EXTENSION OF TRANSIENTLY EVOKED OTOACOUSTIC EMISSION MEASUREMENTS TO COVER THE ENTIRE AUDIOMETRIC FREQUENCY RANGE

by

Nathanael P. Kuehner

A thesis submitted in conformity with the requirements for the degree of Master of Applied Science
Graduate Department of Electrical and Computer Engineering
University of Toronto

© Copyright by Nathanael P. Kuehner (1997)
The author has granted a non-exclusive licence allowing the National Library of Canada to reproduce, loan, distribute or sell copies of this thesis in microform, paper or electronic formats.

The author retains ownership of the copyright in this thesis. Neither the thesis nor substantial extracts from it may be printed or otherwise reproduced without the author’s permission.

L’auteur a accordé une licence non exclusive permettant à la Bibliothèque nationale du Canada de reproduire, prêter, distribuer ou vendre des copies de cette thèse sous la forme de microfiche/film, de reproduction sur papier ou sur format électronique.

L’auteur conserve la propriété du droit d’auteur qui protège cette thèse. Ni la thèse ni des extraits substantiels de celle-ci ne doivent être imprimés ou autrement reproduits sans son autorisation.
ABSTRACT

In this work, we developed methods which extend the range of transiently evoked otoacoustic emission (TEOAE) measurements to cover the entire audiometric range (up to 8 kHz). Currently, TEOAE measurements are limited to frequencies below approximately 5 kHz.

The presence of a stimulus artifact in each TEOAE recording is the primary cause of this limitation. Current methods of removing this artifact yield poor signal to noise ratios for high frequency TEOAE components. We developed a method based on adaptive wavelet packet based subband decomposition of TEOAE recordings. This new method proved better than current methods at cleanly separating high frequency TEOAE components from stimulus artifact contamination.

Inadequate high frequency stimulation of the auditory system is a secondary cause of this limitation. To correct this, we designed and implemented a new stimulus design procedure which was based on estimating parameters in a system model prior to each TEOAE measurement session.
ABSTRACT

LIST OF FIGURES

LIST OF TABLES

TABLE OF TERMS AND ABBREVIATIONS

ACKNOWLEDGMENTS

CHAPTERS

1. INTRODUCTION

2. BACKGROUND

3. SYSTEM MODELING AND TEOAE STIMULI
METHODS

5. SUBBAND DECOMPOSITION BASED ARTIFACT REMOVAL

4. EVALUATION OF CURRENT STIMULUS ARTIFACT CANCELLATION METHODS

4.1 Stimulus Artifacts
   4.1.1 TEOAE Recordings
   4.1.2 Artifact Characteristics

4.2 Simulated TEOAE Recordings and Evaluation Criteria
   4.2.1 Simulated TEOAE Recordings
   4.2.2 Evaluation Criteria

4.3 Windowing Method

4.4 Scaled Subtraction Methods
   4.4.1 Description
   4.4.2 Evaluation

4.5 Adaptive Cancellation Methods

5. SUBBAND DECOMPOSITION BASED ARTIFACT REMOVAL

5.1 Wavelets and Wavelet Packets
   5.1.1 Motivation
   5.1.2 Wavelets
   5.1.3 Wavelet Packets

5.2 Application to Stimulus Artifact Removal
   5.2.1 Selection of a Wavelet Packet Family
   5.2.2 Selection of a Basis Set

5.3 Artifact Removal Procedure
   5.3.1 Raw Data Requirements
   5.3.2 Preprocessing, Basis Selection and Signal Decomposition
   5.3.3 Artifact Removal and TEOAE Preservation
   5.3.4 Reconstruction of the Final TEOAE Signal

5.4 Simulation Results

5.5 Experimental Results
   5.5.1 Raw Data and Preprocessing
   5.5.2 Obtaining the Final TEOAE Signals
   5.5.3 Comparison of Results from the Two Methods

5.6 Practical Implementation and Computational Time

6. CONCLUSIONS AND FUTURE WORK

6.1 Summary of Research

6.2 Directions for Future Work

APPENDIX A. SYSTEM MODEL

REFERENCES
<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.1</td>
<td>Simple schematic showing structure of outer and middle ear regions</td>
<td>5</td>
</tr>
<tr>
<td>2.2</td>
<td>Inner ear structure, showing the cochlea and the semicircular canals</td>
<td>7</td>
</tr>
<tr>
<td>2.3</td>
<td>Cross section through cochlea</td>
<td>8</td>
</tr>
<tr>
<td>2.4</td>
<td>Cross section through a single turn of the cochlea</td>
<td>9</td>
</tr>
<tr>
<td>2.5</td>
<td>Schematic of unrolled cochlea</td>
<td>10</td>
</tr>
<tr>
<td>2.6</td>
<td>Frequency sensitivity within the cochlea as a function of position</td>
<td>11</td>
</tr>
<tr>
<td>2.7</td>
<td>Average of spectrographs of 24 OAEs</td>
<td>14</td>
</tr>
<tr>
<td>2.8</td>
<td>Saturation of TEOAE amplitude with respect to stimulus amplitude</td>
<td>15</td>
</tr>
<tr>
<td>3.1</td>
<td>Overview of TEOAE measurement system</td>
<td>19</td>
</tr>
<tr>
<td>3.2</td>
<td>Signal flow through TEOAE measurement system and subject's ear</td>
<td>20</td>
</tr>
<tr>
<td>3.3</td>
<td>Simulated frequency response of the system at different points</td>
<td>21</td>
</tr>
<tr>
<td>3.4</td>
<td>Frequency responses of microphone and probe input tube</td>
<td>22</td>
</tr>
<tr>
<td>3.5</td>
<td>Measured values of ear drum acoustic impedance</td>
<td>23</td>
</tr>
<tr>
<td>3.6</td>
<td>Stimulus artifacts elicited from four different ears</td>
<td>24</td>
</tr>
<tr>
<td>3.7</td>
<td>Plot of model precision as a function of model order</td>
<td>28</td>
</tr>
<tr>
<td>3.8</td>
<td>Model accuracy vs. length of the parameter estimation noise sequence</td>
<td>29</td>
</tr>
<tr>
<td>3.9</td>
<td>Amplitude response of 31st order FIR filter used for shaping the final stimulus</td>
<td>30</td>
</tr>
<tr>
<td>3.10</td>
<td>Stimulus artifacts obtained with two different probes</td>
<td>31</td>
</tr>
<tr>
<td>4.1</td>
<td>TEOAE recording components and derivation of the final TEOAE signal</td>
<td>33</td>
</tr>
<tr>
<td>4.2</td>
<td>A typical TEOAE recording, and close up view</td>
<td>35</td>
</tr>
<tr>
<td>4.3</td>
<td>Estimated transfer function relationships between stimulus artifacts</td>
<td>37</td>
</tr>
<tr>
<td>4.4</td>
<td>Time and frequency domain representations of a simulated TEOAE</td>
<td>38</td>
</tr>
<tr>
<td>4.5</td>
<td>Windowing method of stimulus artifact removal</td>
<td>40</td>
</tr>
<tr>
<td>4.6</td>
<td>Window which was used for the windowing method</td>
<td>41</td>
</tr>
<tr>
<td>4.7</td>
<td>Simulated results from the windowing method</td>
<td>41</td>
</tr>
<tr>
<td>4.8</td>
<td>Scaled subtraction method of stimulus artifact cancellation</td>
<td>42</td>
</tr>
<tr>
<td>4.9</td>
<td>Results from scaled subtraction simulation</td>
<td>45</td>
</tr>
<tr>
<td>4.10</td>
<td>General structure of adaptive stimulus artifact cancellation methods</td>
<td>47</td>
</tr>
<tr>
<td>4.11</td>
<td>Normalized mean squared error for linear filters of different orders</td>
<td>47</td>
</tr>
</tbody>
</table>
Table 4-1. Power levels and SNRs for the scaled subtraction simulation .......................... 46
Table 5-1 Experimental results, comparing the SNRs of final TEOAE signals from scaled subtraction with those from the subband based method of artifact removal ..................... 67
<table>
<thead>
<tr>
<th>Term or Abbreviation</th>
<th>Meaning</th>
</tr>
</thead>
<tbody>
<tr>
<td>A/D</td>
<td>analog to digital conversion</td>
</tr>
<tr>
<td>D/A</td>
<td>digital to analog conversion</td>
</tr>
<tr>
<td>DNLR</td>
<td>derived non-linear response</td>
</tr>
<tr>
<td>OAE</td>
<td>otoacoustic emission</td>
</tr>
<tr>
<td>RMS</td>
<td>root mean square</td>
</tr>
<tr>
<td>SNR</td>
<td>signal to noise ratio</td>
</tr>
<tr>
<td>SPL</td>
<td>sound pressure level; sound pressure, relative to a reference level of 20 μPa</td>
</tr>
<tr>
<td>TEOAE</td>
<td>transiently evoked otoacoustic emission</td>
</tr>
<tr>
<td>Zwischocki coupler</td>
<td>a standard test cavity which has the acoustic impedance of an average human ear</td>
</tr>
</tbody>
</table>
Many thanks to Professor Hans Kunov for supervising my work. Also, many thanks to Poul Madsen for his patience in providing technical assistance with instrumentation hardware many different times.

Thanks to many in the institute, both staff and students, who’ve been helpful whenever I’ve needed them. In particular, my thanks to the acoustics lab guys - Dave, George, Nic, Peter, and Tony. I’ve enjoyed working with these guys, and I’ll miss them all a tonne. Thanks especially to Dave for his vital assistance in Labview and other computer related headaches, and to Nic for helping sort out a variety of hardware problems. Thanks to George for comic relief, to Tony for lots of proofreading help (and the Jays tickets!), and to Peter for tolerating my silly sense of humor and constant baseball chatter. (And thanks to nobody for giving me a fair chance in those worms matches!).

I’d like to thank my parents for their constant love and support. I thank God immensely for both of them. I also thank God for my church family, who’ve very graciously made me feel at home here in Toronto.

Most of all, I thank God through Jesus Christ my Lord for his constant care these past two years. Without him I can’t do anything, so for whatever I’ve accomplished the credit is entirely his. To him be all glory and honor forever!
1. INTRODUCTION

Twenty years ago David Kemp reported that our ears not only detect sounds - they also emit them (Kemp, 1978). Since then, otoacoustic emissions (OAEs) have sparked broad interest within the entire auditory community. Audiologists hope that they will fulfill their potential as the basis of an efficient, objective neonate hearing test suitable for universal screening. There is hope that they will become a helpful diagnostic tool for patients of any age. Researchers who attempt to model cochlear mechanics use them as a research tool to non-invasively test competing models.

This chapter provides a brief introduction to OAEs and outlines a current problem in the measurement of transiently evoked otoacoustic emissions (TEOAEs). Research objectives for the present work are defined. Finally, an outline of the chapters to follow is provided.

1.1 Otoacoustic Emissions

Of the different classes of OAEs currently known, the two which attract the most interest are distortion product otoacoustic emissions (DPOAEs) and TEOAEs. DPOAEs are evoked with two steady state sinusoidal stimuli. The largest resulting DPOAE for humans is a tone of frequency \(2f_1-f_2\), where \(f_1\) designates the lower of the two stimulus frequencies and \(f_2\) designates the higher. Although this type of emission is the focus of most current research and development efforts, TEOAEs continue to receive substantial attention as well. As their name implies, TEOAEs are evoked using brief stimuli, typically clicks or tone bursts. The response in humans is a low level, broadband, transient sound which begins about 2 ms after the stimulus is applied and typically lasts for between 15 and 30 ms.

No consensus has yet emerged concerning which of these two emission classes is better. It is likely that they will both play an important role for the foreseeable future. Compared to DPOAEs, TEOAEs have a potential advantage because they are broadband. This allows a wide range of frequencies to be tested at once. Sometimes, however, the narrow-banded nature of
results from conventional pure tone audiometry. Also, the signals involved in DPOAE measurements and DPOAE results are simpler than they are for TEOAEs. This can be an advantage when interpreting the measurements. On the other hand, it can be argued that the transient signals used as stimuli for TEOAEs are more similar to natural sounds and are therefore preferable to measurements based on pure tones, even though the results obtained are more complex. In any case, it is clear that further research is needed before the issues involved in the selection of an OAE class can be unequivocally settled.

1.2 The TEOAE Frequency Range Limitation Problem

As just mentioned, one of the potential advantages of TEOAEs is their broad-band characteristic. In theory, a single transient stimulus can test the entire frequency range of interest. In practice, however, TEOAEs are limited to frequencies of about 4 to 5 kHz and below. Since human hearing extends as high as 20 kHz and normal pure tone audiometry extends to 8 kHz, the limitation to 5 kHz and below constitutes a significant drawback to development of a TEOAE based hearing screening test.

The major reason for this frequency limitation is stimulus artifact contamination of TEOAE measurements. The stimulus artifact is a portion of the acoustic stimulus which is recorded at the same time as the TEOAE. The stimulus artifact only affects the early portion of the measurement, but this is the region which contains the higher frequency components of the TEOAE. Effective elimination or reduction of this artifact is therefore necessary before the frequency range of TEOAEs can be extended.

A second contributor to this frequency limitation is difficulty in providing adequate high frequency stimulation. This stems from three sources. Firstly, the frequency range of available miniature speakers is normally limited to around 6 or 7 kHz, although weaker output is possible up to 8 or 9 kHz. Secondly, the acoustics of a typical electroacoustic probe for TEOAE measurements cause the probe to act as a low pass filter. Thirdly, the middle ear's transfer function is bandpass. It peaks around 2 kHz and rolls off at higher frequencies so that less high frequency stimulus energy is conducted to the cochlea.
1.3 Research Objectives

Taken together, the above factors create a formidable barrier to the extension of TEOAEs into higher frequencies. However, the motivation for attempting such an extension is clear. In light of this, the objective of the present work is to extend the frequency range of TEOAE measurements to cover the entire audiometric range (up to 8 kHz). Under this broad objective, three sub-objectives can be defined. They are
1. to gain foundational understanding of the characteristics of key components of the measurement system, thereby understanding the origins of the stimulus artifact,
2. to find a way to provide adequate stimulation of all cochlear regions, including the higher frequency end of the audiometric range, and
3. to assess current approaches to stimulus artifact reduction with regard to their success at rendering the high frequency TEOAE components detectable, and to improve upon these methods.

1.4 Thesis Outline

Chapter 2 provides relevant physiological background, beginning with an overview of the human auditory system, then moving into a general description of OAEs and concluding with a more detailed discussion of TEOAEs. Chapter 3 discusses the characteristics of system components and presents a new way of designing TEOAE stimuli based on system identification. This provides a way to compensate for undesirable frequency response characteristics of the system. In Chapter 4, the stimulus artifact problem is introduced. Strengths and weaknesses of current methods of dealing with stimulus artifacts are assessed. In Chapter 5, a new method based on an adaptive wavelet packet based subband decomposition of the recorded signals to separate TEOAEs from stimulus artifacts is presented and tested. Conclusions and recommendations for continuing work in this area are presented in Chapter 6.
2. BACKGROUND

Basic understanding of the auditory system in general, and otoacoustic emissions in particular, is necessary when considering how to extend the range of TEOAE measurements. This chapter provides an overview of the auditory system with particular reference to those aspects which are necessary to understand something of the origins and properties of OAEs.

2.1 Anatomical Structure and Function

The human ear comprises three separate sections: the outer ear, the middle ear, and the inner ear. Incoming sounds enter the outer ear and are conducted through the middle ear to the inner ear, where they stimulate hair cells. Hair cells generate nervous signals which leave the ear via the auditory nerve and travel toward the auditory cortex in the brain. The entire system works to efficiently match acoustic impedances and keep energy dissipation low while avoiding serious distortion of the original acoustic signal. The entire ear is an amazing creation. Its dynamic range covers six orders of magnitude (20 μPa to 20 Pa), and its bandwidth spans nearly ten octaves (20 Hz to 20 kHz).

2.1.1 Outer Ear

The outer ear consists of the pinna and the ear canal (see Figure 2.1). The pinna is made mostly of cartilage, and the ear canal is mostly bone covered with skin. The ear canal is terminated by the ear drum, or tympanic membrane. The concha (inner central portion of pinna, at its junction with the ear canal) helps collect incoming sounds, funneling them into the ear canal. The outer ear and the head transform the sound field from outside in free space to the eardrum. This transformation depends on both the frequency and the direction of incidence of the sound. It plays an important role in sound localization (Sherwood, 1993, p. 181). Other than this, the major role of the outer ear is to direct sounds to the ear drum, causing it to vibrate and transmit the acoustic signal to the middle ear.
2.1.2 Middle Ear

The ear drum is a conical membrane, stiffened by circular and radial fibers. It is under inward tension because of its attachment to the malleus, the first in a chain of three bones (malleus, incus, and stapes) which conduct sound vibrations through the middle ear. The shape and construction of the ear drum along with the inner tension from the malleus help to stiffen the ear drum without increasing its weight.

In a normal ear, the ear drum completely covers the end of the ear canal, sealing it from the middle ear cavity. This is important in facilitating vibration of the ear drum in response to
The middle ear is an air-filled cavity, typically about 2 cm in volume. The Eustachian tube is a passageway from the middle ear to the nasopharynx. It opens during swallowing to allow ventilation and static pressure equalization between the middle ear and free space.

The three middle ear bones, also referred to as the auditory ossicles, are tightly attached to each other by ligaments. The free end of the stapes is attached to a membrane covering the oval window, an opening into the cochlea. There are two muscles in the middle ear as well, the tensor tympani and the stapedius. The tensor tympani contracts to draw the tympanic membrane further inward, while the stapedius contracts to pull the stapes upward (Møller, 1983, pp. 39-40). At sound pressure levels of approximately 65-70 dB SPL, the stapedius begins to contract in what is called the stapedius reflex. This stiffens the entire chain of ossicles so that it vibrates less easily, especially in the lower range of audio frequencies. As a result, less sound energy is conducted to the cochlea. This helps protect the fragile cochlear structures from prolonged overexposure to high sound pressure levels. For the present work, a significant consequence of this reflex is a change in the acoustic impedance seen when looking into the ear from the ear canal. This causes the acoustic impedance of the ear to vary as a function of incoming sound pressure levels.

The ear drum and the auditory ossicles act as an impedance matching transformer between the air of the ear canal and the fluid of the cochlea. The mechanical advantage is provided mostly by the change in area from the ear drum to the footplate of the stapes. Some additional advantage is gained through the lever action of the ossicles. These two factors provide pressure amplification by a factor of about twenty (Sherwood, 1993, p. 182). Acoustic energy is sent into the cochlea as the stapes moves with a piston-like motion, pulling in and out on the membrane of the cochlea’s oval window.

2.1.3 Inner Ear

2.1.3.1 Structure

The inner ear comprises the cochlea and the semicircular canals (see Figure 2.2). The cochlea relates to hearing and the semicircular canals relate to the vestibular system. This entire structure is located in the temporal bone of the skull, medial to the middle ear.
It has two membrane covered openings adjacent to the middle ear, referred to as the oval window and the round window. A cross section through one turn of the cochlea reveals three fluid filled chambers, the scala vestibuli, the scala media or cochlear duct, and the scala tympani, running longitudinally along the entire length of the cochlea (see Figure 2.3). It is easiest to discuss the internal structure of the cochlea by uncoiling it and referring to the end attached to the middle ear as the basal end, and the far end as the apical end. The oval window to which the stapes is attached is located at the basal end of the scala vestibuli, and the round window is located at the basal end of the scala tympani. These two chambers are actually continuous at the apical end of the cochlea, connected by a small opening called the helicotrema (visible in Figure 2.5). They share a fluid called perilymph which is very similar to extracellular fluid. In contrast, the scala media is isolated from the other two chambers and is filled with endolymph, which is similar to intracellular fluid.

There are three membranes running throughout the length of the cochlea. The vestibular membrane separates the scala media from the scala vestibuli, and the basilar membrane separates the scala media from the scala tympani (see Figure 2.3). These two membranes along with the scala media and everything else in between them are referred to as the cochlear partition because they divide the cochlea in half between the outer two chambers. A third membrane, the tectorial membrane, is located within the scala media between the other

Figure 2.2. Inner ear structure, showing the cochlea and the semicircular canals. Connection of the stapes to the oval window of the cochlea is also visible (adapted from Boys Town National Research Hospital, <http://www.boystown.org/cel/ear.jpg>, who adapted it from Compton’s Interactive Encyclopedia).
two membranes (see Figure 2.4). It covers the organ of Corti, which rests on top of the basilar membrane. These structures play a very important role in the transduction of mechanical energy into neural signals.

2.1.3.2 Mechanical Motion

As the stapes pushes in and out on the oval window, the perilymph within the scala vestibuli is forced to move back and forth. As the oval window is forced in and out by the stapes, the round window is simultaneously forced out and in by the perilymph in the scala tympani. This keeps the overall volume within the cochlea constant, as required by the incompressibility of the cochlear fluids. Through the helicotrema, perilymph can be exchanged between the scala vestibuli and the scala tympani as the force difference between the two windows causes it to vibrate back and forth.

Most of the fluid is not exchanged through the helicotrema, however. For sounds which are within the frequency range of human hearing, the entire cochlear partition is pushed downward against the scala tympani when the scala vestibuli is compressed (see Figure 2.5). In the same manner, when the scala vestibuli is expanded by outward motion of the oval window, the entire cochlear partition is pulled upward toward the scala vestibuli. In the same way that alternating current moves "through" a capacitor by charge being alternately stored on and removed from the two terminals, volume velocity moves through the cochlear partition as
volume is shifted back and forth between the scala vestibuli and the scala tympani by the up and down motion of the cochlear partition.

At its basal end, the basilar membrane is narrow and stiff; at its apical end it is broad and compliant (Sherwood, 1993, p. 184). This design makes its frequency response a function of position along its length, with the basal end tuned to high frequencies and the apical end to low frequencies (see Figure 2.6). When the stapes creates a pressure wave in the perilymph, the portions of the basilar membrane which are tuned to resonate at the frequencies present in the pressure wave vibrate with greater amplitude than the rest of the basilar membrane. As the basilar membrane vibrates up and down, the tectorial membrane shears across the top of the organ of Corti and stimulates hair cells which are embedded along its length. The quality of the basilar membrane's resonance is very high, providing very sharp tuning for sounds near the threshold of hearing.

2.1.3.3 Organ of Corti and Hair Cells

Hair cells are specialized sensory cells which are stimulated by mechanical deformation of the stereocilia ("hairs") located on their tips. In a normal cochlea, there are believed to be approximately 16000 hair cells embedded along the length of the organ of Corti (Allen and Neely, 1992). They are organized into one row of inner hair cells and three rows of outer hair cells (refer back to Figure 2.4). The three rows of outer hair cells are linked to each other, and
the stereocilia of one of the rows are embedded in the tectorial membrane. In contrast, the stereocilia of the inner hair cells do not directly contact the tectorial membrane. Instead, inner hair cells are moved indirectly by the fluid between themselves and the tectorial membrane.

Excitation of hair cells creates impulse activity of afferent nerve fibers. Of all the afferent fibers leaving the cochlea, 95% are from inner hair cells and only 5% are from outer hair cells (Møller, 1983, p. 72). This suggests a primary role other than that of sensory transduction for the outer hair cells. It is believed that this role is to alter the relative motion between the tectorial and basilar membranes, thereby changing the level of stimulation of the inner hair cells. This is consistent with the finding that outer hair cells have more efferent innervation than inner hair cells. It is also supported by evidence that outer hair cells contain actin filaments which allow their stiffness and/or length to be varied. Although the precise mechanism is still unknown, the generation of otoacoustic emissions is believed to be very closely related to the action of these outer hair cells.
2.2 Auditory Pathways

The auditory nervous system is very complex and its function is not yet fully understood. A complete description of the present state of understanding of this system is far beyond the scope of this work, so this description will therefore be relatively brief. The intention is simply to provide an appreciation of the effect that the neural pathways can have on otoacoustic emissions.

2.2.1 Ascending Pathways

Ascending auditory pathways are primarily concerned with signal transformation and the extraction of important features from the incoming signal. Axons from the afferent terminals on hair cells comprise the auditory nerve, a part of the eighth cranial nerve. They conduct signals to the cochlear nucleus, from which there are several pathways to the auditory cortex via a number of different nuclei. It is important to note that between the cochlear nucleus and the auditory cortex there are several nuclei which receive both ipsilateral and contralateral input. The lowest level at which this occurs is that of the superior olivary complex, which receives its inputs from both the ipsilateral and the contralateral cochlear nuclei (Møller, 1983, p. 109). Nuclei which receive inputs from both cochlear nuclei are significant because they provide pathways through which sound received by one ear can influence OAEs from the other ear.
The descending auditory system includes connections to all of the nuclei of the ascending pathways. These are generally believed to aid in the processing of sound information. Of greater interest for the present work is the finding that hair cells receive efferent innervation via the olivocochlear bundle, with outer hair cells receiving much more concentrated innervation than inner hair cells (Spoendlin, 1966; Tamar, 1972). Stimulation of this bundle has been shown to affect otoacoustic emissions (Siegel and Kim, 1982; Mountain, 1980), both ipsilaterally and contralaterally.

2.3 Otoacoustic Emissions

Otoacoustic emissions are defined as sounds which originate in the inner ear. They are believed to be present as a byproduct of normal cochlear function (Kemp et. al., 1986; Whitehead et. al., 1994). Although they are low level, they can normally be detected with a sensitive microphone placed in the ear canal. OAEs fall into a number of different categories, depending on how they are obtained. Broadly speaking, OAEs are considered to be either evoked or spontaneous. Evoked emissions are further classified by stimulus type as either stimulus frequency OAEs, distortion product OAEs, or transiently evoked OAEs.

2.3.1 Spontaneous OAEs

Spontaneous OAEs are produced without the application of any external stimulus. They consist of one or more narrow band sounds, usually between -20 and 20 dB SPL, which are continuously present in the ear canal. It is believed that these emissions are present in about half of all normal ears (Kemp, Ryan, and Bray, 1990; Lonsbury-Martin, Whitehead, and Martin, 1991). This figure could eventually turn out to be much higher however, since the reported prevalence of spontaneous OAEs has steadily increased over time due to improved measurement equipment and methods.

2.3.2 Stimulus Frequency OAEs

Stimulus frequency otoacoustic emissions are evoked by a single pure tone stimulus, usually of low amplitude. This stimulus is usually swept slowly over a range of frequencies (Probst, Lonsbury-Martin, and Martin, 1991). The emission is of the same frequency as the
2.3.3 Distortion Product OAEs

Distortion product OAEs are evoked using two pure tones (designated \( f_1 \) and \( f_2 \)) which are presented simultaneously to the ear. These tones evoke a cochlear response which consists of a number of pure tones at the intermodulation frequencies \( m*f_1-n*f_2 \). The intermodulation tones are named distortion product OAEs because they appear to be the result of nonlinear distortion of the stimulus tones. In humans, the most prominent of these tones is generally the one of frequency \( 2f_1-f_2 \), where \( f_1 \) designates the stimulus tone of lower frequency. A tone of this frequency would be expected from a system with cubic distortion so this is usually called the cubic distortion product.

There is a lot of ongoing study of cubic distortion product OAEs in humans, both for theoretical investigation of hearing mechanisms and for clinical diagnosis. They constitute one of the two major classes of OAEs which are considered to be of great importance. The other major class is that of TEOAEs, the subject of the present work.

2.3.4 Transiently Evoked OAEs

The term transiently evoked otoacoustic emission describes any OAE evoked using a transient stimulus. This stimulus is usually broadband, typically either a click, a tone burst, or possibly a chirp. TEOAEs were the first type of OAE to be discovered, and were first reported by David Kemp in 1978 (Kemp, 1978). Since then a wide body of literature describing various properties of these emissions has developed. No model of their generation mechanism has yet gained wide acceptance, but a number of their qualitative characteristics are now well established.

2.3.4.1 Frequency Content

The frequency content of a TEOAE is limited by the frequency content of the stimulus (Prieve, Gorga, and Neely, 1996). That is, there are not normally any significant emission components at frequencies outside of the frequency band of the stimulus. The emission spectrum is also dependent on the mechanical structure of the individual ear because the middle ear's frequency response affects both the forward transmission of the stimulus and the reverse
methods, TEOAEs are usually limited to between 0.5 kHz and 4 kHz, although this range may extend as high as between 5 and 6 kHz for a few subjects.

2.3.4.2 Time Duration

The duration of TEOAEs is highly variable from ear to ear. In most subjects the emission lasts at least 20 ms, but for approximately a quarter of normal ears, the emission completely dies out in less time than this (Probst et al., 1991). Long lasting TEOAEs typically have one or several very dominant frequency components which decay very slowly.

2.3.4.3 Time-Frequency Structure

From the point in time at which the emission is distinguishable from the stimulus artifact until the point in time at which the emission dies out, a definite time-frequency structure is observed (see Figure 2.7). Higher frequency components in the OAE have shorter duration and occur earlier in the response than lower frequency components. Although this phenomena was observed and commented upon by Kemp (Kemp, 1978), it was investigated and displayed in more detail in work done by Peter Bray in the late 1980s (Bray, 1989).

Bandpass filtering a TEOAE produces a tone burst. By using a number of different bandpass filters with different center frequencies, the location in time of the TEOAE's various frequency components can be observed. The time delay between the application of a stimulus and the center of the bandpass filtered OAE's envelope at a given frequency is referred to as the

![Figure 2.7. Average of spectrographs of 24 OAEs (adapted from Bray, 1989, p. 79).](image)

Figure 2.7. Average of spectrographs of 24 OAEs (adapted from Bray, 1989, p. 79).
Bray's work demonstrated that latency is inversely proportional to the logarithm of a component's frequency for TEOAEs. This relationship appears to be related to the physical layout of the cochlea. The mechanical resonance properties of the basilar membrane create a logarithmically spaced place-frequency map with the highest frequencies located at the base (oval window) of the cochlea and lowest frequencies at the apex of the cochlea. Another factor which affects the latency is the amplitude of the stimulus. Louder stimuli result in lower latencies than quieter stimuli (Neely, Norton, Gorga, and Jestead, 1988).

2.3.4.4 Non-linear Growth of Response with Respect to Stimulus Amplitude

The relationship between the amplitude of a transient stimulus and the amplitude of the corresponding TEOAE shows saturation (see Figure 2.8). At very low stimulus levels (up to about 40 dB SPL peak), there is a nearly linear relationship between stimulus and response amplitudes. Above this, response growth exhibits marked saturation behavior. At high stimulus levels (approximately 85 dB SPL peak), a typical OAE grows only 2 or 3 dB for a 10 dB increase in stimulus level (Bray, 1989).

2.3.4.5 Polarity

It was established by Kemp that the polarity of the evoked response is related to the polarity of the stimulus (Kemp, 1978). Inversion of the stimulus results in inversion of the resulting OAE.
It has been found that spontaneous OAEs affect TEOAE recordings (Kilawiec and Orlando, 1995). This happens despite the fact that many TEOAEs are averaged together to reduce the noise in the final TEOAE recording. The reason appears to be that the spontaneous OAEs of the ear being tested synchronize with the stimulus signals and thus they remain in the final recording rather than being averaged away. This effect can add significantly to the spectrum of the OAE at the frequency of the spontaneous emission.

2.3.4.7 Masking

There are several ways in which TEOAEs can be altered by using other sounds, referred to as maskers. Masking is a well-known phenomenon in psychoacoustics in which certain sounds (tones, or noise) are used to prevent or reduce the perception of other sounds. This can take place due to mechanical interactions on the basilar membrane or to neural interaction in the auditory pathways. It has been shown that the generation of both DPOAEs (Brown and Kemp, 1984; Williams and Brown, 1995) and TEOAEs (Veuillet, Duverdy-Bertholon, and Collet, 1996; Kemp et. al., 1986) can be suppressed. Ipsilateral suppression is more effective than contralateral suppression because it utilizes both mechanical and neural interactions whereas contralateral suppression is dependent on neural effects alone. Suppression can virtually eliminate transiently evoked emissions in the ipsilateral ear, but can only reduce emissions in the contralateral ear by a few dB.

2.3.4.8 Vulnerability

Normally, OAEs do not vary substantially over time (Franklin et. al., 1992). However, OAEs are vulnerable to a number of changes in the cochlea. The common characteristic of all of these changes is that they affect the function of the outer hair cells. The physiological vulnerability of OAEs provides some of the strongest evidence that they are of cochlear origin, and that their production is directly related to the health of outer hair cells.

2.3.4.9 Mechanism of Generation

There is not yet any consensus on the question of the precise origin of OAEs. Several different models have been proposed, each with its own strengths and weaknesses, as a means of explaining the currently available data. Certain points, however, are now becoming clear. A
focused in the region of the basilar membrane which is associated with that particular frequency (Probst et al., 1991). Furthermore, in many instances, time and frequency superposition holds (Prieve et al., 1996). The characteristics of a click response (broadband) can be approximately predicted by combining the characteristics of several tone burst responses (narrow-band) which cover the same range of frequencies as covered by the click response. This suggests that the mechanisms behind click evoked and tone burst evoked transient otoacoustic emissions are very similar. In contrast to this, there is some evidence that the mechanisms behind distortion product emissions and transiently evoked emissions are slightly different, although both appear to be directly related to the outer hair cells (Whitehead, Lonsbury-Martin, and Martin, 1992).

### 2.3.5 Significance of Otoacoustic Emissions

Soon after they were discovered, the clinical potential of OAEs was recognized. There is significant correlation between normal hearing and the presence of normal otoacoustic emissions (Prieve et al., 1993). Because hearing testing based on OAEs is objective and quick, a number of areas for its use are being explored (Whitehead et al., 1994; Kemp et al., 1986). In the United States, the National Institutes of Health have recommended the standardization of neonate hearing screening using OAEs. Besides this, OAE testing has potential as a diagnostic aid when attempting to determine the precise cause of sensorineural hearing loss because OAEs are specific to the cochlea, and to outer hair cells in particular. A third possible use is as a cost effective monitor of hearing changes over time for workers in hazardous noise environments or for patients undergoing treatment involving drugs with ototoxic side effects.

OAEs are also useful as a research tool. They allow noninvasive investigation into certain aspects of cochlear function which would be very difficult or impossible otherwise. No model of the cochlea and hearing transduction is yet satisfactory, and in the ongoing research and debate over various classes of models, evidence provided by OAEs has played a significant role in recent years.
3. SYSTEM MODELING AND TEOAE STIMULI

Specialized equipment is necessary to elicit and record otoacoustic emissions. This equipment, along with the subject’s ear, comprises a complex system which affects both the signals sent into the ear and the signals recorded from the ear. An understanding of this system is helpful when considering how to stimulate the ear and how to process and interpret the resulting recordings. This chapter summarizes some basic characteristics of the system and discusses some consequent issues involved in stimulus selection. Finally, this chapter presents a new way of generating a transient stimulus which partially compensates for undesirable system characteristics.

3.1 System Overview

The entire system is depicted in Figure 3.1. A computer serves as an interface to data acquisition and signal conditioning hardware. From this hardware, electrical signals are sent to a miniature hearing aid speaker in an electroacoustic probe. The probe is inserted into a subject’s external ear canal where its speaker produces sound to evoke otoacoustic emissions from the subject’s cochlea. The probe also contains a miniature microphone to record sound in the ear canal. Figure 3.2 shows how a stimulus signal moves through this system and results in a recording which contains both a stimulus artifact and an otoacoustic emission.

3.1.1 Speaker and Output Path of Electroacoustic Probe

The speaker’s acoustic output is characterized in terms of sound pressure and volume velocity. It would be convenient to represent the speaker as an ideal source of either sound pressure or volume velocity. However, this is not possible because the source impedance of the speaker is comparable in size to the impedance of the acoustic load presented to the speaker by the probe casing and the ear. This makes any transfer function characterization of the speaker valid only for the particular value of external impedance which was used to obtain the transfer
acoustic frequency response of the speaker can only be approximately predicted without knowing each particular subject’s ear canal acoustics.

This situation is completely analogous to the situation in hearing aid development and fitting as described by Gilman et. al. (Gilman, Dirks, and Stern, 1981). There are interactions between the speaker and the ear such that their individual resonant peaks affect each other. The precise form of the receiver frequency response depends on the probe in which it is housed, the ear into which the probe is inserted, and the way in which it is inserted. Figure 3.3 shows the frequency response of the speaker as predicted by an equivalent electrical circuit model using ear drum and ear canal parameters for a commonly used artificial ear. (An explanation of the derivation of this equivalent circuit is available in Appendix A.) This figure illustrates how severely the frequency responses of the speaker and the probe output path affect the stimulus

Figure 3.1 Overview of TEOAE measurement system. (Picture of ear adapted from Sherwood, 1993, p. 178.)
sound pressure spectrum at the eardrum. Different ear canal dimensions and a different ear drum impedance would change the position and sharpness of the spectral peaks.

The results shown in Figure 3.3 are typical for the equipment used in this work. Below 1 kHz the system response rises at a rate of 20 to 40 dB/decade. Between 1 and 2 kHz the spectrum is dominated by a broad peak which is created primarily by the miniature speaker. Above this, in the 2 to 5 kHz range, the spectrum is heavily dependent on the particular ear and the way the probe is inserted into the ear canal. It is usual to have a resonant peak in this region, but its location and sharpness are highly variable. At high frequencies the spectrum falls off rapidly because of the limited dynamic range of the speaker and because of the low pass characteristics of the narrow acoustic passages within the probe.

3.1.2 Microphone and Input Path of Electroacoustic Probe

A second effect of the probe is the way in which its input path acoustics and microphone affect the spectrum of the final recorded signal. Ideally the transfer function
between sound pressure level at the probe tip and voltage out of the microphone would be flat over the entire frequency range of interest. In reality, there are several items which prevent this.

Firstly, as shown in Figure 3.4, the frequency response of the microphone itself is not completely flat. Rather, it slowly rises by about 10 dB between 100 Hz and 5 kHz, and then falls off sharply beyond about 7 kHz. Secondly, sound arrives at the microphone through a long, thin tube. Such a tube acts as an acoustic low pass filter. Finally, transmission line effects are present in this tube at frequencies beyond 2 to 3 kHz. The bottom plot in Figure 3.4 shows the overall effect of these factors in distorting the spectrum of the signal between the probe tip and the microphone output.
Figure 3.4 Frequency responses of microphone (Knowles EM 4068), probe input tube carrying sound from the ear canal to the microphone, and overall from the probe tip sound pressure to the microphone output voltage.

3.1.3 Ear Canal, Ear Drum, and Middle Ear

For frequencies below 10 kHz the ear canal can be modeled as a cylindrical tube down which uniform plane waves propagate (Siegel, 1994). At very low frequencies the occluded ear canal can be represented simply as an enclosed volume. Depending on the insertion depth of the probe, this representation has significant modeling error for signals above several kilohertz because of transmission line effects. For adults, the ear canal can be considered to be a lossless uniform transmission line, while for the softer canals of infants, studies have indicated that the model should include lossiness (Bulen et. al., 1993). The cross-sectional area of the canal mainly affects the overall sound pressure level at the eardrum, and has little effect on the shape of the frequency spectrum (Kates, 1988).

The transmission line representing the ear canal is terminated by an impedance representing the ear drum, middle ear, and cochlea. The value of this impedance varies from
When considering the effect of the system on signals traveling to and from the cochlea, it is useful to have some understanding of the forward and reverse transmission properties of the middle ear. It has been well established that middle ear impedance affects TEOAE measurements (Pröschel and Eysholdt, 1993). For forward transmission, the middle ear acts approximately like a bandpass system with a peak between 1 and 2 kHz, and an upper limit of 10-20 kHz (Hawkins et. al., 1996, p. 43). For reverse transmission, the middle ear affects low frequency TEOAE components (around 1 kHz) in approximately the same way that it affects low frequency stimulus components in forward transmission (Bray, 1989). It is not completely clear what the middle ear’s reverse transmission characteristics are at higher frequencies, but it seems likely that they are similar to the forward transmission characteristics.

3.2 Conventional Stimuli and Stimulus Artifacts

There are several different stimuli which have commonly been used to elicit TEOAEs. In this discussion, unless otherwise indicated, the term “stimulus” refers to the electrical signal which is sent to the speaker in the electroacoustic probe. This should not be confused with the actual acoustic stimulus signal produced by the probe and sent into the ear to evoke an emission. The acoustic stimulus is a function of the electrical stimulus signal in combination with the speaker, probe, and ear system response, whereas the electrical stimulus signal can be

![Figure 3.5 Measured values of ear drum acoustic impedance. (Adapted from Zwislocki, 1980.)(23)]
3.2.1 Broadband Stimuli: Clicks

Brief rectangular pulses (clicks) are very often used as stimuli. They are intended to approximate impulses so that a very brief stimulus can cover a broad frequency range. In this way, a broad region of the cochlea can be stimulated acoustically with a single stimulus. There are at least two drawbacks to this type of stimulus. One is that the amount of energy stimulating any one region of the cochlea is only a small portion of the entire energy of the stimulus. This means that in order to keep the peak level of the stimulus down at a safe level, the energy delivered to any individual region of the cochlea must be lower than would be

![Amplitude Spectrums](image1)

![Time Waveforms](image2)

![Logarithmic Scale - Uncalibrated Units](image3)

![Linear Scale - Uncalibrated Units](image4)

![Frequency (Hz)](image5)

![Time (ms)](image6)

Figure 3.6. Stimulus artifacts elicited from four different ears. All four artifacts shown were obtained using a 0.2 ms pulse as the electrical stimulus signal. Note the variability, especially at higher frequencies, due entirely to the variation of acoustic impedance presented to the probe by different ears. (In this figure, the vertical axis units are not calibrated.)
However, is the variability of the type of stimulation between different ears, and between probes. Different ears and different probes have different acoustic impedances, and the output of the probe speaker depends on the acoustic impedance presented to it. The result is that the same pulse input can yield different acoustic stimuli in two different ears (see Figure 3.6) or when using two different probes (see, for instance, Probe A and Probe B results in Figure 3.10).

3.2.2 Narrowband Transient Stimuli: Tone Pips and Bursts

Tone pips and tone bursts are also used to elicit TEOAEs. A tone burst is a signal of a single frequency modulated by a window, usually either rectangular or sinusoidal. For TEOAEs the tone burst window length is typically several milliseconds. A tone pip is distinguished from a burst by its duration, a pip typically being much shorter than a burst. Both of these signals focus the stimulus on a limited region of the cochlea. Because of its shorter duration, a pip covers a broader frequency band than a burst. Besides this there is no difference between them. These stimuli are appropriate when emissions within a limited band are desired. However, to evoke emissions from the entire cochlea several tests must be performed using bursts or pips of several different center frequencies.

3.3 System Identification and Stimulus Design

In the present work, a new type of TEOAE stimulus has been developed and utilized. Using a simple system identification scheme, a stimulus is designed each time a new testing session starts. The objective of this is to produce an easily controlled amplitude spectrum over the entire frequency range of interest. As mentioned above, the speaker's acoustic output for an electrical click stimulus is heavily dependent on the acoustic impedance presented to the probe speaker. The result of this is unpredictable, uneven stimulus coverage of the frequency range being tested. Low frequencies typically receive much more stimulus energy than higher frequencies, and there is often a deep null somewhere in the spectrum (see Figure 3.6). These characteristics are reflected in the otoacoustic emission, so they affect the quality of any test based on this emission. This is the problem that motivates the search for a better stimulus to evoke TEOAEs.
There are several desirable characteristics for a good transient stimulus design method. The method should be simple and quick enough that it can be of practical use on human subjects. It should be robust and automatic so that operator intervention is not required. The stimuli produced by this method should be highly localized in time to truly be considered transient, and they should provide even energy coverage over the entire frequency range of the test. The method described below has such properties.

3.3.2 System Model and Parameter Estimation

It is necessary to model the system in order to design an appropriate stimulus. Because the system's characteristics depend crucially on the insertion of a particular probe into a particular ear, the model's parameters cannot be properly estimated until the probe is in the particular ear of interest and the test is about to begin. At this point a burst of noise can be sent into the system and the system's response can be recorded through the microphone in the electroacoustic probe. This information can be used to estimate the parameter values of a system model. This model can then be used to design an appropriate stimulus. The entire process should be brief to avoid discomfort to the subject and to minimize the likelihood of subject motion which would affect the parameter values.

A suitable model structure is an all-pole model with a constant numerator term. Parameter estimation and stimulus design are both straightforward for this model, and it gives good results when it is implemented in a real system.

Parameter estimation is accomplished as follows. Designating the input as \( u(k) \) and the output as \( y(k) \), the all-pole model corresponds to the discrete time difference equation

\[
y(k) = bu(k - L) - a_1 y(k - 1) - a_2 y(k - 2) - \ldots - a_N y(k - N),
\]

where \( b \) and \( a_i \) through \( a_N \) are parameters to be estimated, \( L \) is a delay between input and output for the system, and \( N \) is the model order (the number of \( a_i \) coefficients). The parameters can be estimated by recording the system response to a white noise input (Ljung, 1987, p. 373). If we use a sequence of length \( K+1 \), indexed from 0 to \( K \), and let

\[
Y = \begin{bmatrix} y(N) & y(N+1) & \cdots & y(K) \end{bmatrix}^T,
\]

\[
\theta = \begin{bmatrix} a_1 & a_2 & \cdots & a_N & b \end{bmatrix}^T,
\]

26
The parameter estimates can be obtained from the least squares solution to this set of equations, which is (Ljung, 1987, p. 464)

\[
\theta = (\phi^T \phi)^{-1} \phi^T Y .
\] (3.6)

With these parameters, the stimulus is obtained in the following way. In the z-domain the model corresponds to

\[
A(z)Y(z) = b z^{-L} U(z)
\] (3.7)

where

\[
A(z) = 1 + a_1 z^{-1} + a_2 z^{-2} + \ldots + a_N z^{-N} .
\] (3.8)

Designating the system model as \(H(z)\), the model transfer function is

\[
H(z) = \frac{b z^{-L}}{A(z)} .
\] (3.9)

The stimulus \(S(z)\) designed from this model is the inverse of \(H(z)\) without the delay. That is,

\[
S(z) = \frac{A(z)}{b} .
\] (3.10)

This stimulus will compensate for unevenness in the frequency response of the system. In the time domain the stimulus sequence is the inverse z-transform of \(S(z)\), namely

\[
s(k) = \frac{1}{b} \frac{a_1}{b}, \frac{a_2}{b}, \ldots, \frac{a_N}{b} \quad k = 0 \ldots N .
\] (3.11)

This model is convenient because it avoids questions of zeros in the numerator of the transfer function and the consequent issue of whether or not the system is minimum phase. Since the stimulus is the inverse of the system model, any zeros in the model become poles in the stimulus. If the model is not minimum phase so that all of its zeros are inside the unit circle

\[
\phi = \begin{bmatrix}
- y(N - 1) & \ldots & - y(0) & u(N - L) \\
- y(N) & \ldots & - y(1) & u(N + 1 - L) \\
\vdots & \ddots & \vdots & \vdots \\
- y(K - 1) & \ldots & - y(K - N) & u(K - L)
\end{bmatrix},
\]
is guaranteed because with no zeros in the numerator of the model, the inverse is always an FIR system.

The stimulus designed from this model does not need to use the estimated value of the \( b \) parameter. The system being modeled includes amplifiers in both the output path to the probe speaker and the input path from the microphone. The estimated value of \( b \) depends on the settings of these amplifiers. It is more appropriate to scale the sequence \( s(k) \) so that it uses the full dynamic range of the system's digital to analog converter. The amplifier settings then can be used to govern the actual amplitude of the electrical signal, thereby controlling the sound pressure level generated in the ear canal.

To implement this stimulus design procedure, it is necessary to choose values for the model order, \( N \), and the length of the noise sequence, \( K \). To choose a value for \( N \), the response of the system to a long sequence of noise (20000 samples at a 22.05 kHz sampling rate) was obtained. Models were estimated for model orders from 1 to 50. The cost function

\[
V(\theta) = \frac{Y^T (Y - \phi \theta)}{K - N} (1 + 0.01N)
\]

Figure 3.7 Plot of model precision as a function of model order. The most efficient values of \( N \) are in the region of \( N=15 \).
by estimating the correlation between the model's prediction errors and the $y(x)$ values the model is trying to predict. For a good model whose parameters have been properly estimated, this correlation should be very small. In addition, this function incorporates a factor which increases the cost slightly for increasing model orders. The plot suggests a model order in the vicinity of $N=15$. Beyond this there are diminishing returns for increased model size.

The number of noise values which are needed to estimate the parameters depends on the desired accuracy of estimation. To gain insight into the relationship between the accuracy of the model parameters and the number of noise points, the model order was fixed at $N=15$ and the number of noise samples was varied. For different numbers of noise samples, the transfer function amplitude spectrum of the estimated model was calculated. The standard deviation of the amplitude spectrum of each estimated model was then estimated. Figure 3.8 shows the maximum standard deviation of the transfer function spectrum of the estimated model as a function of the number of noise samples which are used for the estimate. There are diminishing returns beyond about 10000 noise samples. Since the number of noise samples affects the computational time, it may be better to accept higher parameter estimation errors than to spend extra computational time. A detailed study of this question is beyond the scope and purpose of the present work. For the present needs, it is clear that a 15th order model estimated using between 2000 and 10000 noise samples is adequate.

![Figure 3.8](image.png)

Figure 3.8 Relationship between model accuracy and length of the noise sequence used for estimating the model parameters.
The electrical stimulus signal designed by the above method is sufficient to produce a stimulus with a flat amplitude spectrum. However, because the frequency range of the probe speakers is limited to about 8 kHz, any electrical stimulus which is designed to produce a flat spectrum over frequencies above 8 kHz would need to be amplified greatly to produce usable sound pressure levels. Such voltage levels cannot be used because they would damage the speaker. Instead, it is necessary to filter the designed stimulus signal to limit its frequency content to the frequency range of the speaker.

It is not necessary to use a filter with a completely flat passband response to do this. The final goal of the stimulus design procedure is to provide even stimulation of the entire cochlea. If the middle ear's low pass transfer function is taken into account, it becomes necessary to emphasize the higher frequencies in the stimulus over the lower frequencies. Since it is straightforward to design FIR filters with arbitrary frequency responses, the same filter that removes frequency components outside of the speaker's dynamic range can be used to emphasize the frequencies in the upper region of the audiometric range.

A 31st order FIR filter was used for the present work. It was designed with the window method, using a hamming window. Frequencies between 5 and 7.5 kHz were emphasized approximately 8 dB more than frequencies below 4 kHz. The region between 4 and 5 kHz was left as a transition region. The filter was designed to sharply attenuate all frequencies above 7.75 kHz. These parameters were chosen to amplify high frequency components as much as possible, given the constraints imposed by the miniature speaker, by matching the high

![Figure 3.9 Amplitude response of 31st order FIR filter which was used to shape the final stimulus for the stimulus design procedure described in the text.]
These modifications do not guarantee evenly distributed cochlear stimulation. Each subject has a different middle ear transfer function and it is very difficult to know exactly how to compensate properly for each individual. Rather, these modifications provide a way of improving the stimulation, on average. They can help in the production of adequate sound pressure levels throughout the audiometric range and if desired, they can simultaneously make approximate compensation for the middle ear’s band pass characteristic. This can be helpful in

Figure 3.10. Stimulus artifacts obtained with two different probes, each inserted normally into a human ear. Probe positions were kept the same between recordings of the click waveforms and recordings of designed stimulus waveforms. These designed stimulus results were obtained with 15th order A(z) polynomials in the model. Parameters were estimated using the system’s response to 100 ms of white noise (22.05 kHz sampling rate). A 31st order FIR filter was used to remove frequencies above 8 kHz in the designed stimuli.
3.3.4 Results

A brief sampling of typical stimulus results which can be obtained with the method described above is presented here. Figure 3.10 illustrates the ability of this method to flatten the spectrum of the acoustic stimulus. These stimuli have been designed to provide equal energy over all frequencies within the audiometric range. For the results shown the spectra for the designed stimuli varied within a 7 dB range between 250 Hz and 7.5 kHz. (Above this, the frequency range of the probe speaker limited the output.) Within this same frequency range a conventional click stimulus resulted in a stimulus spectrum which varied more than 40 dB for probe A, and 50 dB for probe B. In short, extensive experimental use of this method has confirmed that it is very useful for controlling the stimulus.
4. EVALUATION OF CURRENT STIMULUS ARTIFACT CANCELLATION METHODS

The presence of a stimulus artifact in the recorded signal is a very significant reason why the frequency range of TEOAE measurements is limited. In this chapter, the stimulus artifact problem is described in detail, and a way of evaluating stimulus artifact removal algorithms using simulated TEOAE recordings is developed and discussed. Three families of algorithms are described, and the strengths and weaknesses of each are assessed.

4.1 Stimulus Artifacts

4.1.1 TEOAE Recordings

TEOAE recordings are made by snugly inserting an electroacoustic probe into the ear canal and recording the sounds in the ear canal as transient stimuli are applied. To obtain the TEOAE for a specific stimulus, the stimulus is repeated many times at a rate of about 50/second. The recorded signal is then cut into equal length sections with one stimulus and response per section, and all of the sections are averaged together. This “synchronous averaging” reduces the background noise power of the recorded TEOAE signal by a factor of $1/n$, where $n$ represents the number of stimulus/response sections which are averaged (Challis

![Figure 4.1 Components of a TEOAE recording and derivation of the final TEOAE signal.](image-url)
The stimulus artifact is a portion of the acoustic stimulus which is picked up by the electroacoustic probe microphone and recorded along with the TEOAE (see Figure 4.1). The term “stimulus artifact” refers to contamination which is present in the recorded signal because of the acoustic stimulus. The acoustic stimulus itself is really the sound pressure which is applied at the ear drum. The ear canal, the probe, and the microphone all contribute to distortion of the ear drum sound pressure signal before it is recorded. Despite this distinction between the stimulus and the stimulus artifact, it is often assumed that the stimulus artifact is representative of the acoustic stimulus. For instance, when the maker of a prominent TEOAE measuring system refers to a stimulus with a flat spectrum up to 5 kHz (Kemp, Ryan, and Bray, 1990), he is actually referring to the stimulus artifact.

Strictly speaking, the actual acoustic stimulus is rarely, if ever, measured. Only stimulus artifacts are measured because to truly measure the stimulus, a probe would have to be inserted all the way to the ear drum. This is not practical. For practical probe insertion depths, transmission line effects in the ear canal make the sound pressure at the probe tip different from the sound pressure at the ear drum.

4.1.2 Artifact Characteristics

Since the stimulus artifact is the entire system’s response to the electrical stimulus, the shape of the stimulus artifact depends on the entire system’s characteristics. As a consequence, stimulus artifacts are highly variable, depending on the ear drum impedance, ear canal geometry, and probe position. Despite this, it is possible to outline a few general characteristics. The artifact is normally a brief ring containing a broad band of frequencies. There are normally a few frequency components which are more dominant than the rest, but the spectral peaks are broad.

4.1.2.1 Amplitude and decay of the stimulus artifact

The peak amplitude of the stimulus artifact is typically on the order of 100 to 1000 times as great as the peak amplitude of the TEOAE (see Figure 4.2). However, the artifact decays quickly and soon falls below the level of the TEOAE and background noise. This usually happens between 3 and 10 ms after the application of the stimulus, depending on the
system's characteristics and the amplitude of the stimulus. The highest frequency components of the artifact decay most rapidly, so the later portion of the artifact is usually dominated by frequencies in the 1 to 2 kHz range which decay relatively slowly.

4.1.2.2 Variation with stimulus amplitude

The growth of the stimulus artifact with respect to the amplitude of the stimulus is almost linear. It would be completely linear except that the stimulus artifact's shape changes slightly when the stimulus amplitude is changed. The most likely explanation for this is the effect of the stapedius reflex. Recall that the stapedius reflex makes the middle ear's impedance vary as a function of the stimulus amplitude (section 2.1.2, p. 5). The action of the stapedius reflex is slow compared with the rise and fall times of the transient stimuli used for the present work. The onset time of the reflex is approximately 30 to 40 milliseconds, and it takes several hundred milliseconds to reach steady state (Møller, 1983, p. 51). The offset time is approximately double the onset time. Thus, the middle ear does not react to the instantaneous amplitude variation of each individual stimulus. Instead, it reacts to the overall, average level over several stimuli. This means that during the application of a series of stimuli
However, when a second series of stimuli of a different amplitude is applied a few seconds later, the middle ear impedance for the second series is different than it was during the first series. Hence, the stimulus artifact is altered slightly.

This explanation is supported by the data displayed in Figure 4.3, which shows that there is often much more irregularity and variation in artifact shape in recordings made on human ears than in those made on an artificial ear. The artificial ear does not mimic the stapedius reflex, so it is expected to behave very linearly. The figure shows the estimated transfer function which relates a subject’s 67 dB SPL peak stimulus artifact to the same subject’s 77 dB SPL peak stimulus artifact. Since the recordings are normalized so that their peak value is one, the transfer function between these two stimulus artifacts would be 0 dB across all frequencies for a perfectly linear system. Thus, repeatable variations from the 0 dB line indicate non-linear behavior of the system.

Such variations are obvious for most of the plots in Figure 4.3. The Zwislocki coupler results show evidence of a small non-linearity in the region of 1.5 to 2 kHz, but for all other frequencies the results indicate a very linear system. Some of the human subjects show much more non-linearity than others, but all show some degree of non-linearity, and most show more non-linearity than the coupler. The sharp peaks and valleys which often appear right next to each other are worthy of special notice, since they probably indicate a slight shift in a resonant frequency. This would be expected when the mechanical properties of the middle ear ossicular chain change under the influence of the stapedius reflex.

4.2 Simulated TEOAE Recordings and Evaluation Criteria

When considering ways to remove the stimulus artifact, a means of evaluating different methods is needed. Since TEOAEs are different for different ears, there is no such thing as a “standard” TEOAE. This means that for measurements made on real ears, it is not possible to know exactly what the TEOAE should look like after removal of the stimulus artifact. Without knowing exactly what the TEOAE should be, there is no precise way to evaluate the method which is being used to remove the artifact.
One way to circumvent this problem is to create simulated TEOAE recordings. This can be done by adding a simulated TEOAE to a stimulus artifact recorded from an artificial ear.

Figure 4.3 Estimated transfer function relationship between a 67 dB SPL peak stimulus artifact and a 77 dB SPL peak stimulus artifact for five subjects and the Zwislocki coupler. All of the recordings were normalized to a peak stimulus artifact level of 1 V. The upper five subject plots are right ears, and the lower five are left ears. The two plots in each box represent two trials under exactly the same conditions so that noise effects in the transfer function estimates can be assessed. For further explanation of these results, see text.
is to create signals which mimic the time-frequency and amplitude growth characteristics of TEOAEs. This can be done by using available data for TEOAE components up to 5 kHz, and extrapolating the trend in this data up to higher frequencies. In this way, simulated TEOAEs can be created to cover any desired frequency range. This method was found to be adequate for the purposes of the present work, so it was used.

The results from Bray’s work (Bray, 1989) shown in chapter 2 were used to create simulated TEOAEs. Five hundred frequency values were spaced evenly on a logarithmic scale between 500 Hz and 8 kHz. At each of these frequencies, a tone burst with the appropriate latency and duration, according to Figure 2.7 (p. 14) was created. All of the bursts were added together to produce the waveforms of a simulated TEOAE (see Figure 4.4). The amplitude of the TEOAE was chosen according to the growth characteristic curve shown in Figure 2.8 (p. 15). Using this curve, the simulated TEOAE waveform was scaled to the correct level relative

Figure 4.4 Time and frequency domain representations of a simulated TEOAE.
4.2.2 Evaluation Criteria

Two different quantitative measures were used to compare different artifact removal methods. The first measure was the level of the background noise in the final TEOAE signal for a given amount of raw data. This was calculated based on the number of recorded signals that were added or subtracted together, and how much averaging was done. The second measure was the TEOAE signal’s power loss. Every artifact reduction method removes part of the TEOAE signal, and the power loss is a simple way of quantifying this. This measure was used to compare the windowing method, the scaled subtraction method, and the subband decomposition method (chapter 5). Using these two measures together, signal to noise ratios (SNRs) could be calculated. To make the analyses more useful, these quantities were measured in four frequency bands, each 2 kHz wide, between 0 and 8 kHz.

To perform at their best, different artifact reduction methods have slightly different requirements for the raw data they use. It was important to compare these different methods while providing each one with the type of raw data it needed to work its best. Thus the results from different methods had to be obtained using slightly different raw data for each one. Despite this, however, fair comparisons could be made by deriving all results from the same amount of data, or else by correcting results to compensate for using different quantities of raw data to obtain them.

A second way to evaluate stimulus artifact reduction methods is to use real TEOAE measurements. This does not allow the precise evaluation that artificial signals allow, but it is still a very useful way of verifying that a method is working as expected. The results from a Zwislocki coupler can be compared to those from real ears. To help confirm that the final signal obtained from a real ear is really a TEOAE, the amplitude growth saturation characteristic can be checked. Even though the precise form of the TEOAE in a given recording is unknown, it is still known that the TEOAE should grow non-linearly with respect to the amplitude of the stimulus. Checking for this requires a comparison between at least two signals obtained with different stimulus amplitudes. If, after being processed with an artifact removal method, the TEOAEs from two or more different stimulus amplitudes show the normal amplitude saturation characteristic, it is an indication that the attempt to remove the stimulus artifact was successful. On the other hand, if the two final TEOAE signals show strongly linear
Figure 4.5 Windowing method of stimulus artifact removal. The artifact is windowed out by zeros up to time $t_1$, and the transition to the region containing pure TEOAE is smoothed by a sinusoidally rising curve between $t_1$ and $t_2$. For the present work, $t_1=6\text{ms}$, and $t_2=8.5\text{ ms}$.

amplitude growth it is probably an indication that they still contain a lot of stimulus artifact contamination. If there is poor correlation between the final TEOAE signals, it indicates that they are mostly noise and contain very little TEOAE, if any.

4.3 Windowing Method

The simplest way to remove the stimulus artifact is to use a window of zeros to eliminate the portion of the TEOAE recording which is contaminated by stimulus artifact (see Figure 4.5). This amounts to just throwing away the early part of the recording, and keeping the portion of the TEOAE which is completely separated in time from the artifact. Simulated results using this method are shown in Figure 4.7, where power loss values for the TEOAE signal are also shown. For these results, the time window shown in Figure 4.6 was applied to the simulated TEOAE recording.

The most obvious problem with the windowing method is that any portion of the TEOAE which is outside the window is lost. This normally includes all of the high frequency components. A second problem is that it is not easy to know how much of the recorded signal to cut, since the duration of the stimulus artifact varies. Because of these two problems, more sophisticated treatment is necessary.
Figure 4.7 Simulated results from stimulus artifact removal using the windowing method. The top left plot shows a close up view of the original simulated TEOAE recording. The top right plot shows the original simulated TEOAE (upper) and the final TEOAE signal after windowing (lower). The lower four plots compare the original and final TEOAE signals in 2 kHz wide frequency bands. The original TEOAE signal components have been offset upward, and the final TEOAE signal components have been offset downward. The “loss” value in each graph gives the ratio of the power in the original TEOAE to the power in the final TEOAE for that graph’s frequency band.

Figure 4.6 Window which was used for the windowing method. The window is zero for its first 6 ms, and one beyond 8.5 ms; it uses a quarter of a sinusoid in the transition region between 6 and 8.5 ms. These values were chosen because the stimulus artifact was dominant for the first 6 ms of the simulated TEOAE recording, and the TEOAE was dominant beyond about 8.5 ms.
4.4.1 Description

This family of methods is the current standard in TEOAE measurement (Ravazzani and Grandori, 1993; Kemp et. al., 1986). These methods rely on two properties of TEOAE measurements. The first is the strong saturation characteristic of the growth of TEOAEs with respect to stimulus amplitude. The second is the strongly linear growth of stimulus artifacts. As Figure 4.8 illustrates, these methods use two recordings from two different stimulus levels.

If we refer to the lower stimulus level recording as “Recording A”, and the higher stimulus recording as “Recording B”, then the method can be described as follows. Recording A is scaled up so that its stimulus artifact has the same size as the stimulus artifact in Recording B. Recording B is then subtracted from the scaled up version of Recording A. Ideally, their stimulus artifacts completely cancel each other while their TEOAEs only partially cancel. Part of the TEOAE from Recording A remains after subtraction because scaling up Recording A makes its TEOAE component larger than the TEOAE component in Recording B. This is a consequence of the saturation characteristic of the TEOAE’s amplitude. Because of the slight variations in the shape of the stimulus artifacts as a function of stimulus amplitude, there is always some error in the cancellation of the stimulus artifacts. To reduce the effect of this error, a time window must be applied to the scaled subtraction result to produce the final TEOAE signal (see Figure 4.8).

![Figure 4.8: Illustration of the scaled subtraction method of stimulus artifact cancellation. In this illustration, Recording B was obtained using a stimulus twice as large as the stimulus used to obtain Recording A.](image-url)
stimulus levels is the inverse of the ratio of the two stimulus levels. That is, the best SNR is obtained when

\[ \frac{NS1}{NS2} = \frac{S2}{S1} \]  

(4.1)

where \( S1 \) and \( S2 \) represent the two stimulus levels, and \( NS1 \) and \( NS2 \) represent the number of recording sweeps obtained for these two stimulus levels. Furthermore, Bray showed that for the amplitude growth characteristic of typical TEOAEs, it is best to set one of the stimuli at the highest possible level and to set the second stimulus amplitude to about one third of this.

This particular implementation of the scaled subtraction method produces an output which has been termed the “derived non-linear response”. It uses three recordings at the lower stimulus level and one recording at the higher level. The lower three recordings are averaged and scaled up by three, or equivalently, just added together. The higher recording is then subtracted. A window is applied to the resulting signal, yielding the final derived non-linear response TEOAE.

4.4.2 Evaluation

The derived non-linear response method works well enough to eliminate most of the artifact contamination and preserve most of the TEOAE. However, there are two major problems with the scaled subtraction method. One problem is that the background noise from the two signals adds at the same time that the TEOAE level is being reduced by subtraction. This reduces the SNR of the final TEOAE signal. The second problem is that the small differences between the shapes of the two stimulus artifacts for the two different stimulus levels cause artifact cancellation error. This error partially overlaps the TEOAE signal, so when a window is applied to reduce the error contamination, the early portion of the TEOAE signal is affected.

The effect of these two problems can be assessed fairly accurately through the analysis of a simple example. To begin, we suppose that the higher stimulus level, \( S1 \), is set at 77 dB SPL peak and the lower stimulus level, \( S2 \), is set at 67 dB SPL peak. A value of 77 dB was chosen because it was the largest peak amplitude that the probe could produce in a Zwislocki coupler for a stimulus designed in the way which was described in chapter 3. The second
To make the scaled subtraction method work well, equation 4.1 says that approximately three times as many averages should be used for the 67 dB stimulus as for the 77 dB stimulus. Thus, if we use $N_{77}$ to represent the noise power in the 77 dB recording, then the noise power in the 67 dB recording is approximately $N_{77}/3$. To match the stimulus artifact from a 67 dB recording to the artifact from the 77 dB recording, the 67 dB recording must be scaled up by a factor of $\sqrt{10}$. This increases its noise power by a factor of 10, to approximately $3.33N_{77}$. When the 77 dB recording is subtracted from the scaled up 67 dB recording, the noise in the two signals adds to produce a final noise power level of approximately $4.33N_{77}$.

The power level of the TEOAE in the final signal depends on the TEOAE's amplitude growth characteristic and on the stimulus levels which are used. Using $T_{77}$ to represent the amplitude of the TEOAE signal in the 77 dB recording, Bray's work indicates that the TEOAE amplitude at 67 dB is approximately $0.6T_{77}$. When the 67 dB recording is scaled up by $\sqrt{10}$, its TEOAE amplitude becomes $1.9T_{77}$. After subtraction, the TEOAE in the final signal is $0.9T_{77}$. If $P_{77}$ represents the TEOAE power level in the 77 dB recording, then the TEOAE power level in the final TEOAE recording is approximately $0.81P_{77}$. Thus, with perfect scaled subtraction, the SNR of the final TEOAE signal is $0.81P_{77}/4.33N_{77}$, or $0.187(P_{77}/N_{77})$. This translates into a 7.3 dB reduction in the signal to noise (background noise) ratio of the final TEOAE signal when it is compared to the SNR of the original 77 dB recording.

So far, this analysis has ignored the presence of stimulus artifact cancellation error. Now the effects of the window which is used to reduce the cancellation error contamination in the final TEOAE signal must be considered. The effect of this window is different for different frequency components of the TEOAE because of the TEOAE's characteristic time-frequency structure. This analysis has been carried out for the window shown in Figure 4.9, which is typical for established scaled subtraction procedures. It is 0 for 3 ms, rises sinusoidally between 3 and 5.5 ms, and stays at 1 for the duration of the signal. When this window is applied to the simulated scaled subtraction TEOAE signal, the results are as shown in Figure 4.9. The number shown in each window is the reduction in the TEOAE's power because of the application of the window. For the 0-2 kHz frequency band, the window has very little impact on the TEOAE. In fact, because the window removes some of the background noise, it actually

44
Increasingly significant as higher frequency components are considered, until, in the 6-8 kHz band, only 20 percent of the original TEOAE power remains after windowing. This analysis clearly reveals that the scaled subtraction method is at a severe disadvantage when attempting to detect TEOAE frequency components in the higher frequency region.

We conclude with an example using typical signal values. Table 4-1 shows power levels and SNRs for the signals in this example. The example begins with stimulus artifact which is recorded in a Zwislocki coupler using a 77 dB SPL peak stimulus. The overall noise level in this recording is 0.57 nW, distributed between 0 and 8 kHz as shown in the table.

![Scaled Subtraction Window](image)

![Entire Simulated TEOAE](image)

![0-2 kHz band](image)

![2-4 kHz band](image)

![4-6 kHz band](image)

![6-8 kHz band](image)

Figure 4.9 Simulated results from stimulus artifact cancellation using scaled subtraction. The top left plot shows the window which was applied to help suppress the effects of artifact cancellation error after scaled subtraction. The top right plot shows the original simulated scaled subtraction TEOAE and the result after application of the time window. The lower four plots show the original TEOAE (offset upward) and the windowed TEOAE (offset downward) in 2 kHz wide bands. Each plot also shows the power loss caused by the window within each frequency band.
For 77 dB recording:

<table>
<thead>
<tr>
<th>TEOAE power (nW)</th>
<th>Noise power (nW)</th>
<th>SNR (dB)</th>
</tr>
</thead>
<tbody>
<tr>
<td>5.6</td>
<td>0.57</td>
<td>9.9</td>
</tr>
<tr>
<td>3.5</td>
<td>0.34</td>
<td>10.1</td>
</tr>
<tr>
<td>1.3</td>
<td>0.11</td>
<td>10.6</td>
</tr>
<tr>
<td>0.42</td>
<td>0.057</td>
<td>8.7</td>
</tr>
<tr>
<td>0.14</td>
<td>0.031</td>
<td>6.6</td>
</tr>
</tbody>
</table>

For 67 dB recording:

<table>
<thead>
<tr>
<th>TEOAE power (nW)</th>
<th>Noise power (nW)</th>
<th>SNR (dB)</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.0</td>
<td>0.19</td>
<td>10.3</td>
</tr>
<tr>
<td>1.3</td>
<td>0.11</td>
<td>10.8</td>
</tr>
<tr>
<td>0.47</td>
<td>0.034</td>
<td>11.4</td>
</tr>
<tr>
<td>0.15</td>
<td>0.028</td>
<td>7.3</td>
</tr>
<tr>
<td>0.05</td>
<td>0.012</td>
<td>6.1</td>
</tr>
</tbody>
</table>

After scaled subtraction:

<table>
<thead>
<tr>
<th>TEOAE power (nW)</th>
<th>Noise power (nW)</th>
<th>SNR (dB)</th>
</tr>
</thead>
<tbody>
<tr>
<td>4.6</td>
<td>1.8</td>
<td>3.9</td>
</tr>
<tr>
<td>2.9</td>
<td>1.0</td>
<td>4.7</td>
</tr>
<tr>
<td>1.1</td>
<td>0.39</td>
<td>4.3</td>
</tr>
<tr>
<td>0.35</td>
<td>0.21</td>
<td>2.1</td>
</tr>
<tr>
<td>0.11</td>
<td>0.18</td>
<td>-2.1</td>
</tr>
</tbody>
</table>

After windowing:

<table>
<thead>
<tr>
<th>Final TEOAE power (nW)</th>
<th>Final Noise power (nW)</th>
<th>Final SNR (dB)</th>
</tr>
</thead>
<tbody>
<tr>
<td>4.0</td>
<td>1.5</td>
<td>4.3</td>
</tr>
<tr>
<td>2.8</td>
<td>0.81</td>
<td>5.4</td>
</tr>
<tr>
<td>0.84</td>
<td>0.32</td>
<td>4.2</td>
</tr>
<tr>
<td>0.2</td>
<td>0.18</td>
<td>0.4</td>
</tr>
<tr>
<td>0.023</td>
<td>0.15</td>
<td>-8.2</td>
</tr>
</tbody>
</table>

Table 4-1. Power levels and SNRs for the scaled subtraction simulation. Note the decrease in SNR between the original two recordings (at 67 dB and 77 dB) and the scaled subtraction result. This decrease is caused by accumulation of background noise and reduction of TEOAE power. The further decrease in SNR after windowing is caused by the window itself. See text for further explanation. (Power levels are shown in nW, but they are actually in thousands of µPa². The conversion to nW assumes a unit resistance.)

When the simulated TEOAE for a 77 dB SPL peak stimulus is added to the stimulus artifact recording, the overall SNR is approximately 10 dB. However, this varies with frequency as shown in the table. A recording for a 67 dB SPL peak stimulus was also recorded from the Zwislocki coupler. For this recording, three times as many stimuli were averaged as for the 77 dB recording. The noise levels of these recordings are well within the normal range for a random selection of normally hearing human subjects. As the results in the table show, the SNRs for the high frequency bands of the final scaled subtraction TEOAE signal are much worse than in the original 77 dB recording. The most severe loss is in the 6-8 kHz band, where there is a reduction of approximately 15 dB in the SNR.

4.5 Adaptive Cancellation Methods

Like scaled subtraction methods, these methods rely on the nonlinear growth property of TEOAEs. However, they do not rely on perfectly linear stimulus artifact growth. Rather than using a scalar valued multiplier, these methods use a filter to match and cancel the stimulus artifacts in two TEOAE recordings from two different stimulus levels (see Figure 4.10). Many
variations on this theme are conceivable, depending on the size and type of the filter employed. However, the purpose of this section is not to provide an exhaustive study of this family of methods. Only two representative examples are looked at, with a brief assessment of the characteristics of each.

The simplest version of this method uses a linear filter to cancel the stimulus artifact. The filter coefficients are adapted to match a reference artifact to a primary artifact as closely as possible, according to a least squares criterion. (This filter is sometimes referred to as a Wiener filter.) The effectiveness of this method was evaluated with the same simulated signals as for the scaled subtraction simulation. The 77 dB stimulus signal was used as the reference and the 67 dB stimulus signal was used as the primary. Figure 4.11 shows the mean squared error in the final TEOAE signal as a function of the filter order. Figure 4.12 shows the results from the 81st order filter, which has the lowest mean squared error according to Figure 4.11.

![Diagram of general structure of adaptive stimulus artifact cancellation methods](image)

**Figure 4.10** General structure of adaptive stimulus artifact cancellation methods.

![Plot of normalized mean squared error as a function of filter order](image)

**Figure 4.11** Plot of the normalized mean squared error between the simulated TEOAE and the final TEOAE signal from linear filters of different orders.
The first bias is that the filter order was selected using prior knowledge of the TEOAE which was present in the primary. The second bias is caused by the use of stimulus artifacts from a Zwislocki coupler. As shown back in Figure 4.3, the coupler is more linear in its behavior than most human ears. This means that it is easier to use a linear filter to match the reference to the primary for this simulated example than it would be for most real recordings. Therefore, these results should be viewed as an upper bound on the level of performance that could be expected with this method. In most situations, the artifact cancellation error is greater and there is also more distortion of the TEOAE.

The second comment to be made about this example is that the filter places the heaviest

Figure 4.12 Simulation results for the 81st order linear filter. The original TEOAE (offset upward), the TEOAE after filtering (middle), and the stimulus artifact cancellation error and background noise (offset downward) are shown for four different frequency bands. Note that the filter distorts the TEOAE, as the 2-4 kHz band shows most clearly. Note also that some of the stimulus artifact remains in the early portion of the signal, as indicated on three of the four plots. (See text for further explanation.)
not the region of greatest overlap between the stimulus artifact and the TEOAE. Rather, the tail end region of the decaying stimulus artifact overlaps with the TEOAE. The filter does little to improve artifact cancellation in this region, and has sometimes been observed to make the cancellation error worse. Thus, this method is unreliable for separating the early, high frequency TEOAE components from the tail end of the artifact.

A second version of this method uses non-linear neural networks to cancel the stimulus artifact (Yuwaraj, 1995). Many variations are possible, and they cannot be exhaustively evaluated here. A basic approach, as presented by Yuwaraj, is to obtain a TEOAE recording at a high amplitude such as 80 dB, and a second recording at a much lower amplitude, such as 40 dB. The lower amplitude recording is to be such that no TEOAE is detectable in it. It is supposed to contain only a reference stimulus artifact and background noise. In reality, there is still a TEOAE present at such a stimulus level, but it is completely buried in the noise because it is so small. Because it is recorded at such a low stimulus level, this reference signal is very noisy and has to first be denoised using wavelets. The reference and primary are decomposed using a mutual wavelet packet based decomposition, and the corresponding components of the two decomposed signals are used to train non-linear neural networks. The networks "learn" to predict the primary signals based on inputs from the reference signals. The predicted primary signals are subtracted from the actual primary signals to yield the final TEOAE signal components. The final TEOAE signal is reconstructed from these components. The intended result is that all of the stimulus artifact is removed without affecting much of the TEOAE.

As results shown in Figure 4.13 illustrate, the non-linear neural network is able to substantially reduce the artifact, but there is still substantial cancellation error. In fact, the TEOAE in the stimulus artifact region is badly corrupted by artifact cancellation error. To remove this would require application of a window such as in ordinary scaled subtraction methods. This, in turn, would have the same type of effect on high frequency TEOAE components as the scaled subtraction window.

Any implementation of a method from this family suffers from the same type of disadvantages as scaled subtraction methods. The SNR of the final signal is reduced just as much as with scaled subtraction. The advantage is that the filter should give improved artifact cancellation, so the contamination from cancellation error should be reduced. However, this
introduces the danger of distorting the TEOAE itself, since it is also passed through the filter. In some cases, the TEOAE may be almost entirely canceled. In fact, the final results for this family of methods are not significantly better than those of scaled subtraction. In both methods, high frequency components of the TEOAE have a poor SNR. The next chapter will discuss a new method which yields improved SNRs.
5. SUBBAND DECOMPOSITION BASED ARTIFACT REMOVAL

The previous chapter assessed current methods of solving the problem of stimulus artifact contamination in TEOAE recordings. It's conclusion was that none of the present methods are able to remove the artifact cleanly enough to detect the early portions of the TEOAE, and that this restricts the frequency range of the TEOAE components which can be detected. In this chapter, a new method is presented. Results indicate that by using a suitable subband decomposition, it is possible to remove virtually all of the stimulus artifact in a way that preserves most of the TEOAE and extends the frequency range of TEOAE measurements so that they cover the entire audiometric range (Kuehner, Kunov, and Madsen, 1997). In addition, the SNR of lower frequency TEOAE components is improved over previous methods.

5.1 Wavelets and Wavelet Packets

5.1.1 Motivation

Adaptive wavelet packet based subband decompositions have a number of desirable properties for the present application. Firstly, the use of a subband decomposition of the TEOAE recordings is motivated by the fact that different subbands of stimulus artifacts and TEOAEs have different temporal characteristics. Secondly, since a time domain TEOAE should be available after any stimulus artifact removal processing, the subband decomposition should have the property of perfect reconstruction. Thirdly, it is desirable that the subband decomposition be suitable for a wide variety of different stimulus artifacts. All of these properties are provided by a wavelet packet decomposition. In this section, a brief overview of wavelets and wavelet packets is provided, and their application in the present work is described.
Wavelets are orthogonal basis functions which are concentrated locally in both time and frequency. In the same way that a signal can be decomposed onto a set of sinusoidal basis functions by a discrete Fourier transform, a function can be decomposed onto a wavelet basis by a discrete wavelet transform. Unlike the Fourier transform, which yields very high localization in the frequency domain but none in the time domain, wavelets allow a signal to be divided into frequency subbands and then analyzed and processed with a degree of time localization appropriate for the frequency range of each particular subband. They do this by representing the signal as a linear combination of orthogonal basis functions which are all frequency scaled, time shifted versions of a single prototype wavelet. If the prototype wavelet is designated as \( w(t) \), then the signal \( s(t) \) can be represented as

\[
s(t) = \sum_{j,k} b_{j,k} 2^{-j/2} w(2^{-j} t - k),
\]

where \( b_{j,k} \) represents the coefficients of the discrete wavelet transform and \( 2^{-j/2} w(2^{-j} t - k) \) represents time-shifted, frequency-scaled versions of the prototype wavelet, normalized by a factor of \( 2^{-j/2} \). The time shift is controlled by the value of \( k \), and the frequency scale is controlled by \( j \).

Wavelets are related to pairs of FIR filters that meet certain special conditions which make them orthogonal filters (Strang and Nguyen, 1996, p. 109). Loosely speaking, one of these filters is a low pass filter and the other is a high pass filter. Their impulse responses can be designated as \( g(n) \) and \( h(n) \), respectively. If there is a function \( \phi(t) \) which satisfies the "dilation equation",

\[
\phi(t) = \sum_k \sqrt{2} g(k) \phi(2t - k),
\]

then \( \phi(t) \) is called the scaling function. The function \( w(t) \) is obtained by using \( \phi(t) \) and \( h(n) \) in the "wavelet equation". That is,

\[
w(t) = \sum_k \sqrt{2} h(k) \phi(2t - k).
\]

The discrete wavelet transform of a sequence \( s(n) \) is obtained by filtering \( s(n) \) with the two filters \( h(n) \) and \( g(n) \), and downsampling the resulting sequences by a factor of two. That is,

\[
s_{1,0}(n) = \sum_k s(k) g(2n - k) \quad \text{and} \quad s_{1,1}(n) = \sum_k s(k) h(2n - k).
\]

52
Because they come from the first application of the low pass filter, the values of \( s_{1,0}(n) \) are called the first level approximation coefficients. The values of \( s_{1,1}(n) \) are called the first level detail coefficients. The decomposition can be repeated on the approximation coefficients to produce multiple levels of detail coefficients, as diagrammed in Figure 5.1.

### 5.1.3 Wavelet Packets

Wavelet packet based decomposition is a generalization of wavelet based decomposition. With wavelet packets, the decomposition can be repeated for detail coefficients as well as for approximation coefficients. A full, three level wavelet packet decomposition tree

---

Figure 5.1 Wavelet decomposition structure: (a) structure of successive filtering and downsampling operations which yield approximation and detail coefficients, (b) tree diagram showing successive splitting of approximations, and (c) approximate division of frequency space between the different subbands of the wavelet decomposition.
the tree if its nodes have one and only one direct projection to each of the tree's terminal nodes. For example, the nodes [(2,0), (3,2), (3,3), (1,1)] make up one such set. The set [(3,0), (3,1), (2,1), (1,1)], corresponding to the wavelet decomposition tree in Figure 5.1(b), is another example.)

Each basis set corresponds to a particular way of decomposing the signal. For an N-level wavelet decomposition, there is a choice of N+1 different basis sets with which the signal can be represented. For the same sized wavelet packet decomposition, there are $2^N$ basis sets from which to choose. For the present work, it is very helpful to have this increased flexibility so that higher frequency subbands can be split into two narrower subbands when it is appropriate.

![Wavelet Packet Decomposition Tree](image)

**Figure 5.2 Wavelet packet decomposition tree.**

### 5.2 Application to Stimulus Artifact Removal

There is one overriding consideration for the present application. The wavelet packet decomposition which is employed must provide clear time domain separation between TEOAE components and stimulus artifact components within each subband. This requires good separation between different frequency subbands so that long duration, low frequency artifact components are cleanly separated from short duration, high frequency TEOAE components. This requirement places constraints on the selection of a wavelet packet family. It also has implications when choosing the particular basis set to be used from within the selected family.
There are several possible families of wavelets from which to choose. Within each family, there are many possible choices of filter orders. Four standard families were considered for the present work. They were biorthogonal wavelets, Daubechies wavelets, Coiflets, Symlets, and a fifth family, for which there is as yet no standard name. This fifth family actually comprises an entire group of wavelet families whose wavelets are derived from FIR filters with equiripple passband characteristics. For the present work, these wavelets are called equiripple filter wavelets.

Figure 5.3 shows the frequency responses of the high pass and low pass decomposition filters for six different wavelets. The filters with the sharpest cutoffs are those of the equiripple

![Figure 5.3](image-url)
four different bases from four different wavelet families. In this figure, the cleanest separation of frequency bands is produced by the equiripple filter family of wavelets. On the basis of these results, the equiripple family was chosen as the most suitable for the present work.

5.2.2 Selection of a Basis Set

As stated in section 5.1.3, there are many possible sets of orthogonal basis functions for any given wavelet packet type. The choice of which set to use for a given signal can be made manually, or it can be made automatically on the basis of some sort of cost criterion. The type

![Diagram](image)

**Figure 5.4** A typical TEOAE recording decomposed into subbands using wavelets from four of the five wavelet families under consideration. (Results for the Symlet wavelets are very similar to those for the Daubechies wavelets.) For the biorthogonal wavelet, the low pass decomposition filter was $17^{th}$ order and the high pass filter was $11^{th}$ order. For the other three wavelet families shown, all filters were $18^{th}$ order. This figure illustrates the way that the equiripple filter wavelets provide the cleanest separation into subbands.
were not found to be very good for the present work. A good cost function for the present work needs to be sensitive to the time domain duration of different frequency components so that the resulting decomposition separates components of long time duration from components of short duration. The best conventional cost function in this situation is a threshold entropy function which equates the cost of a signal with the number of elements in the signal which exceed a certain threshold. That is,

\[ E(s) = \# \{ i \text{ such that } |s_i| > \text{Thr} \}, \]

where \( E \) is the entropy of the signal \( s \), and \( \text{Thr} \) is the threshold. This worked better than any other conventional entropy function for the present application, but it was found that a small modification improved the decomposition. Better separation between shorter duration frequency components and longer duration components was obtained by using the function

\[ C(s) = (\# \{ i \text{ such that } |s_i| > \text{Thr} \})^p, \]

where \( p \) is an exponent between 1 and 2 and \( C \) is the cost associated with the signal \( s \). The effect of \( p \) is to increase the cost associated with long duration components so that short duration artifact components are more likely to be separated into their own subband. Different values of \( p \) were tested, and a range of values between 1.2 and 1.5 was found to work well for the present application.

5.3 Artifact Removal Procedure

In this section, the general method of stimulus artifact removal based on adaptive subband decompositions is described. This description begin with an explanation of the data requirements of this method. Next, preprocessing is discussed, and followed by the way that the recordings are decomposed. Finally, removal of the artifact and reconstruction of the final TEOAE signal is explained.

5.3.1 Raw Data Requirements

Two TEOAE recordings are required for this method. They need to be made using two different stimulus levels, \( S_1 \) and \( S_2 \), which are far enough apart to make the non-linear growth of the TEOAE show up clearly when the two recordings are compared. The recordings are compared by scaling the \( S_2 \) recording by a factor equivalent to \( S_1 - S_2 \) dB. The method works
Figure 5.5 Plot of the difference between stimulus level and TEOAE level as a function of stimulus level. From this curve, appropriate stimulus levels can be selected for the two TEOAE recordings which are required in order to apply the subband decomposition method of stimulus artifact removal. (Data derived from Figure 2.8, p. 15.)

best if there is a difference of at least 10 dB between the TEOAE amplitudes in the two recorded signals after the $S_2$ signal has been scaled. This can be expressed as

$$|TEOAE_2 + (S_1 - S_2) - TEOAE_1| \geq 10,$$

where $S_1$ and $S_2$ represent the peak stimulus levels in dB, and $TEOAE_1$ and $TEOAE_2$ represent the corresponding TEOAE amplitudes. Rearranging this gives

$$|(S_1 - TEOAE_1) - (S_2 - TEOAE_2)| \geq 10.$$  (5.10)

The value of $S_2$ which is required in order to satisfy this relationship depends on the value of $S_1$. The plot in Figure 5.5 can be used to select $S_2$ appropriately. To use this plot, choose a value for $S_1$ and find that value on the horizontal axis. Find the corresponding vertical axis value from the plot. Now choose an $S_2$ value on the horizontal axis such that the corresponding vertical axis value is at least 10 dB away that of $S_1$.

5.3.2 Preprocessing, Basis Selection and Signal Decomposition

After the raw data has been obtained, the signals need to be decomposed into subbands. As illustrated in Figure 5.6, the core of the procedure is the division of the recordings into frequency subbands, followed by the division of each of these subbands into time segments which separate the stimulus artifact from the TEOAE. It is important that both recordings are
decomposed onto the same basis so that the corresponding subband components from the two recordings can be compared with each other.

Before this is done, however, the signals require a small amount of preprocessing. The recorded signals need to be aligned with one another before doing the subband decomposition. This is easily accomplished by upsampling using a small interpolation filter, lining up the peaks of all recorded signals, and downsampling. For the present work, upsampling by a factor of 20 using an 8th order filter proved sufficient. At this point it is also helpful to remove any offset in the signals so that they have mean of zero, and then to scale them so that they are both the same size. For the results shown in this chapter, all recordings were scaled so that they had a peak value of one.

Because both of the recordings must be decomposed onto the same basis, only one recording needs to be used to select the appropriate basis. (While the cost criterion would usually yield the same decomposition structure for both of the recordings, this is not guaranteed because of the different noise levels in the different signals.) The recording with the higher stimulus amplitude is the better one to use for this because it has the best SNR. By applying the cost function described in section 0 to this signal, the best decomposition structure can be determined and both of the recorded signals can be decomposed onto the selected basis.

5.3.3 Artifact Removal and TEOAE Preservation

Once the recorded signals have been decomposed, the stimulus artifact can be removed from each subband individually. This is done by applying a window to cut out the stimulus
To answer this question, the non-linear growth property of the TEOAE components can be used. Since the recorded signals are scaled so that their stimulus artifacts are both of the same size, the stimulus artifact components in each subband are of the same amplitude for both of the recorded signals. Small variations occur because of noise and because of the slight shape differences between stimulus artifacts from different stimulus levels. For the current purpose, however, these variations are minor. The stimulus artifact components from different recordings line up very closely, as illustrated in Figure 5.7.

This is not the case with the TEOAE components, as Figure 5.8 illustrates. If the two stimulus amplitudes have been selected appropriately, the amplitudes of the TEOAE components in the two signals should differ from each other by a factor of at least three (approximately 10 dB). The key to eliminating the stimulus artifact without affecting the TEOAE is detection of the point in time within each subband where the stimulus artifacts have decayed below the level of the TEOAEs. This can be done using a sliding window, typically

![Figure 5.7 Subband decomposition of two recorded TEOAE signals, recorded at stimulus levels of approximately 85 dB SPL peak and 67 dB SPL peak. The signals have been normalized to a peak value of 1. The stimulus artifact components of the two signals line up so closely in each subband of the decomposition that on the scale shown here, the two signals appear to overlie each other exactly. They are indistinguishable.](image-url)
equal in length to about three cycles of the dominant frequencies of the subband. Moving in steps of approximately one cycle each, the window starts at the beginning of the subband and moves along until the spot where the large amplitude difference between the two signals is detected. This large amplitude difference is the mark which distinguishes the TEOAE region from the stimulus artifact region.

In the present work, this feature has been detected by monitoring the ratio of the mean squared values of the two subband signals within the moving window. (The top value in this ratio is normally the mean squared value of the signal with the larger TEOAE after the sizes of the two stimulus artifacts have been matched.) Throughout the stimulus artifact region, this ratio stays very close to one. At the point where the stimulus artifact dies out and the TEOAE components start, the ratio rises sharply, as illustrated in Figure 5.9. The point at which the ratio crosses a preset threshold is taken as the ending point of the stimulus artifact and the
Figure 5.9 Ratio of the mean squared values of the two signals shown in Figure 5.7. The point in time where the stimulus artifact has decayed below the level of the TEOAE (or possibly the background noise) corresponds to the point in time where the ratio of the mean squared values of the two signals shoots upward.

starting point of the TEOAE. For the data in Figure 5.9, there is a wide range of suitable thresholds. The lower frequency subbands (numbers 1 and 2 in Figure 5.9) are slightly more sensitive to threshold selection than the higher frequency subbands, but this does not seriously affect the detection of the lower subband TEOAE components. To help avoid a sudden step in the final TEOAE signal, the nearest zero crossing after this point can be found and used as the actual starting point.

In addition to starting points which eliminate the stimulus artifact, ending points can be established for the higher frequency subbands at a point after the TEOAE ends for the subband. This helps to improve the SNR in the subbands where the TEOAE components typically last for only a few milliseconds. The final 10 to 15 ms of these subband are just noise. Therefore, for those subbands whose most dominant frequency is between 3.5 kHz and 6 kHz, an ending point can be set at 9 ms. For subbands whose most dominant frequency is higher than 6 kHz, the ending point can be established at approximately 7 ms.
whose peak value does not exceed one percent of the peak value of the original signal.

5.3.4 Reconstruction of the Final TEOAE Signal

Using the starting and ending points for each of the subbands, the final TEOAE signal can be reconstructed. If a group of recordings has been made and two of these recordings have been used to establish a set of starting points for the subband TEOAE components, the TEOAE from each recording can be reconstructed from this set of starting points. To help reduce spectral spreading, a window can be applied to the TEOAE portion of each subband as it is added into the final signal. Conventional windows are not good for this, however, since they are designed for steady state signals. A conventional window would almost completely eliminate the short duration TEOAE components which are located immediately after the starting points in the higher frequency subbands. To avoid this, window shaped as a quarter of a sinusoid with a length equal to approximately two cycles of the dominant frequency in the subband is applied at the beginning and at the end of each subband's TEOAE component. The final TEOAE signal can then be reconstructed simply by adding together the windowed TEOAE components from each of the subbands into which the original recording was decomposed.

5.4 Simulation Results

This method was tested on simulated TEOAE recordings so that it could be compared with the methods which were discussed in chapter 4. The comparison with the derived non-linear response implementation of the scaled subtraction method was particularly emphasized, since this is currently the best established, most dominant of the different artifact cancellation methods. The simulated recording for a 77 dB SPL peak stimulus was used, along with a simulated recording for a 57 dB SPL peak stimulus. The results of the simulation are shown in Figure 5.10.

In comparison to the methods of chapter 4, two points should be noted. First, for the same amount of data, the new method produces much better SNRs. Assume that three recordings are made at 77 dB, and one at 57 dB. (This corresponds to the amount of data in the four recordings which are required for the derived non-linear response version of the scaled
the background noise power in the average of the three recordings is approximately $N/3$. The TEOAE power remains the same, so that if the SNR in one 77 dB recording is $P_{77}/N$, then it is $3P_{77}/N$ in the final signal. This translates into an improvement of about 4.8 dB relative to a single original 77 dB recording. This compares very favorably with the estimated reduction of 7.3 dB relative to a single original 77 dB recording for scaled subtraction and adaptive filter based methods (section 4.4).

Secondly, the losses due to windowing are different for the new method than for the scaled subtraction method. The estimated reduction in TEOAE signal power in different frequency bands for the two methods is shown in Figure 4.9 (p. 45) and Figure 5.10. In the lower frequency bands the new method cuts out slightly more of the TEOAE than the scaled subtraction method does. However, in the higher frequency bands the new method does not cut out nearly as much of the TEOAE as the scaled subtraction method does.

In summary, analysis of this simulated example indicates that, compared to conventional scaled subtraction artifact cancellation, the subband decomposition based method should improve the SNR of the final TEOAE signal by approximately 10 dB in the 0 to 6 kHz frequency bands.
5.5 Experimental Results

TEOAE recordings were made for 10 normally hearing ears from 5 different subjects. Recordings were also made with the probe in an artificial ear cavity (a Zwislocki coupler) to provide a control for comparison with the human subjects. The recordings were processed using the subband based artifact removal method, and the results were compared with those obtained by processing the recordings using scaled subtraction. These results are tabulated in Table 5-1, and shown in more detail in Figure 5.11 through Figure 5.15. They show significant improvement for the subband based results over the scaled subtraction (derived non-linear response) results. The improvement is comparable to that which was predicted by the simulation results.

5.5.1 Raw Data and Preprocessing

For each subject ear, recordings were made for three different stimulus levels. Three recordings were obtained for a stimulus with a peak level of approximately 77 dB SPL. Each of these recordings contained 600 responses which were averaged to reduce the background noise. Two recordings were made at a stimulus level of 67 dB SPL, with 1800 responses averaged for each recording. Finally, a single recording with 600 averaged responses was obtained for a peak stimulus level of 57 dB SPL. Each averaged recording comprised 520 samples at a sampling frequency of 22050 Hz, for a total of 23.58 ms of data.

The mean of each of the recordings was subtracted off to remove any dc offset they might contain. Then, using an eighth order interpolation filter, the recordings were all upsampled by a factor of 20. They were shifted in time so that all of their peaks were lined up, and then they were downsampled by a factor of 10 so that their effective sampling frequency was 44.1 kHz. Finally, they were all normalized so that their peak value was one.

5.5.2 Obtaining the Final TEOAE Signals

The scaled subtraction final TEOAE signal was obtained by using two of the 77 dB recordings along with the two 67 dB recordings. The normalized 77 dB recordings were subtracted from the normalized 67 dB recordings. The resulting pair of signals were both multiplied by a sinusoidally rising window which started at a point 3 ms into the signal and
The subband based final TEOAE signal was obtained using the 57 dB recording and the three 77 dB recordings. Two of the 77 dB recordings were averaged together to make a single signal with 1200 averaged responses. Following the method described in this chapter, the recordings were decomposed into a set of subbands using wavelet packets. Using the 57 dB reference recording and the 77 dB recording with 1200 averages, starting points and ending points were established within each of the subbands. TEOAE signals were made from both of the 77 dB recordings using the starting points which had been selected. A final TEOAE signal was created by averaging these two TEOAE signals, and a noise estimate was obtained by subtracting one of them from the other.

5.5.3 Comparison of Results from the Two Methods

For both methods, the discrete Fourier transforms of both the noise estimate and the final TEOAE signal were obtained. For each of the four frequency bands shown in Table 5-1, the average amplitude of the frequency domain TEOAE signal and the average amplitude of the noise estimate signal were calculated. The ratio of these two, in dB, was computed and used as an estimate of the SNR for the frequency band under consideration. Finally, a correction was made for the purpose of comparison. Since the scaled subtraction results were based on twice as much data as the subband results, 3 dB were subtracted from the scaled subtraction SNRs.

These SNRs are recorded in the table. The scaled subtraction results appear under the heading DNLR, which stands for derived non-linear response. The results from the subband method appear under the heading SBND, for subband. As the averages at the bottom of the table indicate, the subband based method typically achieves SNR improvements of between 10 and 15 dB for a given amount of data. This is in good agreement with expectations based on the analysis and simulation results presented earlier.

To verify that the final TEOAE signals from the subband based method were primarily TEOAE signals and not stimulus artifacts, TEOAEs were recorded from the left ear of subject #1. A total of six TEOAE recordings were made at stimulus amplitudes of 57, 61, 65, 69, 73,
Table 5-1  Experimental results, comparing the SNRs of final TEOAE signals from the standard derived non-linear response (DNLR) stimulus artifact removal method with those from the new adaptive subband based method (SBND). The bottom row of the table shows results for the Zwislocki coupler, which has no TEOAEs. SNRs are in dB.

<table>
<thead>
<tr>
<th>Subject #</th>
<th>DNLK</th>
<th>SBND</th>
<th>DNLK</th>
<th>SBND</th>
<th>DNLK</th>
<th>SBND</th>
<th>DNLK</th>
<th>SBND</th>
</tr>
</thead>
<tbody>
<tr>
<td>1-left</td>
<td>-3.2</td>
<td>3.9</td>
<td>6.5</td>
<td>18.3</td>
<td>4.4</td>
<td>16.8</td>
<td>-1.7</td>
<td>5.5</td>
</tr>
<tr>
<td>2-left</td>
<td>4.0</td>
<td>14.1</td>
<td>-0.1</td>
<td>11.9</td>
<td>-0.7</td>
<td>14.6</td>
<td>-2.6</td>
<td>14.1</td>
</tr>
<tr>
<td>3-left</td>
<td>-4.0</td>
<td>10.5</td>
<td>1.2</td>
<td>12.2</td>
<td>-0.3</td>
<td>13.2</td>
<td>-3.4</td>
<td>16.4</td>
</tr>
<tr>
<td>4-left</td>
<td>-3.9</td>
<td>3.7</td>
<td>-2.6</td>
<td>12.2</td>
<td>5.9</td>
<td>19.8</td>
<td>-0.7</td>
<td>7.7</td>
</tr>
<tr>
<td>5-left</td>
<td>-5.5</td>
<td>7.2</td>
<td>-4.7</td>
<td>8.2</td>
<td>-3.3</td>
<td>14.5</td>
<td>-7.1</td>
<td>3.4</td>
</tr>
<tr>
<td>1-right</td>
<td>-3.5</td>
<td>4.9</td>
<td>8.5</td>
<td>17.1</td>
<td>5.8</td>
<td>21.2</td>
<td>2.2</td>
<td>18.0</td>
</tr>
<tr>
<td>2-right</td>
<td>6.2</td>
<td>14.0</td>
<td>3.0</td>
<td>12.7</td>
<td>-4.5</td>
<td>13.3</td>
<td>-3.7</td>
<td>11.2</td>
</tr>
<tr>
<td>3-right</td>
<td>-6.2</td>
<td>2.9</td>
<td>-5.2</td>
<td>10.1</td>
<td>-5.6</td>
<td>11.1</td>
<td>-3.9</td>
<td>6.5</td>
</tr>
<tr>
<td>4-right</td>
<td>-4.6</td>
<td>5.5</td>
<td>-0.5</td>
<td>16.0</td>
<td>0.4</td>
<td>11.4</td>
<td>-4.0</td>
<td>9.3</td>
</tr>
<tr>
<td>5-right</td>
<td>0.8</td>
<td>12.2</td>
<td>-0.2</td>
<td>14.3</td>
<td>-2.9</td>
<td>13.7</td>
<td>-7.3</td>
<td>5.4</td>
</tr>
<tr>
<td>Average</td>
<td>-2.0</td>
<td>7.9</td>
<td>0.6</td>
<td>13.3</td>
<td>-0.1</td>
<td>15.0</td>
<td>-3.2</td>
<td>9.8</td>
</tr>
<tr>
<td>Zwislocki</td>
<td>-1.3</td>
<td>2.4</td>
<td>-2.6</td>
<td>2.5</td>
<td>1.3</td>
<td>0.6</td>
<td>1.2</td>
<td>1.7</td>
</tr>
</tbody>
</table>

and 77 dB SPL peak. The recordings were processed using the method described in this chapter, and the six final TEOAE signals were obtained. They were filtered into four frequency bands, each 2 kHz wide, between 0 and 8 kHz. The results, plotted in Figure 5.16, clearly show the amplitude growth saturation characteristic. If the final TEOAE signals were actually stimulus artifact components, then their amplitude would grow linearly with respect to the stimulus amplitude. Figure 5.17 shows the actual growth characteristic for the four frequency bands.
Figure 5.11 Amplitude spectra of final TEOAE signals obtained using the subband based processing method. The lower line in each plot is the estimated noise in the signal.
Figure 5.12 Amplitude spectra of final TEOAE signals obtained using the scaled subtraction method. The lower line in each plot is the estimated noise in the signal.
Figure 5.13 Comparison of the SNRs achieved by the scaled subtraction artifact cancellation method (solid line) with those achieved by the subband based method (asterisks) for four different frequency bands (500 Hz - 2 kHz, and from 2 kHz to 8 kHz in 2 kHz increments).
Figure 5.14 Time domain results from the subband based artifact removal method. In each plot, two TEOAE signals are shown overlying one another to show the degree of correlation between repeated trials.
Figure 5.15 Time domain results from the scaled subtraction method. In each plot, two TEOAE signals are shown overlying one another to show the degree of correlation between repeated trials.
Figure 5.16 Final TEOAE signals for different stimulus levels. The stimulus was varied between 57 and 77 dB SPL peak in 4 dB increments. The bottom plot in each group corresponds to the 57 dB stimulus, and the top plot corresponds to the 77 dB stimulus. The signals were filtered into four frequency bands so that the amplitude growth saturation characteristic could be verified throughout the audiometric frequency range.
Figure 5.17 RMS TEOAE level as a function of stimulus level for four different frequency bands. These plots clearly show that the TEOAE amplitude was not growing at the same rate as the stimulus amplitude. This verifies that the signals are actually TEOAEs, and not stimulus artifact components, because stimulus artifact components would grow linearly as a function of stimulus amplitude.

5.6 Practical Implementation and Computational Time

The method described above is suitable for implementation in a practical clinical instrument. Although it requires more computation than the scaled subtraction family of methods, the amount required is not prohibitive. All necessary computations can be performed in less than 30 seconds on a 100 MHz, 80486 processor personal computer using MATLAB. Compiled code running on a faster machine would speed this up considerably.

It is worth mentioning that to keep the wavelet packet computations efficient, data sequence lengths should be in powers of two. When this is the case, the computation is of order N, where N is the length of the data sequence. The best way to ensure this without requiring extra time to be spent obtaining enough data to satisfy this condition is to incorporate it into the preprocessing of the data. After the raw recordings are upsampled and lined up, they need to be downsampled again. They should be downsampled at a rate which yields a signal length which is a power of two.
6. CONCLUSIONS AND FUTURE WORK

6.1 Summary of Research

In this work, the factors which limit the frequency range of TEOAE measurements have been identified and investigated. Solutions to the problems created by these factors have been proposed and tested. We have demonstrated, both by simulation and by real ear measurements, that the proposed solutions yield better results than conventional methods of stimulating and processing TEOAE recordings.

In particular, the characteristics of the system which is used to stimulate and record TEOAEs have been studied, and a way of designing stimuli which compensate for undesirable system characteristics has been developed and tested. This solved the problem of uneven distribution of acoustic energy in the stimulus, and created a way to ensure that as much high frequency stimulation as possible was provided.

A means of objectively evaluating different TEOAE processing methods through simulation was developed. Conventional methods of stimulus artifact reduction were assessed and shown to be unreliable for TEOAE measurements above frequencies of 4 to 5 kHz. Adaptive linear and non-linear filters were also assessed and found to be inadequate.

Finally due to the unsuitability of conventional methods, a new method of separating the TEOAE from the stimulus artifact in TEOAE recordings was developed. The general method, based on subband decomposition of the TEOAE recordings, was explained. It was implemented in a particular way which would make it suitable for hearing screening tests, and compared to the conventional scaled subtraction method using both simulation and real ear results. Test results indicated that this method is capable of separating TEOAE components from stimulus artifact components for frequencies throughout the entire audiometric range. SNRs of final TEOAE signals from this method were shown to be much better than for the scaled subtraction method.
There are at least two major areas for future work which builds upon the present work. One of these relates to continued improvement of the technical methods that have been developed. The other relates to clinical studies using the techniques presented to extend the frequency range of TEOAEs.

A significant advance has been made by pushing TEOAE measurements beyond the 4 to 5 kHz range to which they have been limited since the discovery of TEOAEs. In this regard, the remaining challenge is to make further advances by overcoming the limitations which remain. This will require hardware which is suitable for operation at higher frequencies. The present system has not been designed for operation outside of the normal audiometric range. The current miniature speaker is not capable of providing stimulation outside of this range. In addition, the current probe casing attenuates high frequency sounds. These problems are not easily solved. One possibility is the use of a bone conductor to apply the transient stimuli. This is already being investigated for ultrasonic frequencies and it is possible that it would provide a better means of high frequency audio range stimulation than the speaker presently in use. There is no apparent reason why the algorithms which have been developed in the present work will not prove suitable for the TEOAE recordings with a frequency range greater than that of the present work.

Although the new method has been developed sufficiently to demonstrate its superiority to conventional methods, there is undoubtedly room for future improvement. Further study into ways of more cleanly separating the TEOAE components from stimulus artifact components might be profitable. The present method is limited in that it can only detect and preserve TEOAE components which are separated in time from the stimulus artifact within their own subband. Furthermore, because of this the present method is not entirely free of distortion. Different frequency components of the TEOAE can be affected differently by the set of subband starting points which are chosen to mark the separation point between stimulus artifact components and TEOAE components. Further study to characterize this distortion and investigate ways of reducing it would be profitable.

In the area of clinical testing, there is now an opportunity to use the methods presented in this work to conduct studies which were previously impossible. A screening test such as the one suggested needs to be widely tested in clinical environments before it is clear how much the
which evaluate the effectiveness of TEOAE-based hearing screening use scaled subtraction methods of artifact reduction. They also use electrical click stimuli which are vulnerable to wide variation between subjects in the acoustical stimulus they actually yield. It seems likely that a more standardized, controlled stimulus combined with a wider frequency range of measurement will significantly improve the quality of TEOAE test results.
APPENDIX A. SYSTEM MODEL

This appendix provides the details of a model of the entire system in the form of an equivalent electrical circuit. The models for the miniature speaker and microphone were provided by Knowles Electronics, Inc. The models for the ear canal, ear drum, and middle ear were obtained from the literature which addresses this topic, and from Knowles. The probe's equivalent electrical circuit was derived by approximating the probe's geometric features with shapes for which equivalent electrical circuit elements are well established. The entire model was implemented using an evaluation version of the electrical circuit simulation package, PSpice.

A.1 Miniature Speaker (Receiver)

The model for the receiver has two parts. One part represents the electrical coil and the other part represents the mechanical and acoustical components. The following segment of PSpice code was provided by Knowles Electronics, Inc., as an appropriate model for its EF 1933 receiver between 100 Hz and 10 kHz. PSpice code for the electrical driver circuitry is also shown below. It consists of a constant current source (designated IKE501) and a parallel resistor and capacitor combination (designated Rhif_lift and Chif_lift, respectively) which serves to slightly increase the high frequency output capacity of the receiver.

* Constant current electrical source,  
* with high frequency lift.  
.WIDTH OUT=132  
.AC LIN 100 100HZ 10KHZ  
IKE501 501 502 AC .850E-3

Rhif_lift 401 501 .100E+5  
Chif_lift 401 501 .100E-7  
*
*KKEEF1, 0.85 mA, 195 OHM DC COIL  
* (electrical coil of EF 1933 receiver)  
RK501 505 518 .121E+5  
RK502 518 0 .121E+5  
RK509 501 503 .100E+9  
RK510 503 502 .100E+9

RK511 503 518 .100E+9  
GK501 501 503 505 518 .919E-3  
GK502 505 518 501 503 -.919E-3  
GK503 503 502 518 0 .919E-3  
GK504 518 0 503 502 -.919E-3  
*
*KKEEF1, GENERIC EF TRANSDUCER  
* (mechanical and acoustical components of  
* EF 1933 receiver)  
RK503 516 517 .100E+3  
RK508 0 515 .100E+6  
RK512 0 519 .100E+9  
RK513 0 513 .100E+9  
RK514 0 517 .100E+9  
LK501 517 504 .600E-1

78
A.2 Probe Output Path

Figure A.1 shows the layout of the electroacoustic probe. Acoustically, the output path can be analyzed in four sections, labeled A, B, C, and D in the figure.

![Diagram of probe output path](image)

Figure A.1 Longitudinal section of electroacoustic probe, showing the output path and the input path. The Knowles EF 1933 hearing aid receiver, and the Knowles EM 4068 microphone are labeled. As shown, eight narrow tubes conduct sound from the receiver out and around the enclosure containing the microphone.

Section A has length of 0.6 cm and a volume of 0.186 cm$^3$. This is approximately equivalent to a rectangular solid with a length of 0.6 cm and a cross sectional area of 0.31 cm$^2$. The cross sectional area at the outlet of the speaker is 0.03 cm$^2$ (see Figure A.2). At the junction between section A and section B, the cross sectional area of the eight tubes which conduct sound around the microphone housing totals 0.04 cm$^2$. This arrangement can be approximated as two unenclosed air masses in series, each 0.3 cm in length. The first air mass has a cross sectional area of 0.03 cm$^2$, and the second has an area of 0.04 cm$^2$. These masses have volumes of 0.009 cm$^3$ and 0.012 cm$^3$ respectively. The rest of the volume of section A is considered to be an enclosed volume of 0.167 cm$^3$. 
an enclosed volume of air can be modeled as a capacitance, and an unenclosed air mass acts as an inductance. These are evaluated as follows. (Values of constants are displayed in Table A-1 below, and symbols for dimensions of acoustic elements are shown in Figure A.3).

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Density</td>
<td>0.001223 g·cm⁻³</td>
</tr>
<tr>
<td>Viscosity</td>
<td>184×10⁶ dyn·s·cm⁻²</td>
</tr>
<tr>
<td>Speed of sound</td>
<td>34400 cm·s⁻¹</td>
</tr>
</tbody>
</table>

Figure A.3 Symbols used to represent dimensions of acoustic elements.

Table A-1. Values of constants for air at 21°C and 760 mm Hg.

\[
L_{A1} = \frac{\rho l}{A} = \frac{(0.001223 \, g \cdot cm^{-3})(0.3 \, cm)}{0.031 \, cm^2} = 11.8 \, mH
\]

\[
L_{A2} = \frac{\rho l}{A} = \frac{(0.001223 \, g \cdot cm^{-3})(0.3 \, cm)}{0.04 \, cm^2} = 9.2 \, mH
\]

\[
C_A = \frac{V}{pc^2} = \frac{0.167 \, cm^3}{(0.001223 \, g \cdot cm^{-3})(34400 \, cm \cdot s^{-1})^2} = 0.115 \, \mu F
\]

Section B comprises eight narrow tubes in parallel, each 1.15 cm long and 0.08 cm in diameter. Each tube can be represented by a circuit section like the one in Figure A.5. The values of the circuit elements are frequency dependent, and can be determined from the approximations developed by Zuercher et al., (Zuercher, Carlson, and Killion, 1988). The
Table A-2. Frequency dependent circuit element values for section B.

<table>
<thead>
<tr>
<th>Frequency</th>
<th>$R_{B,t}$ ($\Omega$)</th>
<th>$L_B$ (mH)</th>
<th>$R_{B,p}$ (k$\Omega$)</th>
<th>$C_B$ (pF)</th>
</tr>
</thead>
<tbody>
<tr>
<td>100 Hz</td>
<td>27.2</td>
<td>44.5</td>
<td>330</td>
<td>40.9</td>
</tr>
<tr>
<td>350 Hz</td>
<td>33.5</td>
<td>43.5</td>
<td>130</td>
<td>37.0</td>
</tr>
<tr>
<td>1000 Hz</td>
<td>47.4</td>
<td>41.8</td>
<td>73.8</td>
<td>35.0</td>
</tr>
<tr>
<td>3500 Hz</td>
<td>80.1</td>
<td>38.6</td>
<td>37.5</td>
<td>33.7</td>
</tr>
<tr>
<td>10 kHz</td>
<td>128</td>
<td>36.5</td>
<td>21.4</td>
<td>33.1</td>
</tr>
</tbody>
</table>

Section C is treated in the same way as section A (see Figure A.6). It is 0.4 cm long and has a volume of 0.161 cm$^3$. The cross sectional area at the junction of sections B and C is 0.04 cm$^2$, and at the junction of sections C and D the area is 0.053 cm$^2$.

\[
L_{c1} = \frac{\rho l}{A} = \frac{0.001223 \text{ g cm}^{-3} \times (0.2 \text{ cm})}{0.04 \text{ cm}^2} = 6.1 \text{ mH}
\]

\[
L_{c2} = \frac{\rho l}{A} = \frac{0.001223 \text{ g cm}^{-3} \times (0.2 \text{ cm})}{0.053 \text{ cm}^2} = 4.6 \text{ mH}
\]

\[
C_C = \frac{V}{\rho c^2} = \frac{0.142 \text{ cm}^3}{(0.001223 \text{ g cm}^{-3} \times (34400 \text{ cm s}^{-1})^2} = 0.098 \mu \text{F}
\]

Section D is the region between the microphone tube and the outer casing of the probe's end section. It has the shape of a narrow slit wrapped around the microphone tube. According to Rossi (1988, p. 276), a slit can be modeled by a resistance and an inductance in series, as shown in Figure A.7. The values of $R_D$ and $L_D$ are as follows.

\[
R_D = 12 \left( \frac{\eta l}{k^3 b} \right) = \frac{12 \times (184 \times 10^{-6} \text{ dyn cm s}^{-2}) \times (1.7 \text{ cm})}{(0.055 \text{ cm})^3 \times (0.958 \text{ cm})} = 23.5 \Omega
\]

\[
L_D = 12 \frac{\rho l}{A} = \frac{(1.2) \times (0.001223 \text{ g cm}^{-3}) \times (1.7 \text{ cm})}{0.053 \text{ cm}^2} = 47.1 \text{ mH}
\]

The corresponding PSpice code is shown below.

*PROBE OUTPUT PATH
Lleak 504 0 4.00E-1
Rrec 405 406 2.20E+3
La1 406 407 11.7E-3
Ca 407 0 0.13E-6
La2 407 408 9.20E-3

RB3 410 0 10G
GRbs 408 409 FREQ (V(408,409))

+ (100, -2.86E+1, 0)
+ (350, -3.05E+1, 0)
+ (1000, -3.35E+1, 0)
+ (3500, -3.81E+1, 0)
+ (10000, -4.21E+1, 0)
A.3 Ear Canal, Ear Drum, and Middle Ear

The ear canal, ear drum, and middle ear can most easily be modeled using the equivalent circuit provided by Knowles Electronics, Inc. for an occluded ear simulator (provided below in PSpice code). If direct control over the ear drum and middle ear impedance is desired, then the ear canal can be modeled as a lossless transmission line with a characteristic impedance of 95 g·s⁻¹·cm⁻⁴. A typical length is approximately 1.5 cm, but it can be varied as needed to simulate different probe insertion depths. The ear drum/middle ear impedance is connected as the terminating impedance on the ear canal transmission line. This impedance can be varied to gain insight into the impact of different ear drum/middle ear impedances on the overall system. (Typical values of ear drum impedance are plotted in Figure 3.5, p.23.)

*OCCLUDED EAR SIMULATOR
LK561 569 560 .520E-2
RK561 562 565 .120E+4
RK562 561 565 .650E+3
RK563 561 567 .970E+3
RK564 561 569 .340E+3
LK561 562 563 .580E-0
LK562 565 566 .109E+0
LK563 567 568 .700E-1
LK565 569 560 .520E-2
LC561 563 0 700E-7
LC562 566 0 120E-6
LC563 568 0 120E-6
LC564 560 0 150E-6
LC565 417 0 100E-7
LK561 414 0 561 0 ZO=.949E+2 TD=.311E-4
LK562 561 0 417 0 ZO=.949E+2 TD=.584E-5

A.4 Probe Input Path

The probe input path is a tube, 2.4 cm long and 0.1 cm in diameter. Because of its length, it needs to be modeled as a transmission line for frequencies above approximately 2.5 kHz. At lower frequencies it acts as an inductance of approximately 370 mH \((L=\mu l/A=(.001223)(2.4)/(.0079)=0.372)\). The parameters of the transmission line can be approximated using the results of Zuercher et. al., (Zuercher, Carlson, and Killion, 1988). The transmission line is a series connection of circuit segments such as the one in Figure A.5, each representing no more than 0.4 cm of tube length. (This is the ideal case, and creates more
A.5 Miniature Microphone and Input Filter

The microphone model is shown below, as provided by Knowles Electronics, Inc. The model is preceded by the 2.2 kΩ resistor (designated Rrmike) which is placed in front of it in the probe. The microphone model is followed by a first order high pass RC filter (designated Rhp, Chp) with a 3 dB frequency of 165 Hz. This filter reduces the low frequency noise level in the microphone output signal.
For the present work, only an evaluation version of PSpice was available. This version has limited capacity and was not able to handle the full model of the system without compromising the model's accuracy at frequencies above about 4 to 5 kHz. If there were no limitations on the circuit size, then transmission lines would have been used to represent the eight parallel tubes in section B, and the output tube in section D. These lines would have been built up with circuit segments, just as the microphone input tube was. This does not affect the structure of the model, nor the results for lower frequencies. It only causes errors from neglecting transmission line effects at higher frequencies.


Salt, A.N. (1997), Cochlear Fluids Research Laboratory, Department of Otolaryngology, Washington University School of Medicine, St. Louis, Missouri, <http://oto.wustl.edu/cochlea/>.


