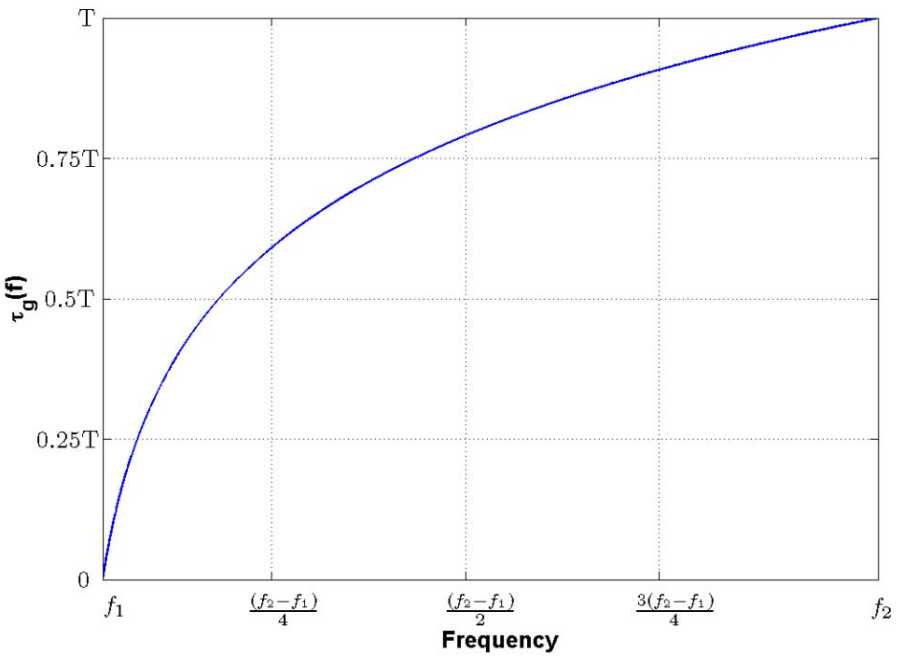

Figure 13. The group delay of the swept-tone increases exponentially from f_1 to f_2 with delays varying from 0 (for the starting frequency) up to T (for the ending frequency).

Where $a = T/\log(f_2/f_1)$ and $b = -\log f_1$

$$\varphi(f) = -\frac{2\pi}{f_2 - f_1}\int \big(a(\log f + b)\big)\mathrm{d}f \tag{2.18}$$

The constant can be pulled out of the integral, and the sum can be separated

$$\varphi(f) = -\frac{2\pi}{f_2 - f_1} a\left(\int \log f \,\mathrm{d}f + b\int \mathrm{d}f\right) \tag{2.19}$$

Solving for the integral

$$\varphi(f) = -\frac{2\pi}{f_2 - f_1} a(f\log f - f + bf + C) \tag{2.20}$$

After some factoring and back-substitution

$$\varphi(f) = -\frac{2\pi Tf}{(f_2 - f_1)\log\frac{f_2}{f_1}}\left(\log\frac{f}{f_1} - 1 + C\right) \tag{2.21}$$

Again, solving for C using the initial condition that $\varphi(f_1) = 0$

$$C = \frac{f_1}{f} \tag{2.22}$$

Finally, the phase can be written as a function of frequency as

$$\varphi(f) = -\frac{2\pi Tf}{(f_2 - f_1)\log\frac{f_2}{f_1}}\left(\log\frac{f}{f_1} - 1 + \frac{f_1}{f}\right) \tag{2.23}$$

A generic representation of the phase is shown in Figure 14, indicated by its bounding frequencies and cycles given in radians.

The total complex solution can be written by combining the magnitude and phase into a complex exponential form

$$S(f) = A(f)e^{j\varphi(f)} \tag{2.24}$$

$$S(f) = \sqrt{\frac{f_1}{f}}\, e^{-j\frac{2\pi Tf}{(f_2 - f_1)\log\frac{f_2}{f_1}}\left(\log\frac{f}{f_1} - 1 + \frac{f_1}{f}\right)} \tag{2.25}$$

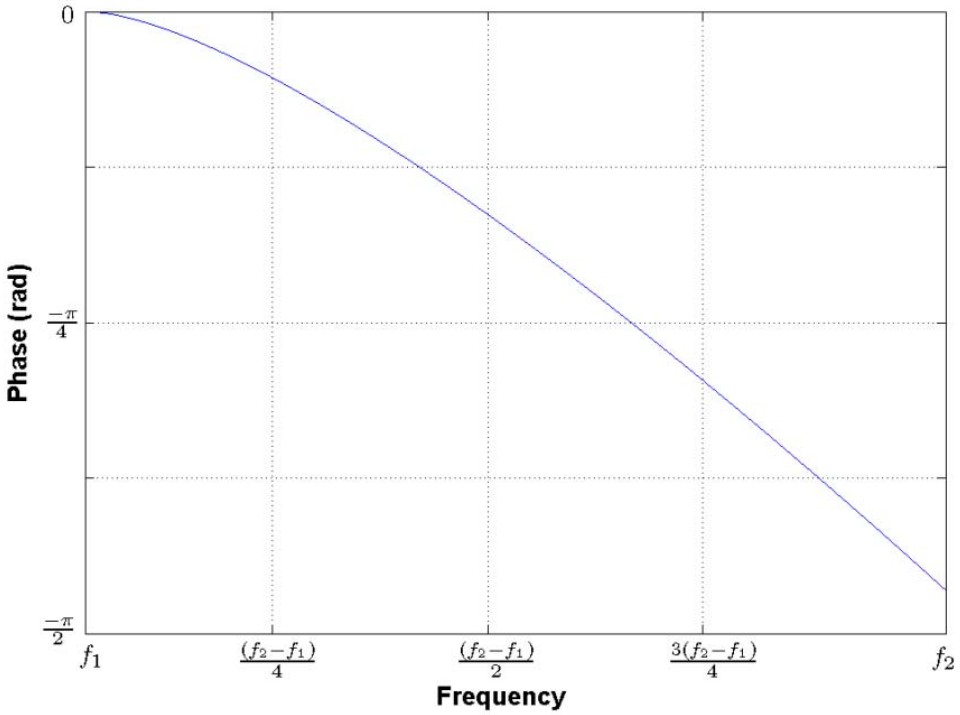

Figure 14. The phase of the swept-tone stimulus decreases nonlinearly from 0 to π/2 over the region of f1 to f2. The phase is constructed from the prescribed group delay function and is used in constructing the swept-tone stimulus.

The time-domain signal can be constructed by taking the inverse Fourier transform

$$s(t) = \mathcal{F}^{-1}\{S(f) \cdot S^*(-f)\} \tag{2.26}$$

An example swept-tone stimulus is shown in Figure 15 with f_1=0.3 kHz, f_2=9.6 kHz, and T=0.1 s.

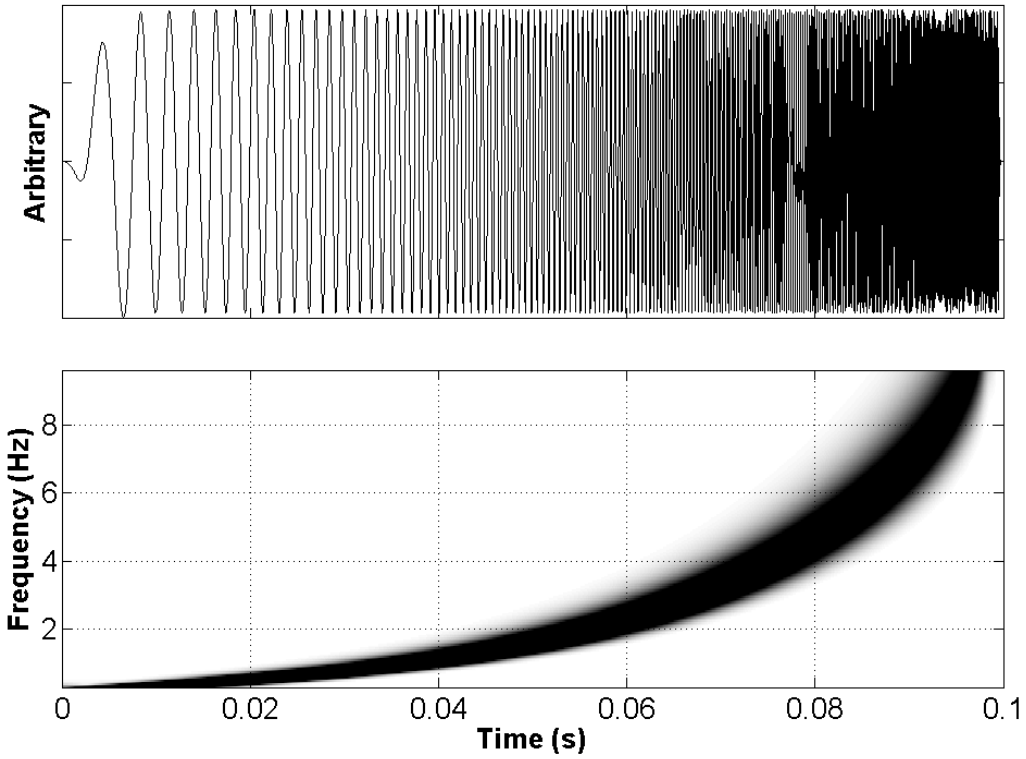

Figure 15. The swept-tone stimulus is an exponentially increasing swept-frequency sinusoid. The upper panel shows a particular stimulus with start and end frequencies of 0.3 and 9.6 kHz over a duration of 100 ms. The lower panel shows a time-frequency representation of the stimulus, which takes an exponential form.

2.2.2.2 Inverse Sweep

The inverse sweep can also be generated in the frequency domain and converted to the time domain via an inverse Fourier transform. The inverse sweep is defined simply as

$$S^{-1}(f) = \frac{BPF(f)}{S(f)} \tag{2.27}$$

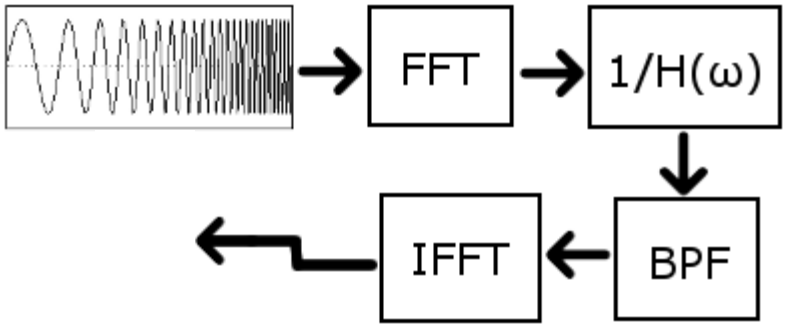

Figure 16. The inverse swept-tone, which is used in post-processing, is obtained in the frequency domain by inverting the swept-tone stimulus, applying a band-pass filter at the f_1 and f_2 frequencies, and transforming to the temporal domain.

Where $BPF(f)$ is a band-pass filter with f_1 and f_2 as its lower and upper cutoff frequencies (shown in Figure 16). The band-pass filter prevents instability that may be induced by the division outside the band-limits of the generated swept-tone signal. The time-domain signal can be constructed by performing an inverse Fourier transformation on $S^{-1}(f)$. The inverse sweep, along with its magnitude and phase response is shown in Figure 17. The magnitude and phase responses are complementary to those of the stimulus. The instantaneous frequency of the inverse sweep decreases logarithmically from f_2 down to f_1 over the same time interval, T.

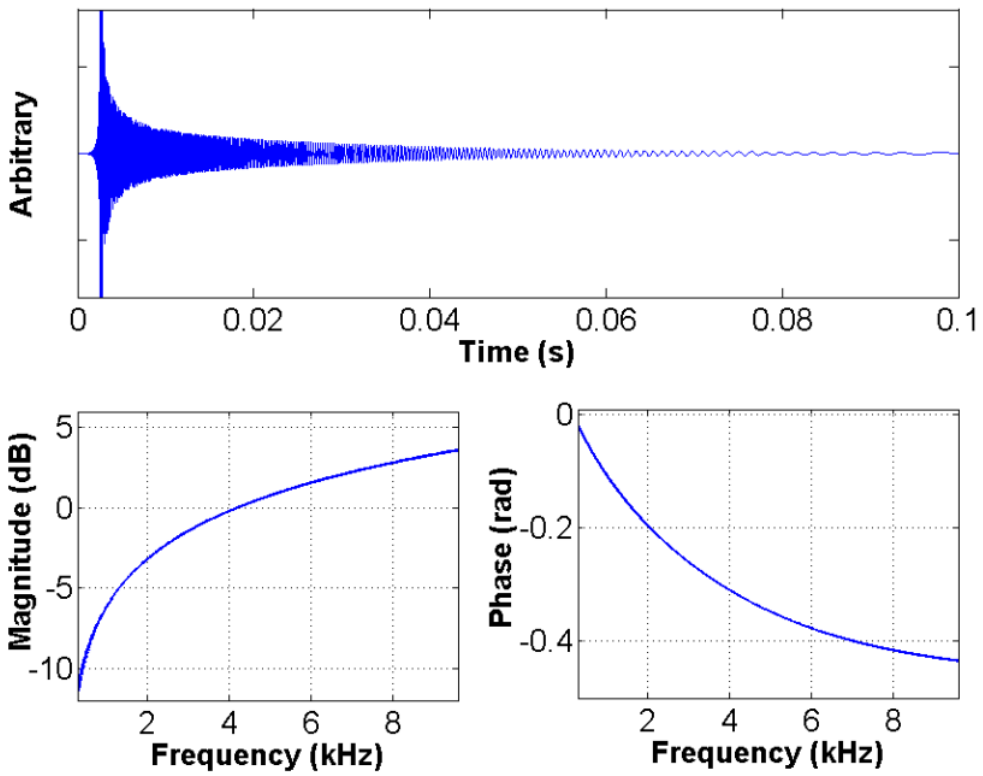

Figure 17. The inverse sweep signal (which is used in post-processing for analysis) is not flat envelope like the stimulus signal, but rather has a -10 dB/decade envelope. The magnitude and phase responses are complementary to those of the swept-tone stimulus.

The inverse sweep, when multiplied with the swept-tone stimulus (in the frequency domain) yields a flat-spectrum, linear-phase signal. The resulting magnitude is a product of the magnitudes of the swept-tone stimulus and inverse sweep, and should be identically 0 dB within the band of $[f_1, f_2]$, as shown in Figure 18. The resulting phase response is linear, with a slope equivalent to $T/2$. Multiplication of $S(f)$ with $S^{-1}(f)$ in the frequency domain is equivalent to the convolution of $s(t)$ with $s^{-1}(t)$ in the time domain, the result of which is a near impulse.

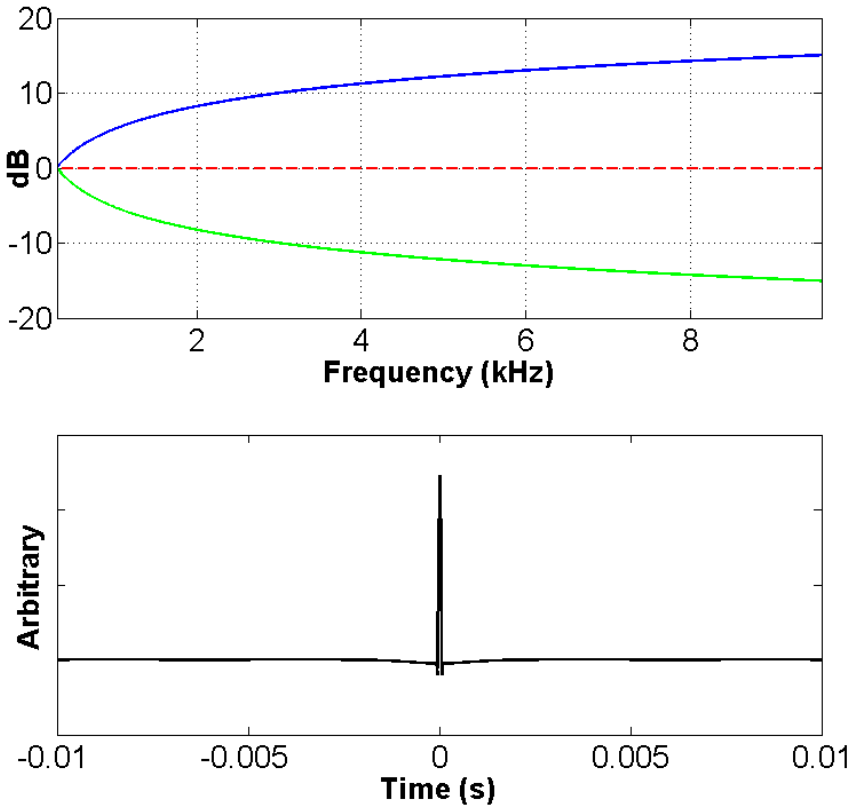

Figure 18. The upper panel shows the magnitude responses of $S(f)$ (green) and $S^{-1}(f)$ (blue) and the resulting magnitude response of their product (red dash). When the product is transformed to the time-domain, a near impulse is generated, as shown in the lower panel with a range of ±10 ms.

2.2.3 Impulse Response Using Swept-Tone Analysis

The impulse response of an acoustical system can be extracted using swept-tone analysis. Similar to how convolution of $s(t)$ with $s^{-1}(t)$ yields an impulse, the convolution of the swept-tone response to some Hammerstein system, $r(t)$, with $s^{-1}(t)$ yields the impulse response of that system. Furthermore, if the system is weakly nonlinear, as described in

Section 2.2.1 then the impulse response of the distortion components can also be obtained.

When the swept-tone stimulus, $s(t)$, is passed through an LTI system, the swept-tone the response, $s(t) * h(t)$, is produced, where $h(t)$ is the impulse response of the LTI system. But if this response is then convolved with the inverse sweep, the delayed impulse response can be obtained.

$$h(t-T) = s(t) * h(t) * s^{-1}(t) = h(t) * \delta(t-T) \tag{2.28}$$

Furthermore, when the swept-tone stimulus is passed through a Hammerstein system, response contains not only the LTI impulse response, but distortion component impulse responses as well. Consider first the group delay of the inverse sweep, which can be simply represented as the negative of the group delay of the swept-tone, as $-\tau_g(f)$. During the process of convolution, the group delay of the response and the group delay of the inverse sweep undergo the process of summation (phase resides in the argument of the exponential, and multiplication of two exponentials results in addition of their phases). The inverse sweep group delay is well defined as simply the negative, or inverse, of Eq. (2.15). The group delay of the response, $r(t)$, can be at the same stimulus frequency, f, or at some integer multiple of the fundamental (eg., $2f$, $3f$, etc.), as in the case of a weakly nonlinear, or Hammerstein system. So the group delay of the response can be generalized as

$$\tau_r(f) = \frac{T}{\log\frac{f_2}{f_1}}\left(\log\frac{pf}{f_1}\right) \tag{2.29}$$

Notice the difference between Eqs. (2.15) and (2.29) is the inclusion of the constant p scaling the frequency variable f. Since the system under consideration is a Hammerstein system, then its harmonic components are limited to positive integers, whereby $p \in \mathbb{N}$. The impulse response group delay is the summation of $\tau_r(f)$ and $-\tau_g(f)$.

$$\tau_{ir}(f) = \tau_r(f) - \tau_g(f) \tag{2.30}$$

Again the laws of logarithms are applied, giving

$$\tau_{ir}(f) = \frac{T}{\log\frac{f_2}{f_1}}[(\log p + \log f - \log f_1) - (\log f - \log f_1)] \tag{2.31}$$

$$\tau_{ir}(f) = \frac{T}{\log\frac{f_2}{f_1}}\log p \tag{2.32}$$

So the group delay of the impulse response of the analyzed response is no longer a function of frequency, and instead is just a function of the order of the system. This implies that each harmonic element of the response will have linear phase of unique slope. And depending on the duration of the impulse response, there will be varying amounts of overlap between the responses of the various orders. The impulse responses become more compressed as p increases. The overlap of the impulse responses can be decreased by either increasing the duration of the swept-tone stimulus, T, or by decreasing the ratio of the starting and ending sweep frequencies, f_1/f_2. In the case that f_1=0.3 kHz, f_2=9.6 kHz, and T=0.1 s, the 2nd order impulse response will arise at t=-20 ms (giving 20 ms of no overlap) and the 3rd order response will arise at t=-32 ms (giving 12 ms of no overlap).

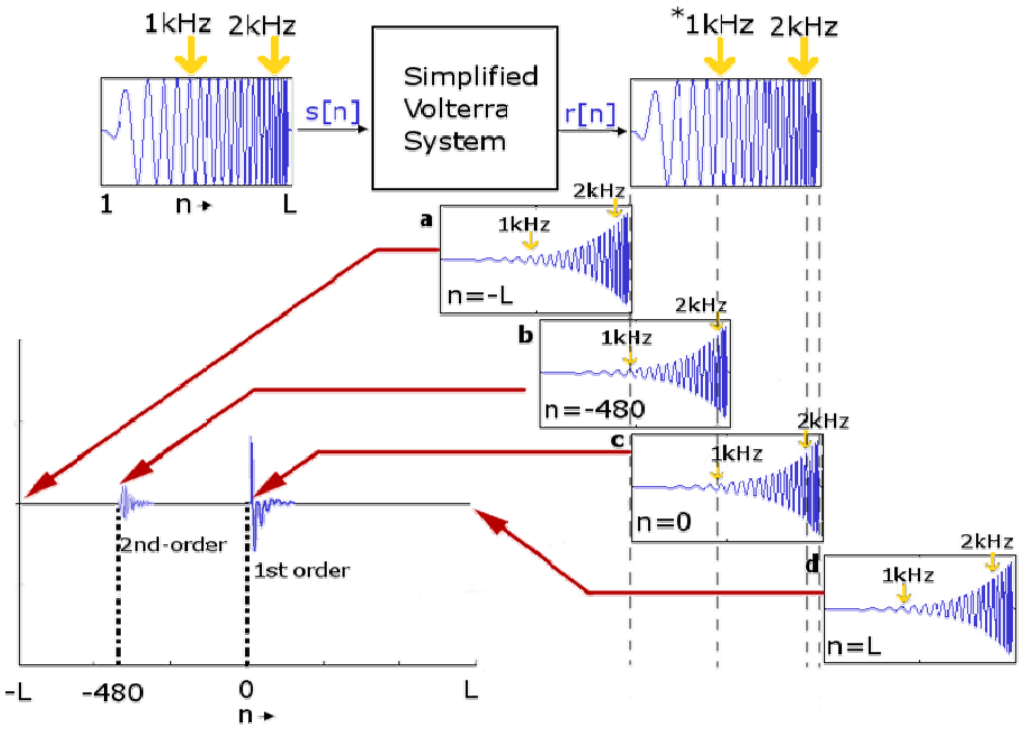

Figure 19. In this example, a swept-tone stimulus, *s*[*n*], of length *L* is passed through a Hammerstein system, resulting in *r*[*n*]. The 1 and 2 kHz locations are identified with yellow arrows. Assume that this system introduces a 2nd-order harmonic at 1 kHz. This distortion component is labeled with an asterisk (*). The inverse sweep is convolved with the response, and so is first time-reversed, and then shifted. At each shift, a summation of the dot product of the overlapped regions is the result of the convolution. At step (a), *n*=-*L*, only 1 sample overlaps, this marks the beginning of the convolved response. At step (b), the 2 kHz component of the inverse sweep overlaps with the 2 kHz distortion component of the response. This marks the beginning of the location of the 2nd-order impulse response. At step (c), full overlap occurs, which marks the beginning of the 1st-order impulse response. Shifting, multiplication, and summation continue until the two signals no longer overlap.

2.2.4 SIGNAL TO NOISE RATIO USING SWEPT-TONE ANALYSIS

When recording a system response, it is ideal to reduce the noise that may be recorded by the microphone. Ambient noise can be reduced by taking the recordings in a quiet environment. Further hardware considerations, such as a clean power supply or using balanced analog lines can also reduce unwanted noise. Additive noise can either be coherent, in other words phase-locked to the onset of the stimulus, or non-coherent. With synchronous averaging, coherent noise cannot be reduced through the process of increasing the number of epochs and averaging since it is reinforced by virtue of being phase-locked with the stimulus. With swept-tone analysis both types of additive noise are reduced when the phase is compressed using the inverse sweep.

2.2.4.1 Variance of Noise for an Averaged Click Response

The noise from each sweep is given by η and is assumed to be white Gaussian. There are K total epochs in the averaging scheme, which calculates the noise variance, given by $var(\eta)$. If the mean of the noise is assumed to be zero, then the variance of the averaged noise, $\bar{\eta}$, can be written as

$$var(\overline{\eta_{clk}}) = E\left\{\left[\frac{1}{K}\sum_{k=1}^{K}\eta_k\right]^2\right\} \tag{2.33}$$

Then the squared term is factored, and re-written with indices i and j.

$$var(\overline{\eta_{clk}}) = E\left\{\frac{1}{K}\sum_{i=1}^{K}\eta_i\frac{1}{K}\sum_{j=1}^{K}\eta_j\right\} \tag{2.34}$$

When $i \neq j$ then $E\{\eta_j\eta_i\} = 0$, so

$$var(\overline{\eta_{clk}}) = E\left\{\frac{1}{K^2}\sum_{k=1}^{K}\eta_k^2\right\} \tag{2.35}$$

A $1/K$ term can be pulled outside the expectation since it is a constant,

$$var(\overline{\eta_{clk}}) = \frac{1}{K}E\left\{\frac{1}{K}\sum_{k=1}^{K}\eta_k^2\right\} \tag{2.36}$$

What remains in the expectation is the definition of the variance of the noise for a single epoch,

$$var(\overline{\eta_{clk}}) = \frac{var(\eta_{clk})}{K} \tag{2.37}$$

Rewritten in the frequency domain, using the transform $var(\eta) \Leftrightarrow var(N)$

$$var(\overline{N_{clk}}) = \frac{var(N_{clk})}{K} \tag{2.38}$$

2.2.4.2 Variance of Noise for an Averaged Swept-Tone Response

The average noise variance is calculated comparably when using the swept-tone, however, each epoch the noise is filtered by the inverse sweep s^{-1},

$$var(\overline{\eta_{swp}}) = E\left\{\frac{1}{K^2}\sum_{k=1}^{K}(\eta_k * s_k^{-1})^2\right\} \tag{2.39}$$

If rewritten in the frequency domain using the transforms $\eta \Leftrightarrow N$ and $s^{-1} \Leftrightarrow S^{-1}$,

$$var(\overline{N_{swp}}) = E\left\{\frac{1}{K^2}\sum_{k=1}^{K}|N_k|^2 \cdot \left|S_k^{-1}\right|^2\right\} \tag{2.40}$$

Since K and the inverse sweep are constant terms, they can be pulled out of the expectation function.

$$var(\overline{N_{swp}}) = \frac{|S^{-1}|^2}{K} E\left\{\frac{1}{K}\sum_{k=1}^{K}|N_k|^2\right\} \quad \textbf{(2.41)}$$

What remains within the expectation is the variance of noise for a single epoch.

$$var(\overline{N_{swp}}) = \frac{|S^{-1}|^2 var(N_{swp})}{K} \quad \textbf{(2.42)}$$

The average noise variance using a click can be substituted in. It can be seen that the average noise variance using a swept-tone is noise-shaped, governed by the spectral characteristics of the inverse sweep,

$$var(\overline{N_{swp}}) = var(\overline{N_{clk}}) \cdot |S^{-1}|^2 \quad \textbf{(2.43)}$$

2.2.4.3 Noise Shaping Using Swept-Tone Responses

The noise is clearly affected by the properties of the inverse sweep, as shown in Eq. (2.42). However the broadband magnitude response of the inverse sweep does not contain any inherent noise reduction properties. Rather, the inverse sweep, $S^{-1}(f)$, contains noise shaping properties, as shown in Figure 20, and this is how the noise reduction is obtained.

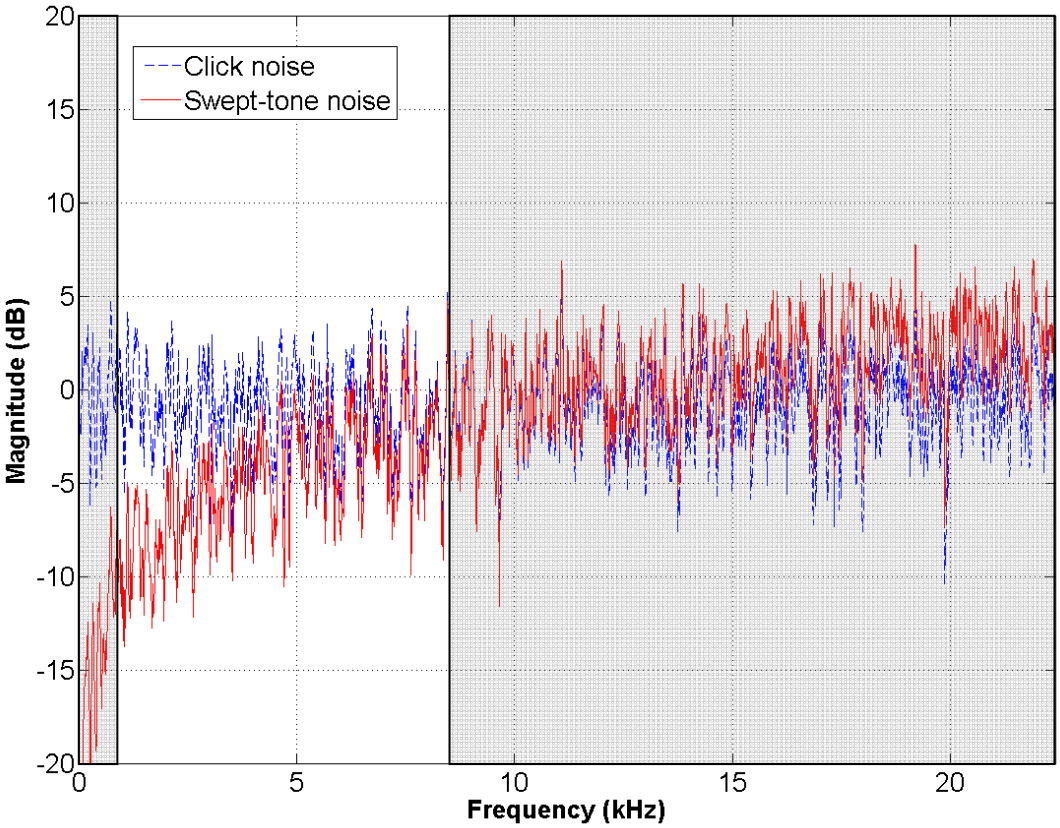

Figure 20. A simulation of Gaussian noise is presented under two acquisition paradigms (click and swept-tone). The click method does not shape the noise frequency characteristics with synchronous averaging. However, the swept-tone acquisition shifts the noise to the high-frequencies, and below 1/3-band, the noise is suppressed. In this case, the sampling rate is 48 kHz, and the analysis band is 0.5 to 8.0 kHz, so the effective noise level is reduced.

2.2.5 Distortion Component Analysis

Distortion components from a simplified Volterra system exist in the form of

$$y = \sum_{p=2}^{\infty} h_p[n] x^p[\theta] \tag{2.44}$$

Note first, that p begins at 2, this is to explicitly consider higher-order distortion components separately from first order, or linear, components. Also it should be noted that in the case of a swept-tone stimulus, x is a sinusoid, and its argument, θ, is defined in the frequency domain, as described in Section 2.2.2.1. However, from the swept-tone impulse response method, what is instead recovered is

$$y = \sum_{p=2}^{\infty} \overline{h_p}[n]x[p\theta] \tag{2.45}$$

First note that the order of the system is now inside the argument instead of as an exponent. Also note the $\overline{\text{bar}}$ above h_p, indicating that this is the measured impulse response from the swept-tone method. But it is possible to reconcile these two equations. Since the input is a sinusoid, the Power Reduction Theorem can be applied, which is given as

$$\begin{aligned} \sin^n\theta &= \frac{2}{2^n}\sum_{k=0}^{\frac{n-1}{2}}(-1)^{\left(\frac{n-1}{2}-k\right)}\binom{n}{k}\sin\big((n-2k)\theta\big), \text{odd } n \\ \sin^n\theta &= \frac{1}{2^n}\binom{n}{n/2} + \frac{2}{2^n}\sum_{k=0}^{\frac{n}{2}-1}(-1)^{\left(\frac{n}{2}-k\right)}\binom{n}{k}\cos((n-2k)\theta), \text{even } n \end{aligned} \tag{2.46}$$

The first few orders of Eq. (2.476) are shown below for the purposes of developing an explicit equation in terms of the measured impulse responses.

$$\begin{aligned} \sin^2\theta &= \frac{1}{2} - \frac{1}{2}\cos 2\theta \\ \sin^3\theta &= \frac{3}{4}\sin\theta - \frac{1}{4}\sin 3\theta \end{aligned} \tag{2.47}$$

$$\sin^4\theta = \frac{3}{8} - \frac{1}{2}\cos 2\theta + \frac{1}{8}\cos 4\theta$$

$$\sin^5\theta = \frac{5}{8}\sin\theta - \frac{5}{16}\sin 3\theta + \frac{1}{16}\sin 5\theta$$

The reduced power expressions can be substituted into an expanded form of Eq. (2.44), utilizing the first five terms and reorganizing, the substitution gives

$$\begin{aligned} y = {} & \left(\frac{1}{2}h_2 + \frac{3}{8}h_4\right) + \left(h_1 + \frac{3}{4}h_3 + \frac{5}{8}h_5\right)\sin(\theta) + \left(-\frac{1}{2}h_2 - \frac{1}{2}h_4\right)\cos(2\theta) \\ & + \left(-\frac{1}{4}h_3 - \frac{5}{16}h_5\right)\sin(3\theta) + \left(\frac{1}{8}h_4\right)\cos(4\theta) \\ & + \left(\frac{1}{16}h_5\right)\sin(5\theta) \end{aligned} \tag{2.48}$$

A Fourier transform is applied to Eq. (2.52) to give

$$\begin{aligned} Y = {} & \left(\frac{1}{2}H_2 + \frac{3}{8}H_4\right) + \left(H_1 + \frac{3}{4}H_3 + \frac{5}{8}H_5\right)X(\omega) + \left(-\frac{1}{2}H_2 - \frac{1}{2}H_4\right)jX\left(\frac{\omega}{2}\right) \\ & + \left(-\frac{1}{4}H_3 - \frac{5}{16}H_5\right)X\left(\frac{\omega}{3}\right) + \left(\frac{1}{8}H_4\right)jX\left(\frac{\omega}{4}\right) + \left(\frac{1}{16}H_5\right)X\left(\frac{\omega}{5}\right) \end{aligned} \tag{2.49}$$

Where X is the Fourier transform of the swept-tone sinusoid, and ω is the Fourier transform of the argument θ. The cosine functions are represented as simply a shift of a sine to the imaginary axis via a multiplication by the complex constant j. Next the Fourier transform is applied to Eq. (2.45), with the series expanded and truncated to the first five terms, which gives

$$Y = \overline{H_1}X(\omega) + \overline{H_2}X\left(\frac{\omega}{2}\right) + \overline{H_3}X\left(\frac{\omega}{3}\right) + \overline{H_4}X\left(\frac{\omega}{4}\right) + \overline{H_5}X\left(\frac{\omega}{5}\right) \tag{2.50}$$

A linear system of equations can be solved by equating the corresponding terms of Eqs. (2.49) and (2.50), which can be written after some algebra as the following system of equations

$$\begin{aligned} H_1 &= \overline{H_1} + 3\overline{H_3} + 5\overline{H_5} \\ H_2 &= 2j\overline{H_2} + 8j\overline{H_4} \\ H_3 &= -4\overline{H_3} - 20\overline{H_5} \\ H_4 &= -8j\overline{H_4} \\ H_5 &= 16\overline{H_5} \end{aligned} \qquad \textbf{(2.51)}$$

Obviously, if more than five terms of the sinusoidal power series were considered, then this system of equations in (2.51) would extend potentially to infinity and beyond. However, for the particular case of transient evoked OAEs, recent research would suggest that OAE distortion components exist only in the 2nd and 3rd orders, and certainly not beyond the 5th order (Thornton *et al.*, 2001). The symmetric inverse Fourier transform of Eq. (2.51) gives the Volterra kernel diagonal slices, $h_1 ... h_5$, which can be formulated in terms of the measured impulse responses $\overline{h_1} ... \overline{h_5}$ which are found using the swept-tone stimulus.

2.2.6 Comparison to IR and MLS Methods

2.2.6.1 Impulse Response Method

The impulse response (IR) can be obtained from an acoustical system by creation of an electrical click which is presented by a speaker to the system, and a microphone is used to record the response of the system. Unless the system noise is extremely low, multiple iterations, or epochs, are needed to reduce the noise floor and improve SNR. This is

standard in biomedical signal processing through a process known as synchronous averaging. Synchronous averaging is robust against wide sense-stationary (WSS) or Gaussian noise, but is susceptible to time variance and also to coherent noise. However, IR-based methods can be made robust against time variance by the use of time delay estimation, such as Woody's method (Woody, 1967). However, an IR-based method is still susceptible to nonlinearities, so system distortions will appear in the impulse response. Furthermore, IR-based methods are rate-limited, in that the duration of the IR determines the minimal acquisition period, otherwise convolutional overlap of the responses will occur. The final caveat to an IR-based method is the selection of the stimulus. The electrical stimulus fed to the speaker must be a rectangular pulse in order to induce excursion from the driver. If the pulse width is too narrow, then the speaker's bass response will be attenuated. However, the pulse width comes with a trade-off, as pulse width is related to a high-frequency roll-off and notch, whereby as the pulse width gets larger, the high-frequency notch moves towards baseband. This is an engineering trade-off that must be accounted for in stimulus design.

2.2.6.2 Maximum Length Sequences (MLS)

Maximum length sequences (MLS) are a type of IR-based method, but they can be acquired at very high rates, such that overlap of the responses does occur. The MLS method uses a prescribed pseudo-random stimulus pattern which is highly jittered. The convolved response can be correlated with the input stimulus to obtain the transient response. And due to the higher acquisition rate, the MLS method has an improved SNR over a standard IR-based method. Also, a complete Volterra kernel describing the nonlinearities can be derived using the MLS method (Reed and Hawksford, 1996).

However, unlike the IR-based method, an adaptive time-delay estimation, like the Woody's method, cannot be used. And in fact, the MLS-method is extremely susceptible to time-variance since the stimulus timing is strictly prescribed. A MLS is a specific duration of 2^n-1, where $n \in \mathbb{N}$, and so a priori knowledge regarding the length of the IR is needed to formulate the sequence. And while the pseudo-random sequence itself has a white spectrum, the individual stimulus pulses are still subject to the high-frequency/low-frequency trade-off as the IR-based methods.

2.2.6.3 Continuous Loop Averaging and Deconvolution (CLAD)

A generalized form of the MLS method is the so-called continuous loop averaging and deconvolution (CLAD) method. The CLAD method uses a specially designed jittered sequence as well but it has the advantage of designing for either a maximally jittered sequence (which is the same class as MLS) or minimally jittered sequence. Transient responses are obtained via frequency-domain division of the response by the stimulus. However, unlike MLS, noise becomes a serious consideration, and a sequence must be specially designed to avoid amplifying noise in a particular frequency region.

2.2.6.4 Swept-Tone Method

The swept-tone method uses an extended sinusoid with a flat-top envelope in the time domain and a pink noise magnitude response in the frequency domain. The SNR properties, as shown in Equation (2.43), are significantly improved over standard click methods. This is because the stimulus energy lasts much longer (on the order of 1000 times) than the signal energy of a click. Furthermore, the swept-tone method is resilient to additive coherent noise (e.g., electro-magnetic interference or other artifacts which are phase-locked to the onset of the stimulus) which can readily disrupt responses under IR-

based methods. Also, like the MLS method, the swept-tone method can separate nonlinear responses from linear ones in a weakly nonlinear system. However, while the MLS method is capable of extracting the full Volterra kernel, the swept-tone method can only identify the diagonals of the Volterra kernel. Finally, the swept-tone method is robust against slight time variance, which can disrupt the acquired response using the MLS method. In general, it is expected that the swept-tone method will:

1) provide improved SNR over any IR-based method;
2) separate linear from nonlinear responses;
3) remove correlated, or phase-locked, noise from the response;
4) remain relatively robust against time variance in dynamic systems.

2.3 TIME-FREQUENCY ANALYSIS

The wavelet transform allows for a relatively short signal to be decomposed into its elementary frequency constituents with high frequency and time resolution. The selection of a mother wavelet is important in the subsequent analysis. For these experiments, a Morlet mother wavelet was chosen. The mother wavelet is scaled and shifted to the proper analysis frequency and latency, which in this study ranged from 0.8 to 8.0 kHz in 0.05 kHz steps from 0 to 20 ms post stimulus onset. The Morlet wavelet, given by Equation (2.52) below, is a complex Gaussian shaped function in both the time and frequency domains. It is also symmetric on the real and imaginary axes.

$$w(t, f_0) = \frac{e^{\left(-t^2/2\sigma_t^2\right)} e^{2j\pi f_0 t}}{\sqrt{\sigma_t \sqrt{\pi}}} \quad \textbf{(2.52)}$$

The mother wavelet, $w(t, f_0)$, is a function of the center frequency, f_0, and time, t, with σ_t representing the standard deviation of the signal in the time domain.

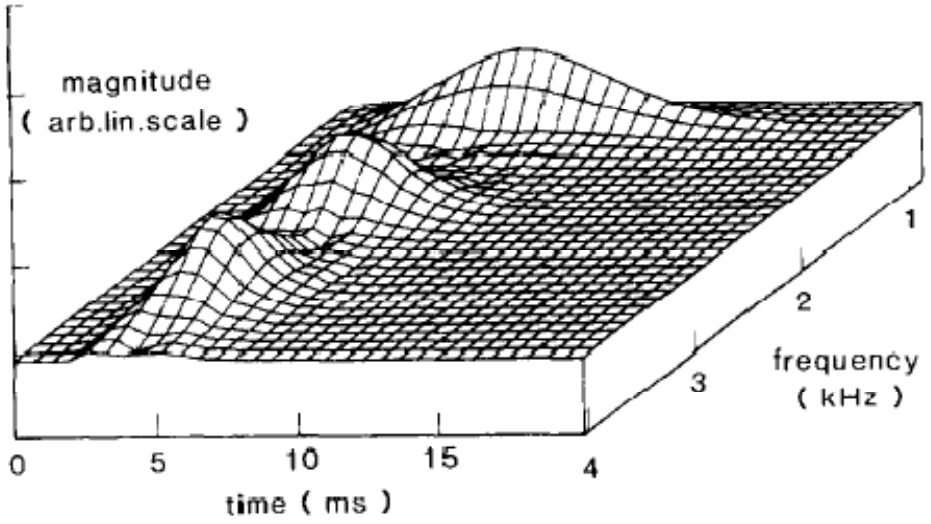

Figure 21. This figure shows a wavelet decomposition of a synthesized TEOAE signal. Note that high-frequency components occur in the early latencies, and low-frequency components occur in the late latencies. From (Wit *et al.*, 1994).

Wavelet transformation is a useful post-processing tool to decompose TEOAE into its analysis contributing frequency components (Wit *et al.*, 1994; Tognola *et al.*, 1997). The point of maximal energy at a specific frequency is related to the latency of that component. Since the basilar membrane has a frequency-dependent topology from basal to apical end, a frequency-dependent TEOAE latency is expected as well as shown in the wavelet constructed time-frequency displays. In such analyses, however, the high-

frequency early-latency components generally are obstructed by the extent of the meatal artifact.

Chapter 3 Compensation for the Meatus

In acquiring TEOAEs, as mentioned previously, electrical rectangular pulses (henceforth referred to as clicks) are fed to an acoustical receiver, or speaker. This speaker resides within a probe assembly, which is also coupled with a microphone for recovery of the response from within the outer auditory canal. However, the transfer function of the speaker, the probe assembly, the ear canal, and the microphone all conspire to distort the click, and what results acoustically within the ear canal is actually far from a click. One such distortion is the acoustical ringing of the meatus. If thought of as a single open-ended Helmholtz resonator 27 mm in length, the meatus will have a half-wave resonance frequency around 3 kHz. The acoustic ringing primarily occurs at this resonant frequency.

This acoustical ringing has two primary artifacts. In the frequency domain, a peak resonance is created with roll-off on either side, creating an unequal distribution of energy across frequency, obfuscating the actual signal reaching the tympanic membrane. Secondly, acoustic ringing causes a prolonged impulse response in the ear canal known as the meatal response (MR). An ideal click would be short and transient in duration generating no oscillations afterwards. In reality, however, after the initial burst of energy, smaller pressure variations linger in the meatus for up to 6 ms. This creates a problem in the recovery of early latency TEOAEs. Both of these artifacts negatively impact high frequency OAEs; a high-frequency roll-off due to the resonant response of the meatus supplies less energy to high frequencies, and the residual ringing in the meatus obscures the high frequency OAEs, which are early latency emissions.

3.1 THEORY AND METHODS

The objective is to design an equalization filter whose magnitude and phase responses will compensate that of any meatal response. Given a recorded ear response $h(t)$, the corresponding frequency-domain transfer function will be $H(f)$. Then some acoustic input $x(t)$ will be filtered by the meatus such that the recorded output, $y(t)$, will be

$$y(t) = x(t) * h(t) \Leftrightarrow Y(f) = X(f) \cdot H(f) \tag{3.1}$$

The objective is to find a linear time-invariant (LTI) filter that can be used to reshape the stimulus, resulting in $y(t)$ that is more like a desired acoustic click after being convolved with the auditory canal. In this experiment, the swept-tone is used to find the transfer function of the ear. Then, this function will be used to find the compensation filter that will in turn generate the compensated click. The delivery of the compensated click to the ear canal will result in an impulsive acoustic input which will be similar to an ideal impulse function with reduced acoustic ringing.

3.1.1 MAGNITUDE EQUALIZATION

First, a magnitude equalization process is derived from the meatal impulse response obtained with the swept-tone technique. Few epochs (on the order of 100) are needed to obtain the impulse response of the meatus. Magnitude equalization will dampen the main resonance of the meatus, which in turn shortens the duration of the impulse response. The equalization is applied to the stimulus, as a type of pre-emphasis, or compensation, which will flatten the magnitude response.

The solution is to create an inverse filter, $H^{-1}(f)$, that has peaks and notches that complement those of the meatal response, $H(f)$. However, certain elements of $H(f)$ may

be very close to zero, particularly at the specification limits of the probe microphone, or generally very low and very high frequencies. If inverted, these elements will create extremely high values. For this reason, parametric band-limits were used to limit the inversion algorithm to the frequencies of interest. Typical TEOAE protocols define an analysis bandwidth of 0.8 to 4.0 kHz (Hatzopoulos *et al.*, 2003), but an extended upper limit is desired for this study, so a bandwidth of the decade spanning 0.8 to 8.0 kHz is used.

The quasi-inverse of $|H(f)|$ defined on $[0,\pi]$ is given as $|H'(f)|$ in the subset $[f_1, f_2]$, where f_1, f_2 are the upper and lower limits of the swept-tone stimulus. Therefore, $|H'(f)|$ is defined piecewise as

$$|H'(\omega)| = \begin{cases} \dfrac{1}{|H(\omega)|}, & \omega \in [\omega_1, \omega_2] \\ 1, & \omega < \omega_1, \omega > \omega_2 \end{cases} \quad \textbf{(3.2)}$$

A transition region must be created to eliminate ringing in the time-domain at any created edge frequencies that arise from the piecemeal construction of the magnitude response. One of the stated objectives of this compensation method is to reduce ringing of the meatus; therefore special attention must be given to the edge frequencies to eliminate the possibility of introducing new process-related ringing. Windowing is applied at the bands between baseline to ω_1 and between ω_2 to $\omega_2 + \Delta\omega_2$, or some arbitrary higher frequency. A Hann window function was chosen to transition smoothly from the compensation region to the unity regions to eliminate time-domain edge frequency ringing.

3.1.2 PHASE EQUALIZATION

A transient signal should have constant group delay across all frequencies; this implies that the phase is linear. In the construction of the compensation filter, a compensation phase must be applied to linearize the phase of the meatal response. First an ideal phase response must be computed which is given by a line connecting $H(0)$ to $H(\pi)$. The phase is given by the argument of the complex variable H, and will be denoted by $\angle H(f)$. Then the idealized phase as a function of ω is

$$\angle H_{ideal}(f) = \big(\angle H(\pi) - \angle H(0)\big) \cdot f + \angle H(0) \tag{3.3}$$

If we let the phase at DC be zero, then the ideal phase simplifies to

$$\angle H_{ideal}(f) = \angle H(\pi) \cdot f \tag{3.4}$$

When a system of two LTI filters are in series, their phases sum linearly, so the phases of the compensation filter and the meatus should sum to our constructed idealized phase. So the phase of the compensation is the difference of the phase of the meatus and the idealized phase, whereby

$$\angle H'(f) = \angle H(\pi) \cdot f - \angle H(f) \tag{3.5}$$

3.1.3 OVERALL COMPENSATION FILTER

The compensation filter can now be constructed using the desired magnitude and phase responses, such that

$$h'(t) = \mathcal{F}^{-1}\{H'(f)\} = \mathcal{F}^{-1}\left\{|H'(f)| \cdot e^{j \cdot \angle H'(f)}\right\} \tag{3.6}$$

The input click stimulus, or in this case a rectangular pulse denoted by x(t), is filtered with the compensation filter, $h'(t)$, whereby the output is given as $y(t) = x(t)*h'(t)*h(t) \approx$

$x(t)$. In other words, the stimulus in the ear canal with compensation filter is a close approximation to the actual desired stimulus than is the non-compensated stimulus. An overview of the compensation method is diagramed in Figure 22.

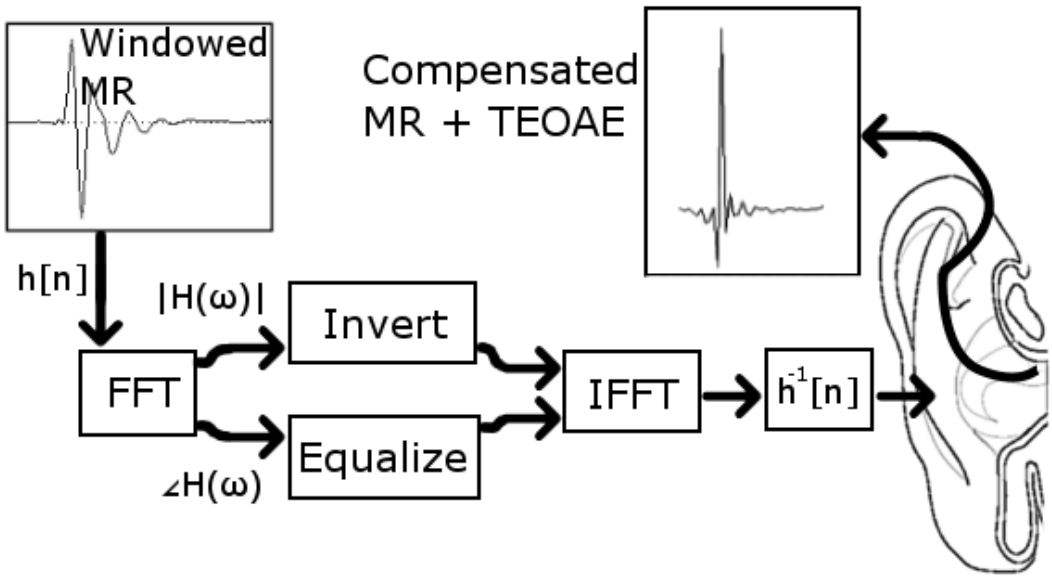

Figure 22. After obtaining the impulse response of the ear via the swept-tone technique, time windowing is performed to obtain the MR signal, $h[n]$. Frequency-domain inversion is performed by deriving the auditory canal compensation filter. The phase of $H^{-1}(\omega)$ is designed to sum with the phase of $H(\omega)$ to create a linear phase response (constant group delay). The magnitude of $H^{-1}(\omega)$ is designed such that the product with $H(\omega)$ is unity. The resulting time-domain transform yields the electrical stimulus, $h^{-1}[n]$.

3.1.4 SUBJECTS

Data were acquired from 12 ears of 6 young adult subjects (4 male and 2 female) of ages ranging from 18 to 28, all in accordance with an IRB-approved protocol. All subjects

had normal audiograms with thresholds better than 15 dB HL at frequencies 0.5, 1.0, 2.0, and 4.0 kHz, and bone-air gaps of less than 15 dB. For TEOAE acquisition, the subjects sat or lied down in an acoustically attenuated and electro-magnetically shielded environment. An Etymotic Research ER-10D OAE probe fitted with a rubber tip was inserted into the meatus to form an acoustic seal.

3.1.5 RECORDING PROTOCOL

For the standard click recordings, 75 μs clicks were used. The compensated stimuli were mean-adjusted to prevent unnecessary speaker excursion and digitally filtered forwards and backwards to prevent phase distortion using a 4-pole Butterworth filter with cutoff frequencies of 0.8 and 8 kHz. The swept-tone stimulus used for acquiring the MR had beginning and end frequencies of 0.256 and 16.384 kHz, and were swept over 2,400 samples. With a sampling rate of 48 kHz, the total duration of the swept-tone was 50 ms. A raised half cosine window was applied to the first and last millisecond.

The compensated stimulus is expected to have complementary magnitude and phase characteristics to the transfer function of the auditory canal. The magnitude response of the compensated stimulus when multiplied with the magnitude response of the meatus created a flat spectrum, and the phase response of the compensated stimulus when summed with the phase of the meatus created a constant group delay. The phase response was unwrapped and detrended to emphasize the small fluctuations in the response, which can have a large impact on the group delay, or the ringing of the time-domain response.

TEOAE responses were recorded at 86 dB peak sound pressure level (pSPL), or 400 mPa, down to 50 dB pSPL (6 mPa) in 6 dB steps. TEOAE recordings were obtained by synchronous mean averaging of 2000 sweeps. The averaging window was 25 ms, or 1200 samples at a 48 kHz sampling rate. The stimuli were presented to the subjects at a rate of 39.1 Hz in order to reduce the impact of electromagnetic interference through the process of averaging. Artifact rejection of 5 mPa (48 dB pSPL) was used during the time window of 5 to 25 ms post-stimulus onset.

3.1.5.1 Windowing Criterion

An objective method for analyzing the duration of the MRs was selected in order to compare TEOAEs acquired via standard and compensated stimuli. The MPEG-7 standards delineate a specific temporal descriptor, which is intended to identify the effective duration of a transient (Peeters and Deruty, 2008). This measure first derives the temporal envelope estimation, *e*, by the process of moving a selectable window across the calculated energy of the signal, given by

$$e[n] = \sqrt{\frac{1}{N} \sum_{i=n-N+1}^{n} h^2[i]} \tag{3.7}$$

where *h* is the impulse response of the auditory canal, *N* is the length of the averaging window, and *n* is the discrete time index. The temporal envelope estimation is based on the short-time average energy, where the 1 ms window (N=48, Fs=48 kHz) acts as a low-pass filter of the energy. The effective duration is the amount of time that the envelope is above a given threshold. A threshold of 0.75 mPa was chosen so as not to exclude any

TEOAEs, and to choose a level low enough such that the MR will not obstruct the TEOAE response significantly.

3.1.5.2 Spectral Flatness Measure

In order to objectively compare the spectra of the standard click with the compensated click, a spectral flatness measure (SFM), adopted from the MPEG-7 standards of low level audio descriptors was used (Hellmuth *et al.*, 2001). The spectral flatness is a scalar that is computed by the ratio of the geometric mean to the arithmetic mean of the spectrum of interest, and is given by

$$SFM = \frac{\left(\prod_{k=k_1}^{k_2} H(k)\right)^{1/(k_2-k_1)}}{\frac{1}{(k_2-k_1)}\sum_{k=k_1}^{k_2} H(k)} \quad \textbf{(3.8)}$$

$H(k)$ is the value of the magnitude spectrum at bin index k, and specific indices k_1 and k_2 are the FFT bin locations of the upper and lower band limits f_1 and f_2. In this study, a 4096-point FFT was used, so f_1 and f_2 correspond to bins k_1=44 and k_2=2,796. Using the SFM formula, tonal signals will have values closer to 0, and broadband signals will have values closer to 1.

3.2 RESULTS

Electrical and acoustical characteristics of the standard and compensated clicks are characterized as recorded from one subject (SUB01R) in Figure 23. The rectangular click waveform displayed in the top row shows the expected slowly decreasing magnitude response with flat phase. The corresponding acoustic waveform plots in the second row display the acoustic ringing (the meatal response) resulting from the ear canal

characteristics. The corresponding magnitude and phase plots show the expected magnitude gain around 3 kHz and the phase shifts at high frequencies. The waveforms of the electrical compensated click, in the third row, show the complementary magnitude and phase characteristics to the plots shown in the second row. The electrical compensated click is comprised of several oscillations starting about 0.7 ms prior to the onset of the click and lasts about 2.7 ms. The resulting acoustic waveform of the compensated click is plotted in the fourth row (leftmost) and shows a more impulsive stimulus as expected. The magnitude response plot of the acoustic compensated click shows an improvement in the desired flatness (spectral flatness index for this case improved from 0.8819 to 0.9997), most notably in a boost to high-frequencies, and a smoothing of the low-frequency ripple (in-band ripple decreased in this case from 17.3 dB to 6.7 dB, a 10.6 dB reduction in ripple). Furthermore, the phase plot shows near zero phase characteristics, indicating a constant group delay.

As expected the compensated stimuli varied from ear to ear, as each is unique to the individual transfer characteristics. However, some similarities among the stimuli existed; each stimulus contained 0.7 ms of pre-onset energy and 2.7 ms of post-onset energy for total stimulus duration of 3.4 ms. The equalization is performed on $h(t)$ that is normalized for digital processing to [-1, 1], the electrical compensated stimulus was then scaled to its proper output level to match the standard click in terms of dB pSPL as described in (Burkard, 1984). Alternative measurements of sound pressure level are peak to peak (ppSPL) and peak equivalent (peSPL), and the latter can be subdivided into baseline-to-peak or peak-to-peak peSPL. For a brief acoustic transient, the difference

between pSPL and peSPL is about 3.5 dB, which can be assumed as the margin of error in the intensity of the output level of the compensated click.

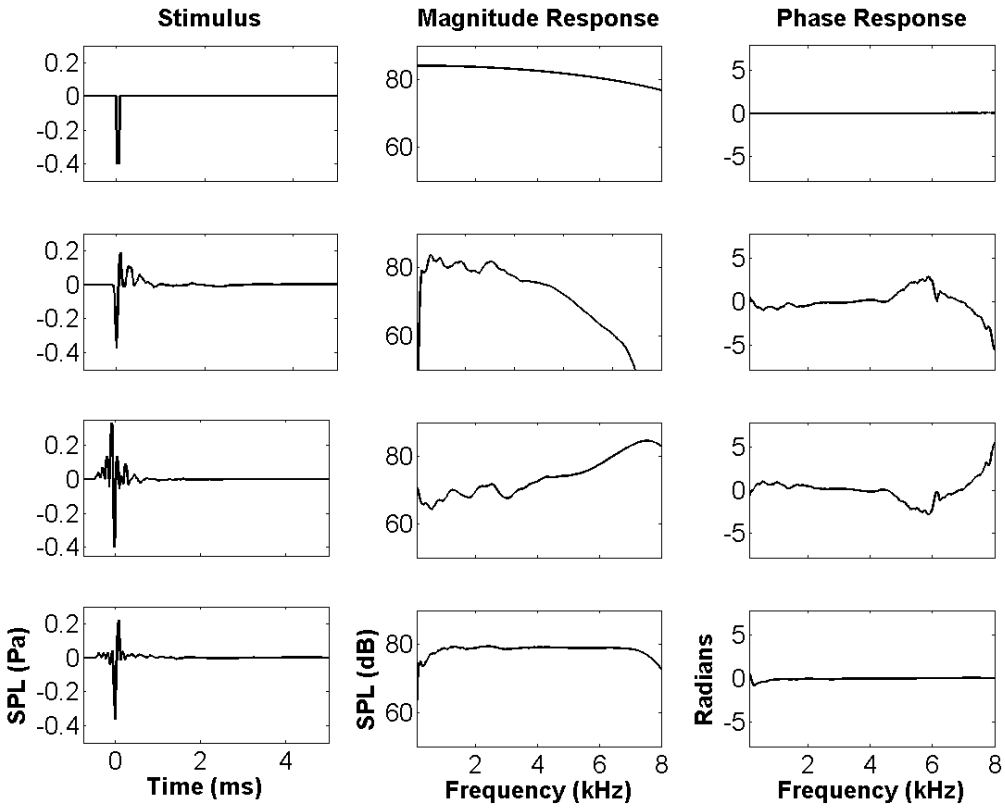

Figure 23. Time and frequency characteristics of the signals of the OAE system: (left column) time-domain signals, (middle column) corresponding magnitude response, and (right column) phase response. The top two rows show the 75 μs rectangular pulse electrical stimulus (1st row) and corresponding acoustic response (2nd row) as recorded from the auditory canal. The bottom two rows show the electrical compensated stimulus (3rd row), and its acoustic compensated meatal response as recorded in the auditory canal (4th row), respectively.

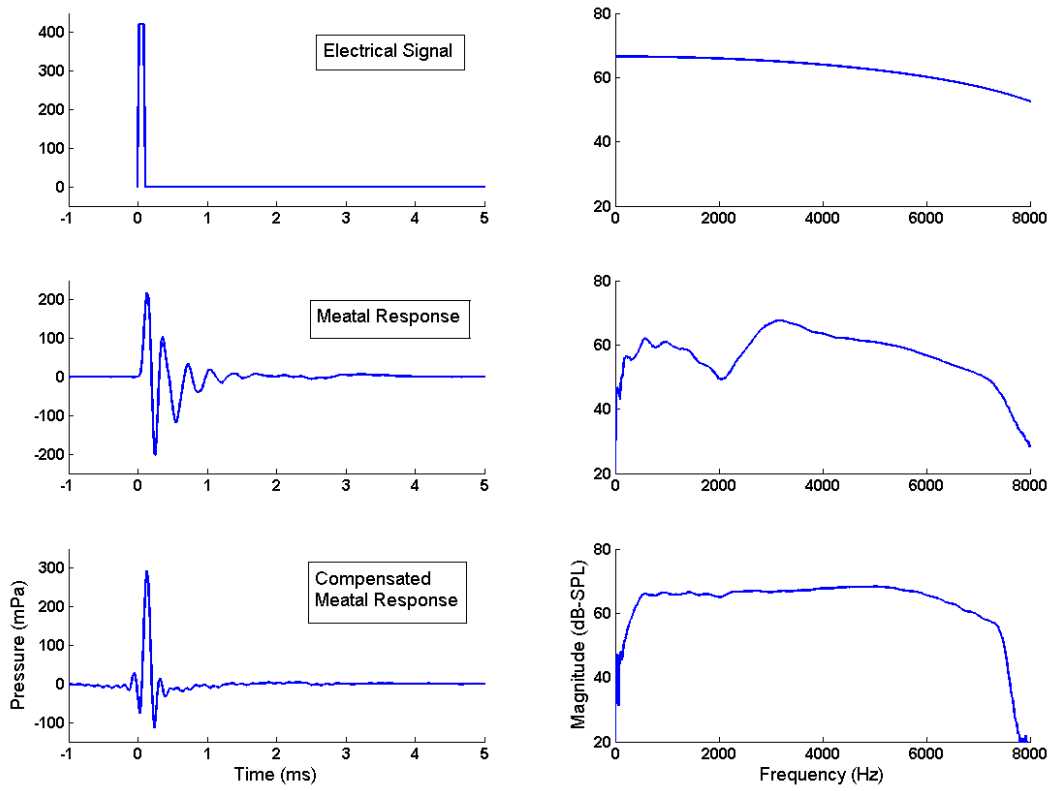

Figure 24. Transfer Functions (SUB02R). The electrical stimulus is short in duration, and relatively flat in magnitude response. A standard click exhibits time-domain ringing and an uneven magnitude response. A compensated click produces an idealized waveform with flat magnitude response and little time-domain ringing.

The use of the compensated click results in improved auditory canal responses. For example, Figure 24 shows an ear canal recording taken from one subject. The desired stimulus is short in duration and generally flat in magnitude response. However, an uncompensated acoustic click stimulus will tend to have resonant ringing, resulting in an uneven magnitude spectrum. However, with the use of a compensated click, a short-duration, flat magnitude response can be obtained. This also has implications on the recovery of TEOAEs. Figure 25 shows an example of the TEOAE acquired with a

standard click, exhibiting large amounts of ringing and with a compensated click, exhibiting reduced ringing. If the temporal ringing is reduced, then early-latency high-frequency TEOAEs can be recovered more readily.

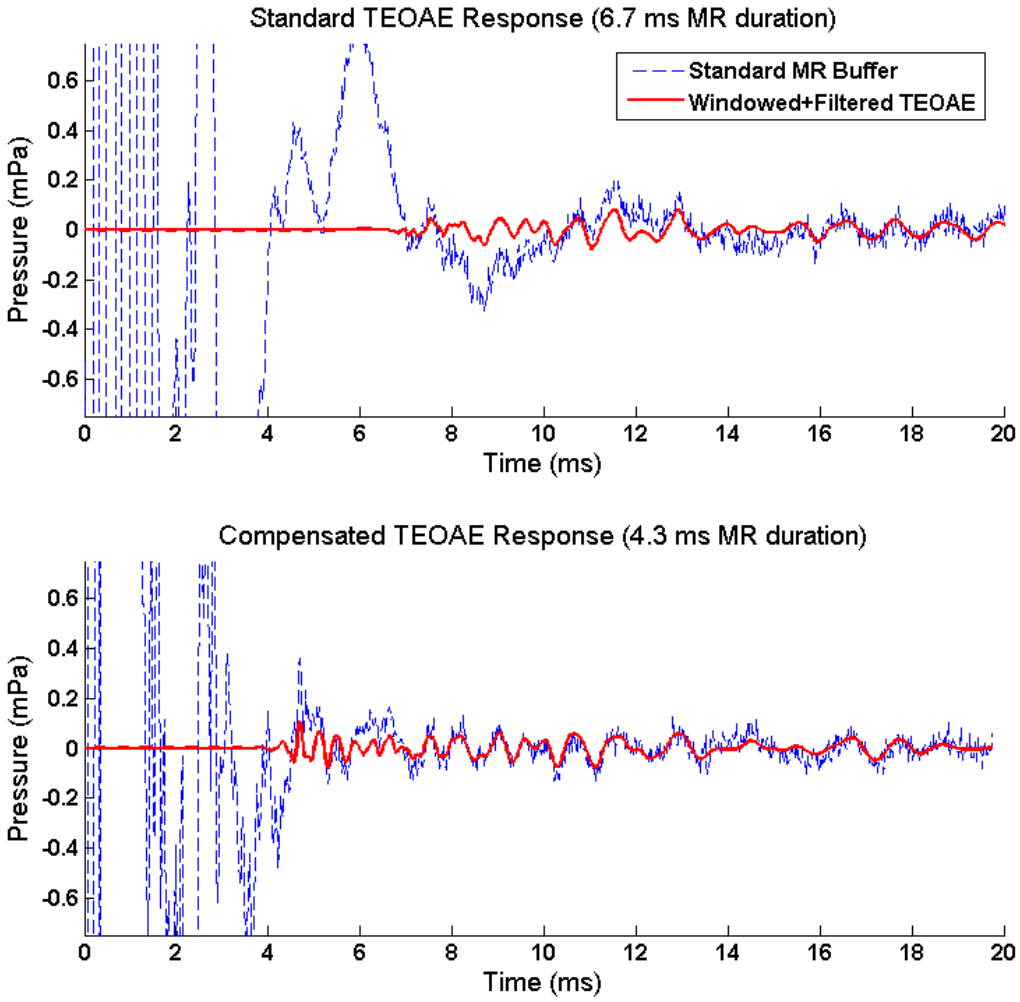

Figure 25. Standard and Compensated Click TEOAE Responses (SUB09R). The ringing of the standard click (upper) prevents the acquisition of TEOAEs prior to 6.9 ms. However, by utilizing a compensated click (lower), TEOAEs as early as 4.7 ms may be obtained. The "raw" and otherwise unfiltered buffer (- - -) and the windowed and band-pass (0.5 to 6.0 kHz) TEOAE response (—) are displayed.

A collection of all of the individual ear canal responses recorded at 86 dB pSPL is shown in Figure 26 in time- and frequency-domain representations. The effective duration, as defined in Section 3.1.5, of the MRs are reduced at this intensity level by an average of 1.2 ms. Furthermore, as a result of the compensated click being represented as a true acoustic transient, the magnitude and phase responses are flattened between the bands of interest. The magnitude flatness, as measured by the spectral flatness index, improves on average from 0.7626 to 0.9918, which is close to a true acoustic transient, which has a theoretical spectral flatness index of 1.0. The flattening of the magnitude and phase responses is merely the frequency-domain representation of shortening the impulse response of the acoustic stimulus. This is especially noticeable for the phase response, whose derivative directly gives the frequency-dependent group delay, such that a flatter phase response with no fluctuations will have a more constant group delay than a phase with many fluctuations.

When considering the removal of the meatal artifact, the ringing must be reduced to at least the level of TEOAEs, if not lower in order to classify the data as meaningful. So observation on a reduced pressure scale reveals how the compensated click truly reduces the meatal artifact across all ear canal recordings. Figure 27 shows the ear canal recordings at the highest-intensity recording (86 dB pSPL) for all subjects in both ears. The vertical axis is scaled for observation of the ringing at intensities near the windowing point. As described in a previous section, the window point is selected as time at which the effective duration of the ear canal recording falls below 0.75 mPa. It is clear from Figure 27 that the compensated click will have an effective duration that is shorter than

the standard rectangular pulse. In fact, at 86 dB pSPL, the average effective duration for the standard click is 6.0 ms and for the compensated click is 4.8 ms.

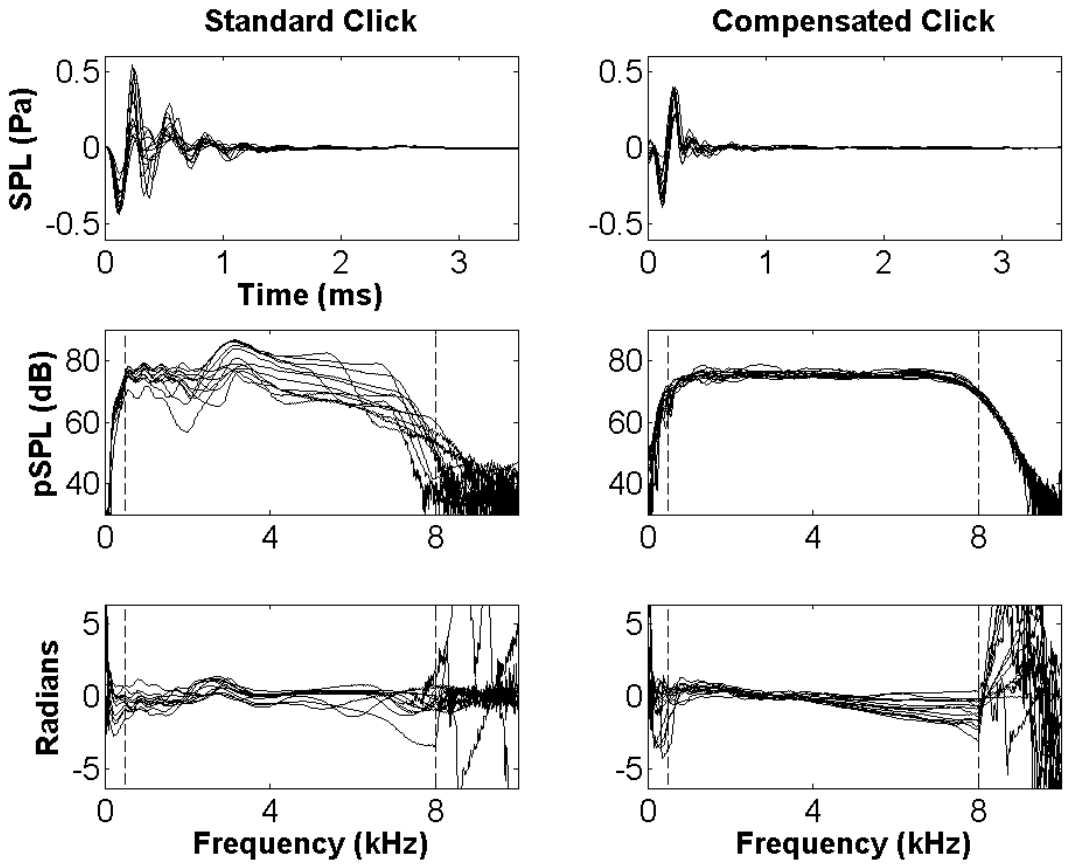

Figure 26. The impulse response (top row), magnitude response (middle row), and phase response (bottom row) for a standard rectangular pulse (left column) and the ear canal compensated click (right column). The effective duration for the compensated click is much shorter, has a flatter magnitude response, and also displays a more linear phase response than its rectangular pulse counterpart.

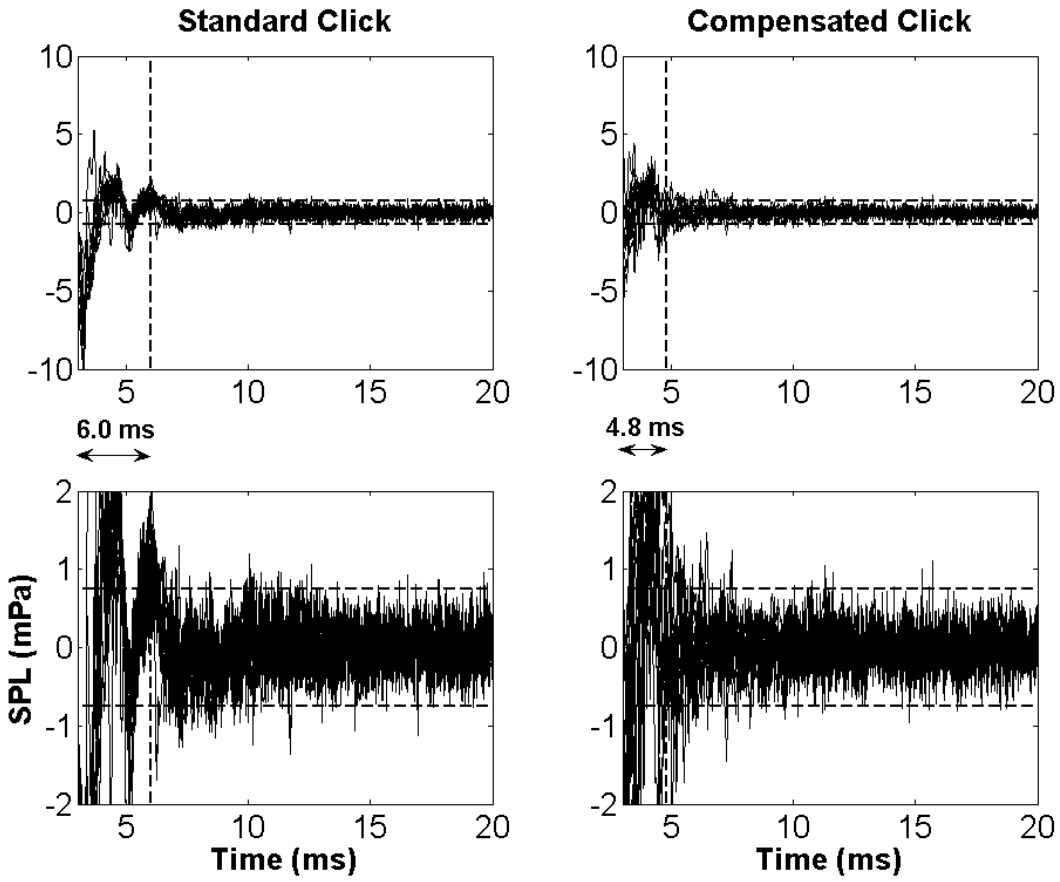

Figure 27. Composite acoustic waveforms from all ears to 86 dB-SPL stimulus showing the standard (left column) and compensated (right column) clicks from the time period of 3 ms to 20 ms. The bottom row shows y-axis zoomed version of the first row. The advantage of the compensated click becomes apparent at closer zoom intervals, whereby ringing in the time-domain is reduced by several ms.

The effective durations of MR for both standard and compensated clicks in linear and nonlinear mode are plotted for all intensities in Figure 28. The horizontal axis indicates all of the intensity levels which were recorded, and the vertical axis indicates the effective duration, in ms, of the mean inter-subject recordings. The effective duration is an indication of the duration of ringing in the ear canal recordings. In both linear and

nonlinear mode the standard rectangular pulse click has a significantly longer effective duration than the compensated click for moderate to high intensity stimuli. In linear mode, the smallest improvement is 0.22 ms at 50 dB pSPL and the greatest improvement is 1.5 ms at 68 dB pSPL.

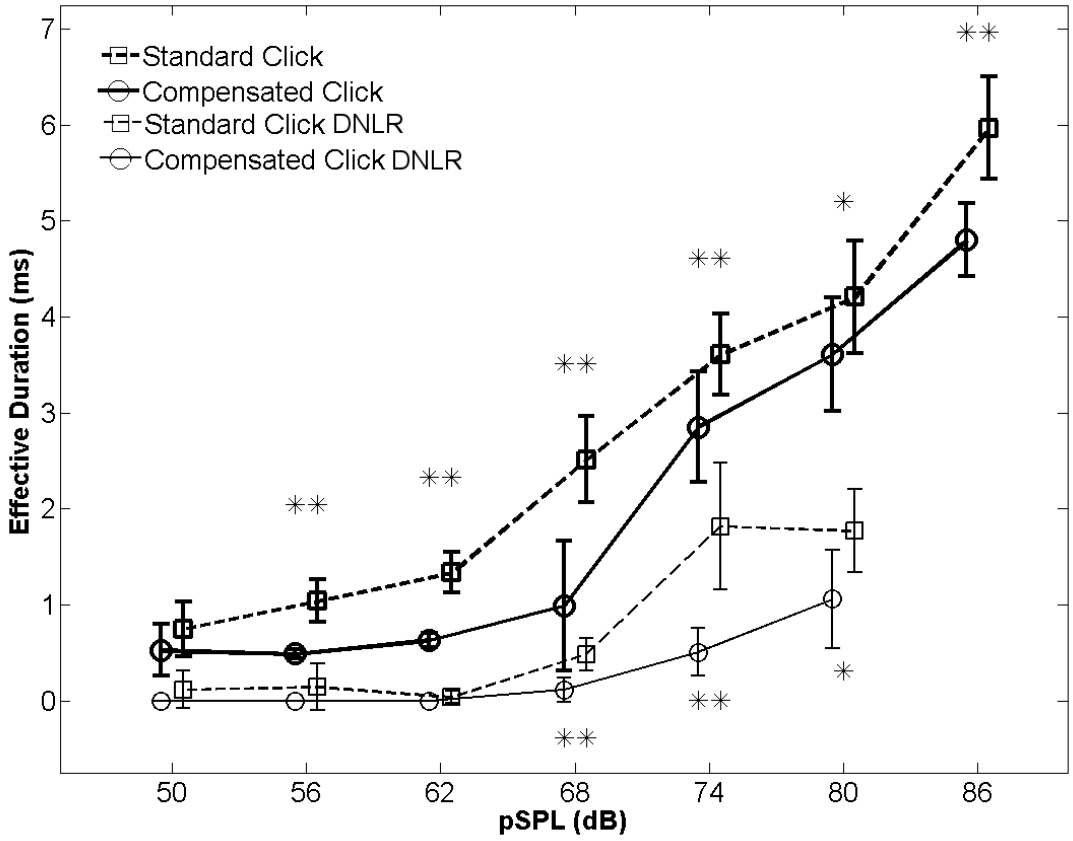

Figure 28. The mean effective durations (± 1 standard deviation) of the MR at each recorded intensity level for both the standard (dashed) and the compensated (solid) clicks are shown for linear acquisition (in bold) and for DNLR acquisition (in thin). An average of about 1 ms improvement is obtained by using an auditory canal compensated click. Significance (* p<0.05; **p<0.01) is found at moderate to high intensity levels for both modalities, and even at low intensities for linear mode.

The use of the compensated click serves primarily to create an ideal acoustic click in the ear canal, of which the two most salient features are a shortened impulse response and a flattened magnitude response. The impulse response was objectively determined using its effective duration as the modality of measurement, and similarly the magnitude response, or spectrum, is objectively determined with the use of the spectral flatness measure (SFM). As shown in Figure 29, the compensated click has a significantly flatter spectrum ($p<0.01$) compared to a rectangular pulse click at all intensity levels. This indicates that, much like an ideal acoustic click, the compensated click has a magnitude response that is nearly flat, and has a SFM that is 10-15% higher than the SFM of a standard click response.

3.3 DISCUSSION

This experiment explored the improvements of using a compensated click that mimics an idealized acoustic click in the auditory canal. This is accomplished by acquiring the transfer function of the meatus via a swept-tone stimulus. From this transfer function, an inverse filter is created by equalizing both the magnitude and phase responses within the band limits of interest. The inverse filter is used to create the electrical stimulus, which when distorted by the meatus will produce an ideal acoustic click, meaning it will have a very short effective duration and a flat magnitude response.

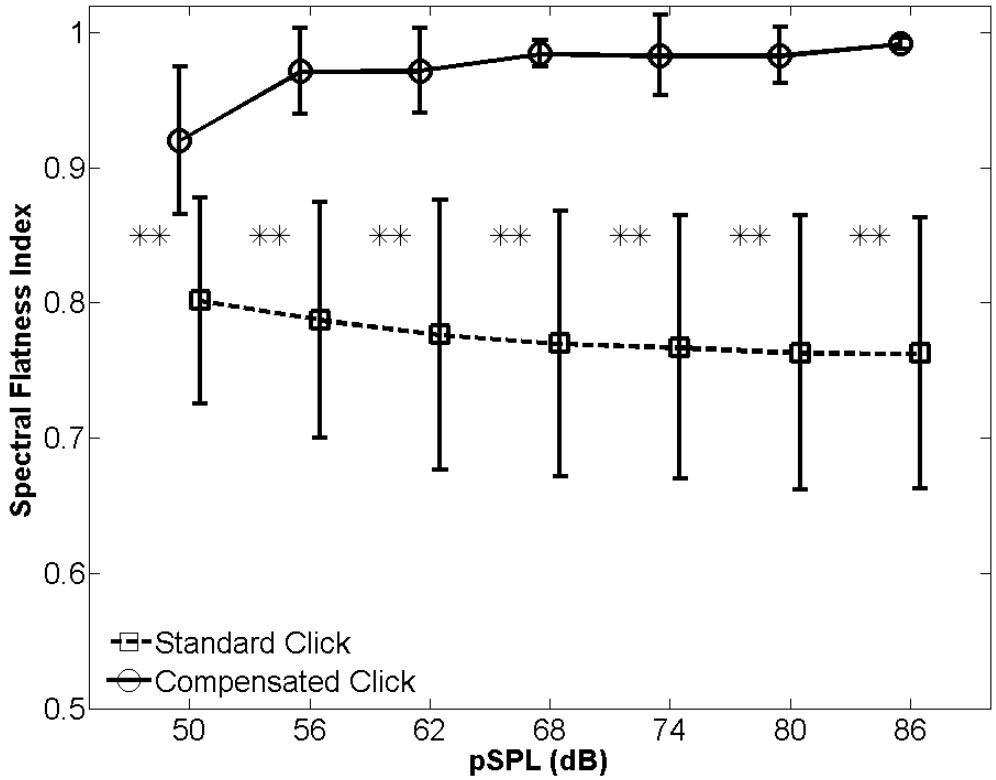

Figure 29. The mean (± 1 standard deviation) spectral flatness index plotted at each recorded intensity level for linear modality. The spectral flatness measure is a scalar from 0 (tonal) to 1 (broadband) indicating how flat the magnitude response is within a given band limit (0.8 – 8.0 kHz). Strong significance (p<0.01) is found at all click levels. Furthermore, the inter-subject variance is lower for the compensated click at all intensities.**

Tests were done on 12 individual ears at multiple intensities from 50 to 86 dB pSPL in 6 dB steps with both the compensated click and a standard rectangular pulse click. Objective measures of the effective duration and the spectral flatness were determined by using low level audio descriptors from the MPEG-7 standard. It was found that use of the compensated click significantly reduces the effective duration of the ear canal MR. In turn, this reduces the meatal artifact, revealing early latency TEOAEs, which are high-

frequency emissions. In fact, at these early latencies, even a small shift of the windowing point towards the stimulus onset results in a large increase in the frequency acquisition. Many of the TEOAE time-frequency models use a power equation in the general form of $t = af^b$ whereby the TEOAE frequency, *f*, increases exponentially (according to constants, $b \approx -0.5$ and $a \approx 10$) and with a large gradient near early latency times, *t*. Furthermore, it was found that use of the compensated click significantly flattens the magnitude spectrum of the meatal response, presumably providing to the middle ear a more equal energy distribution in the TEOAE band of frequencies (0.8 to 8.0 kHz). Specifically, it was found that the compensated click removes a peak near 3 kHz (the Helmholtz resonant point of a 27 mm cavity) and boosts the high-frequency roll-off.

The combination of revealing early latency emissions while providing an increased output level to high-frequencies yields a TEOAE response that has an observably improved high-frequency response. Figure 30 (top panel) shows the recordings from a single ear at 86 dB pSPL using a standard click and a compensated click, respectively. The window point was selected when the effective duration of the meatal response reached a threshold of 0.75 mPa. Time-frequency displays shown below the time plots reveal the concentrations of TEOAE energy for both types of clicks. Additionally, offline DNLR processing was performed, shown in Figure 30 (bottom panel). The compensated click has a shorter effective duration than a standard rectangular click. Both linear and nonlinear modalities for the same subject show a reduction in effective duration, and in linear mode an increase in high-frequency responses can be observed. The reduction of the acoustic ringing exposes early latency, high-frequency TEOAEs for the linear modality. In fact, this trend is seen generally across all of the subjects.

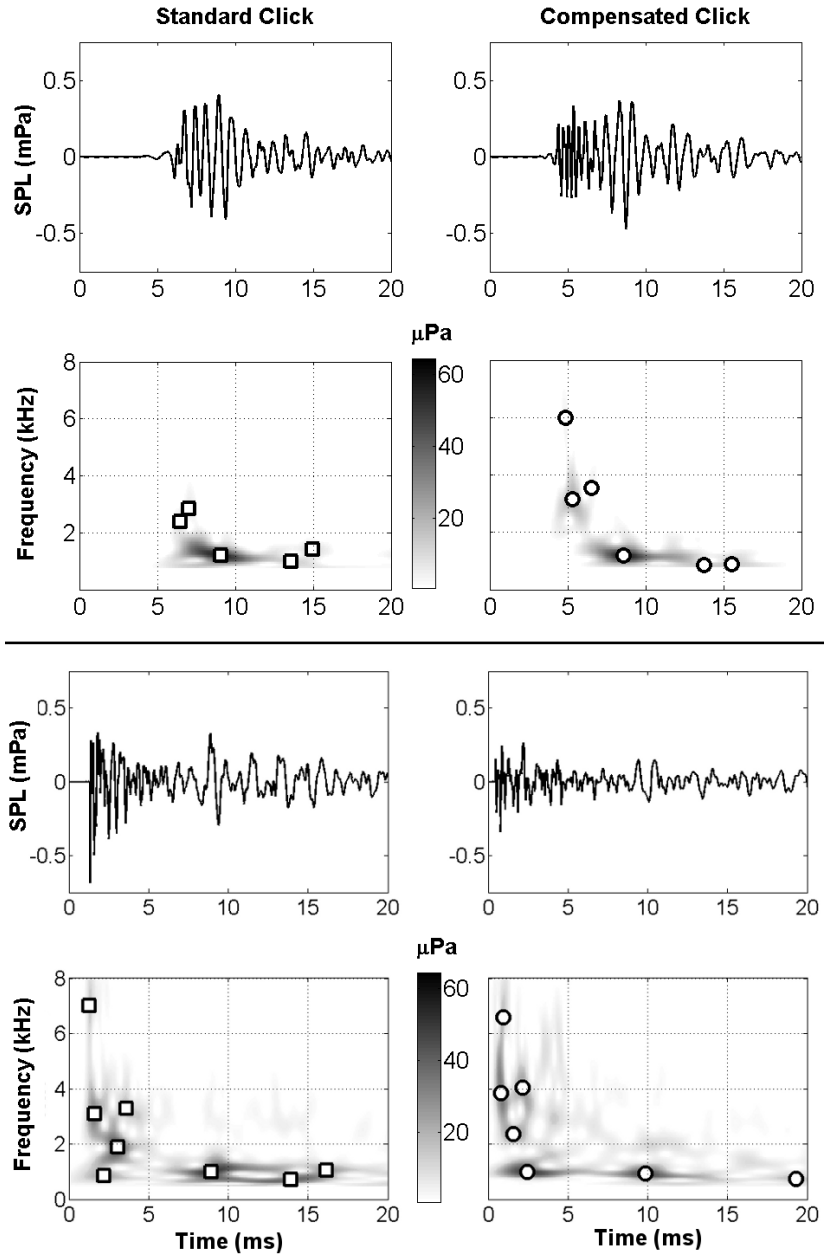

Figure 30. Standard and compensated click recordings are shown from one subject (SUB01R) for linear (top 4 panels) and DNL modes (bottom 4 panels). The upper rows show the TEOAEs, and the lower rows show the corresponding TFA. An increase in high-frequency responses can be seen with the compensated click. A decrease in the effective duration of the acoustic click in both modalities is observed with the compensated click. Time-frequency peaks are indicated with square (standard) and circle (compensated) markers.

The time-frequency wavelet plots tend to reveal energy formed in blobs, which are groups of contiguous pixels with continuous frequency and time and mostly exhibiting an ellipsoid shape. The regional maxima for each of these blobs are determined for a neighborhood of at least 8 connected pixels. Peaks below 30 μPa were removed from the blob peak analysis. A composite of all the blob peaks from all tested ears at 86 dB pSPL is shown in Figure 31 for both click types. It can be seen that there is an improvement in high-frequency information using the compensated click at all intensity levels. This suggests that the use of compensated clicks yields improved TEOAE information above the traditional upper band limit of 4.0 kHz.

As in the case in Figure 30, it was generally seen that in linear mode, there was no significant change in low-frequency content of the TEOAE response, but in nonlinear mode a reduction of TEOAE response was found in the 1.0 to 3.0 kHz components. These results confirm those obtained by (Chertoff and Guruprasad, 1997) whom found a general reduction in TEOAE amplitudes when calibrated clicks were used. The Chertoff studies used the DNLR acquisition mode which relies on the response heavily saturating at the high-intensity click. As the authors suggest, if the TEOAEs elicited from energy in the spectral peak drop out of saturation when the peak is reduced, then the reduction in response will be magnified due to the nature of nonlinear residual subtraction. As a confirmation of this hypothesis, in this study a similar reduction in nonlinear mode was found around the 1.0 to 3.0 kHz components.

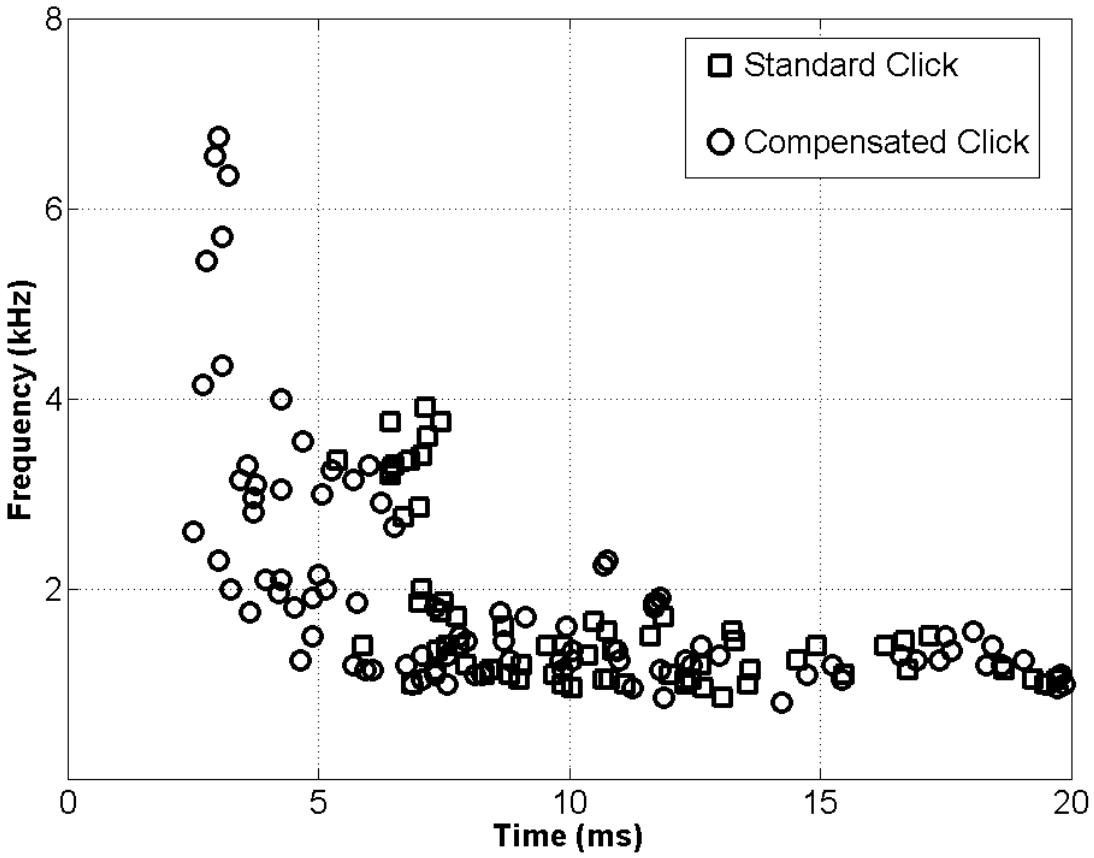

Figure 31. The time-frequency peaks of the energy blobs obtained at 86 dB pSPL for all the ears using linear modality. All local maxima were found, and this number was reduced to only relevant peaks by applying a threshold (30 μPa). For high intensity stimuli, the use of the compensated click (circles) reveals early-latency, high-frequency emissions better than a standard click (squares).

Overall, the advantages of the compensated click described above may improve clinical OAE testing, especially in hearing screening. Time savings offered by the compensated click with the expanded information gained may reduce the number of false positives and reduce costs. In research, better modeling of TEOAEs can be achieved with spectrally compensated clicks. This could be especially important in animal models where high quality TEOAEs could not be obtained easily due to short durations and small

amplitudes. Finally, insight into early latency emissions may provide valuable information for TEOAE generation theories and time-frequency models often used in research.

Chapter 4 Swept-Tone TEOAE

In this experiment, an interpretation of SFOAEs is explored in which TEOAE-like responses are derived from a swept-tone single frequency stimulus. The response can be considered as an SFOAE because, although the stimulus frequency is constantly rising, at any instantaneous moment there is only a single pure tone being presented to the subject. However, the swept-tone response can be presented as a response to an impulse through some post-processing techniques. This derived impulse response closely resembles a click-evoked OAE in phase and time-frequency characteristics. Through this sort of analysis, direct comparisons between SFOAE and TEOAE responses can be made.

This swept-tone method can be used for the acquisition of transient OAEs (as shown in Figure 32), which will be referred to as swept-tone TEOAE, or sTEOAE, and OAEs evoked from a traditional click transient will be referred to as cTEOAE. Swept-tone analysis has several benefits, first in its ability to analyze weakly nonlinear systems, second in its ability to separate linear components from higher order distortion harmonic components, third in its improvement in signal-to-noise ratio (SNR) by reducing additive (coherent or random) noise, and fourth in its ability to remove synchronized spontaneous OAEs (SSOAEs) from a recording. In this Chapter, the theoretical treatment of the proposed swept-tone OAE method is first given. Then recordings from normal hearing ears using a high resolution real-time system are presented. The above advantages of the sTEOAE over cTEOAE are then demonstrated using these recordings.

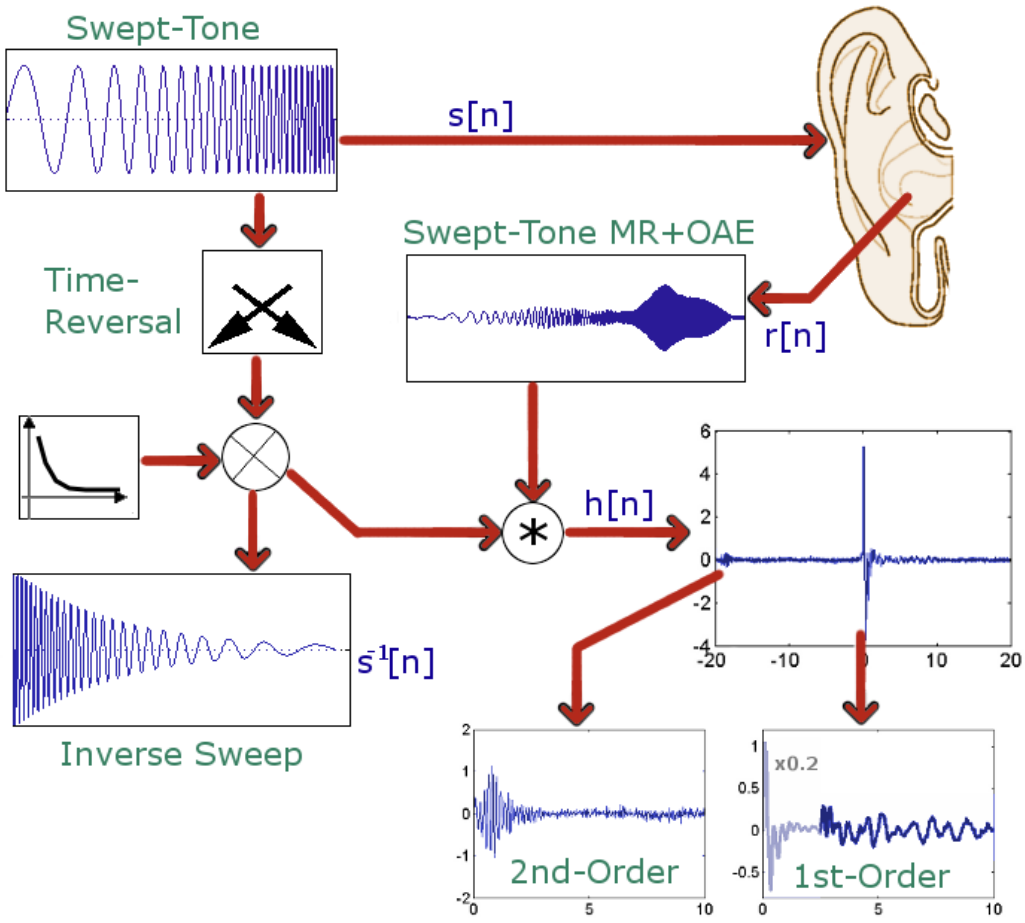

Figure 32. The stimulus, $s[n]$ is created according to Eq. (2.26). The inverse sweep, $s^{-1}[n]$, is generated by a time-reversal and a -10 dB per decade envelope. The stimulus is presented to the subject and the response, $r[n]$, contains the stimulus with the OAE response embedded. Convolution of $r[n]$ with $s^{-1}[n]$ results in the transient meatal response and TEOAE response. Windowing according to Eq. (2.32) can be used to separate high-order distortion responses from the 1st order response.

4.1 Theory and Methods

The objective of this experiment is to record the cochlear response to a long (100 ms) swept-tone stimulus. The response is comparable to an SFOAE recording since only 1

instantaneous frequency is presented to the ear. However, this response can be processed and presented as a response to in impulse. In this way, the swept-tone OAE, or sTEOAE, can be compared directly to a click TEOAE, or cTEOAE, with regard to magnitude, phase, time-frequency, and amplitude growth characteristics.

4.1.1 SUBJECTS

Data were acquired from both ears of 13 normally hearing adult subjects (26 ears; 6 male and 7 female; ages 18-48; median age 22) in accordance with an IRB-approved protocol. All subjects had normal audiograms with thresholds better than 20 dB hearing level (HL) at frequencies 0.5, 1.0, 2.0, and 4.0. For TEOAE acquisition, the subjects sat down in an acoustically attenuated and electro-magnetically shielded environment. An Etymōtic Research (Elk Grove Village, IL) ER-10D OAE probe fitted with a rubber tip was inserted into the meatus to form an acoustic seal. The custom acquisition board, described in Section 2.1 was used to gather recordings from the subjects. The subjects were tested at various low to moderate intensities for spontaneous activity. SSOAEs are defined as long lasting single frequency (typically in the 1 to 2 kHz range) components that are triggered by an external stimulus, but persist for up to 80 ms. SSOAEs, if present, will dominate a click TEOAE recording with a single strong sinusoid, and will last beyond the typical 20 ms acquisition window. In a time-frequency display, SSOAEs are characterized by a horizontal line and in the frequency domain by a frequency spike. Of the 26 ears tested, six showed SSOAEs at some intensity. Except where specified within the text, subjects with SSOAEs were considered separately from those without SSOAEs.

4.1.2 RECORDING PROTOCOL

For the standard click recordings, 75 μs duration rectangular electrical pulses were used. The swept-tone stimulus was used with parametric values of T=0.1 s, f_s=48 kHz, f_1=0.3 kHz, and f_2=9.6 kHz. A raised cosine window was applied to the first and last millisecond of the swept-tone signal to prevent undesirable clicking. Both the click and the swept-tone were calibrated according to the baseline-to-peak sound pressure level (pSPL), as described in (Burkard, 1984). It is inherently difficult to directly compare the perceptual loudness of a transient with a periodic sound, but the pSPL of the swept-tone response when time-compressed to an impulse response matches the pSPL of the click response. For this reason, other loudness measures such as peak-to-peak or peak-equivalent SPL determinations were not considered.

The click and the time-compressed swept-tone responses were windowed from 0 ms to 25 ms (1200 samples) post stimulus onset. A second window from 2.5 ms to 20 ms post stimulus onset was applied to each response in order to eliminate the MR artifact and residual noise. Finally, the responses were band-pass filtered (BPF) using a 4^{th} order digital Butterworth filter with cutoffs of 0.5 and 8.0 kHz. Forward and reverse filtering was used to maintain proper phase response.

Responses from both click and swept-tone stimulus sets were acquired using the derived nonlinear response (DNLR) residual method. Under the traditional click stimulus paradigm, the DNLR method utilizes a repeating train of four stimuli (Kemp *et al.*, 1986). The first three stimuli are of the nominal intensity and polarity. The fourth stimulus is three times higher intensity and of inverted polarity. Upon averaging the entire train of four responses, linear elements of the responses will be subtracted out, and only nonlinear

elements will remain. This is useful for TEOAE responses which saturate when elicited by high intensity stimuli. The same technique can be employed for the swept-tone stimulus. Although the swept-tone stimulus is bipolar, the DNLR method can be used simply by inverting the polarity of the fourth stimulus in the train and increasing its peak amplitude by a factor of three.

In this study, the experiments were conducted at fixed rates of 39.088 clicks per second for the click stimulus and 9.772 sweeps per second for the swept-tone stimulus. These rates were chosen to reduce the effects of electromagnetic interference on the signals. The experiments were designed to observe time-frequency and waveform characteristics of each stimulus set. The experiments were conducted at several durations in order to track the noise and SNR characteristics of the swept-tone and click stimuli. Durations of 32 s, 16 s, 8 s, and 4 s were used corresponding to a total of 1252, 626, 312, and 156 epochs respectively for the click stimulus, and 312, 156, 80, and 40 epochs for the swept-tone stimulus. The number of epochs was rounded to the nearest multiple of 4 in order to ensure that the DNLR train, which consists of four grouped stimuli, was not interrupted. Also, experiments were conducted at several intensities, from 76 dB pSPL down to 46 dB pSPL in 6 dB steps. The intensity study was performed to characterize the input/output growth function of the sTEOAE compared to cTEOAE.

4.2 RESULTS

Examples of swept-tone TEOAE (sTEOAE) and click TEOAE (cTEOAE) from six ears at 76 dB pSPL are plotted in Figure 33. These results are indicative of the type of phase and amplitude similarities between the two types of responses. A strong relationship

between the phase characteristics of the two responses was generally found in all subjects. The magnitude characteristics, however, tended to diverge at low-frequencies. As observed, most of the responses of the late-latency sTEOAEs were lower in amplitude than those of the cTEOAE responses.

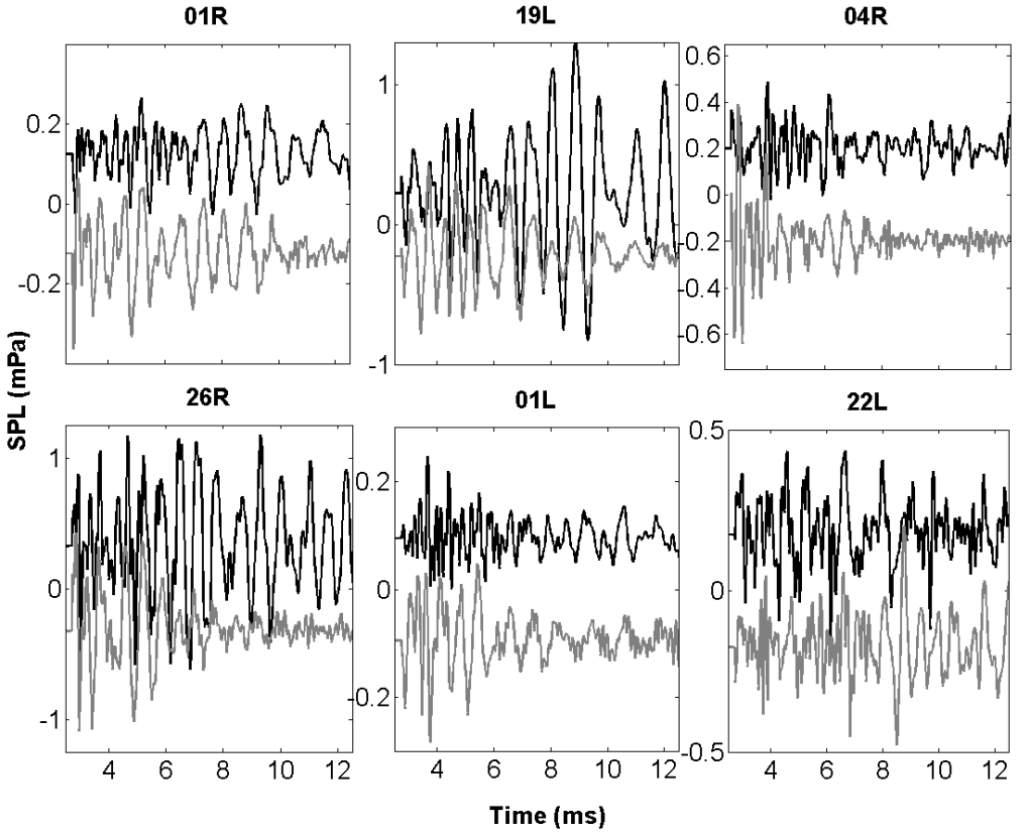

Figure 33. cTEOAE (black, top) and sTEOAE (gray, bottom) responses to 76 dB pSPL stimuli from 6 selected subjects (01R, 19L, 04R, 26R, 01L, 22L) are shown here from 2.5 to 12.5 ms, offset vertically for visualization. It was found that sTEOAEs were generally phase and time aligned with traditional cTEOAE responses across all frequencies. However, some differences arise in the magnitude of these frequency components, with early latency (high-frequency) components matching strongly in amplitude, and late latency (low-frequency) components decaying in amplitude to the point of cessation in most cases by 10-15 ms in the sTEOAE responses.

4.2.1 CORRELATION COEFFICIENT

If SFOAEs and TEOAEs both arise from the same fundamental generation mechanisms, then it is believed that when the SFOAE is transformed into a TEOAE through phase compression that the two responses should be generally aligned temporally. This alignment can be measured simply with the use of a cross-correlation coefficient in order to measure the reproducibility of an acquired response. Correlation coefficients can be characterized in qualitative terms as trivial, small, moderate, and large corresponding to coefficients of $\rho=0.0$, $\rho=0.1$, $\rho=0.3$, and $\rho=0.5$ respectively (Cohen, 1988).

The correlation coefficient for the cTEOAE and the sTEOAE was computed for the 32 s duration, 76 dB pSPL experiment for every ear between 2.5 and 12.5 ms post-stimulus onset. A leading or lagging shift of up to 0.25 ms (12 samples) was allowed in either direction in order to maximize the correlation coefficient, but the mean and median shift across all ears was less than 0.05 ms (2 samples). The correlation between the sTEOAE and cTEOAE responses was 0.38 ± 0.17. The subject correlation coefficient data can be found in Table 1 in the Appendix. Out of the 26 ears tested, those with a p-value of $p<0.05$ are considered significant. The percentage of ears with a significant correlation between sTEOAE and cTEOAE responses was 81%. That is, 21 out of 26 ears showed a high correlation between sTEOAE and cTEOAE responses.

4.2.2 SIGNAL TO NOISE RATIO

As described in Section 2.2.4, it was expected that the swept-tone response will have a reduced noise floor. This was experimentally confirmed in 95 out of 104 trials (91.3%), which led to an improved SNR. Noise floor levels for each ear tested is presented in

Table 2 in the Appendix. The noise floor at different recording durations at the fixed stimulus intensity of 76 dB pSPL was calculated for each ear and averaged and is shown in Figure 34 (upper panel). Not only was the noise floor for the sTEOAE recordings significantly lower than in the cTEOAE recordings, but they also had a smaller standard deviation, indicating a more consistent noise floor. As expected for both stimuli, the noise floor decreased as a function of the square root of the number of epochs, but the noise floor for the sTEOAE was consistently below that of the cTEOAE at all durations. It is worth noting that the difference between the average noise floor for a 4 s sTEOAE recording (74.4 ± 39.1 μPa) and the average noise floor for a 32 s TEOAE recording (63.5 ± 46.6 μPa) was statistically insignificant ($p=0.36$).

The average signal RMS value was also computed for the click and swept-tone responses (see Figure 34 middle panel). Unlike the noise floor, the signal power is not theoretically subject to be altered with an increasing number of epochs. It can be seen that the RMS value of the sTEOAE responses were generally consistent at multiple recording durations, but the cTEOAE responses had a significantly higher RMS values at shorter durations. This is because the cTEOAE responses were severely corrupted with noise at these short durations, as evidenced by the large standard deviations. Only at 32 s did the cTEOAE response average RMS value become aligned with that of the sTEOAE, where the difference in average RMS values was statistically insignificant ($p=0.19$). It is for this reason that SNR calculations for the cTEOAE could only be reliably calculated at the longest duration.

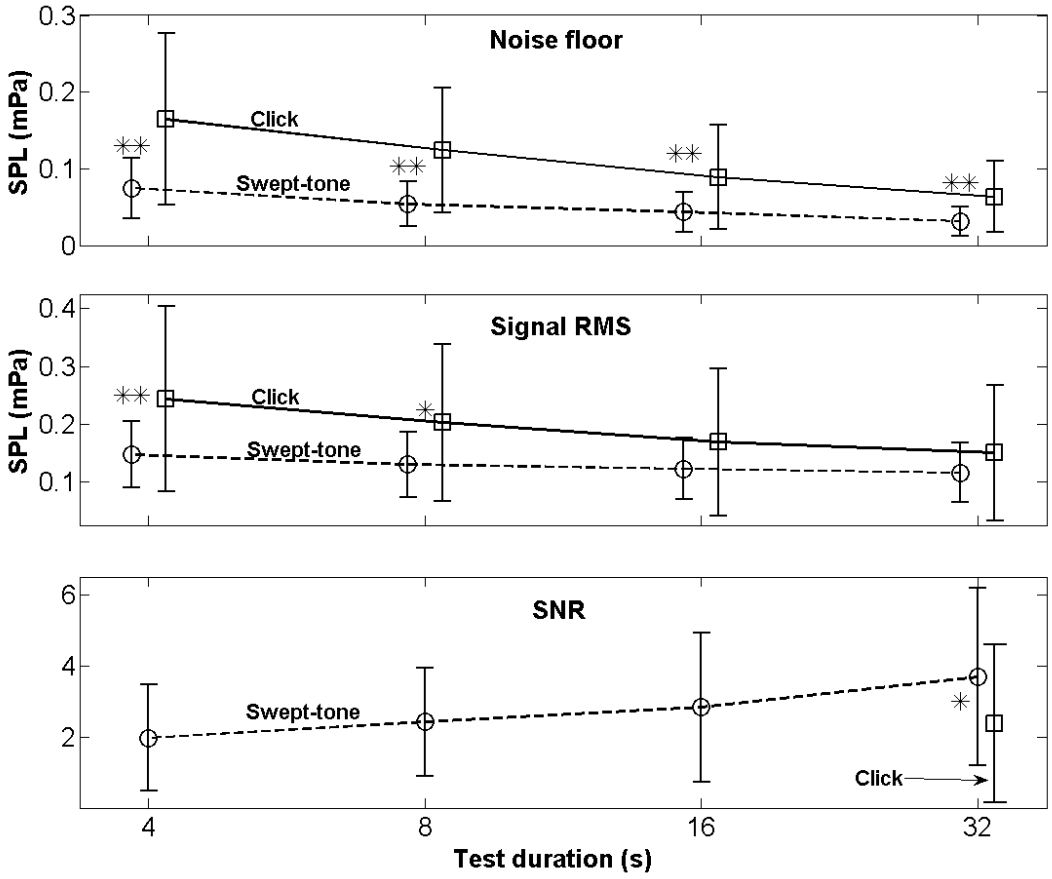

Figure 34. Noise floor calculations were made using the even/odd buffer paradigm for both click and swept-tone stimuli (upper panel). The noise floor for the swept-tone recordings was significantly ($p<0.01$) lower at all recording durations. Furthermore, the standard deviation (indicated by vertical bars) of the noise floor levels was much lower in the swept-tone recordings than in the click recordings. Also, the average signal RMS value was calculated (middle panel). The average signal power should remain constant for all durations, as seen in the swept-tone recordings. The click signal RMS value was very high at short durations because the signal was corrupted with noise. The reduced noise floor translated into improved SNR (bottom panel) which could only be reliably calculated for the click at 32 s. It was found that the swept-tone recordings had a significant (* $p<0.05$) improvement of about 6 dB SNR.**

The SNR, as shown in Figure 34 (lower panel) gives an indication of the relative RMS level of the signal compared to the noise floor. It can be seen that the improvement in SNR for the sTEOAE recordings was 6.1 dB at the 32 s test duration, which is comparable to the theoretical SNR improvement of 7.1 dB, as described in Section 2.2.4. The improvement in SNR using the sTEOAE for a test duration of 32 s was significant ($p<0.05$). This has beneficial implications in clinical recordings. Since an improvement of SNR is generally obtained by increasing the number of epochs, and if SNR is used as the stopping criterion for an OAE recording, then the overall duration of the OAE acquisition can be reduced using a swept-tone stimulus. Furthermore, since the SNR improves as a square root of the number of epochs, then the factor of 6.1 dB equates to a reduction in recording duration by better than a factor of four. If a swept-tone stimulus was used to obtain sTEOAEs rather than a click to obtain cTEOAEs, then the recording protocol could be shortened while maintaining constant SNR.

4.2.3 TIME-FREQUENCY PROPERTIES

The responses (both cTEOAE and sTEOAE) were analyzed at all intensities in time-frequency displays (TFD) using Morlet wavelet analysis, as described in Section 2.3. Only 32 s recordings were used in order to make the comparisons with the low noise floors. Three example recordings taken with 76 dB pSPL stimuli are shown in Figure 35 for both sTEOAE and cTEOAE responses. The three columns delineate temporal, spectral, and time-frequency representations. Temporally, the cTEOAE and sTEOAE plots were phase-aligned; however the cTEOAE responses had much higher amplitudes in the late-latency, low-frequency components. Conversely, the sTEOAE responses tended to lose their power after 10 to 15 ms. Spectrally, the noise floor of the sTEOAE

responses was considerably lower compared to the noise floor of the cTEOAE responses, and this provided a higher SNR for the sTEOAE responses. Considering the TFDs, the high-frequencies (2-4 kHz) were more prominent for the sTEOAE responses.

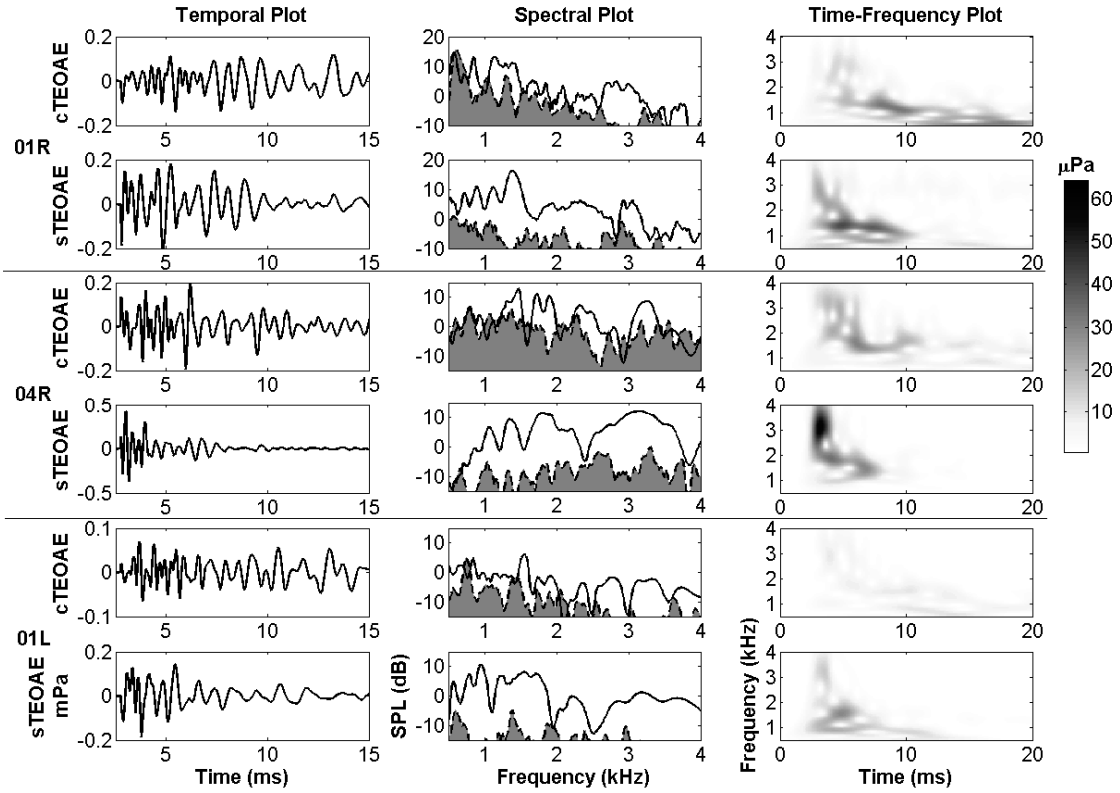

Figure 35. cTEOAE and sTEOAE recordings from 3 ears (01R, 04R, 01L) are shown using traditional click stimuli (upper panel) and swept-tone stimuli (lower panel) obtained at 76 dB pSPL. This is visualized in 3 modalities, time domain plot (left column), magnitude spectrum (middle column), and time-frequency displays (right column). The noise floor in the spectrum plots is shown by filled in gray. The signal is represented by the continuous dark lines. The gray level calibration of the TFDs are shown on the right side.

Population averaging was possible using time-frequency representation since each time-frequency block following the wavelet transform was represented by a scalar value,

regardless of phase. Since phase was not a consideration, then averaging of the responses was possible in the time-frequency domain, as compared with the temporal domain in which averaging was not possible due to large variations of phase and group delay among subjects. The population averaged time-frequency plots of cTEOAEs and sTEOAEs are shown in Figure 36.

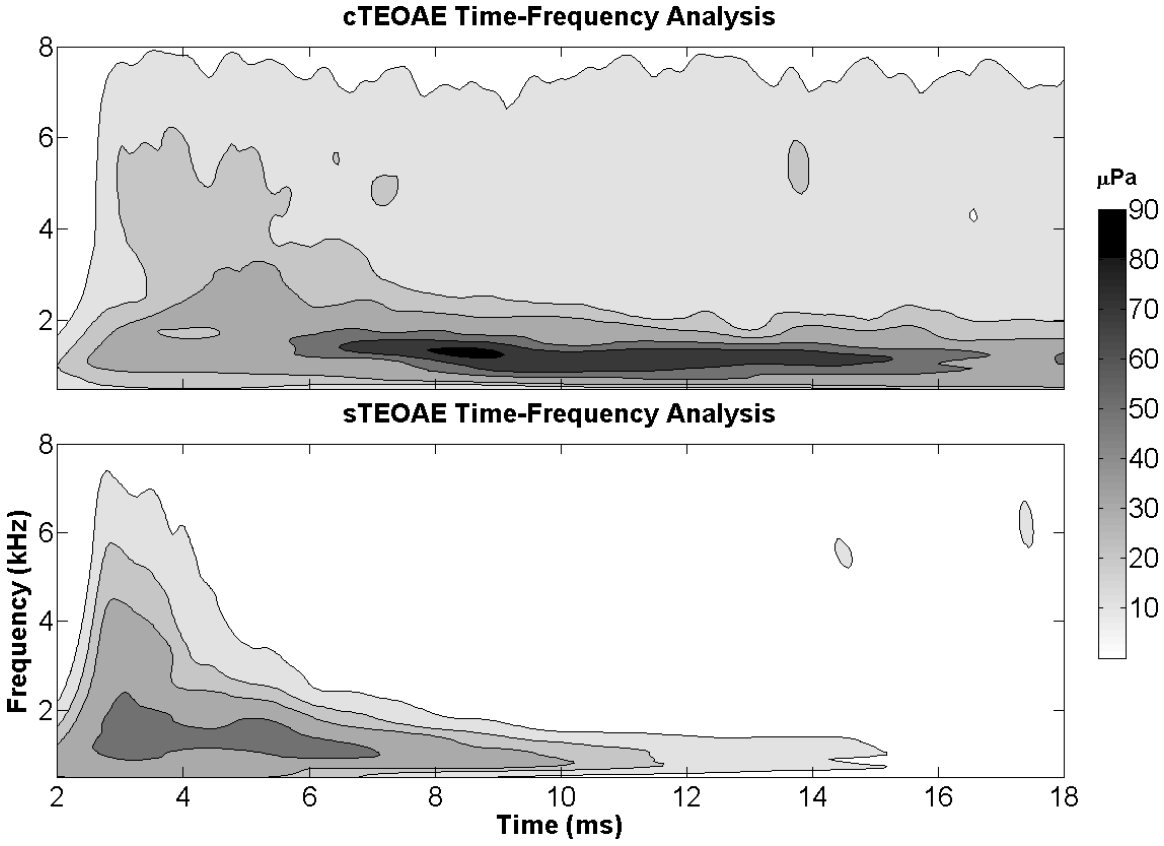

Figure 36. Time-frequency analysis was performed for the swept-tone and click recordings at an intensity of 76 dB pSPL (test duration of 32 sec) from ears with no spontaneous activity. The responses from each ear were averaged together for both stimuli to generate a population averaged response. The click recordings are dominated by late-latency, low-frequency components, which is not due to SSOAEs, since the 2 ears exhibiting SOAEs were excluded from this analysis. The response from the swept-tone recordings shows improved definition in the early latencies.

The first notable difference between the two TFDs is the noticeable presence of background noise in the cTEOAE. The presence of this background noise is discussed in Section 2.2.4. A second notable difference was the presence of high-energy late-latency components in the cTEOAE population average. Such long lasting low-frequency components were not present in the sTEOAE population average. As observed in sample plots in Figure 35, there was generally no OAE response beyond 15 ms in sTEOAEs.

4.2.4 Removal of Spontaneous OAE

An additional benefit to swept-tone analysis was the removal of coherent, additive noise. Non-coherent additive noise would include any sort of random processes, such as electrical or ambient noise, and this type of noise is reduced typically by increasing the number of epochs in synchronous averaging. Coherent additive noise could include 60 Hz electromagnetic interference (depending on the rate) or, in the case of TEOAE acquisition, synchronized spontaneous OAEs (SSOAEs). Typically, coherent additive signals are not cancelled by synchronous averaging because they may become phase-locked with the onset of the stimulus. However, with swept-tone analysis, coherent noise was also reduced through two processes; first a frequency-dependent magnitude reduction, and also through phase smearing that is a result of the time-compression process. Convolutional processes (both linear and weakly nonlinear) are reconstructed through the time compression provided by the inverse sweep, but additive processes are disrupted temporally.

As a result, the obstruction due to SSOAEs was significantly reduced with the use of swept-tone analysis. This incidence of SSOAEs varied with intensity level, but considering all intensities together, the combined incidence of SSOAEs was 23%. In all cases where SSOAEs were observed using a click stimulus, the use of the swept-tone stimulus and subsequent phase reconstruction removed the spontaneous activity to reveal sTEOAEs with characteristic time-frequency dispersions. Examples from two selected subjects with SSOAEs are shown in Figure 37. These recordings were elicited by 76 dB pSPL click and swept-tone stimuli. In each of these three individual cases, strong SSOAEs are observed with the click stimulus. These appear as sinusoids in the time-domain, a peak in the spectral-domain, and a constant zero-slope line in the time-frequency domain. Contrasting these results are the sTEOAEs, which display characteristic high- to low-frequency dispersion in the time and time-frequency domains, and show a more broadband response in the spectral domain. Even though the presence of strong SSOAEs is considered as passing criterion in hearing screening, useful diagnostic information is obscured when this spontaneous activity is present. These spontaneous emissions were removed through the use of swept-tone TEOAEs.

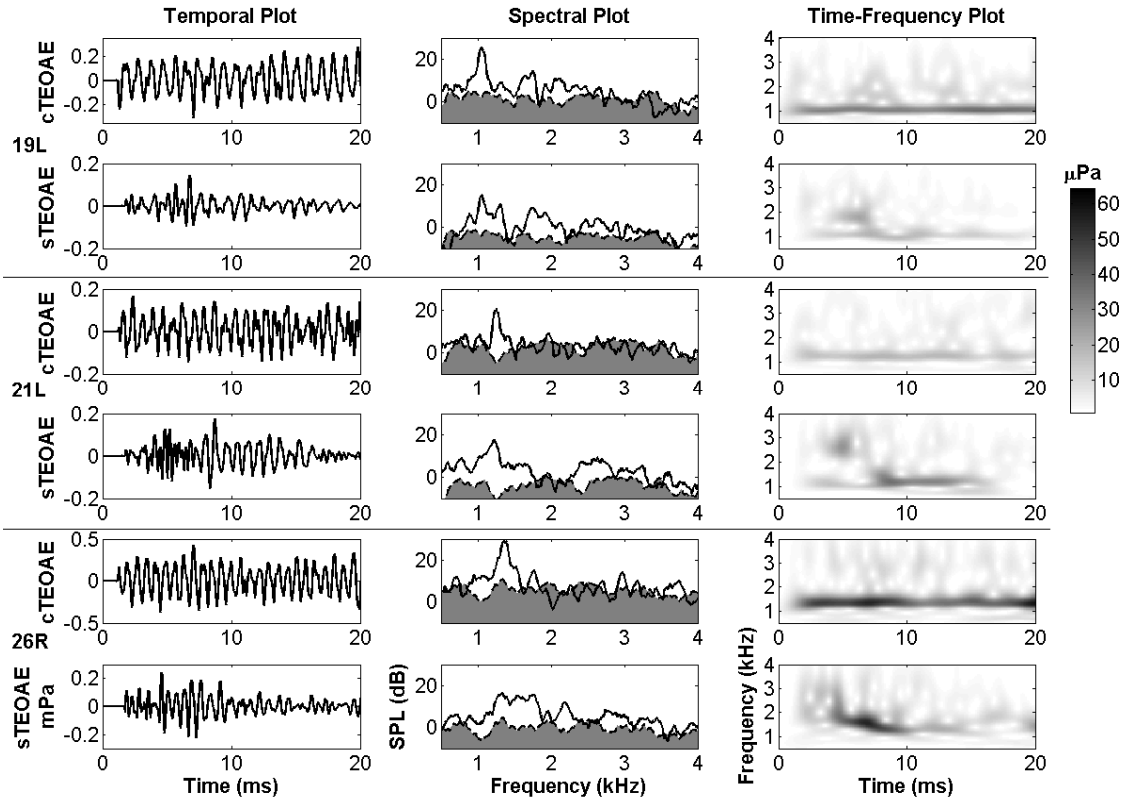

Figure 37. sTEOAE and cTEOAE recordings from 3 ears are shown using traditional click stimuli (upper panel) and swept-tone stimuli (lower panel) obtained at 76 dB pSPL. The SSOAEs are characterized by one long-lasting, low-frequency component that masks other frequency components. This is visualized in 3 modalities, time domain plot (left column), magnitude spectrum (middle column), and time-frequency plot (right column). sTEOAE responses do not exhibit any strong SSOAEs, and have typical time-frequency dispersion characteristics that would be expected in the absence of SSOAEs.

4.2.5 COMPUTATION OF COCHLEAR DELAY

In order to get an understanding of the time-frequency latencies of sTEOAEs, cochlear delays were computed using the TFDs. This was done by finding the latency of the point

of maximal energy for each of the frequency components, which were separated by 50 Hz steps. Careful attention was given to the early-latency, low-frequency components, which are thought to be related to the stimulus artifact. These time-frequency components were excluded from the analysis.

The time-frequency peak latencies found in frequency slices of the TFDs were then fitted to a power model in the form $\tau(f) = a \cdot f^{-b}$. This equation is used by many researchers (Tognola *et al.*, 1997; Jedrzejczak *et al.*, 2004; Notaro *et al.*, 2007; Sisto and Moleti, 2002; Harte *et al.*, 2009) to quantify cochlear delay by different physiological measures. This model was also used in the present study. For the sTEOAE responses constructed in this study, values of a=10.4 and b=0.4 were found. For the cTEOAE responses, values of a=12.0 and b=0.51 were computed.

The fitted model parameter values were compared to models from previous literature (see Figure 38). The b value for both stimulus types was consistent with several other reported values which had a range of 0.37 to 0.57. The a value for the swept-tone responses was also consistent with previous findings, which were in the range of 8.6 to 11.1; however, the a value for the click responses was slightly above this range.

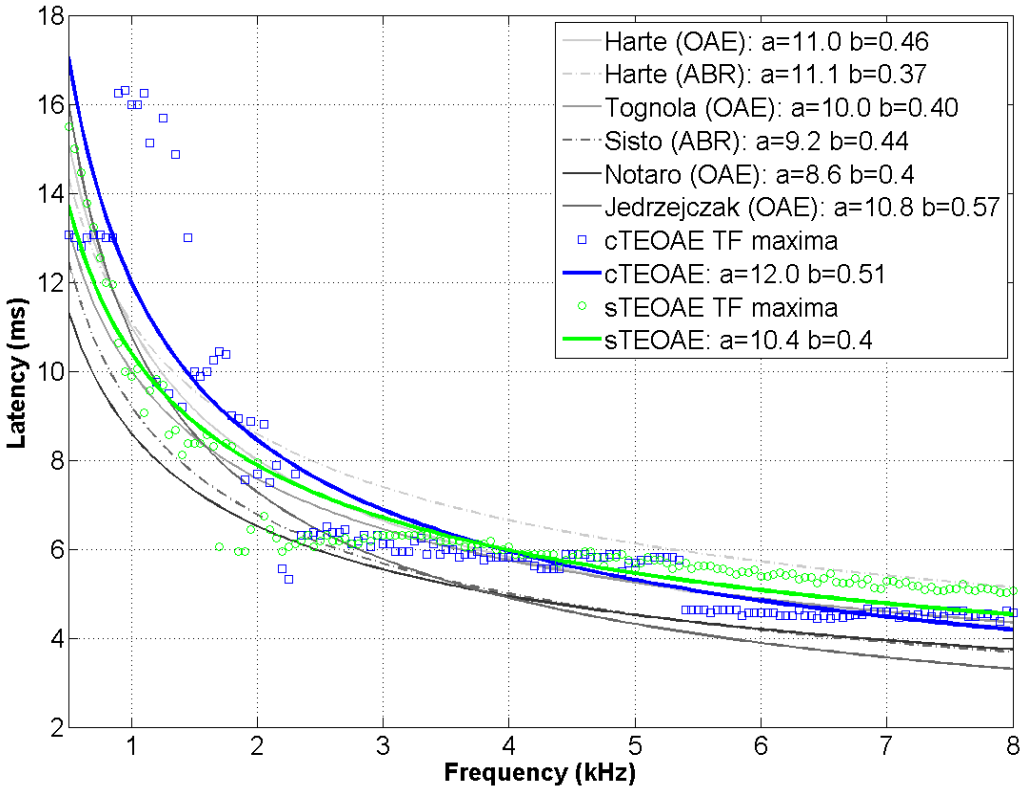

Figure 38. Time-frequency latencies were determined by finding the latency corresponding to the response-related peak amplitude of each frequency component in the wavelet transform. These time-frequency latencies were then fitted to a power law in the form of $\tau(f) = a \cdot f^{-b}$. The sTEOAE (circles) and cTEOAE (squares) were fitted with the dashed and solid thick lines, respectively. For comparison to previous literature, fitted models from other researchers are shown as indicated in the legend.

4.2.6 Input/Output Growth Function

The input/output growth function gives an indication of the amount of nonlinearity of a particular response. The axes are normally of the same unit (in this case, intensity in dB SPL) but differ because the horizontal axis gives the input value and the vertical axis

gives the output value. If a response is perfectly linear, then the input/output growth function will be a straight line. Several nonlinearities may exist (eg., expansive nonlinearities or hysteresis), but in the case of compressive nonlinearity, the input/output growth function will take more of a classic sigmoidal shape. TEOAEs evoked with a click stimulus will have a compressive nonlinearity, typically with a breakpoint from linearity to saturation somewhere around 60 dB SPL input intensity. The slope of the saturated region is not popularly reported, but is generally around 0.2 to 0.3 dB/dB.

Since the amplitude envelope of the sTEOAE responses was not consistent across the entire response duration, the responses were divided into time chunks to separately compare the input/output function at different latencies. The sTEOAE and cTEOAE filtered responses were framed into overlapping 5 ms windows, beginning at a center location of 2.5 ms and ending with a center location of 15.0 ms. The frames have 50% overlap, giving a center location spacing of 2.5 ms between consecutive frames. A Hann window (essentially a sine-squared envelope) was applied to each frame. The RMS value of each windowed frame was calculated, and averaged across all subjects and at varying intensities from 46 to 76 dB pSPL in 6 dB steps. The results are shown in Figure 39.

Subject data relating to intensity-dependent RMS values are reported in Table 3 in the Appendix. Table 3 reports RMS values for the entire response, which differs from Figure 39, which reports RMS values for windowed segments. Even considering the entire response together, the click responses show compressive growth (which is expected) and the swept-tone responses show less limited, although still compressive, growth.

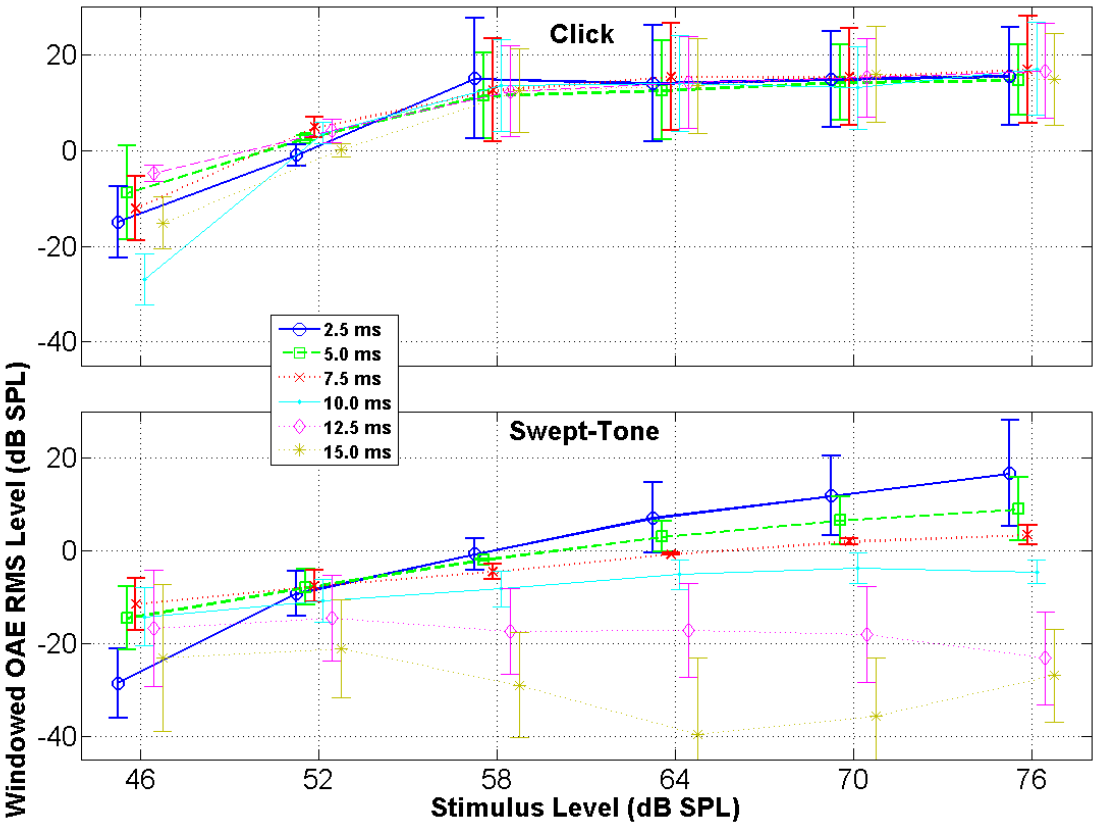

Figure 39. Each click and swept-tone TEOAE response was framed, and the RMS value was calculated. The average (±1 standard deviation) RMS level for all 26 ears is shown. The cTEOAE responses saturate more compressively than the sTEOAE responses. Also, the reduced late-latency (12.5+ ms) low amplitudes are manifest, as those RMS levels do not rise above -15 dB pSPL at any stimulus intensity.

In the click responses, a break from linearity to saturation can be observed around 55 dB pSPL, which is consistent with previous findings (Grandori and Ravazzani, 1993) and with cochlear models (Sisto and Moleti, 2008). The sharp knee from linearity to saturation is not readily apparent in the swept-tone responses, which show a more gradual transition into saturation.

The level of saturation can be given as a dimensionless in terms of dB/dB. If the output response grows at a rate of 1 dB/dB, then it is linear, and any growth rate between 0 and 1 is compressively nonlinear. The click growth rate in the saturation region is about 0.2 dB/dB, which is generally in line with literature values (Kemp, 1978; Ryan and Kemp, 1996). However, the swept-tone responses show a more gradual transition into saturation with an input-output growth rate of 0.3 dB/dB. Less compressive growth rates are generally found among SFOAE and DPOAE responses, with typical values being in the range of 0.2 to 0.6 dB/dB (Schairer *et al.*, 2003; Withnell and Yates, 1998).

As previously reported, the late-latency sTEOAE responses are relatively lower in amplitude as compared to the cTEOAE responses (on the order of 10-100 μPa compared to 100-1000 μPa). This phenomenon is also manifest in Figure 39, which shows that the late-latency windows have RMS values that are 20 dB or more lower than the early latency windows.

4.2.7 Alternative Computation of Time-Frequency Latencies

Time-frequency latency computations give important information about the OAE responses. The latency associated with a particular frequency is related to the round-trip travel time, including forward transmission of the stimulus to tonotopic place, and backwards transmission of the OAE to the ear canal, where it is recorded by the microphone. Since the cochlea is frequency-dispersed, the tonotopic place for each frequency component varies. Assuming the travel velocities and other conditions are constant, the location of the tonotopic place is the only variable which changes the

duration of the round-trip travel time. Therefore, time-frequency latencies give an indication of the frequency mapping of the cochlea.

The most common method for deriving time-frequency latencies is to somehow divide the OAE response into frequency bands, and then determine the time delay associated with the region of maximal energy. For example, if the response were filtered by a bank of bandpass filters, and then each filtered response was enveloped, it would be expected that the high-frequency bandpassed response would have an envelope peak in the early latencies and that the low-frequency bandpasssed response would have an envelope peak in the late latencies. Typically a more sophisticated method than a simple bandpass filter would be employed, and Section 1.2.3 covers some of these methods, but the notion remains the same. In fact, time-frequency latency estimates were calculated in this manner and were reported in Section 4.2.5.

The unifying concept behind these time-frequency latency computation methods is to analyze the amplitude (or magnitude) of the response, and look for the peak of the response. These methods do not readily consider phase, which can give quite valuable time-frequency information. It should be noted that in fact the derivative of the phase gives the group delay, which is an indication of time-frequency latencies.

For this analysis, the group delay was considered in tandem with the magnitude response in order to calculate the time-frequency latencies of the OAE response. The procedure was straight forward. First the complex discrete Fourier transform was computed for the OAE response. Next, the phase angle at each frequency bin was calculated, and the

derivative of the phase response was computed. This gives an approximate, albeit noisy, estimate of time-frequency latencies.

The noise was controlled by utilizing the magnitude response. If the magnitude of each phasor component was not above the threshold of -10 dB below the RMS value of the magnitude response, then it was excluded from the analysis. An example of this type of segmentation is shown in Figure 40. Of the included data points, an average was computed for each frequency bin, and the results are plotted in Figure 41.

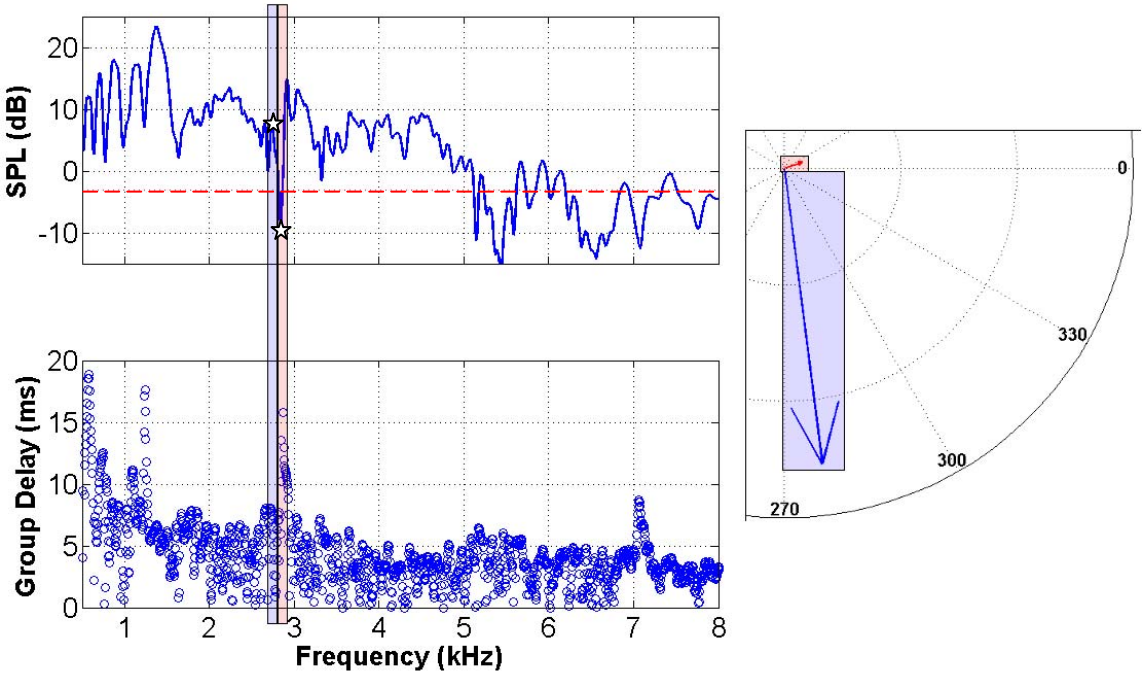

Figure 40. An example from SUB01R is shown in this figure displaying magnitude (left top) and group delay (left bottom). Here, the RMS value of the magnitude plot is 6.9 dB SPL, and so the threshold is -3.1 dB SPL. Two adjacent regions are considered, 2735 and 2835 Hz, with their phasors shown (right). The group delay elements of the 2735 Hz bin are kept since they are above the noise threshold and the elements of the 2835 Hz bin are discarded.

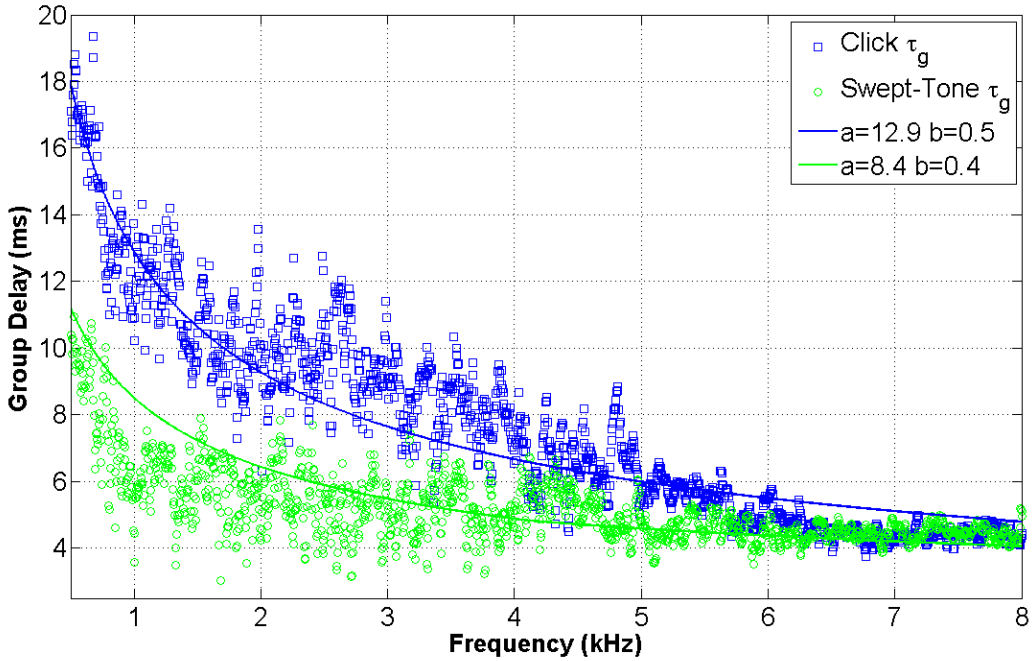

Figure 41. Group delay averages were computed from the phase responses of individual ears. Noisy components were excluded from the average by only considering phasors with significant magnitudes. The data were fitted to a model of the form $\tau(f) = a \cdot f^{-b}$. a and b values are comparable to literature values for the cTEOAE responses. Time-frequency latency information gives indications about the topology of the cochlea with respect to cochlear travel times.

Overlaid on the data is a curve-fit model representation of the form $\tau(f) = a \cdot f^{-b}$. Refer to Section 4.2.5 for details and references regarding this particular model. Typical literature values from both OAE responses and auditory brain responses (ABR) indicate values of $a \approx 10$ and $b \approx 0.5$. The best fit parameters for the cTEAOE responses were $\{a, b\} = \{12.9, 0.5\}$, and $\{a, b\} = \{8.4, 0.4\}$ for the sTEOAE responses. These parameter

values are comparable to the values obtained using the wavelet method in Section 4.2.5 (cTEOAE: $\{a, b\} = \{12.0, 0.5\}$; sTEOAE $\{a, b\} = \{10.4, 0.4\}$).

4.2.8 NONLINEAR DISTORTION COMPONENT ANALYSIS

The nonlinear components of the OAE responses can be recovered through a two-step process. First, the harmonic responses should be isolated using Eq. (2.32), which describes the locations of each harmonic component. Through simple windowing, each component is extracted for analysis. Next those measured harmonic components must be transformed into nonlinear distortion components, utilizing Eq. (2.51), which the linear transformation process to convert from measured harmonic impulse response to distortion component.

The nonlinear components are not analyzed in a rigorous context, as it is outside the scope of this manuscript. However, some individual observations have been made regarding the presence and morphological features of the nonlinear components. Just like the first order responses, the nonlinear components appear to consist of some stimulus contamination as well as physiological response, at least at high stimulus intensity levels. An example of possible stimulus-related probe distortion is shown in Figure 42.

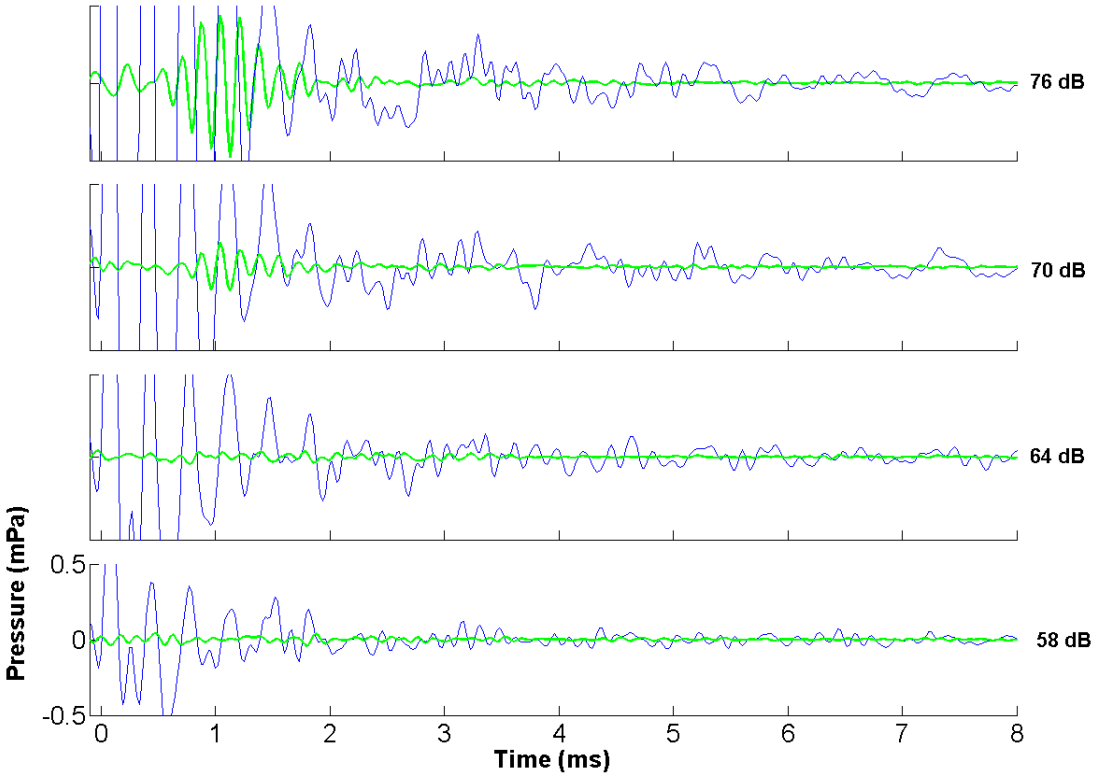

Figure 42. The first-order (thin blue) and second-order (bold green) responses are shown overlaid in decreasing stimulus intensities from SUB03L. The second order response grows expansively at higher intensities, and seems to be a stimulus-related artifact.

However, other example examples exist indicating that some of the nonlinearities may in fact be physiologically related. Due to the stimulus parameters (start and end frequencies and duration), the higher order impulse responses are shorter than would be desirable. Some of the portions of the higher order responses exhibit characteristic high to low frequency dispersion at appropriate latencies, indicating a cochlear origination. An example is shown in Figure 43, which contains very little stimulus-related probe distortion since the stimulus intensity is moderately low (58 dB pSPL or 16 mPa).

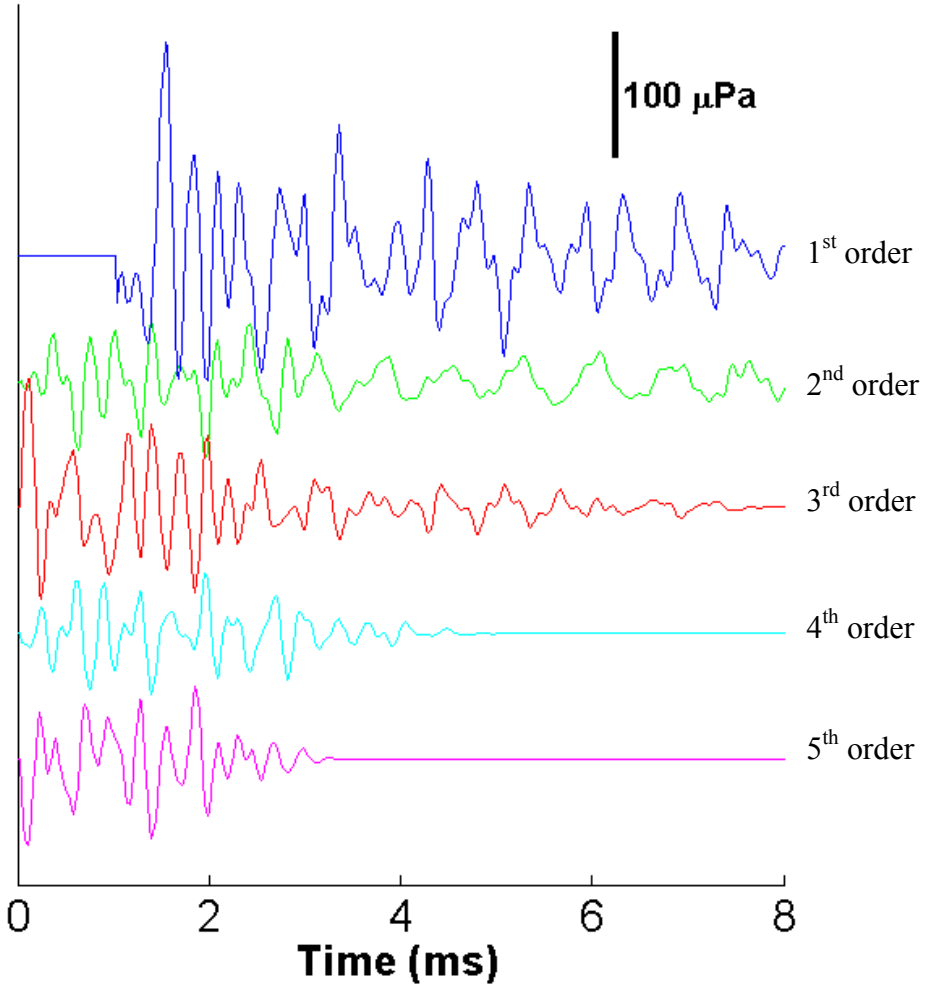

Figure 43. The first four nonlinear components (2nd: green, 3rd: red, 4th: cyan, 5th: magenta) are plotted in ascending order, along with the 1st-order response (on top in blue) from SUB23R at an intensity of 58 dB pSPL. Similar morphological features such as the characteristic frequency dispersion can be seen in the nonlinear components. Some elements may be stimulus related and some may be physiologically related.

Finally, some further evidence, with respect to the input-output growth function of the nonlinear components, indicates that some of the response is indeed physiological. OAEs grow compressively with respect to stimulus input intensity. While the growth of the

apparent stimulus related distortions grow expansively with the stimulus, (as the stimulus increases by a factor of 2 for every 6 dB increase), there are portions of the nonlinear response that grow compressively, which would indicate that those portions are physiologically related, as opposed to stimulus-related.

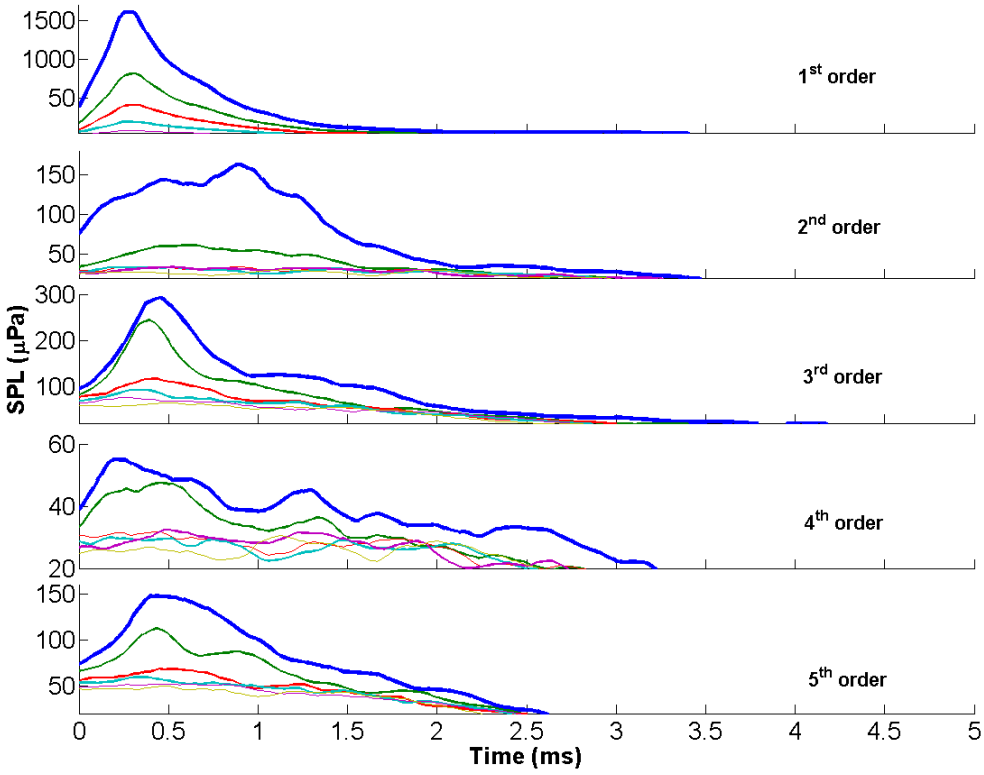

Figure 44. A population average of enveloped distortion responses is shown (76 dB: blue, 70 dB: green, 64 dB: red, 58 dB: cyan, 52 dB: magenta, 46 dB: yellow). The 1st and 2nd order responses grow at the same rate as the stimulus intensity, indicating that they are related to the stimulus artifact response. However, the higher ordered responses grow compressively, possibly indicating that they are related to the physiological OAE response.

Figure 44 shows the average envelopes for the 1st order response (mostly comprised of stimulus artifact) and the following 4 higher order nonlinear responses at multiple stimulus intensity levels. The 1st order response grows expansively (as expected), and so does the 2nd order response, indicating that those distortions may be stimulus-related. However, the higher order responses grow compressively, which is a characteristic of OAE response, indicating that those portions of those responses may be physiologically related. The growth of the higher order responses can be compared to Figure 45.

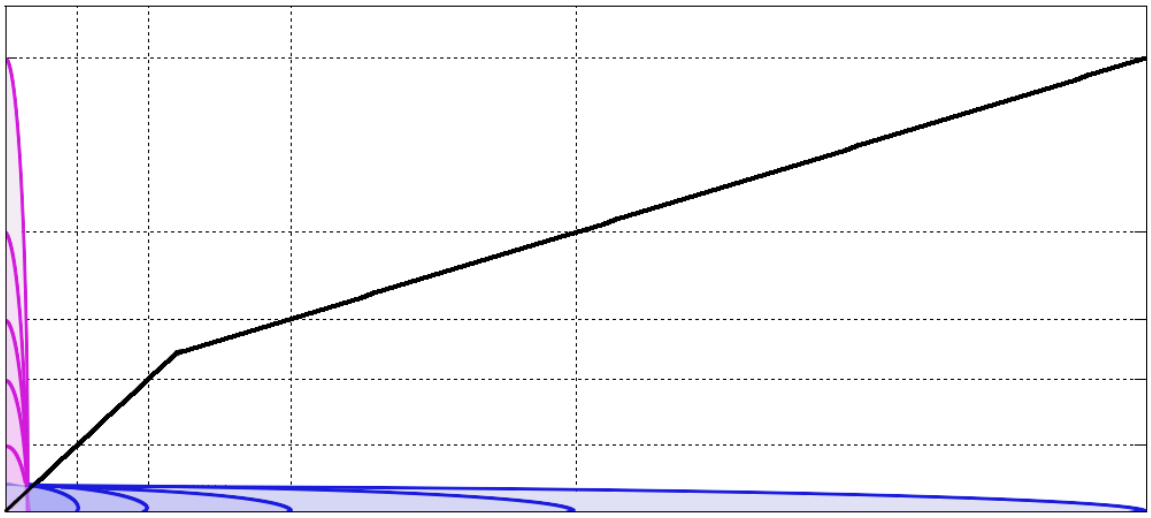

Figure 45. This shows a contrived example of a nonlinear growth function. The input (along the x-axis) doubles in intensity, but the output (along the y-axis) begins to saturate at a nonlinearity of 0.3 dB/dB (which is characteristic of OAE growth). The compressive growth can be compared to Figure 44, which also displays compressive growth in the higher ordered responses.

Chapter 5 Discussion

This dissertation examines three primary improvements in the acquisition of TEOAEs: development of high-resolution instrumentation, improvement of existing click methods, and the formulation of a novel method to acquire TEOAEs using swept-tone stimuli. In Chapter 2, the rationale for a high bit-depth system was examined, primarily finding that in order to adequately characterize both the meatal response (or stimulus artifact) and also the corresponding OAE response, 24 bits (or over 16 million levels of resolution) are required. Furthermore, in Chapter 2 the theoretical formulation of a swept-tone stimulus, analytic filter, and the post-processing method of acquiring an impulse response and nonlinear impulse responses was developed. Chapter 3 discussed the application of the swept-tone method to characterize the transfer properties of the auditory canal. A magnitude and phase equalization method was developed in order to produce an acoustic click within the auditory canal. Chapter 4 presented a novel modality for the acquisition of TEOAEs using the swept-tone as a stimulus. Responses were compressed and analyzed in terms of SNR, time-frequency latencies, and observations about the nonlinear components were made.

5.1 Improvements in Instrumentation

Instrumentation can play an important role in the meaningfulness and veracity of acquired bio-signals. In Section 2.1, the instrumentation design parameters of an OAE acquisition device are considered. Although not reported in this dissertation, improvements in TEOAE response are possible through improvements in instrumentation alone (Bennett and Özdamar, 2009). In the cited proceedings article, it is shown that the

DNLR method can be improved simply by using a system that operates with 24 bits as opposed to 16 bits. This can intuitively be considered by recalling that the DNLR mode retains nonlinear processes and eliminates linear processes. Under this paradigm, the stimulus is considered to be linear and the response nonlinear. However, if the stimulus is saturating the ADC (due to a deficit of bits), then that portion of the stimulus will be nonlinear, and consequently will be retained with the response. As a result, the ear canal artifact lasts longer with a 16-bit system than in a 24-bit acquisition system. So even by simply improving instrumentation alone, higher frequency TEOAEs can be recovered.

5.2 IMPROVEMENTS IN EXISTING CLICK METHODS

Under the click modality, additional high-frequency information can be recovered with the use of a compensated click. Some discussion is included in Section 3.3 regarding the usage of a compensated click in the acquisition of TEOAEs. The hypothesis is, although intuitive, two-fold: (1) if the in-ear acoustic click can be improved (by reducing ringing), then the ear canal artifact can be reduced, and (2) if the ear canal artifact is reduced, then higher-frequency TEOAEs can be recovered. Both hypotheses are confirmed and reported by the Bennett and Özdamar (2010a). The results confirm other assumptions and findings from literature that the cochlear response to a click is frequency-dispersive from high to low frequency and follows an exponential shape (Goodman *et al.*, 2009; Hatzopoulos *et al.*, 2000; Wit *et al.*, 1994; Tognola *et al.*, 1997; Zhang *et al.*, 2008; Cheng, 1995; Blinowska *et al.*, 1997; Jedrzejczak *et al.*, 2004; Dau *et al.*, 2000; Heidelberg and Frankfurt, 2007). The experimental findings in the very early latency regions concur with expectations that very high frequency emissions should be observable.

5.3 Development of a Novel Acquisition Method

The swept-tone stimulus was used in Chapter 3 to characterize the transfer properties of the auditory ear canal for the purposes of developing a compensatory filter that equalizes magnitude and phase distortions of the probe and ear canal system. The swept-tone was further adapted in Chapter 4 to develop a novel OAE modality, as reported in (Bennett and Özdamar, 2010b). The swept-tone stimulus contains only one frequency component at any instantaneous moment, and was swept from low to high frequency over a duration of 100 ms. Once the swept-tone response was obtained, it was "compressed" using a post-processing convolution method, resulting in a derived impulse response. The physiological portion of the response (as opposed to the stimulus related portion) exhibited characteristic high to low frequency-dispersion. Furthermore, when compared to a click response from the same subject, the swept-tone and click responses align temporally, leaving little doubt as to their similitude, at least in the early latencies (see Figure 46). Furthermore, when tested in a 2 cc cavity or in an ear which lacks OAE (e.g., SUB06R due to presbycusis), an absence of response with a click stimulus is verified by an absence of response with the swept-tone stimulus, as well. These examples provide evidence that the swept-tone responses are physiologically related, and are not due to processing artifacts, which is a legitimate question since this is a novel modality of acquisition.

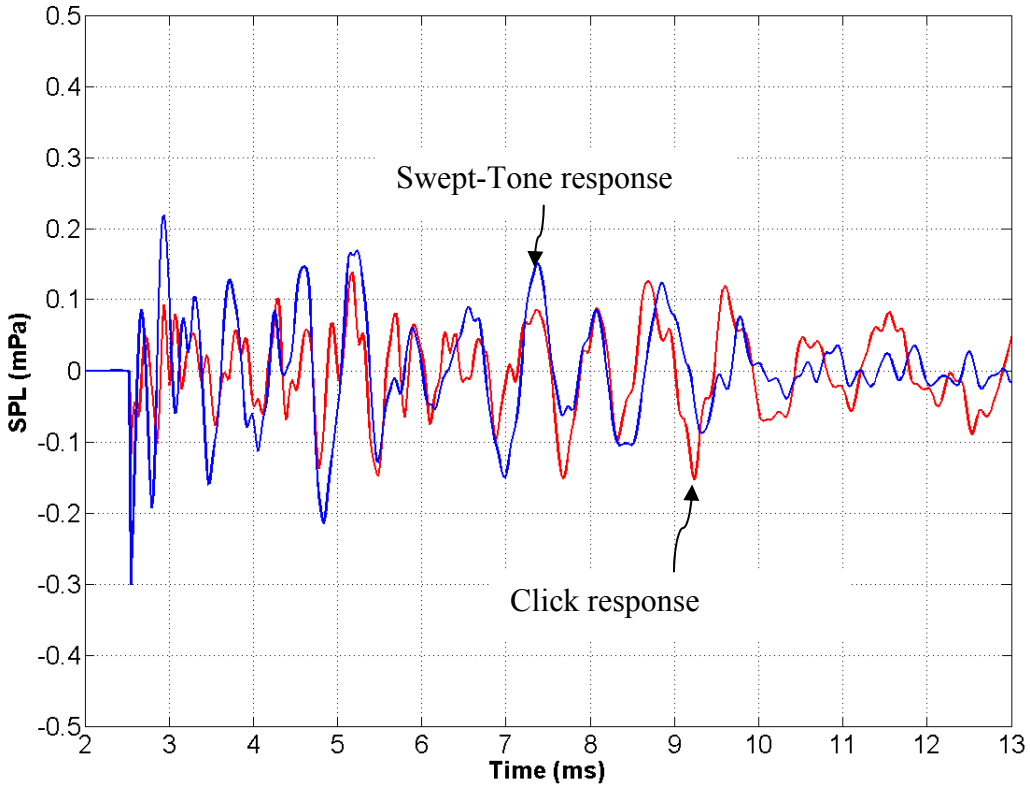

Figure 46. The swept-tone (blue) and click (red) responses at 76 dB pSPL stimulus intensity is shown for a single subject (SUB01R). Strong similarities in phase and time-alignment can be seen.

The results from Chapter 4 indicate that there are some strong similarities and some divergences as well, between swept-tone TEOAE and click TEOAE responses. The similarities are in the phase alignment, with consistent zero-crossings and polarity between the two responses. The divergence is in the late latency amplitude, with the swept-tone responses losing signal power beyond 10 – 15 ms. The point at which signal loses amplitude seems to be subject dependent, but basically the final third of the response (in the late latencies) is highly attenuated. This could give some indication as to the generation mechanisms of click responses versus stimulus-frequency responses.

5.3.1 GENERATION MECHANISMS REVISITED

There are some common elements in the generation sources of SFOAEs and TEOAEs. It is thought for either modality, that a response will be comprised of two primary constituent signals, namely linear coherent reflection at low stimulus intensities (Talmadge *et al.*, 1998; Shera and Guinan Jr., 1999), or a combination of linear coherent reflection at and nonlinear distortion at high stimulus intensities (Shera and Guinan Jr., 1999; Yates and Withnell, 1999; Long *et al.*, 2001; Talmadge *et al.*, 2000).

However, click responses may be comprised of additional physiological sources. Because of the nonlinear characteristics of the basilar membrane, and the fact that a click stimulus is comprised of several frequency components, intermodulation distortion is thought to arise as these frequency components interact along a nonlinear medium (Yates and Withnell, 1999; Withnell *et al.*, 2000). Under this theory, high levels of stimulus-related nonlinear interactions occur at basal regions of the basilar membrane, causing a preponderance of low-frequency intermodulation distortions, which appear as emissions. So in addition to linear coherent reflection sources and nonlinear distortion sources, an intermodulation distortion source may exist in click responses, but not in stimulus-frequency responses.

Additionally, a click stimulus is known to evoke synchronized spontaneous emissions. Spontaneous emissions can be thought of as a free-rotating phasor at a particular frequency location (Sisto and Moleti, 1999). Synchronized spontaneous emissions can have that phasor reset at the onset of a click stimulus, allowing for their observation within a typical click recording modality. It has been suggested that click responses may be comprised of such synchronized spontaneous activity (Kulawiec and Orlando, 1995;

Probst *et al.*, 1986). Spontaneous OAEs generally occur in the 1 to 2 kHz region (Hall, 2000), and if contributing to the click response, would seemingly manifest within that frequency region.

Click OAE responses are a bit more complex than SFOAEs, in that they are comprised of possibly two additional sources: intermodulation distortion and/or spontaneous OAEs. One or both of these additional sources can be attributed to the discrepancy in the late-latency (low-frequency) amplitudes between the click and swept-tone OAEs. For both theories, the TEOAE should have an amplified low-frequency response. However, it is not entirely clear which source is the more likely contributor. The swept-tone stimulus has a single instantaneous frequency, thus preventing intermodulation distortion. But also the swept-tone post-processing method attenuates (or does not evoke) spontaneous activity. So it is not entirely clear which source primarily contributes to the enhanced low-frequency amplitude in click responses.

5.3.2 TIME-FREQUENCY LATENCIES VIA GROUP DELAY

Time-frequency latencies measure the time from stimulus onset to a particular frequency group in a transient response. They are commonly used to give some indication regarding cochlear delay, and are generally computed as the time difference between the stimulus onset and the peak magnitude of some frequency region. The peak magnitude can be computed by several analysis techniques, many of which are discussed in Sections 1.2.3, 2.3, and 4.2.5. But an alternative method was developed (described in Section 4.2.7), which computes cochlear delay using phase response instead of magnitude response. There was a strong agreement in the calculation of cochlear delays between the

magnitude and phase methods. For example, recall that in a typical time-frequency model $t = af^{-b}$, the literature and experimental values obtained with the magnitude and phase methods were $\{a_{\text{lit}}, a_{\text{mag}}, a_{\text{phase}}\}$ = {10.0, 12.9, 12.0} and $\{b_{\text{lit}}, b_{\text{mag}}, b_{\text{phase}}\}$ = {0.5, 0.5, 0.4} for the click response and $\{a_{\text{mag}}, a_{\text{phase}}\}$ = {10.4, 8.4} and $\{b_{\text{mag}}, b_{\text{phase}}\}$ = {0.4, 0.4} for the swept-tone response. Calculation of cochlear delays via phase methods (or more specifically via the negative derivative of the phase, with respect to frequency, otherwise known as group delay) may provide an alternative method to characterize cochlear responses.

5.3.3 Considering Alternative Nonlinear Models

The assumption made in the analysis of swept-tone OAEs is that the probed ear can be characterized as a Hammerstein model, which consists of a static nonlinear system followed by a linear system. Other such models exist, for example a Wiener model, which consists of a linear system followed by static nonlinear system. The Hammerstein model uses a swept-tone stimulus to obtain the response, and the Wiener model uses a noise stimulus to obtain the response. Wiener models have been used in OAE analysis in the past to study the bullfrog ear (van Dijk *et al.*, 1997) and the human ear (Maat *et al.*, 2000), but without success. In the two aforementioned studies, responses were found in the first order Wiener kernel, but only noise was recovered in the higher ordered kernels. Some studies, in fact, suggest that in the case of a fixed-dynamic process, the Hammerstein model will outperform a Wiener model (Aguirre *et al.*, 2005; Novák *et al.*, 2009), which may argue for the selection of a Hammerstein model in OAE analysis. Future studies may consider a hybrid model, known as a Wiener-Hammerstein cascade

model, which consists of a linear system followed by a static nonlinearity followed by another linear system, with the intention of fully characterizing the probe-ear system.

The limitation of either a Hammerstein or Wiener model is that they can only characterize the diagonals of the higher order kernels (in other words, they only consider the instantaneous inputs without regard to previous inputs). Alternatively, a method was developed by Thornton and colleagues (Thornton, 1997; Thornton *et al.*, 2001; Slaven *et al.*, 2003; de Boer and Thornton, 2006) using a maximum length sequence, which is analogous to a noise stimulus, and is discussed in Section 1.2.2. Use of a maximum length sequence allows for the characterization of the entire Volterra kernel at higher orders. As such, nonlinearities due to current or previous (or future, for that matter) inputs can be modeled. The results in the current study concur with those studies' findings with regards to the higher order responses in that their duration was only 12-15 ms long, they exhibited high to low frequency dispersion in their response, and their amplitudes were lower than first order responses. This confirms that the selection of a Hammerstein model is appropriate for OAE analysis.

5.4 FUTURE IMPROVEMENTS

While this dissertation discusses a novel method for acquiring OAEs, it perhaps raises more questions than it answers. Some of these questions include: What would extended nonlinear responses (i.e., longer than 5-10 ms) tell us, especially with regard to the late latency, low frequency regions? Are spontaneous emissions or intermodulation distortions the primary source of low-frequency energy in click responses, and can a modified swept-tone study help determine this? Can the swept-tone method be extended

to operate under DPOAE modality? Can the swept-tone be compensated for flat envelope inside the auditory canal? Can the swept-tone method be extended to operate under bone conduction modality?

Effects of Sampling Rate on SNR

During the convolution process with the inverse sweep, the additive noise is shaped by the spectrum of the inverse sweep, as discussed in Section 2.2.4.3. As a result, noise shaping occurs, pushing the noise into the high-frequencies. If the sampling rate is increased, then theoretically the noise would be even further reduced, resulting in an even better SNR. In fact, for every doubling of the sampling rate (assuming a fixed upper cutoff frequency) the SNR can be further improved by +3 dB. In terms of synchronous averaging, an improvement of +3 dB corresponds to a halving of acquisition time.

Extended Swept-Tone

In the current study, a 100 ms swept-tone stimulus was chosen. However, the length of the stimulus could be (seemingly) arbitrarily long. If the stimulus length is increased, then the duration of the higher ordered responses is also increased. If one were to design the stimulus to have a full 20 ms at the 5^{th} order response, then that would correspond to a stimulus length of 311 ms, which can be calculated simply by using Eq. (2.32). Furthermore, extending the stimulus length even more would allow for a more direct comparison to current SFOAE or DPOAE techniques, which typically employ continuously swept primaries which are anywhere from 5 to 20 s in duration. With an extended duration, a more complete analysis of the nonlinear components could be made, especially with regard to input/output growth function and time-frequency latencies.

Alternate Methods for Removal of Stimulus Artifact

There are other methods for removing the stimulus artifact in the SFOAE modality. In this study, the DNLR method was employed, which helped reduce the artifact by one to two orders of magnitude. This is very similar to the "compression" method for reducing the artifact from SFOAE responses. Radha Kalluri and Chris Shera rigorously examine the compression method, as well as two other methods, including the suppression method and spectral smoothing (Kalluri and Shera, 2007a). Employing an alternative method, such as the suppression or spectral smoothing method, would allow for linear acquisition of the swept-tone response, while also reducing the stimulus artifact. This would be an improvement over the DNL or compression techniques, which summarily eliminate linear components of the OAE response. The linear components may be of interest in the development of an OAE model, or with regard to the generation mechanisms.

Ear Canal Compensation and Bone Conduction

In Chapter 3, a routine for the compensation of the ear canal is introduced. In this dissertation, the compensation, as applied to a click stimulus, is explored. However, this technique could further be applied to the swept-tone stimulus in order to achieve a flat envelope stimulus within the ear canal. This may help reduce the stimulus artifact, recover higher frequency OAEs, and provide equalized stimulation along the cochlear partition.

Bone conduction may provide another method for equalizing the stimulus and reducing the ear canal artifact. Bone conduction has been used for the acquisition of DPOAEs and TEOAEs and also for the suppression of those OAEs (Schiopu, 1997). Bone conduction

allows for very high stimulation frequencies to be achieved, above those that can be achieved via air conduction, since interference patterns are not established (Sheffield *et al.*, 2010).

Swept-Tone OAE Extended to DPOAE Acquisition Modality

Perhaps the most intriguing future possibility involves extending the swept-tone from a single frequency SFOAE modality to a two-tone DPOAE modality. This would actually require the formulation of three swept-tones. Under the SFOAE modality, there is a single instantaneous frequency, and which evokes an OAE of the same frequency. However, under the DPOAE modality, there are two continuously sweeping primaries, but the response occurs at a separate DP frequency (typically the inner, or cubic distortion product is considered, which is given by $2f_1$-f_2). The third swept-tone would be at this DP frequency, and would be used to compress the DPOAE response to resemble a click response. It would be expected that the nonlinear components would be more heavily populated since DPOAE responses contain a strong nonlinear distortion source, whereas SFOAE responses consist mostly of a linear coherent reflection source. Continuously sweeping primaries have been used in DPOAE responses (Long *et al.*, 2008), and so a future method would instead use the swept-tone stimuli, as shown in Figure 47.

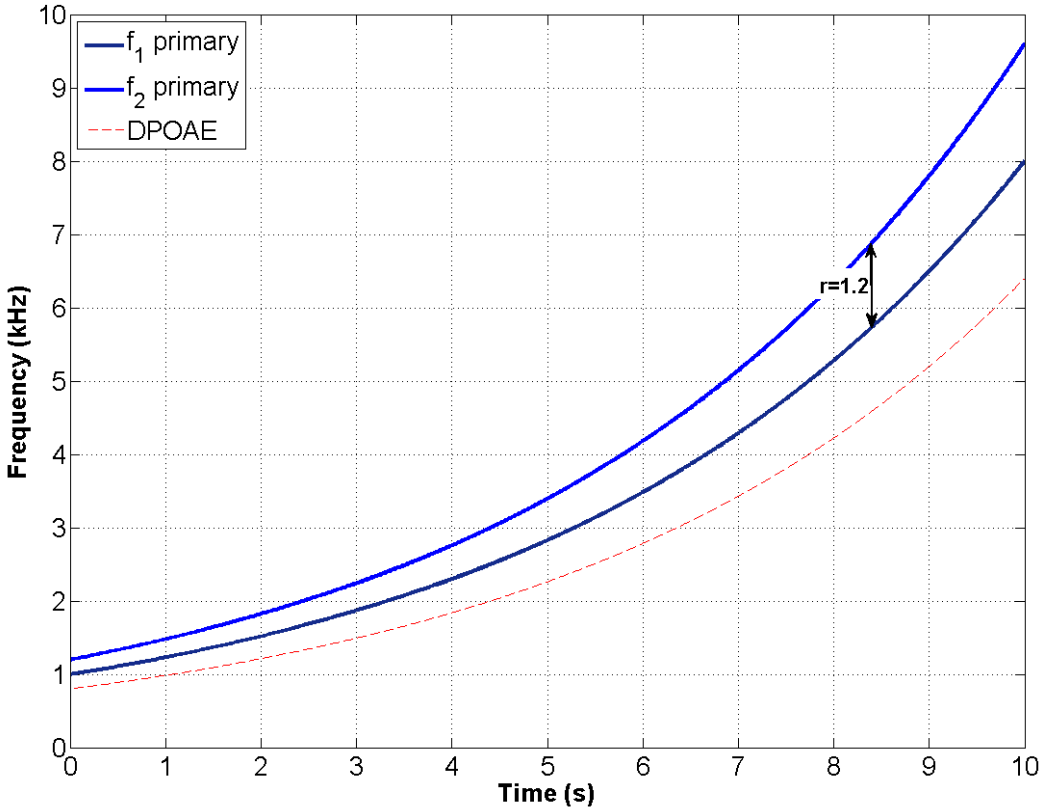

Figure 47. In a proposed DPOAE design, the continuously sweeping primaries would consist of swept-tone stimuli, separated by a ratio of 1.2. The cubic DP occurs at the $2f_1$-f_2 frequency, indicated by the dashed red curve. The f_1 primary (dark blue) sweeps from 1 kHz to 8 kHz, and the f_2 primary (blue) from 1.2 kHz to 9.6 kHz. The DPOAE response could be compressed to a click-like response using the swept-tone post-processing method.

Under current processing methods, the DPOAE is separated from the stimuli by a least-squares-fit method developed by Carrick Talmadge and reported in the Appendix of (Long and Talmadge, 1997). This method could be employed to separate the DPOAE response from the swept-tone stimuli. Furthermore, in current processing methods, the

nonlinear distortion component is separated from the linear coherent reflection source by a employing a recursive exponential window (Kalluri and Shera, 2001). The idea being that the DPOAE will contain a mixture of the distortion and reflection sources, but that the distortion will have a very slow-rotating phase (implying a short latency) and that the reflection source will have a fast-rotating phase (implying a longer latency). This generally has to do with the fact that the distortion source is fixed to the travelling wave, whereas the reflection source is fixed to its tonotopic place on the Basilar membrane. The purpose of the source unmixing is to isolate the reflection source, which may give a better indication of cochlear functioning than the distortion source. An alternative to this technique may be employable with the swept-tone post-processing method, in which the distortion sources may be readily placed within the nonlinear impulse responses. Finally, the problem of determining whether the late latency click emissions are a manifestation of intermodulation distortion could be approached. The DP paradigm necessarily involves intermodulation distortion (of only two frequency components), so the comparison of a compressed DP response to a click response should match if the low-frequencies in a TEOAE are a product of intermodulation distortion. Conversely, if the low-frequencies in a TEOAE are a product of synchronized spontanesous activity, then the compressed DPOAE response should resemble that of an SFOAE, and not a TEOAE.

Appendix

Table 1. Correlation coefficient values between cTEOAE and sTEOAE responses for individual ears at 76 dB pSPL are shown. Correlation values are separated into short-, medium-, and long-latency time windows.

	2.5-6.0 ms	6.0-9.5 ms	9.5-13.0 ms	2.5-13.0 ms
SUB01L	0.602	0.516	0.394	0.433
SUB01R	0.697	0.427	0.407	0.579
SUB03L	0.584	0.146	0.713	0.323
SUB03R	0.550	0.159	0.147	0.243
SUB04L	0.731	0.464	0.494	0.546
SUB04R	0.625	0.421	0.519	0.524
SUB18L	0.275	0.256	0.490	0.118
SUB18R	0.216	0.223	0.597	0.226
SUB19L	0.693	0.843	0.814	0.612
SUB19R	0.789	0.347	0.479	0.586
SUB20L	0.432	0.162	0.115	0.242
SUB20R	0.276	0.106	0.221	0.166
SUB21L	0.330	0.142	0.886	0.304
SUB21R	0.428	0.369	0.620	0.306
SUB22L	0.731	0.474	0.307	0.392
SUB22R	0.173	0.408	0.174	0.203
SUB23L	0.603	0.450	0.405	0.444
SUB23R	0.620	0.157	0.442	0.345
SUB24L	0.676	0.721	0.814	0.638
SUB24R	0.210	0.198	0.452	0.185
SUB25L	0.229	0.551	0.725	0.447
SUB25R	0.259	0.522	0.680	0.425
SUB26L	0.596	0.462	0.117	0.519
SUB26R	0.593	0.514	0.789	0.473
SUB27L	0.196	0.431	0.404	0.119
SUB27R	0.822	0.516	0.418	0.597
Mean (ms)	0.50±0.21	0.38±0.19	0.49±0.23	**0.38±0.17**

Table 2. RMS noise levels for individual ears. The even/odd buffering method was used to calculate noise levels for each recording duration. The swept-tone method consistently outperforms the standard IR method with respect to noise floor.

	cTEOAE (mPa)				sTEOAE (mPa)			
	4 s	**8 s**	**16 s**	**32 s**	**4 s**	**8 s**	**16 s**	**32 s**
SUB01L	0.046	0.030	0.025	0.015	0.017	0.015	0.009	0.008
SUB01R	0.050	0.036	0.033	0.032	0.030	0.019	0.029	0.014
SUB03L	0.125	0.084	0.055	0.043	0.048	0.031	0.035	0.018
SUB03R	0.134	0.083	0.059	0.043	0.049	0.032	0.028	0.019
SUB04L	0.110	0.074	0.051	0.035	0.053	0.032	0.025	0.018
SUB04R	0.115	0.077	0.053	0.042	0.047	0.036	0.025	0.023
SUB18L	0.556	0.386	0.282	0.213	0.201	0.136	0.103	0.071
SUB18R	0.340	0.244	0.184	0.122	0.152	0.126	0.106	0.066
SUB19L	0.125	0.084	0.063	0.045	0.069	0.039	0.030	0.020
SUB19R	0.126	0.087	0.060	0.068	0.052	0.038	0.032	0.030
SUB20L	0.168	0.110	0.078	0.053	0.067	0.058	0.033	0.024
SUB20R	0.118	0.075	0.064	0.037	0.037	0.035	0.024	0.020
SUB21L	0.151	0.210	0.074	0.050	0.062	0.046	0.036	0.027
SUB21R	0.142	0.242	0.076	0.056	0.079	0.057	0.032	0.026
SUB22L	0.163	0.113	0.087	0.051	0.044	0.037	0.120	0.091
SUB22R	0.484	0.295	0.211	0.136	0.139	0.104	0.082	0.065
SUB23L	0.047	0.037	0.021	0.020	0.095	0.062	0.042	0.027
SUB23R	0.148	0.114	0.068	0.052	0.058	0.036	0.030	0.023
SUB24L	0.061	0.060	0.026	0.019	0.053	0.023	0.023	0.014
SUB24R	0.117	0.073	0.060	0.033	0.104	0.071	0.061	0.033
SUB25L	0.175	0.124	0.096	0.183	0.078	0.065	0.035	0.023
SUB25R	0.186	0.124	0.081	0.056	0.076	0.055	0.035	0.025
SUB26L	0.170	0.137	0.096	0.071	0.070	0.053	0.039	0.029
SUB26R	0.202	0.147	0.267	0.085	0.089	0.065	0.043	0.037
SUB27L	0.181	0.136	0.090	0.061	0.074	0.052	0.040	0.030
SUB27R	0.185	0.124	0.078	0.067	0.087	0.046	0.035	0.035
Mean	**0.170**	**0.127**	**0.090**	**0.065**	**0.074**	**0.053**	**0.044**	**0.031**
	±0.119	**±0.085**	**±0.069**	**±0.048**	**±0.040**	**±0.030**	**±0.028**	**±0.020**

Table 3. RMS values for every subject are shown for both click and swept-tone responses at each stimulation intensity from 46 to 76 dB pSPL. Responses grow compressively, as expected, and are somewhat higher (by a few dB) for the click responses on average.

	cTEOAE (mPa)						sTEOAE (mPa)					
	46 dB	**52 dB**	**58 dB**	**64 dB**	**70 dB**	**76 dB**	**46 dB**	**52 dB**	**58 dB**	**64 dB**	**70 dB**	**76 dB**
SUB01L	0.02	0.03	0.03	0.02	0.03	0.04	0.01	0.01	0.02	0.03	0.04	0.06
SUB01R	0.04	0.03	0.04	0.05	0.05	0.06	0.02	0.02	0.03	0.04	0.06	0.08
SUB03L	0.06	0.06	0.05	0.06	0.06	0.14	0.03	0.03	0.04	0.05	0.07	0.09
SUB03R	0.05	0.06	0.12	0.11	0.17	0.09	0.04	0.03	0.04	0.05	0.07	0.09
SUB04L	0.06	0.05	0.06	0.06	0.06	0.07	0.03	0.03	0.04	0.05	0.07	0.10
SUB04R	0.06	0.06	0.05	0.06	0.07	0.08	0.03	0.03	0.04	0.06	0.08	0.10
SUB18L	0.09	0.13	0.10	0.10	0.14	0.33	0.04	0.05	0.07	0.09	0.10	0.13
SUB18R	0.12	0.13	0.50	0.54	0.32	0.33	0.06	0.07	0.09	0.11	0.16	0.19
SUB19L	0.10	0.18	0.26	0.32	0.38	0.42	0.09	0.11	0.13	0.16	0.19	0.22
SUB19R	0.05	0.12	0.14	0.18	0.19	0.21	0.04	0.05	0.06	0.07	0.08	0.10
SUB20L	0.08	0.08	0.07	0.07	0.07	0.09	0.04	0.04	0.04	0.05	0.06	0.09
SUB20R	0.08	0.07	0.07	0.06	0.08	0.13	0.05	0.05	0.04	0.04	0.06	0.08
SUB21L	0.08	0.09	0.41	0.33	0.27	0.32	0.06	0.07	0.08	0.11	0.12	0.15
SUB21R	0.09	0.08	0.12	0.16	0.19	0.28	0.05	0.06	0.08	0.10	0.12	0.16
SUB22L	0.07	0.07	0.07	0.09	0.09	0.11	0.03	0.04	0.04	0.05	0.06	0.11
SUB22R	0.09	0.11	0.12	0.13	0.15	0.15	0.03	0.04	0.04	0.05	0.06	0.11
SUB23L	0.04	0.04	0.04	0.06	0.09	0.09	0.03	0.04	0.05	0.07	0.09	0.11
SUB23R	0.07	0.07	0.08	0.08	0.09	0.09	0.03	0.04	0.04	0.04	0.05	0.07
SUB24L	0.06	0.05	0.05	0.03	0.03	0.04	0.01	0.02	0.02	0.03	0.04	0.06
SUB24R	0.07	0.06	0.06	0.06	0.06	0.06	0.02	0.03	0.03	0.03	0.05	0.07
SUB25L	0.08	0.09	0.09	0.09	0.09	0.18	0.04	0.04	0.05	0.06	0.07	0.09
SUB25R	0.09	0.08	0.09	0.10	0.10	0.10	0.05	0.05	0.05	0.06	0.08	0.10
SUB26L	0.09	0.09	0.08	0.09	0.09	0.12	0.04	0.08	0.05	0.07	0.08	0.11
SUB26R	0.10	0.19	0.20	0.21	0.31	0.40	0.09	0.10	0.14	0.18	0.22	0.24
SUB27L	0.09	0.08	0.09	0.09	0.10	0.11	0.04	0.05	0.05	0.06	0.09	0.13
SUB27R	0.09	0.09	0.08	0.09	0.12	0.11	0.04	0.04	0.05	0.07	0.09	0.13
Mean (dB SPL)	**11.2**	**12.5**	**15.4**	**15.8**	**16.3**	**18.0**	**6.0**	**7.4**	**8.7**	**10.6**	**12.7**	**15.1**

References

Aguirre, L.A., Coelho, M.C.S. and Correa, M.V. (**2005**). "On the interpretation and practice of dynamical differences between Hammerstein and Wiener models," IEE Proc. Contr. Theor. Ap. **152**, 349-356.

Arslan, R.B., Özdamar, Ö. and Ulgen, Y. (**2001**). "Digital subtraction method for transient evoked otoacoustic emissions recording with ipsilateral noise suppression: An application to stimulus artifact reduction," Audiology **40**, 55-62.

Bennett, C.L. and Özdamar, Ö. (**2010b**). "Swept-tone transient evoked otoacoustic emissions," J. Acoust. Soc. Am. **In Press**.

Bennett, C.L. and Özdamar, Ö. (**2010a**). "High-frequency transient evoked otoacoustic emissions acquisition with auditory canal compensated clicks using swept-tone analysis," J. Acoust. Soc. Am. **127**, 2410-2419.

Bennett, C.L. and Özdamar, Ö. (**2009**). "High resolution system for improved transient-evoked otoacoustic emission acquisition," Conf. Proc. IEEE Eng. Med. Biol. Soc. 2009, 6263-6266.

Blinowska, K.J., Durka, P.J., Skierski, A., Grandori, F. and Tognola, G. (**1997**). "High resolution time-frequency analysis of otoacoustic emissions," Technology and Health Care **5**, 407-418.

Bray, P.J. (**1989**). *A study of the properties of click evoked otoacoustic emissions and development of a clinical otoacoustic hearing test instrument*, a PhD Dissertation, University of London.

Burkard, R. (**1984**). "Sound pressure level measurement and spectral analysis of brief acoustic transients," Electroencephalogr. Clin. Neurophysiol. **57**, 83-91.

Cheng, J. (**1995**). "Time-frequency analysis of transient evoked otoacoustic emissions via smoothed pseudo Vigner distribution," Scand.Audiol. **24**, 91-96.

Chertoff, M.E. and Chen, J. (**1996**). "An in-situ calibration procedure for click stimuli," J. Am. Acad. Audiol. **7**, 130-137.

Chertoff, M.E. and Guruprasad, S.N. (**1997**). "Application of a stimulus spectral calibration routine to click evoked otoacoustic emissions," J. Am. Acad. Audiol. **8**, 333-341.

Choi, Y.S., Lee, S.Y., Parham, K., Neely, S.T. and Kim, D.O. (**2008**). "Stimulus-frequency otoacoustic emission: Measurements in humans and simulations with an active cochlear model," J. Acoust. Soc. Am. **123**, 2651-2669.

Cohen, J. 1988, *Statistical Power Analysis for the Behavioral Sciences,* 2nd edn., Lawrence Erlbaum Associates, Hillsdale, New Jersey.

Craven, P.G. and Gerzon, M.A. (**1992**). "Practical adaptive room and loudspeaker equalizer for hi-fi use," *92nd Audio Eng Soc Convention,* Vienna, Austria.

Dau, T., Wegner, O., Mellert, V. and Kollmeier, B. (**2000**). "Auditory brainstem responses with optimized chirp signals compensating basilar-membrane dispersion," J. Acoust. Soc. Am. **107**, 1530-1540.

Daubechies, I. (**1988**). "Orthonormal bases of compactly supported wavelets," Comm. Pur. Appl. Math. **41**, 909-996.

de Boer, J. and Thornton, A.R.D. (**2006**). "Volterra slice otoacoustic emissions recorded using maximum length sequences from patients with sensorineural hearing loss," Hear. Res. **219**, 121-136.

Dreisbach, L.E., Long, K.M., and Lees, S.E. (**2006**). "Repeatability of high-frequency distortion-product otoacoustic emissions in normal-hearing adults," Ear Hearing. **27**, 466-479.

Durrant, J.D. & Lovrinic, J.H. 1995, *Basis of hearing science,* 3rd edn., William & Wilkins, Baltimore, MD.

Farina, A. (**2007**). "Advancements in impulse response measurements by sine sweeps," *122nd Audio Eng Soc Convention,* Vienna, Austria.

Farina, A. (**2000**). "Simultaneous measurement of impulse response and distortion with a swept-sine technique," *110th Audio Eng Soc Convention,* Paris, France.

Fobel, O. and Dau, T. (**2004**). "Searching for the optimal stimulus eliciting auditory brainstem responses in humans," J. Acoust. Soc. Am. **116**, 2213-2222.

Goodman, S.S., Fitzpatrick, D.F., Ellison, J.C., Jesteadt, W. and Keefe, D.H. (**2009**). "High-frequency click-evoked otoacoustic emissions and behavioral thresholds in humans," J. Acoust. Soc. Am. **125**, 1014-1032.

Grandori, F. and Ravazzani, P. (**1993**). "Non-linearities of click-evoked otoacoustic emissions and the derived non-linear technique," Br. J. Audiol. **27**, 97-102.

Greenwood, D.D. (**1961**). "Critical bandwidth and the frequency coordinates of the basilar membrane," J. Acoust. Soc. Am. **33**, 1344-1356.

Griesinger, D. (**1996**). "Beyond MLS - Occupied hall measurement with FFT techniques," *101st Audio Eng Soc Convention,* Los Angeles, California.

Hall, J.W. (**2000**). "Spontaneous otoacoustic emissions (SOAEs)" in *Handbook of otoacoustic emissions.* Singular Publishing Group, San Diego, CA, pp. 67-93.

Harte, J.M., Pigasse, G. and Dau, T. (**2009**). "Comparison of cochlear delay estimates using otoacoustic emissions and auditory brainstem responses," J. Acoust. Soc. Am. **126**, 1291-1301.

Hatzopoulos, S., Cheng, J., Grzanka, A., Morlet, T. and Martini, A. (**2000**). "Optimization of TEOAE recording protocols: A linear protocol derived from parameters of a time-frequency analysis," Scand. Audiol. **29**, 21-27.

Hatzopoulos, S., Petrucelli, J., Morlet, T. and Martini, A. (**2003**). "TEOAE recording protocols revised: Data from adult subjects." Int. J. Audiol. **42**, 339-347.

Heidelberg, G. and Frankfurt, G. (**2007**). "TEOAE amplitude growth, detectability, and response threshold in linear and nonlinear mode and in different time windows," Int. J. Audiol. **46**, 407-418.

Hellmuth, O., Allamanche, E., Herre, J., Kastner, T., Cremer, M. and Hirsch, W. (**2001**). "Advanced audio identification using MPEG-7 content description," *110th Audio Eng Soc Convention,* New York, NY, USA.

Janušauskas, A., Marozas, V., Engdahl, B., Hoffman, H.J., Svensson, O. and Sörnmo, L. (**2001**). "Otoacoustic emissions and improved pass/fail separation using wavelet analysis and time windowing," Med. Biol. Eng. Comput. **39**, 134-139.

Jedrzejczak, W.W., Blinowska, K.J., Konopka, W., Grzanka, A. and Durka, P.J. (**2004**). "Identification of otoacoustic emissions components by means of adaptive approximations," J. Acoust. Soc. Am. **115**, 2148-2158.

Kalluri, R. and Shera, C.A. (**2007a**). "Comparing stimulus-frequency otoacoustic emissions measured by compression, suppression, and spectral smoothing," J. Acoust. Soc. Am. **122**, 3562-3575.

Kalluri, R. and Shera, C.A. (**2007b**). "Near equivalence of human click-evoked and stimulus-frequency otoacoustic emissions," J. Acoust. Soc. Am. **121**, 2097-2110.

Kalluri, R. and Shera, C.A. (**2001**). "Distortion-product source unmixing: A test of the two-mechanism model for DPOAE generation," J. Acoust. Soc. Am. **109**, 622-637.

Kapadia, S. and Lutman, M.E. (**2000**). "Nonlinear temporal interactions in click-evoked otoacoustic emissions. I. Assumed model and polarity-symmetry," Hear. Res. **146**, 89-100.

Keefe, D.H. (**1998**). "Double-evoked otoacoustic emissions. I. Measurement theory and nonlinear coherence," J. Acoust. Soc. Am. **103**, 3489-3498.

Keefe, D.H. and Ling, R. (**1998**). "Double-evoked otoacoustic emissions. II. Intermittent noise rejection, calibration and ear-canal measurements," J. Acoust. Soc. Am. **103**, 3499-3508.

Kemp, D.T. (**1978**). "Stimulated acoustic emissions from within the human auditory system," J. Acoust. Soc. Am. **64**, 1386-1391.

Kemp, D.T., Bray, P., Alexander, L. and Brown, A.M. (**1986**). "Acoustic emission cochleography--practical aspects," Scand. Audiol. Suppl. **25**, 71-95.

Kemp, D.T. and Chum, R. (**1980**). "Properties of the generator of stimulated acoustic emissions," Hear. Res. **2**, 213-232.

Kruglov, A., Artamasov, S., Frolenkov, G. and Tavartkiladze, G. (**1997**). "Transient evoked otoacoustic emission with unexpectedly short latency," Acta. Otolaryngol. **117**, 174-178.

Kulawiec, J.T. and Orlando, M.S. (**1995**). "The contribution of spontaneous otoacoustic emissions to the click evoked otoacoustic emissions," Ear Hear. **16**, 515-520.

Lineton, B., Thornton, A.R.D. and Baker, V.J. (**2006**). "An investigation into the relationship between input–output nonlinearities and rate-induced nonlinearities of click-evoked otoacoustic emissions recorded using maximum length sequences," Hear. Res. **219**, 24-35.

Long, G.R. and Talmadge, C.L. (**1997**). "Spontaneous otoacoustic emission frequency is modulated by heartbeat," J. Acoust. Soc. Am. **102**, 2831-2848.

Long, G.R., Talmadge, C.L. and Lee, J. (**2008**). "Measuring distortion product otoacoustic emissions using continuously sweeping primaries," J. Acoust. Soc. Am. **124**, 1613-1626.

Long, G.R., Talmadge, C.L. and Thorpe, C.A. (**2001**). "Experimental measurement of level dependence of stimulus frequency otoacoustic emission fine structure," 24th Midwinter Meeting, Assoc. Res. Otolaryngol.

Maat, B., Wit, H.P. and van Dijk, P. (**2000**). "Noise-evoked otoacoustic emissions in humans," J. Acoust. Soc. Am. **108**, 2272-2280.

Medri, E. and Özdamar, Ö. (**2000**). "Simultaneously recorded chirp evoked otoacoustic emissions and auditory brainstem responses," 23rd Midwinter Meeting, Assoc. Res. Otolaryngol.

Muller, S. and Massarani, P. (**2001**). "Transfer-function measurement with sweeps," J. Audio Eng. Soc. **49**, 443-471.

Notaro, G., Al Maamury, A.M., Moleti, A. and Sisto, R. (**2007**). "Wavelet and matching pursuit estimates of the transient-evoked otoacoustic emission latency," J. Acoust. Soc. Am. **122**, 3576-3585.

Novák, A., Simon, L., Kadlec, F. and Lotton, P. (**2009**). "Nonlinear System Identification Using Exponential Swept-Sine Signal," IEEE T. Instrum. Meas. **In Press**.

Peeters, G.G. and Deruty, E. (**2008**). "Automatic morphological description of sounds," Proc. Acoustics '08, Paris, France, 5783-5788.

Prieve, B.A., Gorga, M.P. and Neely, S.T. (**1996**). "Click-and tone-burst-evoked otoacoustic emissions in normal-hearing and hearing-impaired ears," J. Acoust. Soc. Am. **99**, 3077-3086.

Probst, R., Coats, A.C., Martin, G.K. and Lonsbury-Martin, B.L. (**1986**). "Spontaneous, click-, and toneburst-evoked otoacoustic emissions from normal ears," Hear. Res. **21**, 261-265.

Probst, R., Lonsbury-Martin, B.L. and Martin, G.K. (**1991**). "A review of otoacoustic emissions," J. Acoust. Soc. Am. **89**, 2027-2067.

Reed, M.J. and Hawksford, M.O.J. (**1996**). "Identification of discrete Volterra series using maximum length sequences," IEE Proc. Circuits Devices Syst. **143**, 241-248.

Rosowski, J.J. (**1996**). "Models of external and middle ear function" in *Auditory computation.*, eds. H.L. Hawkins, T.A. McMullen, A.N. Popper & R.R. Fay, Springer, New York, pp. 15-61.

Ryan, S. and Kemp, D.T. (**1996**). "The influence of evoking stimulus level on the neural suppression of transient evoked otoacoustic emissions," Hear. Res. **94**, 140-147.

Schairer, K.S., Fitzpatrick, D. and Keefe, D.H. (**2003**). "Input–output functions for stimulus-frequency otoacoustic emissions in normal-hearing adult ears," J. Acoust. Soc. Am. **114**, 944-966.

Schiopu, N. (**1997**). *Transient Evoked Otoacoustic Emissions Elicited by Bone-Conducted Ultrasonic Stimuli*, a Master's Thesis, University of Toronto.

Sheffield, B., Martin-Roff, J., Sokolich, G., Dreisbach, L. and and Zeng, F.G. (**2010**). "DPOAEs suppressed by bone-conducted ultrasound in humans," Assoc. Res. Otolaryngol., Anaheim, CA.

Shera, C.A. and Guinan Jr., J.J. (**2003**). "Stimulus-frequency-emission group delay: A test of coherent reflection filtering and a window on cochlear tuning," J. Acoust. Soc. Am. **113**, 2762-2772.

Shera, C.A. and Guinan Jr., J.J. (**1999**). "Evoked otoacoustic emissions arise by two fundamentally different mechanisms: A taxonomy for mammalian OAEs," J. Acoust. Soc. Am. **105**, 782-798.

Shera, C.A. and Zweig, G. (**1993**). "Noninvasive measurement of the cochlear traveling-wave ratio," J. Acoust. Soc. Am. **93**, 3333-3352.

Sisto, R. and Moleti, A. (**2008**). "Transient evoked otoacoustic emission input/output function and cochlear reflectivity: Experiment and model," J. Acoust. Soc. Am. **124**, 2995-3008.

Sisto, R. and Moleti, A. (**2002**). "On the frequency dependence of the otoacoustic emission latency in hypoacoustic and normal ears," J. Acoust. Soc. Am. **111**, 297-308.

Sisto, R. and Moleti, A. (**1999**). "Modeling otoacoustic emissions by active nonlinear oscillators," J. Acoust. Soc. Am. **106**, 1893-1906.

Slaven, A., Lineton, B. and Thornton, A.R.D. (**2003**). "Properties of Volterra slices of otoacoustic emissions in normal-hearing humans obtained using maximum length sequences of clicks," Hear. Res. **179**, 113-125.

Stavroulake, P., Apostolopoulos, N., Segas, J., Tsakanikos, M., and Adamopoulos, G. (**2001**). "Evoked otoacoustic emissions--an approach for monitoring cisplatin induced ototoxicity in children." Int. J. Pediatr. Otorhinolaryng. **59**, 47-57.

Stevens, S.S., Volkmann, J. and Newman, E.B. (**1937**). "A scale for the measurement of the psychological magnitude pitch," J. Acoust. Soc. Am. **8**, 185-190.

Swabey, M.A., Chambers, P., Lutman, M.E., White, N.M., Chad, J.E., Brown, A.D. and Beeby, S.P. (**2009**). "The biometric potential of transient otoacoustic emissions," Int. J. Biometrics. **1**, 349-364.

Talmadge, C.L., Long, G.R., Tubis, A. and Dhar, S. (**1999**). "Experimental confirmation of the two-source interference model for the fine structure of distortion product otoacoustic emissions," J. Acoust. Soc. Am. **105**, 275-292.

Talmadge, C.L., Tubis, A., Long, G.R. and Piskorski, P. (**1998**). "Modeling otoacoustic emission and hearing threshold fine structures," J. Acoust. Soc. Am. **104**, 1517-1543.

Talmadge, C.L., Tubis, A., Long, G.R. and Tong, C. (**2000**). "Modeling the combined effects of basilar membrane nonlinearity and roughness on stimulus frequency otoacoustic emission fine structure," J. Acoust. Soc. Am. **108**, 2911-2932.

Tavartkiladze, G.A., Frolenkov, G.I., Kruglov, A.V. and Artamasov, S.V. (**1994**). "Ipsilateral suppression effects on transient evoked otoacoustic emission," Br. J. Audiol. **28**, 193-204.

Thornton, A.R. (**1997**). "Maximum length sequences and Volterra series in the analysis of transient evoked otoacoustic emissions," Br. J. Audiol. **31**, 493-498.

Thornton, A.R.D., Shin, K., Gottesman, E. and Hine, J. (**2001**). "Temporal non-linearities of the cochlear amplifier revealed by maximum length sequence stimulation," Clin. Neurophysiol. **112**, 768-777.

Tognola, G., Grandori, F. and Ravazzani, P. (**1997**). "Time-frequency distributions of click-evoked otoacoustic emissions," Hear. Res. **106**, 112-122.

Van Campen, L.E., Grantham, D.W. and Hall III, J.W. (**1994**). "Motivation and Methods for Manipulating Acoustic Ringing of Earphones." Ear Hear. **15**, 461-466.

van Dijk, P., Maat, A. and Wit, H.P. (**1997**). "Wiener kernel analysis of a noise-evoked otoacoustic emission," Br. J. Audiol. **31**, 473-477.

Wilson, J.P. (**1980**). "Evidence for a cochlear origin for acoustic re-emissions, threshold fine-structure and tonal tinnitus." Hear. Res. **2**, 233-252.

Wit, H.P., van Dijk, P. and Avan, P. (**1994**). "Wavelet analysis of real ear and synthesized click evoked otoacoustic emissions," Hear. Res. **73**, 141-147.

Withnell, R.H., Hazlewood, C. and Knowlton, A. (**2008**). "Reconciling the origin of the transient evoked otoacoustic emission in humans," J. Acoust. Soc. Am. **123**, 212-221.

Withnell, R.H. and Yates, G.K. (**1998**). "Onset of basilar membrane nonlinearity reflected in cubic distortion tone input-output functions," Hear. Res. **123**, 87-96.

Withnell, R.H., Yates, G.K. and Kirk, D.L. (**2000**). "Changes to low-frequency components of the TEOAE following acoustic trauma to the base of the cochlea," Hear.Res. **139**, 1-12.

Woody, C.D. (**1967**). "Characterization of an adaptive filter for the analysis of variable latency neuroelectric signals," Med. Biol. Eng. Comput. **5**, 539-554.

Yao, J. and Zhang, Y.T. (**2000**). "From otoacoustic emission modeling to bionic wavelet transform," *Proc. 22nd Eng. Med. Biol. Soc. Ann.* Chicago, Illinois.

Yates, G.K. and Withnell, R.H. (**1999**). "The role of intermodulation distortion in transient-evoked otoacoustic emissions," Hear. Res. **136**, 49-64.

Zhang, Z., Zhang, V., Chan, S., McPherson, B. and Hu, Y. (**2008**). "Time–frequency analysis of click-evoked otoacoustic emissions by means of a minimum variance spectral estimation-based method," Hear.Res. **243**, 18-27.

Zheng, L., Zhang, Y.T., Yang, F.S. and Ye, D.T. (**1999**). "Synthesis and decomposition of transient-evoked otoacoustic emissions based on an active auditory model," IEEE T. Bio.-Med. Eng. **46**, 1098-1106.

Zwicker, E. (**1979**). "A model describing nonlinearities in hearing by active processes with saturation at 40 dB," Biol. Cybern. **35**, 243-250.

Zwicker, E. (**1961**). "Subdivision of the audible frequency range into critical bands," J. Acoust. Soc. Am. **33**, 248.