Spring 2011

The effect of Kalman weighted filtering and in-situ pre-amplification on the accuracy and efficiency of ABR threshold estimation

Julie Kathleen Wheeler
James Madison University

Recommended Citation

Wheeler, Julie Kathleen, "The effect of Kalman weighted filtering and in-situ pre-amplification on the accuracy and efficiency of ABR threshold estimation" (2011). *Dissertations*. 42.
https://commons.lib.jmu.edu/diss201019/42

The Effect of Kalman Weighted Filtering and In-situ Pre-amplification on the Accuracy and Efficiency of ABR Threshold Estimation

Julie K. Wheeler

A dissertation submitted to the Graduate Faculty of
James Madison University
In
Partial Fulfillment of the Requirements
for the degree of
Doctor of Audiology

Communication Sciences and Disorders

May 2011

Dedication

To my husband who offered unconditional support throughout this entire process. You were an amazing rock without whom I would not have made it this far. Not only did you endure being my practice participant a time or two and help me cut and paste a million different waveforms into the flip books, you encouraged me and helped me to see the light at the end of the tunnel. Since the moment I met you, you have been my biggest fan, unwavering in your support of me and I want you to know that you are loved and appreciated.

To my family who has always stood behind me and encouraged me to reach for my dreams, to be successful, and to figure out what I love and then figure out how to get paid for it. You have nurtured my love of Audiology since I was eight years old and have encouraged me along the long (long!) path. I am blessed to have such a supportive family.

In Memory of

Dr. Roger Ruth

This dissertation is a tribute to you from the students you so tirelessly dedicated your life to educating. I count myself among the very luckiest of students that I had the privilege of taking your electrophysiology classes (although I am not going to lie and say it was easy!). This dissertation was one of the last projects you began and I am pleased to bring it to fruition. Your impact on the field of audiology and on the audiology program at James Madison University is profound.

Acknowledgements

To my advisor and mentor, Dr. Brenda Ryals, who deserves to be Advisor of the Year. You must have revised my dissertation a hundred times, offering invaluable advice (and “suggestions”). I am eternally grateful for the long hours you put in on this project, your meticulous attention to detail, and your friendship. I hope I can adequately convey my appreciation for all that you have done for me not only with my dissertation, but throughout my entire four years at James Madison University.

To my committee members, Dr. Lincoln Gray and Dr. Dan Halling, who have helped immensely throughout this entire project. Whether it was meeting after work, on the weekends, or over the phone you were each willing to provide expert advice and guidance. I appreciate the integral input you have given that has helped mold this research into a worthy dissertation. And, of course, thank you Dr. Gray for being my statistics guru.

To my Reviewers, Dr. Claire Jacobson, Dr. Donna Mallory, and Dr. Brenda Ryals who spent many long hours reviewing waveforms. Thank you for donating your time and expertise to further audiologic research. You each played an integral role in this research and I am eternally grateful for each one of you.

Finally, to Dr. Aaron Steinman and Dr. Isaac Kurtz at Vivosonic Incorporated. Thank you for loaning the Vivosonic equipment to James Madison University for this dissertation and for helping to divide the waveforms obtained by the Vivosonic equipment into useable chunks. Your help was invaluable and it was a pleasure working with you both.

Table of Contents

List of Tables

List of Figures

Abstract

Auditory brainstem responses (ABRs) are important for acquiring frequency specific information for determination of the degree and type of hearing loss for infants and difficult-to-test populations when behavioral audiometry cannot be carried out. This study investigated the effects of Kalman weighted filtering and in-situ pre-amplification employed by the Vivosonic Integrity V500 ABR system on threshold accuracy and efficiency in an environment of high physiologic noise in comparison to a conventional ABR system which employs a standard artifact rejection paradigm. Auditory brainstem responses were collected using the Vivosonic ABR system and a conventional ABR system both in quiet and in noise using tonal stimuli at 500 and 4000 Hz (eight total conditions). ABRs were administered to twenty adult participants with normal hearing acuity (behavioral thresholds better than 20dB HL). Physiologic noise was created by having the participant chew gum to emulate the movement of an infant sucking on a bottle or pacifier.

Results indicated that there was a statistically significant main effect for equipment when examining all data (both quiet and noisy) with the exception of "No Responses" at 4000 Hz indicating that the Vivosonic measured significantly lower, more accurate, ABR thresholds than the conventional system regardless of activity level. There was no significant main effect for equipment noted when examining all data (both quiet and noisy) with the exception of "No Responses" at 500 Hz indicating that each system measured similar thresholds at this frequency. When dividing the data into subsets by frequency, no statistically significant differences were found for threshold accuracy measurements between the Vivosonic and the conventional systems in quiet or in noise at either 500 or 4000 Hz. At 4000 Hz, the Vivosonic equipment was found to be significantly more efficient at acquiring threshold than the conventional ABR system, but again no difference between systems was noted at 500 Hz. Findings suggest that neither system was particularly accurate or efficient at 500 Hz as it appears that physiologic noise is problematic at this frequency with either traditional artifact rejection of with Kalman weighted filtering and in-situ pre-amplification.

Further exploration into the effects of Kalman weighted filtering and in-situ pre-amplification are warranted based on the findings of this study. Trends indicated in this study suggest that Kalman weighted filtering and in-situ pre-amplification may lead to more accurate and more efficient ABR acquisition without the need for sedation, at least for higher frequencies.

Chapter I

Introduction

With the advent of newborn hearing screening mandates, Auditory Brainstem Response (ABR) testing has become an increasingly integral component of accurate testing of infants. While otoacoustic emissions are often used for newborn hearing screenings because they are quicker and easier than ABR measurements, they are limited in their ability to detect degree, configuration, and type of hearing loss. In order to achieve accurate threshold estimation in infants, children, and difficult-to-test populations, an objective measure with a high degree of sensitivity and specificity is required. Conventional ABR systems use a standard artifact rejection protocol in which an artifact rejection level is pre-determined, allowing only sweeps that do not exceed the pre-set voltage to be included in the final waveform. If the artifact rejection level is set too high, a large amount of artifact- physiologic and electrophysiologic- can easily enter and degrade the quality of the waveform. Conversely, if the artifact rejection level is set too conservatively, the acquisition time to acquire a waveform can be excessive. The Vivosonic Integrity V500 is an ABR system which employs Kalman weighted filtering and in-situ pre-amplification as an alternative to a conventional artifact rejection paradigm. In-situ pre-amplification is carried out by mounting the pre-amplifier directly on the ground electrode, eliminating electromagnetic noise that would traditionally enter the system through long lead wires. The pre-amplifier mounted on the forehead also reduces physiologic noise as the pre-amplifier moves in conjunction with the participant’s head. Kalman weighted filtering is a means of averaging sweeps by assigning a weighting to each sweep depending on the degree of noise measured in the electroencephalogram- noisier sweeps are assigned a weight closer to zero while quieter sweeps are assigned a weight closer to one. With Kalman weighted filtering, each sweep is used to some degree in an attempt to significantly reduce acquisition time while still measuring a waveform with clear morphology. This study will examine the effects of Kalman weighted filtering and in-situ pre-amplification on the accuracy and efficiency of ABR measurements in comparison to a conventional system using standard artifact rejection.

Chapter II

LITERATURE REVIEW

Auditory Evoked Responses

An auditory evoked response (AER), according to Hall (2007), is “activity within the auditory system (the ear, the auditory nerve, or auditory regions of the brain) that is produced or…evoked by sounds” (p. 1). AERs are evoked by an acoustic stimulus and measured by electrodes placed on the scalp. The electrode is connected via long leads to a pre-amplifier, filter, analog to digital converter, and eventually to a computer where the measured signal is processed (Hall, 2007). Optimally, these electrodes are placed over the temporal lobes of the brain between the ear and the midline of the head; however, in a clinical setting the electrodes are most commonly placed at the vertex where they can pick up activity from both hemispheres as well as on the earlobe, mastoid, or inside the ear canal (Burkard, Don, & Eggermont, 2007). A ground electrode is also adhered somewhere on the body, typically the low forehead. While it is important to have low impedance values for each electrode (<5 kΩ), it is crucial for common mode rejection that there be good balance between the electrodes as even small imbalances can greatly reduce the effectiveness of common mode rejection (Burkard et al., 2007).

Once the acoustic stimulus is delivered, activity ranging from 0.1 µV to over 1000 µV in amplitude is measured within the auditory structures (Hall, 2007). The latency of the response allows for determination of the generator site along the auditory pathway (Hall, 2007). The resulting brainwave will be a direct representation of the stimulus intensity i.e. loud sounds produce more clearly defined, larger amplitude responses with shorter latencies (Hall, 2007). Even under optimal conditions, AERs have very small voltages- often as small as .1 µVs. Activity arising from the inner ear, auditory nerve, and brainstem involves relatively few neurons compared to the ongoing EEG and occurs at a relatively great distance from the electrodes (far field response) and, therefore, only about .1 to 1µV in amplitude is often recorded (Hall, 2007). Because of the small amplitudes of these responses various techniques are used to extract the signal from the underlying noise, or electroencephalogram, including pre-amplification, differential amplification, artifact rejection, filtering, and signal averaging.

In order to amplify the AER effectively, the amplifier must have a common mode rejection function which is accomplished through differential amplification. The theory behind common mode rejection is that when two electrodes are placed on the head, noise arrives at each electrode at the same time and in phase and is cancelled by subtracting the voltage at the inverting electrode from that recorded at the non-inverting electrode, while a response that occurs between the two electrodes is amplified. The differential amplifier amplifies the differences between the voltages recorded at one electrode (inverting electrode) from those recorded at the other electrode (non-inverting electrode) (Hall, 2007). This, in effect, causes the noise to be cancelled when the inverting and non-inverting recordings are added, but does not cancel the response of interest. This is because the electrodes are far enough apart that they are not measuring the response in the same way. This serves to increase the amplitude of the response while decreasing the amplitude of the noise; thus, increasing the signal to noise ratio.

In a conventional system, artifact rejection is a setting that specifies the sensitivity of the amplifier so that it rejects any signal that is larger than the designated range of acceptable amplitude voltages. In a conventional system, sweeps that exceed the acceptable noise level are completely eliminated.

The objective of filtering in the frequency domain is to "eliminate unwanted nonresponse activity (electrical and muscle interference or artifact) while preserving the actual response" (Hall, 2007, p. 90). Changes in filter settings can serve to make waveforms more pronounced and remove artifact; however, if a filter is too narrow, it can cause significant distortion or eliminate the response altogether (Hall, 2007). Unwanted low-frequency signals arise from the brain, eyes, and heart. These signals occur at frequencies less than 50 Hz and are easily removed with a conventional band pass filter; however, noise arising from the skeletal muscle occurs in the frequency range from 30-500 Hz and can be as large as 100-200µV (Sokolov, Kurtz, Steinman, Long, & Sokolova, 2006). This noise is particularly problematic as it falls in the same frequency range as the ABR and is, therefore, difficult to filter with a conventional artifact rejection setting (Sokolov et al., 2006). An ABR typically has three frequency bands- one at approximately 125 Hz or below, another at about 552.5 Hz, and the final one at approximately 967. 5 Hz making a traditional band-pass filter of 30 to 1500 Hz highly effective at removing external noise as long as the patient is relaxed (Hall, 2007).

Signal averaging is also used to pull the response out of the background noise. The signal of interest is imbedded in a background noise that is not related to changes in auditory stimulus input (Burkard et al., 2007). This background noise- or the electroencephalogram- is a combination of a much larger number of cortical neurons that are continuously firing and are not necessarily arising from acoustic stimulation (Burkard et al., 2007). The background noise occurs randomly without regard to the stimulus while the AER occurs consistently with approximately the same waveform each time it is recorded. Over the course of hundreds or thousands of sweeps, the evoked potentials are added together while the electroencephalogram is slowly cancelled, thus increasing the signal to noise ratio (Burkard et al., 2007).

Auditory Brainstem Responses

The Auditory Brainstem Response (ABR) is a type of AEP. First described in 1971 by Jewett and Williston, it is a neurologic test of auditory brainstem function. It is a transient response to either a broadband click stimulus or a frequency specific tone burst (Hall, 2007). ABR testing is a far field test and is conventionally recorded with a minimum of three electrodes adhered to the participant's scalp. Although there are many possible electrode configurations, Stapells (1998) recommends a single channel montage for air conduction ABR measurement with the non-inverting electrode placed at Cz (vertex) and the inverting electrode placed on the ipsilateral mastoid (M1/M2). Additionally, a ground electrode is required which can be adhered anywhere on the body-generally the low forehead, shoulder, or non-test ear (Hall, 2007). The optimal stimulus for evoking an ABR is a very brief broadband click or frequency specific tone burst (~100µs) designed to increase the synchronous firing of the eighth cranial nerve. The response it evokes normally occurs within 5-6 ms after the onset of the stimulus and produces a distinct waveform with five prominent waves which are labeled with Roman numerals I-V (Hall, 2007). The distal portion of the eighth nerve where it is surrounded by bone and the proximal portion of the eighth nerve where it is no longer surrounded by bone as it enters the brainstem are the generator sites for wave I and wave II, respectively

(Hall, 2007). Waves III, IV, and V are thought to be generated at the level of the cochlear nucleus, the superior olivary complex, and the inferior colliculus, respectively (Hall, 2007). In adults, wave I normally appears at approximately 1.5 ms after the presentation of the stimulus and each subsequent wave occurs at roughly 1ms intervals after wave I; thus Wave V latency is about 5.5msec. The time interval between Wave I and Wave V is on the order of 4msec and reflects the time it takes for the signal to travel from the distal portion of the eighth nerve to the higher portion of the brainstem (Burkard et al., 2007).

ABR and Auditory Threshold Estimation

According to the American Speech-Language-Hearing Association (1991), ABR is the preferred method for testing infants less than six months of age. While otoacoustic emissions are often used for newborn hearing screenings because they are quicker and easier than ABR measurements, they are limited in their ability to detect degree, configuration, and type of hearing loss (Stapells, 1998). Other types of AEP testing, according to Stapells et al. (1998), such as the middle latency responses and slow cortical responses are unreliable in children and are extremely sensitive to patient state of arousal and sedation (as cited in Stapells, 1998).

Auditory brainstem responses (ABRs) are robust to changing patient states as they can be performed under sedation, during sleep, or in a relaxed patient. They provide important frequency specific information for determining the degree and type of hearing loss in infants, young children, and difficult-to-test populations when traditional behavioral audiometry cannot be carried out or when ear specific information is required.

Auditory brainstem response testing is employed to differentiate between conductive, cochlear, and retrocochlear pathologies and has become increasingly useful with the advent of universal newborn hearing screening mandates. ABR is an effective means of objectively determining hearing thresholds and provides the necessary information to establish appropriate amplification strategies long before the infant is able to be tested behaviorally.

ABR wave V is the most prominent wave in an ABR recording; therefore, it is used in threshold determination. ABR threshold is defined as the lowest intensity where there is an identifiable wave V (Beattie, Kenworthy, & Vanides, 2005). According to Sininger and Cone-Wesson (2002), electrophysiologic thresholds are highly correlated with behavioral hearing thresholds. In fact, the behavioral pure tone averages and thresholds obtained using click stimuli in an ideal recording (i.e. relaxed patient, quiet environment) have a .979 correlation coefficient (Sininger & Cone-Wesson, 2002). Stapells and Oates (1997) conducted a study examining the correlations between ABR thresholds and behavioral thresholds at 500, 2000, and 4000 Hz for normal hearing participants and participants with sensorineural hearing which revealed a .94 (73 ears), .95 (96 ears), and .97 (51 ears) correlation coefficient, respectively, indicating that ABRs can accurately predict behavioral thresholds under ideal recording conditions. A study published by Beattie et al. (2005) examined the accuracy of ABR threshold estimation when compared with behavioral thresholds in twenty-six participants with gradually sloping sensorineural hearing loss. The findings of this study indicated that ABR thresholds obtained at 500 and 1000 Hz were within 16 dB of behavioral thresholds in 85% of the participants and ABR thresholds obtained at 2000 and 4000 Hz were within 9 dB of behavioral thresholds in 85% of participants. In 2000, Stapells conducted a meta-analysis of 32 different studies examining ABR threshold estimation using tone burst stimuli. Stapells indicated that results were consistent across the studies with tone burst ABR thresholds typically measured at 10-20 dBnHL in normal hearing participants and, in adult participants with sensorineural hearing loss, ABR thresholds are measured approximately 5-15 dB higher than behavioral thresholds. Stapells (2000) indicates that the studies included in the meta-analysis agreed that, as a whole, ABRs evoked by a 500 Hz tone burst are not as reliable as those evoked by a 4000 Hz tone burst. In fact, he indicates that, on average, thresholds obtained using a 500 Hz tone burst are about 7 dB higher than those obtained using a 4000 Hz tone burst. Each of these studies indicate a strong relationship between behavioral and tone burst evoked auditory brainstem response when measurements are obtained in ideal conditions (i.e. relaxed patient, quiet, electrically shielded environment); however, ABRs are often not obtained under these ideal conditions.

ABR- Factors That Influence Reliability

As previously stated, ABR amplitudes are very small, ranging from 0.1 to 1µV (Hall, 2007) and various forms of background noise can significantly degrade the quality of the recording, making threshold determination unreliable. Artifact, as it pertains to auditory evoked responses, is “electrical activity that is not part of the response and should not be included in analysis of the response” (Hall, 2007, p.89). Artifact can be a result of such activity as “patient movement, neuromuscular activity, [and] electrical interference” (Hall, 2007, p. 89). There are three main categories of background noise that can degrade the integrity of the waveform: fundamental noise, electromagnetic interference, and endogenous noise (Cutmore & James, 1999). Fundamental noise is created by the amplifier and circuits. It can be a result of overheating equipment (thermal noise) as well as noise that is created at junctions in the circuit (shot noise) i.e. between the skin and electrode or between electrical components of the amplifier (Cutmore & James, 1999). Electromagnetic interference occurs as a consequence of electromagnetic signals created in the environment by objects such as computers, fluorescent lights, or electric motors. In an environment with AC electrical current, electrodes act as miniature antennae picking up the fluctuating magnetic and electrical interferences (Sininger & Cone-Wesson, 2002). These interferences become noise as they are sent to the bio-amplifier and amplified along with the signal of interest. A type of electromagnetic noise called line noise is the result of undesirable signals transmitted through cables and wires (Sokolov et al., 2006). Conventional ABR equipment has wires as long as 3ft, making the system extremely susceptible to line noise (Kurtz & Sokolov, 2004).   Line noise is reduced by keeping the leads equal in length and as short as possible (Cutmore & James, 1999). Finally, endogenous noise is physiologic artifact which arises from within the human body. The amount of physiologic noise is influenced by activity in the head, neck and trunk muscles causing a disruption in the electrode contacts (McCall & Ferraro, 1991). According to Sokolov et al. (2006), physiologic artifact is derived from many sources including the brain, eyes, heart, and skeletal muscles. Particularly of interest for ABR measurements is the electromyogram or skeletal muscle movement (Sokolov et al., 2006). Skeletal muscle movement- especially in the face and neck- causes noise in the ABR recording measuring 100-200 µV, much larger than any portion of the ABR waveform. Furthermore, because a typical band pass filter for an ABR is 30-1500Hz, skeletal muscle movement is problematic as it produces signals in the 30-500 Hz range and, therefore, these unwanted signals would be included in the ABR recording with a traditional analogue filter (Sokolov et al., 2006). In most cases, this will cause an erroneous overestimation of hearing thresholds (Hall, 2007).

According to a survey conducted by Tannenbaum (2005), 50% of respondents indicated that they would prefer to test in outpatient suites and emergency rooms if electromagnetic noise were not a problem. This survey of 60 clinical audiologists also revealed that 84% found noise to be a major frustration in auditory evoked potential (AEP) testing with endogenous noise ranking number one and electromagnetic noise ranking number two (Tannenbaum, 2005). The respondents also indicated that the cumbersome wires and the length of time it took to acquire the ABR recordings in noisy environments were particularly frustrating. An ABR system in which these various interferences do not affect the recordings would be highly advantageous.

Endogenous and electromagnetic noise can often be controlled by meticulous set-up and patient instruction and preparation. In optimal settings, a clear waveform will be easily discernable; however, many settings do not allow for optimal measurement. An optimal setting for ABR testing would be a Faraday cage which is made from conducting material designed to block out electrical fields (Sokolov, 2007). Unfortunately there is often a need to perform ABRs in suboptimal settings such as ICUs, operating rooms, nurseries, and hospital wards where electromagnetic interference is high. According to the National Institute of Occupational Safety and Health, the intensive care unit generally has magnetic field exposure of .1-220 mG, an MRI produces a magnetic field of .5-280 mG, power cables have magnetic fields of 15-170 mG, and computers have magnetic fields of .4-6.6 mG (as cited in Kurtz and Sokolov, 2005). The vast amounts of electromagnetic artifact found in these environments make ABR measurement virtually impossible.

For conventional ABRs, patients must be relaxed, preferably with their eyes closed, in order to reduce physiologic artifact; however, in many cases, such as with infants and difficult-to-test populations, it is not practical that a patient will be quiet and relaxed throughout the collection process (Sokolov, 2007). In a study conducted by McCall and Ferraro (1991), it was found that the likelihood of a child passing an ABR screening is much lower if he/she is awake versus if the child is asleep. That study showed that the fail rate when awake was 67% while the fail rate when asleep was less than 10%. In cases like this where the child is awake and unsettled, sedation or anesthesia is often required in order to identify the signal of interest that is imbedded in the background noise (Reich & Wiatrak, 1996). Medical personnel must be involved in the administration of sedation to ensure safety and for compliance with standard-of-care due to health risks involved with the use of sedation. The cost of sedated ABR is very high. According to Hall (2007), sedated ABRs are extremely time/labor intensive in that numerous personnel are required for sedation including a physician, a registered or licensed nurse, and a professional who is skilled in airway management and cardiopulmonary resuscitation. Furthermore, Reich and Wiatrak (1996) reported in their survey study that most respondents indicated an extensive amount of time- 30 to 60 minutes- for proper sedation and difficulty keeping the patient sedated during the ABR collection process. In their study, they also examined the specific costs of anesthesia at the University of Alabama at Birmingham Hospital which included $180 for the anesthesia itself, $193 for an oximeter, and $16 for a cardiac monitor. If the ABR requires anesthesia it is generally done in an operating room which Reich and Wiatrak indicate costs an additional $431.70 and $180.20 for the recovery room. Outpatient processing fees cost approximately $121.70. The grand total for acquiring an ABR in an operating room under anesthesia is, therefore, $1122.60, not including the cost for the doctors and other personnel required. This reinforces the need for an ABR system that is not affected by the patient's muscle movement so that it can be carried out without the need for anesthesia, even in difficult to test populations.

While the risks of sedation to the patient are low, they do still exist. The most common agent used for sedation according to Reich and Wiatrak (1996) is oral chloral hydrate, but they indicate that it is cause for concern that the amount administered is not standardized. The American Academy of Pediatrics sites that the risks associated with infant sedation include hypoventilation, airway obstruction, apnea, laryngospasm, and cardiopulmonary impairment (as cited in Sokolov, 2007). Reich and Wiatrak (1996) also sited oxygen desaturation, rashes, respiratory arrest, and difficulty waking the child from sedation as possible problems associated with sedation. Furthermore, Hall (2007) states that anesthesia causes prolonged ABR wave latencies as a result of "interrupting transmission of neural impulses at the skeletal neuromuscular junction" (p. 310). In mice, it has been found that when using anesthesia in order to obtain accurate ABR recordings significant increases in absolute latencies of peaks I-III do occur (Van Looij et al, 2004). In this study, Van Looij et al. hypothesized that this increase in latency is due to the anesthesia slowing down the conduction time by paralyzing the eighth nerve. They also found that, under anesthesia, the data were more highly variable than when anesthesia was not used and that hearing thresholds were an average of 8 dB nHL worse under sedation. The increased risks and costs associated with sedation make a system for which sedation would not be required extremely beneficial to healthcare professionals.

Other factors besides noise can affect the reliability of threshold determination in ABR recordings. While ABR measurements require little participation on the part of the patient, the actual determination of whether wave V is present or absent is a subjective assessment on the part of the audiologist. Stapells (1998) indicates that audiologists often make the determination of whether a response is “absent” or “present” without sufficient information. He suggests that in order to determine that a waveform is present it must be replicable across the entire waveform and that in order for a waveform to be absent it must be flat. Noise, as discussed above, can degrade the quality of the waveform and can often be misinterpreted as a reliable waveform. If neither of these conditions are satisfied, Stapells indicates that more replications or a greater number of sweeps are required to accurately assess the response. Further, measures such as the correlation co-efficient between the two waveforms add objectivity to threshold determination with a correlation co-efficient of .5 or higher being indicative of a response. The Fsp measure provides further objective information which tells the audiologist when enough sweeps have been obtained to indicate a present or absent response. It is a statistical approach for determining the probability of a response being present. The Fsp value is an estimate of unaveraged noise and is calculated as the ratio of the variance across the average to the variance of a single point from sweep to sweep (Sininger, 1993). This technique not only provides for more objective measurement of ABR thresholds, but also stops the collection process as soon as a true neural potential is present as determined by the Fsp value. This can greatly reduce the length of time needed to collect ABR measurements. A higher Fsp is favorable and having an Fsp of greater than 3.1 is indicative of a response.

ABR- Kalman Weighted Filtering and In-Situ Pre-Amplification

The need for an ABR system that is resistant to electromagnetic and endogenous noise led to the introduction of the Vivosonic Integrity V500 (Vivosonic Incorporated; Toronto, Ontario). The Vivosonic Integrity V500 system is a wireless system that is designed to eliminate the need for sedated auditory brainstem response (ABR) testing by utilizing Kalman weighted filtering and in-situ pre-amplification. It employs the use of wireless technology, body-mounted pre-amplifiers, and short, shielded wires to reduce the effects of electromagnetic and muscular artifact.

The Vivosonic Integrity V500's amplifier (Amplitrode®) is placed directly on the ground electrode in an attempt to eliminate electromagnetic noise within the system. The Amplitrode® "integrates a preamplifier and electrode clip that snaps directly onto the ground electrode" (Sokolov et al., 2006). The fact that the preamplifier is directly attached to the ground electrode clip, unlike with a conventional ABR system where the preamplifier is connected to the electrodes via wires, eliminates some of the possibility of line noise (Sokolov et al., 2006). Having in-situ pre-amplification, with the lead wires, electrodes, and preamplifier mounted directly onto the head, reduces physiologic artifact as all components move as a unit (Kurtz & Sokolov, 2004).

Furthermore, the wires leading to the inverting and non-inverting electrodes are significantly shorter for the Vivosonic Integrity than the 3ft electrode leads used with conventional ABR systems because the preamplifier, as stated above, is mounted directly on the head. All of the aforementioned wires for the Vivosonic are electrically shielded, which further reduces line noise contamination (Kurtz & Sokolov, 2004).

Vivosonic Integrity uses wireless Bluetooth® technology to transmit the recordings from the interface mounted on the person's head to the computer (Sokolov et al., 2006). In a conventional ABR system, the interface is connected to the computer via an interfacing cable which introduces further electrical noise into the system. The reduction of electromagnetic noise through components such as wireless Bluetooth® technology, the Amplitrode®, and shielded wires is designed to make measuring in unshielded environments more feasible.

Finally, the Vivosonic Integrity V500 filters using Kalman-weighted filtering which is intended to dramatically reduce artifact caused by muscular and ocular movement allowing for clear, reliable results regardless of patient activity i.e. sedation will no longer be necessary for patients. For conventional ABR systems, sedation or anesthesia may be necessary if the patient is not very relaxed as physiologic artifact originates from muscular movement particularly in the head, neck and trunk (Sokolov, 2007). As previously stated, the objective of filtering is to "eliminate unwanted nonresponse activity (electrical and muscle interference or artifact) while preserving the actual response" (Hall, 2007, p. 90). A conventional ABR system uses a band pass filter to remove noise in the system after it has already been sent through the pre-amplifier, hence allowing in many different types of artifact. The pre-amplifier does not differentiate between noise and the response which results in amplification of both signals. Once the signals are sent to the preamplifier together, no analog or digital filter will be able to separate them. In a conventional ABR system, artifact rejection causes portions of a signal to be extracted from a recording altogether if the noise is determined to be too great (Kurtz & Steinman, 2005). Alternatively, Kalman-weighted filtering occurs *before* the signal gets to the preamplifier. This type of filtering assigns a greater weight to sweeps with less noise and less weight to sweeps where more noise is present. Markovsky, Amann, and Van Huffel (2008), describe Kalman filtering as a way of analyzing the quietest period during a given recording and giving it the largest weighting. It then compares each of the subsequent sweeps to that "quiet" benchmark and assigns an appropriate weight to that sweep in relation to the quietest sweep. This allows the system to come up with a linear model for assigning weights to sweeps. According to Sokolov (2007), Kalman filtering does not reject any sweeps; rather, it extracts a response from each sweep. The most information is extracted from recordings with less noise, but even the noisy sweeps are able to be used. With Kalman filtering, less time should be required to attain recordings as all sweeps are used to some extent. This, of course, assumes that all waveforms in a recording are of similar activity levels because a period of extreme quiet which would be assigned a weighting closer to 1 paired with other noisy sweeps which would be assigned a weighting closer to 0 would actually require a longer acquisition time because all of the noisy sweeps would be assigned a very small weighting given the short period of quiet with which they would be compared. The Kalman filter works most efficiently when activity level is consistent throughout acquisition.

There are many benefits to acquiring ABR recordings without the need for sedation and electrically shielded rooms. First, ABRs can be administered more frequently because the risks are drastically reduced thus ABRs can be used to regularly monitor hearing thresholds of infants who are being submitted to ototoxic treatments. Furthermore, the cost of acquiring ABRs will be drastically reduced if sedation is no longer needed.

The purpose of this study is to examine the accuracy and efficiency of Kalman weighted filtering and in-situ pre-amplification on obtaining reliable auditory brainstem response thresholds in the presence of physiologic noise without the need for sedation. The Vivosonic Integrity V500 will be utilized as the experimental equipment as it uses both Kalman weighted filtering and in-situ pre-amplification. In order to determine the accuracy and efficiency of thresholds obtained with the non-traditional amplifier placement and filtering techniques, auditory brainstem responses using tone burst stimuli at 500 Hz and 4000 Hz will be administered to a group of adults with normal hearing sensitivity (as defined by behavioral thresholds better than 20dB HL) using both a conventional system as well as the Vivosonic Integrity. Accuracy in this study means lower threshold (closer to 20dBnHL) as it is already established that the participants in this study had normal hearing. Efficiency in this study refers to how much time was required to attain a sensitive (low) measure of threshold. There were three main research questions examined in this study. First, in terms of threshold accuracy, does the system which uses Kalman weighted filtering and in-situ pre-amplification (Vivosonic Integrity) measure more accurate thresholds in quiet than the conventional system? Does it measure more accurate thresholds in the presence of physiologic noise than the conventional system? Finally, does the Vivosonic Integrity measure thresholds more efficiently than the conventional system in the presence of high physiologic noise? It is hypothesized that there will be no differences in threshold or in the efficiency with which threshold is acquired between the conventional ABR system and the Vivosonic Integrity V500 system when the participant is quiet and relaxed and that the Vivosonic Integrity V500 will predict more accurate thresholds, more efficiently than the conventional ABR system when both are measuring in the presence of muscular artifact.

Chapter III

METHODS

Participants

Twenty adults (16 females and 4 males) ages 21-27 ($\bar{x}$= 24yrs), served as participants in this study. Participants included college students obtained via word of mouth recruiting within the James Madison University Department of Communication Sciences and Disorders. In order to participate in the study, participants were required to have normal hearing acuity (air conduction thresholds at 500 and 4000Hz better than 20dB HL; DPOAE response >6dB above noise floor) and no middle ear pathology (Type A tympanograms). Both ears for each participant were initially tested and the better of the two ears was chosen as the test ear for the experimental ABR recordings. If both ears were tested and considered equivalent, the left ear was chosen as the test ear. Final results are thus reported on 18 left ears and 2 right ears. All participants provided informed consent before testing and could withdraw from the study at any time without penalty.

ABR Equipment

All participants were screened for normal hearing using behavioral testing (Grason-Stadler v.61 audiometer with ER 3A or 3B foam inserts depending on external auditory canal diameter), tympanometry (Grason-Stadler TympStar v.2) and distortion product otoacoustic emissions (Bio-logic Scout OAE system). Auditory brainstem thresholds were determined using both the Biologic Navigator and the Vivosonic Integrity V500.

ABR Test Parameters

Each system was set to use default parameters as can be seen in Table 1 below. All settings were identical between the two ABR systems with the exception of artifact rejection. The purpose of this study was to examine the differences in ABR threshold accuracy and efficiency using a conventional system (the Bio-logic) and a system that uses Kalman weighted filtering in place of artifact rejection (the Vivosonic).

Table 1.

ABR Test Parameters

	Bio-logic	*Vivosonic*
Filter	*100-1500*	*100-1500*
Channels	*One Channel*	*One Channel*
Electrode Montage	*Fpz/A1 or Fpz/A2*	*Fpz/A1 or Fpz/A2*
Artifact Rejection	*23.8 µV*	*Kalman filtering*
Intensity	*70dBnHL down to 20dBnHL*	*70dBnHL down to 20dBnHL*
Polarity	*Rarefaction*	*Rarefaction*
Stimulus	*Tone Burst*	*Tone Burst*
Rise/Plateau/Fall (Cycles)	*2-0-2*	*2-0-2*
Frequency	*500 and 4000 Hz*	*500 and 4000 Hz*

Filter Skirts	*12 dB/octave*	*12 dB/octave*
Transducer	*ER-3A*	*ER-3A*
Rate	*39.9*	*39.9*
Window	*25msec*	*25msec*
Windowing	*Blackman*	*Blackman*

Table 1. ABR Test Parameters. The only difference between the conventional system and the Vivosonic ABR system can be seen in the highlighted portion- Artifact Rejection. The conventional system uses a standard artifact rejection of 23.8 µV while the Vivosonic uses Kalman weighted filtering and in-situ pre-amplification.

ABR Test Procedure

Auditory brainstem response (ABR) thresholds were obtained from one ear in each participant for a 500 and 4,000 Hz tone burst both in quiet and under conditions of high physiological noise using both the Biologic Navigator and the Vivosonic Integrity V500 systems. Physiological noise was created by asking the participant to actively chew a regular size piece of Wrigley's Doublemint gum. Pilot data suggested that this condition was sufficient to provide clear differences in physiologic noise levels. Other means of creating physiologic noise were experimented with when collecting pilot data including having the participant put a puzzle together or read aloud during acquisition both of which created very little artifact. The eight different conditions are shown in Table 2 below along with the code for each condition.

Table 2

Eight Measurement Conditions

	Equipment	Activity	Frequency	Coding
1	Bio-logic Navigator	Quiet	500 Hz	b5Q
2	Bio-logic Navigator	Quiet	4000 Hz	b4Q
3	Bio-logic Navigator	Noise	500 Hz	b5N
4	Bio-logic Navigator	Noise	4000 Hz	b4N
5	Vivosonic Integrity	Quiet	500 Hz	v5Q
6	Vivosonic Integrity	Quiet	4000 Hz	v4Q
7	Vivosonic Integrity	Noise	500 Hz	v5N
8	Vivosonic Integrity	Noise	4000 Hz	v4N

Table 2. Eight Measurement Conditions.

Participants were seated in a recliner in a sound proof booth. For the Bio-logic Navigator, the electrodes were plugged into the head box (the electrodes for the Vivosonic are permanently attached to the head box). The skin was prepped and electrodes adhered to the skin at Fpz, Fz, and A1 or A2 as described above. Zinc disc electrodes were used for the Bio-logic Navigator device and standard disposable electrodes were used for the Vivosonic Integrity device. Prior to collection, impedance values were measured. The standard for impedance is that each individual electrode must be less than 5 kΩ and the difference between any two electrodes must be less than 2 kΩ. For the purposes of this study, impedance values were required to be less than 5kΩ with no greater than 1kΩ difference in impedance values between electrodes regardless of the type of electrode used by the system. If impedances were not found to be within these ranges, the electrode site was re-scrubbed until adequate impedance values were obtained. The EA-3 inserts were inserted into the participant's ear so that the lateral edge of the insert was flush with the opening of the external auditory canal.

For each trial, intensity began at 70 dBnHL and decreased to 40 dBnHL, 30 dBnHL, and finally to 20 dBnHL. The 20dBnHL intensity level was repeated on each participant as a measure of replicability. Testing lasted approximately two and one half hours per participant and each participant was tested under all eight conditions (Table 2) in a single test session. During the first trial, participants were asked to relax with their eyes closed and to remain relaxed throughout testing. This trial was coded as “quiet.” For the second trial, participants were asked to create muscle activity by chewing gum when undergoing measurement to create potential myogenic interference with ABR measurement. This trial was coded as “noisy.” The order of testing (both equipment and condition) was randomized by flipping a coin prior to testing. Heads indicated that the Vivosonic equipment would be used first and tails indicated that the Bio-logic equipment would be used first. The coin was then flipped again to determine if the noisy or quiet condition would be conducted first. Heads indicated that the quiet condition would be tested first and tails indicated that the noisy condition would be tested first. Further, the noisy and quiet conditions were rotated (i.e. quiet 4000 Hz, noisy 4000 Hz, quiet 500 Hz, noisy 500 Hz) throughout testing so as to give the participants’ jaws a chance to relax between noisy conditions to reduce carryover effects.

Data Collection and Analysis

Data analysis was slightly different for the two machines due to the limitations of each system. Due to the fact that the Bio-logic Navigator cannot be set to stop acquisition at a given number of sweeps when using both accepted and rejected sweeps, a time interval rather than absolute sweeps was used to evaluate efficiency in order to make each ABR system equivalent. For the Bio-logic, during the quiet trials, averages were taken in 15 second intervals and then summed sequentially post hoc to obtain composite waveforms at 15 seconds, 30 seconds (first 15 seconds + second 15 seconds, etc.), 45 seconds, and 60 seconds in order to determine the fewest number of sweeps needed to obtain a reliable waveform. During the noisy trials, averages were taken in 30 second intervals and then added together post hoc to obtain composite waveforms at 30 seconds, 60 seconds, 90 seconds, and 120 seconds. Time intervals were increased for the noisy trials based on pilot data indicating that no results were obtained for either set of equipment within the sixty seconds allotted in the quiet trials.

Waveforms recorded via the Vivosonic Integrity are unable to be summed following acquisition; therefore, the waveform was fully obtained and then divided post hoc into 15 second, 30 second (first 15 seconds + second 15 seconds, etc.), 45 second, and 60 second intervals for the quiet trial and into 30 second, 60 second, 90 second, and 120 second intervals for the noisy trial. This post hoc deconstruction of waveforms could not be done by the investigator using the Vivosonic Integrity unit on site. Therefore all waveforms were sent via a secure FTP server to Vivosonic Incorporated headquarters so that they could be divided into time intervals. Waveforms were deconstructed to appropriate time intervals by Dr. Aaron Steinman, VP for Research, and transmitted to the investigator by placing them on a secure FTP server. Dr. Steinmen and the Research Unit of Vivosonic Incorporated agreed to this arrangement prior to initiation of testing. To ensure accuracy of the post-hoc waveforms received from Vivosonic, screen captures were taken randomly on site prior to sending waveforms at several points throughout acquisition and then compared to the processed waveforms that were returned to the investigator by Vivosonic Incorporated via the FTP server. For example, a screen shot taken at 500 Hz during a quiet condition following 30 seconds of acquisition time was compared to the post-hoc waveform at 30 seconds which was returned by Vivosonic Incorporated. They were analyzed by the investigator to ensure that the waveforms had equivalent latencies for the major peaks, equivalent amplitudes for the major peaks, and identical morphologies. No discrepancies in returned waveforms were detected. This methodology for post hoc data analysis allowed data collected from each ABR system to be comparable in the time domain thus allowing efficiency comparison between each machine under each condition.

Threshold Determination.

Three experienced audiologists served as independent Reviewers of the ABR data. ABR data were compiled into "flipbooks" so that Reviewers could view all waveforms under all conditions independently and without knowledge of condition. Flipbooks were composed of printouts of each 15 second (30 seconds in noise) sequentially added waveform for each condition. Each of the books had eight different subsections corresponding to each of the eight conditions (Table 2). The flipbooks were created so that within each of the subsections there were four smaller sections which corresponded to each of the intensities (70, 40, 30, and 20 dBnHL) and each of *those* sections had four sections corresponding to each of the summed waveforms (i.e. 15, 30, 45, and 60 seconds in quiet or 30, 60, 90, and 120 second in noise. Reviewers were instructed to begin at 70 dB HL on panel A (15/30 second recording). They were to make a determination of whether or not a wave V was observed. If it was not observed, they were instructed to flip to panel B (30/60 second recording). If wave V was still not observed, the Reviewer was instructed to flip to panel C (45/90 second recording) and so forth. If a wave V was observed on panel B, the Reviewer was instructed to circle “B” on the Reviewer response sheet. Once a wave V was observed, the Reviewer was instructed to move to 40 dB HL, then 30 dB HL, and finally to 20 dB HL. If a Reviewer did not see a wave V at a given intensity, the last intensity where a wave V was observed was considered to be threshold. The Reviewer was then instructed to simply flip to the next condition in the book. An average of the thresholds from each Reviewer was used as true threshold. The length of time was obtained from all three Reviewers and then averaged to obtain the mean length of time. Finally, each Reviewer was also asked to indicate whether or not they felt the threshold chosen was reliable enough to be used in clinical practice by indicating a simple ‘yes’ or ‘no’ on the data sheet. The Reviewers were blind to the patient identity, the system used to perform the test, and whether the patient was moving or quiet. They were given the frequency of the tone burst used to obtain the waveforms and normative data to facilitate more accurate selection of threshold.

An example of the first page of the flip book can be found in Figure 1 (left panel) below. This is an example of the first page the Reviewer was shown where all four panels arranged vertically down the page are showing panel A. In quiet, panel A always corresponded to 15 seconds (in noise, 30 seconds), panel B to 30 seconds, etc. Figure 1 (right panel) below is an example of how a Reviewer might finish with a section. The top set of panels (70 dB) is set on C which corresponds to 45 seconds, the set of panels below that one (40 dB) is set on D which corresponds to 60 seconds, the third set of panels (30 dB) is set to A which corresponds to 15 seconds, and the fourth (bottom) set of panels (20 dB) is set to B which corresponds to 30 seconds. If this were an actual judgment made by a Reviewer, the Reviewer would have indicated a threshold of 20 dBnHL and an acquisition time of 2 minutes and 30 seconds. Figures two and three below represent the first page a Reviewer would see for 4000 Hz in noise and 500 Hz in quiet (left panel) and in noise (right panel), respectively.

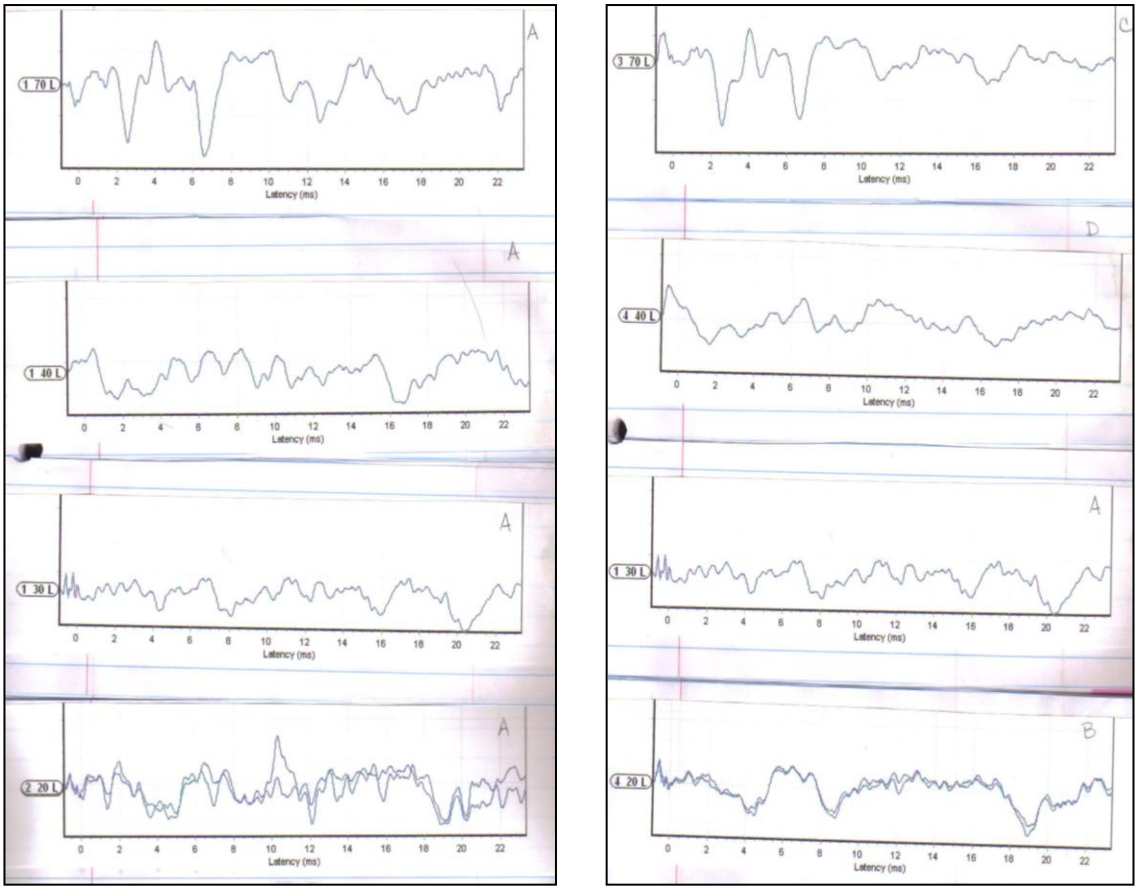

Figure 1. Sample ABR Waveforms. Left: 4000 Hz quiet, represents the first page a Reviewer would see. Right: 4000 Hz quiet, hypothetical final panel a Reviewer may choose corresponding to an acquisition time of 2 ½ minutes.

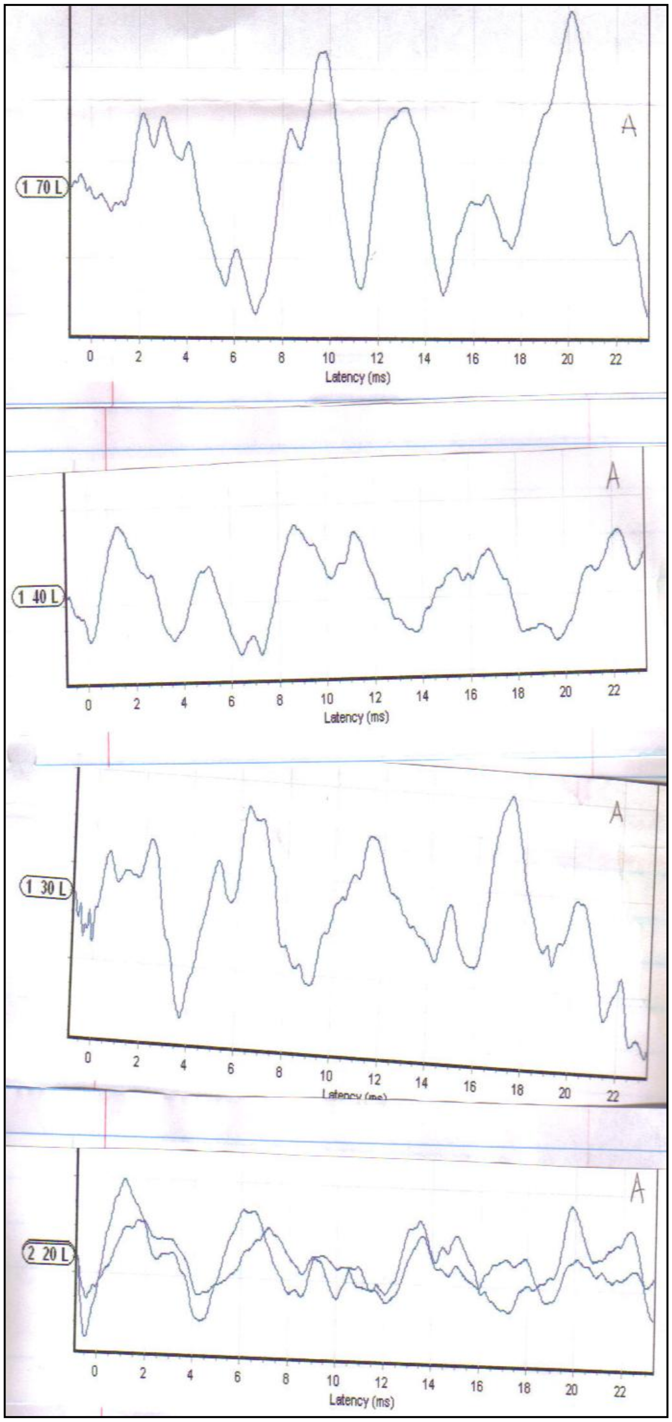

Figure 2: Sample ABR Waveforms. 4000 Hz Noise

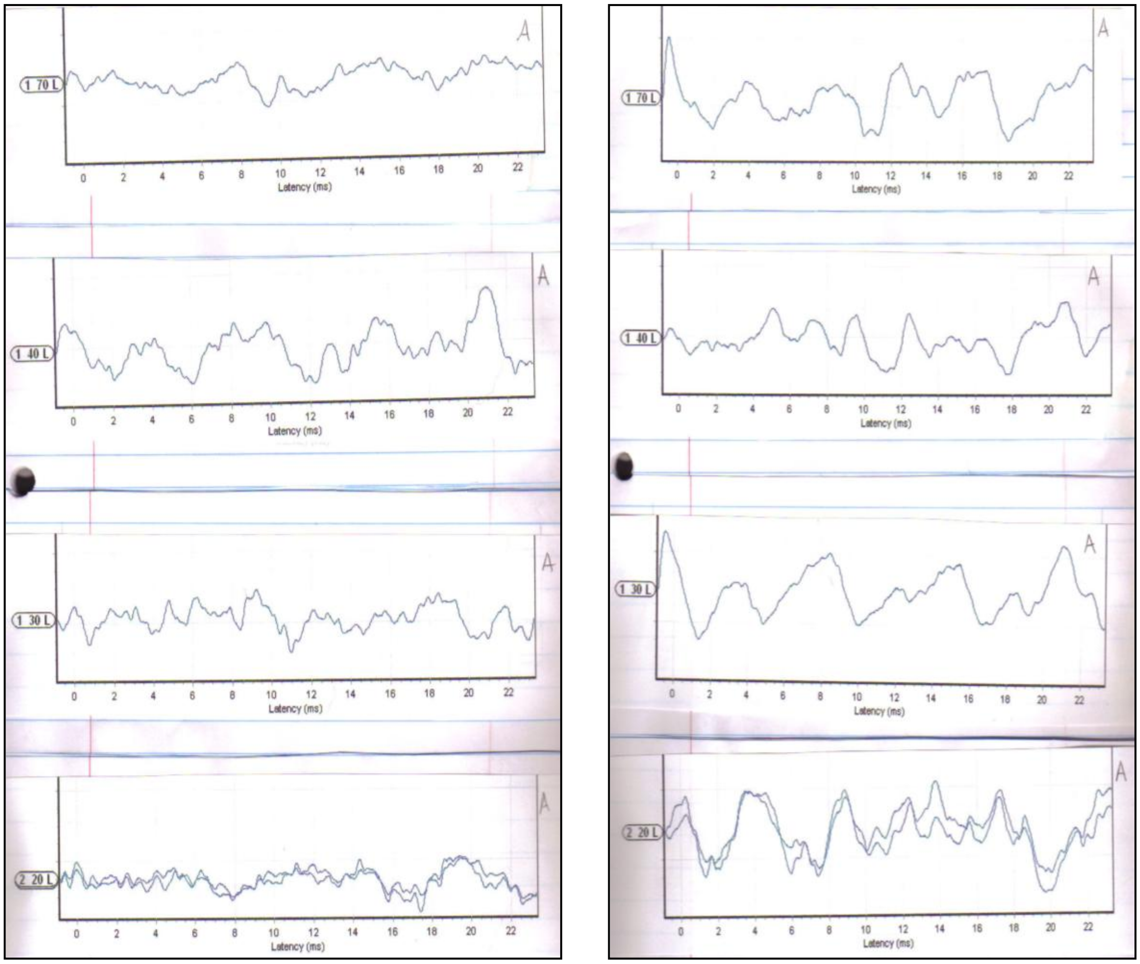

Figure 3: Sample ABR Waveforms. Left: 500 Hz quiet. Right: 500 Hz noise.

Chapter IV

Results

One hundred and sixty ABR's (8 for each of twenty subjects) were evaluated by a panel of three Reviewers. All Reviewers were blind to condition (quiet/noise/ Biologic/Vivosonic). Each Reviewer determined the presence of wave V in ABR waveforms that had been sequentially added in 15 second (quiet) or 30 second (noise) intervals using Vivosonic Integrity and Biologic Navigator evoked potential units in order to determine threshold. Results were analyzed for both threshold accuracy and efficiency and quantitative reliability of responses.

Qualitative Evaluation of Reliability

Reviewers selected the intensity and 15 (or 30) second interval necessary to obtain threshold. Threshold was defined as the lowest intensity at which an ABR wave V was present. If the Reviewers could not determine threshold they were asked to indicate "ABR not present". For statistical analysis, in cases where the Reviewer indicated that an ABR was not present at the highest level of stimulation (70dBnHL) a maximum value of 80dBnHL was assigned and a maximum time of 480 seconds in noise or 240 seconds in quiet was assigned. A total of 38 responses were coded as No Response and, thus, assigned a maximum value (Figure 4). There were 25 "No Responses" for the Bio-logic Navigator and 13 "No Responses" for the Vivosonic Integrity. A chi-square analysis was run to determine if these differences in "No Response" data by equipment were considered significant. There was a significant difference between the equipment, $\chi^2(1, N=480) = 4.115$, $p = .042$ with the Vivosonic yielding significantly fewer "No Responses" than the Bio-logic. The most notable differences in "No Response" data were at 4000 Hz in noise where the Vivosonic Integrity had zero "No Responses" and the Bio-logic Navigator yielded ten "No Responses" and at 500 Hz in noise where the Vivosonic Integrity had five 'No Responses' and the Bio-logic Navigator had eleven "No Responses". Chi-square analysis was conducted for each paired condition (i.e. Bio-logic at 500 Hz in Noise vs. Vivosonic at 500 in Noise, etc.). There were no significant differences between any of these conditions with the exception of Vivosonic at 4000 Hz in Noise and Bio-logic at 4000 Hz in Noise, χ^2 (1, *N*=240) = 10.44, *p* = .001.

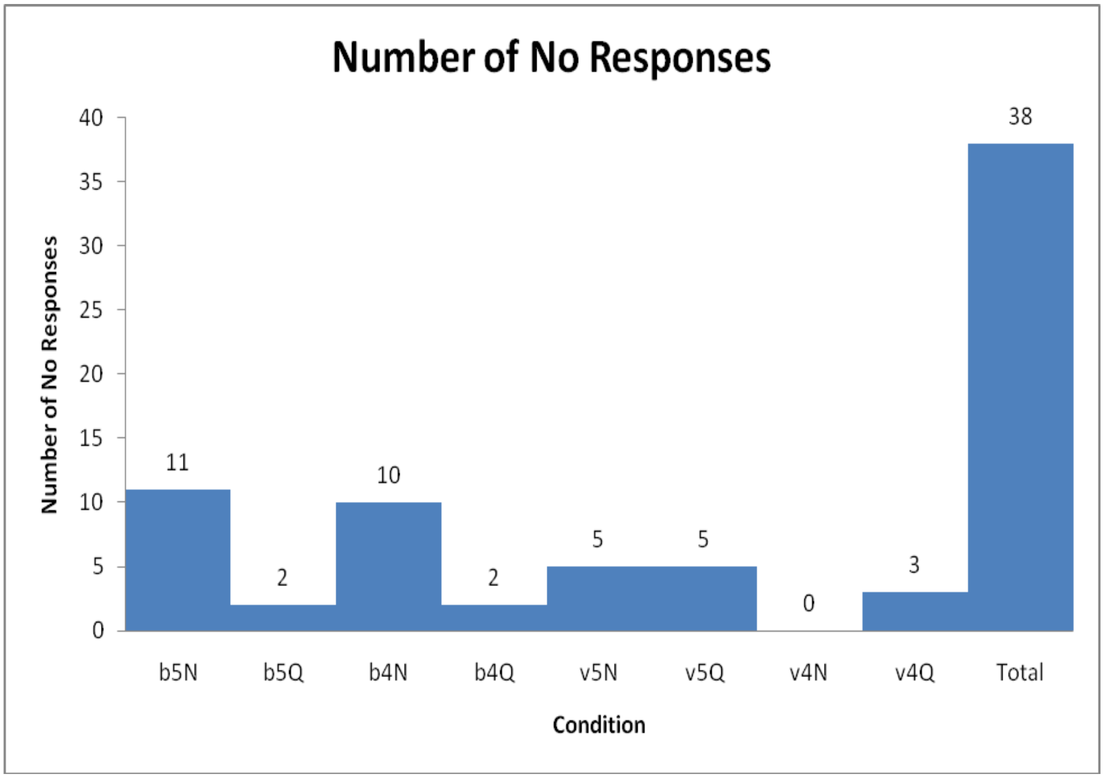

Figure 4. Number of No Responses by Condition. Of the total 38 No Responses by the Reviewers, 25 were for the Bio-logic Navigator equipment and 13 were for the Vivosonic Integrity equipment.

The “No Response” data were further divided to examine the difference between Reviewers. In order to determine whether there were significant differences in Reviewer judgments for the presence of an ABR response, the number of “No Response” were determined as a function of Reviewer. Below, in Figure 5, the number of “No Responses” can be seen by Reviewer. A chi-square analysis was run to determine if these differences in Reviewers were considered significant. There was a significant difference between the Reviewers, χ^2 (2, *N*=480) = 9.317, *p* =.009. There is a clear trend for Reviewer 3 (with 21 “No Responses”) to be far more conservative in judging the presence of Wave V than Reviewers 1 and 2 who indicated 7 and 10 No Responses, respectively.

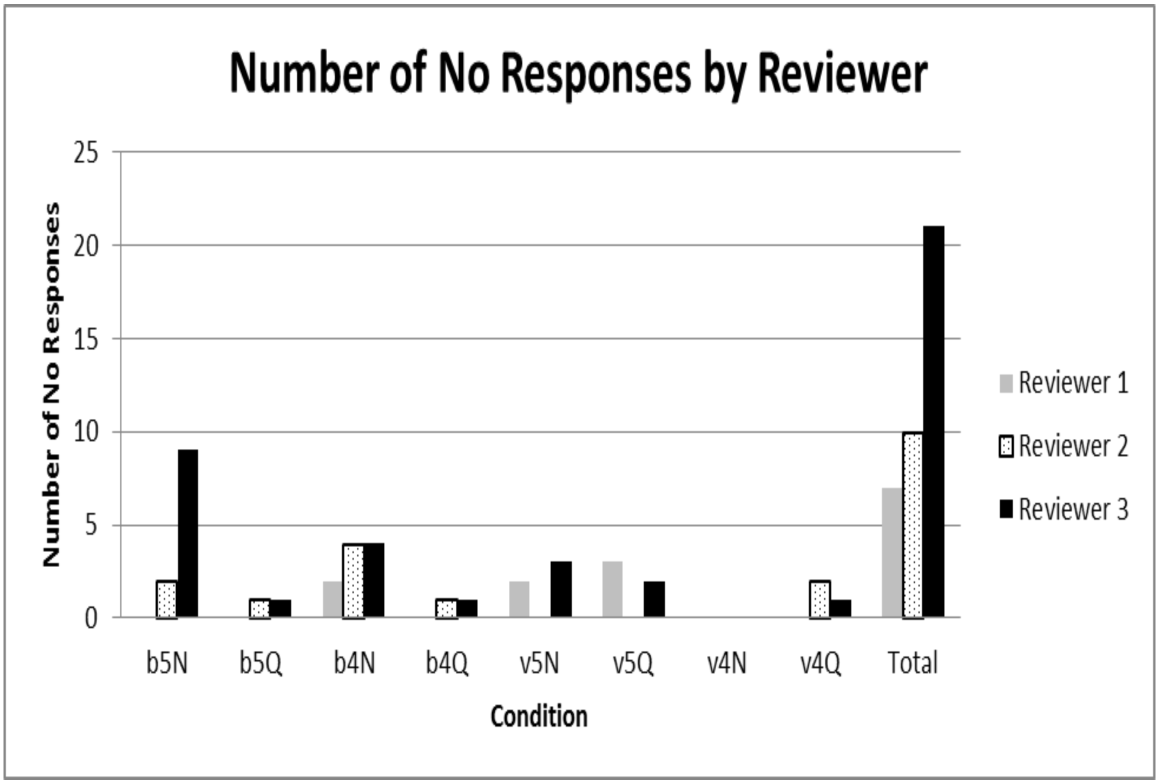

Figure 5. Number of “No Responses” by Reviewer. Reviewer 3 is significantly more conservative than Reviewers 1 and 2. (Blue=Reviewer 1, Red=Reviewer 2, Green=Reviewer 3)

Reviewers were asked during evaluation to rate each waveform as reliable i.e. “I’d use this in my practice” or unreliable i.e. “I’d want to do more testing.” Of the 120 responses for each category (480 total responses)- Bio-logic in Quiet, Bio-logic in Noise, Vivosonic in Quiet, and Vivosonic in Noise- 75 were deemed reliable for the Bio-logic in Quiet, 54 were deemed reliable for the Bio-logic in Noise, 82 were deemed reliable for the Vivosonic in Quiet, and 78 were deemed reliable for the Vivosonic in Noise.

A chi-square test was run to evaluate the significance of the difference between the reliability ratings of these different conditions. There was a significant difference in reliability judgment as a function of activity level for the Bio-logic Navigator. For Bio-logic in Quiet and the Bio-logic in Noise, χ^2 (1, N=240) = 7.392, p = .007; however, there was no significant difference in judgment as a function of activity level for the Vivosonic Integrity, χ^2 (1, N=240) = .300, p = .584.

For Bio-logic in Noise and Vivosonic in Noise, χ^2 (1, N=240) = 9.70, p = .002 and for Bio-logic in Quiet and Vivosonic in Quiet, χ^2 (1, N=240) = .902, p = .342. There was a significant difference between the two sets of equipment in noise indicating that Reviewers rated the Vivosonic waveforms significantly more reliable than waveforms obtained using the Bio-logic. Reviewers indicated that both systems were equally reliable in quiet.

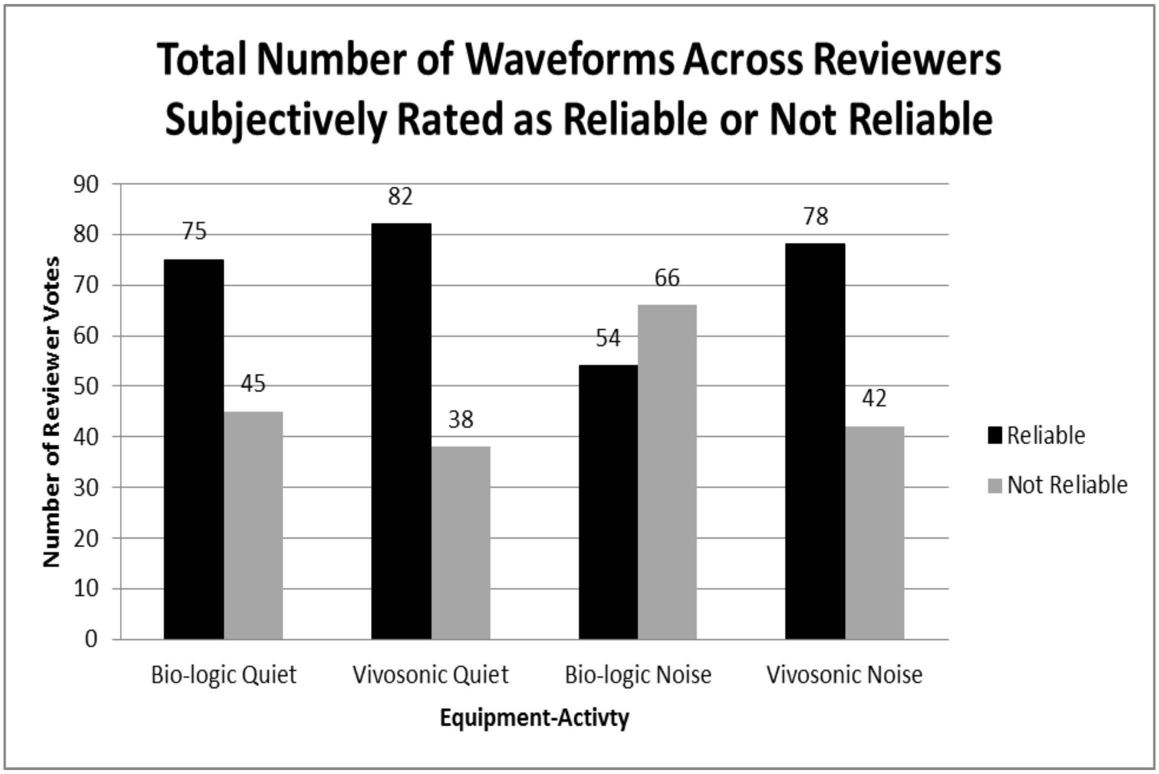

Figure 6. Total Number of Waveforms across Reviewers Subjectively Rated as Reliable or Not Reliable. Figure shows that the quiet conditions were rated more reliable regardless of equipment and that the thresholds measured using the Vivosonic were rated more reliable than those obtained with the Bio-logic regardless of activity level.

Results were further divided by Reviewer in order to determine any significant inter-Reviewer differences. Reviewer 3 consistently indicated that thresholds were unreliable more often than Reviewer 1 and 2, regardless of condition. In all conditions, except Reviewer 3 in quiet, Reviewers indicated that waveforms obtained using the Bio-logic equipment were less reliable than those thresholds obtained with the Vivosonic equipment. For example, Reviewer 1 indicated 15 Unreliable waveforms for the Bio-logic in quiet and only 12 for the Vivosonic in quiet. Likewise, Reviewer 1 indicated 14 Unreliable waveforms for the Bio-logic in Noise and only 8 for the Vivosonic in Noise. These results can be found in Figure 7 below.

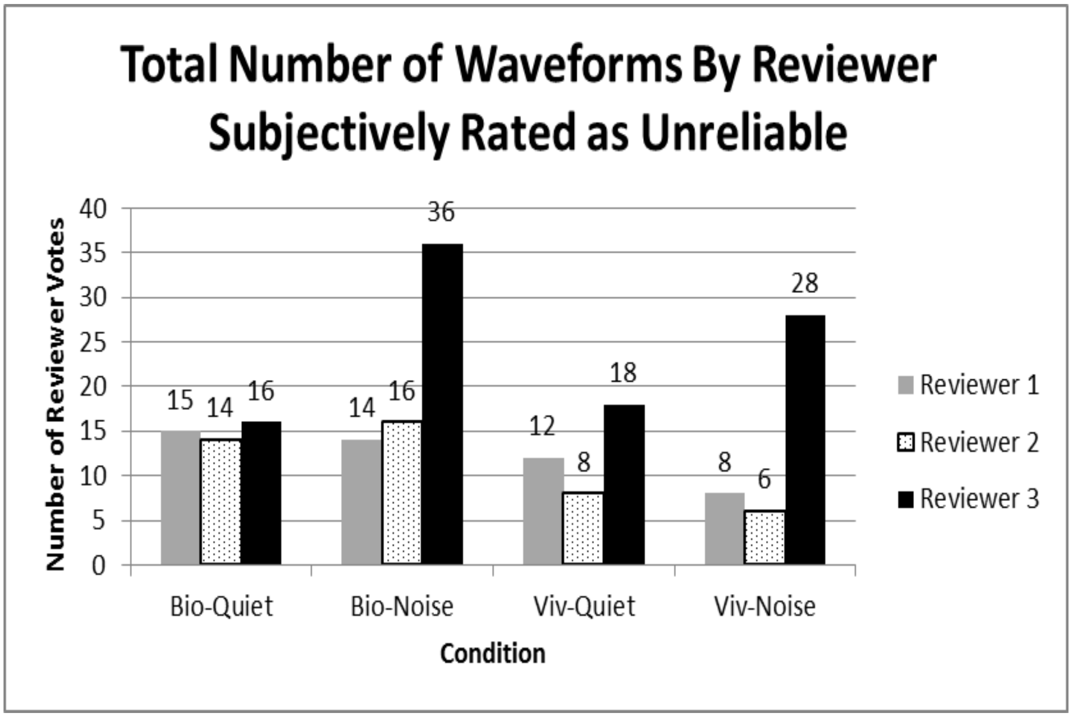

Figure 7. Total Number of Waveforms by Reviewer Subjectively Rated as Unreliable. Figure indicates that Reviewer 3 was consistently more conservative in judging ABR waveforms.

Behavioral threshold vs. ABR threshold

The average observed ABR thresholds for each condition as well as the difference between the average ABR threshold and the average behavioral threshold at that frequency can be found in Table 3 when No Responses are included (top section) and when No Responses are excluded (bottom section). It is expected that thresholds obtained using a 4000 Hz tone burst will be closer to true behavioral threshold than thresholds obtained using a 500 Hz tone burst (Beattie et al., 2005). Likewise, Stapells (2000) indicates that participants with normal hearing should have ABRs within 10-20 dB of behavioral thresholds and that ABR thresholds measured with a 4000 Hz tone burst are, on average, 7 dB better (lower) than those obtained with a 500 Hz tone burst. As the lowest obtainable threshold in this study is 20 dB HL, the ABR thresholds are not expected to match Beattie et al. (2005) predicted values because there are no thresholds below 20 dBnHL to bring the average down; however, they are expected to follow the same *trend*. As previously mentioned, Beattie et al. found that 85% of ABR thresholds obtained at 500 Hz are within 16dB of the behavioral threshold and 85% of ABR thresholds obtained at 4000 Hz are within 9 dB of the behavioral threshold. Beattie et al. (2005) allowed for acquisition at intensities below 20 dBnHL (they allowed intensity levels as low as necessary to obtain threshold); therefore, it is expected that their findings would more closely correlate with behavioral thresholds than the present study. In the present study, average behavioral thresholds at 500 Hz were 4.75 dB HL (SD=4.36) and the average ABR threshold (excluding No Responses) for the Bio-logic in quiet was 42.41 dBnHL and for the Vivosonic in quiet was 41.64 dBnHL. Behavioral thresholds at 4000 Hz were obtained on average at 5.5 dB HL (SD=4.76) and the average ABR threshold found for the Bio-logic in quiet was 34.83 dBnHL and for the Vivosonic in quiet was 27.89 dBnHL. The data shows a similar trend as that found by Beattie et al. (2005) and by Stapells (2000) where thresholds obtained with a 4000 Hz tone burst are lower (more accurate) that those obtained with a 500 Hz tone burst. In quiet, the difference between average ABR thresholds at 500 and 4000 Hz for the Bio-logic was 7.4 dB HL and for the Vivosonic it was 14.3 dB HL. In noise, the difference between average ABR thresholds at 500 and 4000 Hz for the Bio-logic was 1.7 dB HL and for the Vivosonic it was 9 dB HL. This indicates that in all conditions, thresholds obtained with the 4000 Hz tone burst were more accurate that those obtained with the 500 Hz tone burst.

Table 3

Comparison of ABR and Behavioral Thresholds

Including No Responses	**b5Q**	**b4Q**	**b5N**	**b4N**	**v5Q**	**v4Q**	**v5N**	**v4N**
Average ABR threshold	43.7	36.3	47.0	45.3	44.8	30.5	40.3	31.3
SD	19.2	17.4	21.8	21.0	22.4	18.5	22.3	18.5
ABR threshold-Behavioral threshold	38.9	31.6	42.3	39.8	40.1	25.0	35.6	25.8
Excluding No Responses								
Average ABR threshold	42.4	34.8	39.6	38.4	41.6	27.9	36.7	31.3
SD	18.3	15.6	16.7	15.4	20.6	15.0	19.6	18.5
ABR threshold-Behavioral threshold	37.7	29.3	34.8	32.9	36.9	22.4	32.0	25.8

Table 3. Mean ABR thresholds obtained for each condition. Top: ABR and Behavioral threshold averages with No Response data included. Bottom: ABR and Behavioral threshold averages with No Response data excluded.

In the present study, the specificity of each system for each condition was obtained by determining the percentage of ABR thresholds within 20 dB of behavioral thresholds (Table 4). Only 22% of ABR thresholds obtained at 500 Hz in quiet using the Bio-logic Navigator are within 20 dB HL of the behavioral threshold and 35% of ABR thresholds obtained at 500 Hz in quiet using the Vivosonic are within 20 dB HL of the behavioral threshold. In noise at 500 Hz, only 20% of ABR thresholds obtained using the Bio-logic Navigator are within 20 dB HL of the behavioral threshold and 45% of ABR thresholds obtained using the Vivosonic Integrity are within 20 dB HL of the behavioral thresholds. At 4000 Hz in quiet, only 38% of ABR thresholds measured using the Bio-logic Navigator are within 20 dB HL of the behavioral threshold and 73% of ABR thresholds obtained using the Vivosonic Integrity are within 20 dB HL of the behavioral threshold. At 4000 Hz in noise, only 23% of ABR thresholds measured using the Bio-logic Navigator are within 20 dB HL of the behavioral threshold and 63% of ABR thresholds obtained using the Vivosonic Integrity are within 20 dB HL of the behavioral threshold. The intensity 20 dB HL was used as a bench mark because the present study did not permit a stimulus level below 20 dB HL; therefore, it would skew the data to determine the percentage of ABR thresholds obtained within 9 dB HL of the behavioral threshold as Beattie et al. (2005) reported.

Table 4

Percentage of ABR Thresholds Within 20 dB of Behavioral Thresholds (Specificity)

	Bio-logic	**Vivosonic**
500 Hz Quiet	22%	35%
500 Hz Noise	20%	45%
4000 Hz Quiet	38%	73%
4000 Hz Noise	23%	63%

***Table 4*. % of ABR thresholds within 20 dB HL of Behavioral Thresholds**

Correlations.

Correlations were calculated between behavioral and ABR thresholds for each condition. Very weak correlations were noted for all behavioral and ABR thresholds. Weak negative correlations were noted for all Vivosonic conditions with the exception of 500 Hz in quiet which indicated a weak positive correlation (r=.064, p=.682).

ABR threshold was determined as the lowest intensity at which wave V continued to be present. In order to determine the presence of wave V, Reviewers were asked to mark the first flipbook panel at each intensity where they felt certain a wave V was present. Because the lower threshold would require a longer acquisition time, a negative correlation between the two variables- threshold and efficiency- was expected. No significant correlation, *r*(480)= -.073, *p*=.131, was found between time and threshold for all conditions when all collected data was used in the analysis (including No Response data). The Vivosonic data alone yielded a slightly stronger negative correlation, *r*(240)= -.290, *p*<.001, while the Bio-logic data alone yielded a very weak positive correlation, *r*(240)= -.098, *p*=.131. Further consideration of these correlations revealed that the inclusion of “No Response data” likely skewed the results as a “No Response” was assigned a high threshold (80dB) and a long acquisition time (four minutes for quiet conditions, eight minutes for noisy conditions). In fact, this response should have resulted in a short time to acquire threshold. The “No Response” data skews the correlation as it is the opposite of the rest of the data i.e. it is expected that a low threshold will have a longer acquisition time and a high threshold will have a shorter acquisition time; however, “No Responses” are coded as a high threshold (80 dBnHL) and a long acquisition time (four minutes in quiet, eight minutes in noise). This causes the correlations to be deceivingly low. The Vivosonic only had 13 “No Responses”; therefore, it still indicates a negative correlation, albeit a weak one. The Bio-logic had 25 “No Responses” and, therefore, was much more affected in the above calculations by the “No Response” data.

Due to the weak correlations found with all of the data included, an analysis was completed which removed all of the “No Response” data and the results were recalculated. These results were analyzed with all of the “No Response” data removed due to the low correlations between time and threshold obtained in all conditions with the “No Response” data included. For the following data, the “No Responses” have all been removed from analysis.

A strong negative correlation (low threshold, long acquisition time or high threshold, short acquisition time) was found for all data combined, $r(440) = -.607$, $p<.001$ (Figure 8 below). This means that 36.8% of the acquisition times can be explained by the threshold obtained.

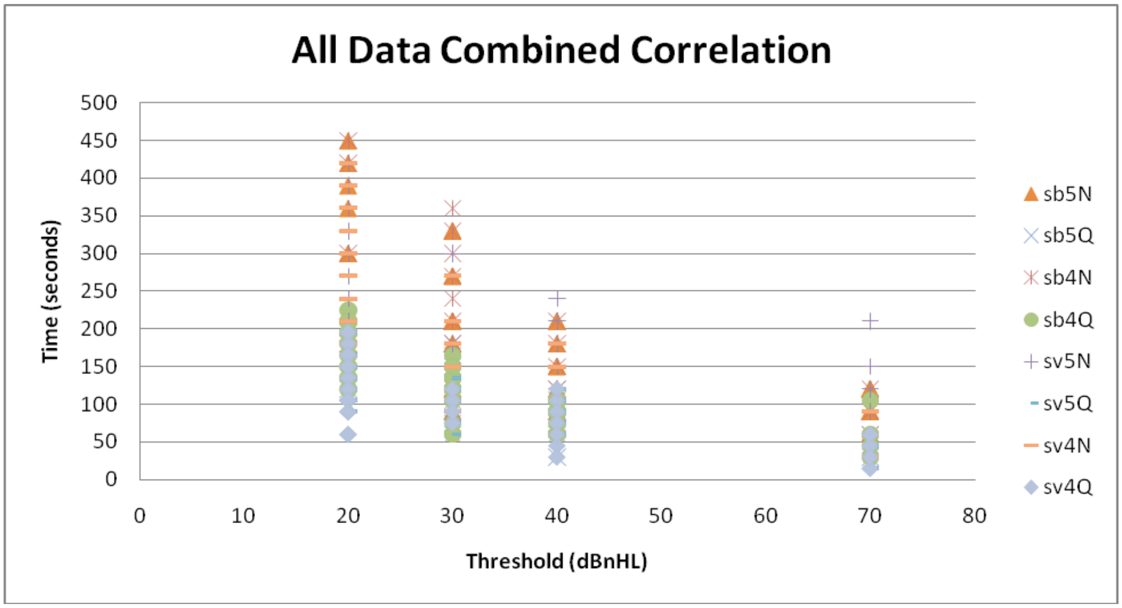

Figure 8. Scatterplot of all data combined (except No Responses). Figure shows a strong, negative correlation. In most cases the longest time to threshold was obtained for a 500Hz tone pip, in noise using the Biologic Navigator evoked potential unit.

Further, when sorted for equipment, a strong negative correlation was found for threshold and time for the Bio-logic Navigator, *r*(215)= -.587, *p*<.001, as well as for the Vivosonic Integrity, *r*(227)= -.628, *p*<.001 (Figures 9 and 10, respectively). The correlation of threshold and time for the Bio-logic in quiet, *r*(97)= -.728, *p*<.001, is nearly identical to the same correlation for the Vivosonic in quiet, *r*(113)= -.784, *p*<.001. An even stronger negative correlation is noted in noise than in quiet for the Bio-logic, *r*(114)= -.796, *p*<.001, and the Vivosonic, *r*(110)= -.799, *p*<.001. A strong negative correlation is found when analyzing the Bio-logic at 500 Hz, *r*(105)= -.620, *p*<.001, and at 4000 Hz, *r*(106)= -.542, *p*<.001. Similarly strong negative correlations were found for threshold and time when analyzing the Vivosonic at 500 Hz, *r*(108)= -.660, *p*<.001, and at 4000 Hz, *r*(115)= -.582, *p*<.001.

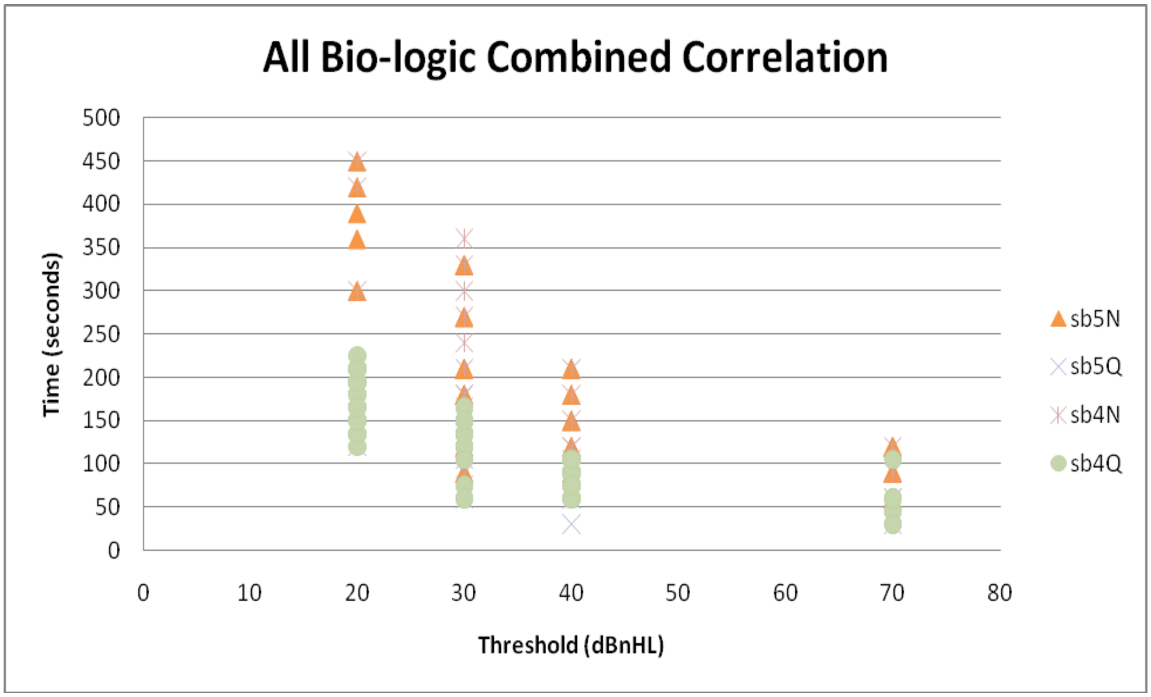

Figure 9. Scatterplot of all Bio-logic data (except No Responses). Figure shows a strong negative correlation (r=-.587) for threshold and time.

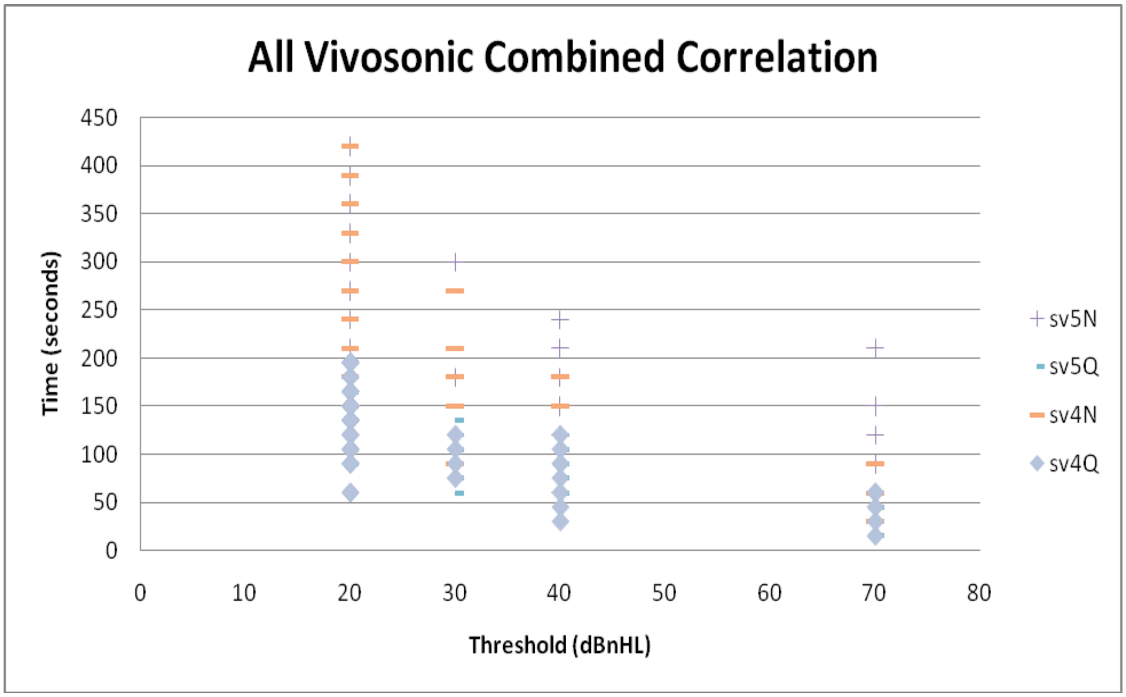

Figure 10. Scatterplot of all Vivosonic data (except No Responses). Figure shows a strong negative correlation (r=-.628) for threshold and time.

ABR Threshold: Accuracy and Efficiency

A repeated measures ANOVA was used to determine whether there was a significant difference in ABR threshold as a function of equipment (excluding "No Response" data). A significant main effect for equipment was found between all of the test conditions collectively performed with the Bio-logic Navigator and all of the test conditions collectively performed with the Vivosonic Integrity, *F*=(1, 59)=11.55, *p*=.001. The Vivosonic Integrity ($\bar{x}$ =36.75 dBnHL) was found to measure significantly lower thresholds than the Biologic Navigator ($\bar{x}$ =43.08 dBnHL) (Figure 11).

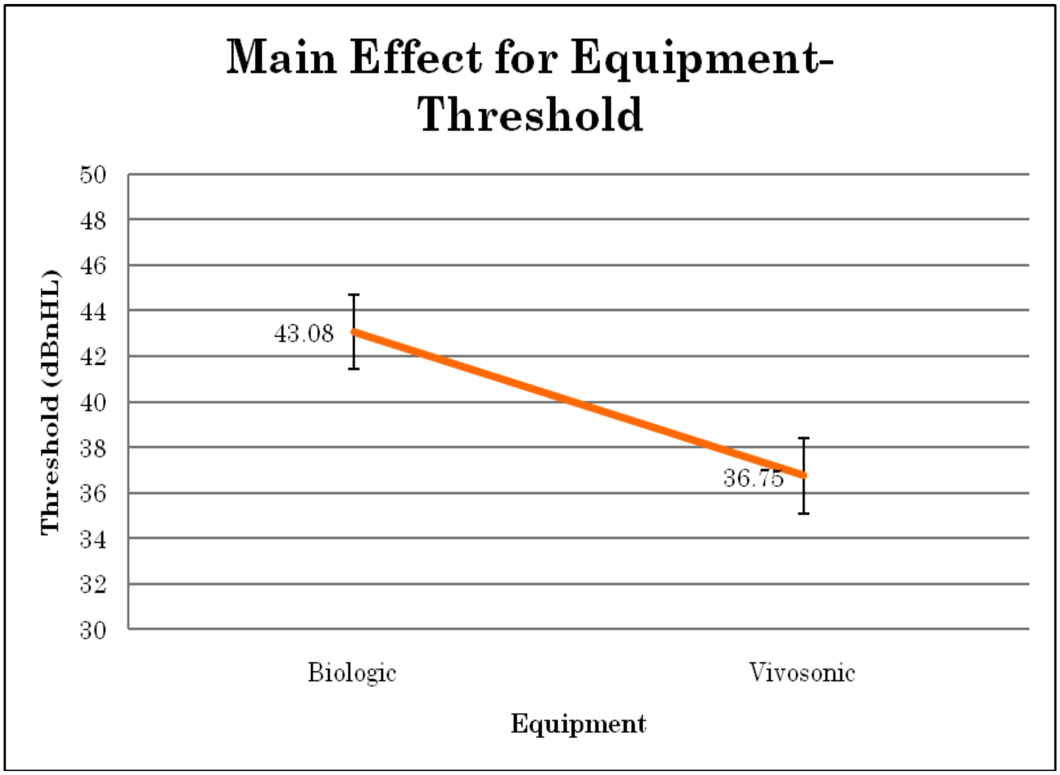

Figure 11. Main Effect for Equipment (Threshold Only). Figure shows that the Vivosonic Integrity measured significantly lower (more accurate) thresholds than the Bio-logic Navigator system.

This main effect, however, must be cautiously interpreted due to the fact that a significant interaction was found between equipment and Reviewer, $F(2,57)=13.046$, $p<.001$. The effect of equipment on threshold depended on the Reviewer. Reviewer One's thresholds indicated an opposite trend (with the Bio-logic Navigator measuring lower thresholds overall than the Vivosonic Integrity) than the other two Reviewers who each indicated trends for the Bio-logic Navigator to measure higher thresholds overall than the Vivosonic Integrity. Because of this interaction, further exploration was needed of the various variables- frequency, equipment, and activity level- to determine if a true effect was observed. The variability of Reviewer threshold estimation was expected as ABR threshold estimation is a subjective task. The fact that the Reviewers vary should not be viewed as a weakness of the study, rather it is a weakness of ABR measurements as a clinical utility. Because the purpose of this study was to assess how each system functions in a clinical setting where, inherently, there are clinicians with varying levels of clinical skills, the authors felt it was not clinically accurate to remove any of the Reviewers from the present study.

Below, Figure 12 depicts the average thresholds obtained for each of the four conditions- Bio-logic in Noise, Bio-logic in Quiet, Vivosonic in Noise, and Vivosonic Quiet. Of note, the intra-rater variability is consistent across conditions and equipment types; however, inter-rater variability is greater for the Bio-logic equipment. Noted in Figure 12 is the strong Reviewer-by-condition interaction which makes the main effect difficult to interpret. Although a strong main effect for equipment was noted when the "No Response" data were included, these results should be interpreted cautiously due to the strong Reviewer interaction.

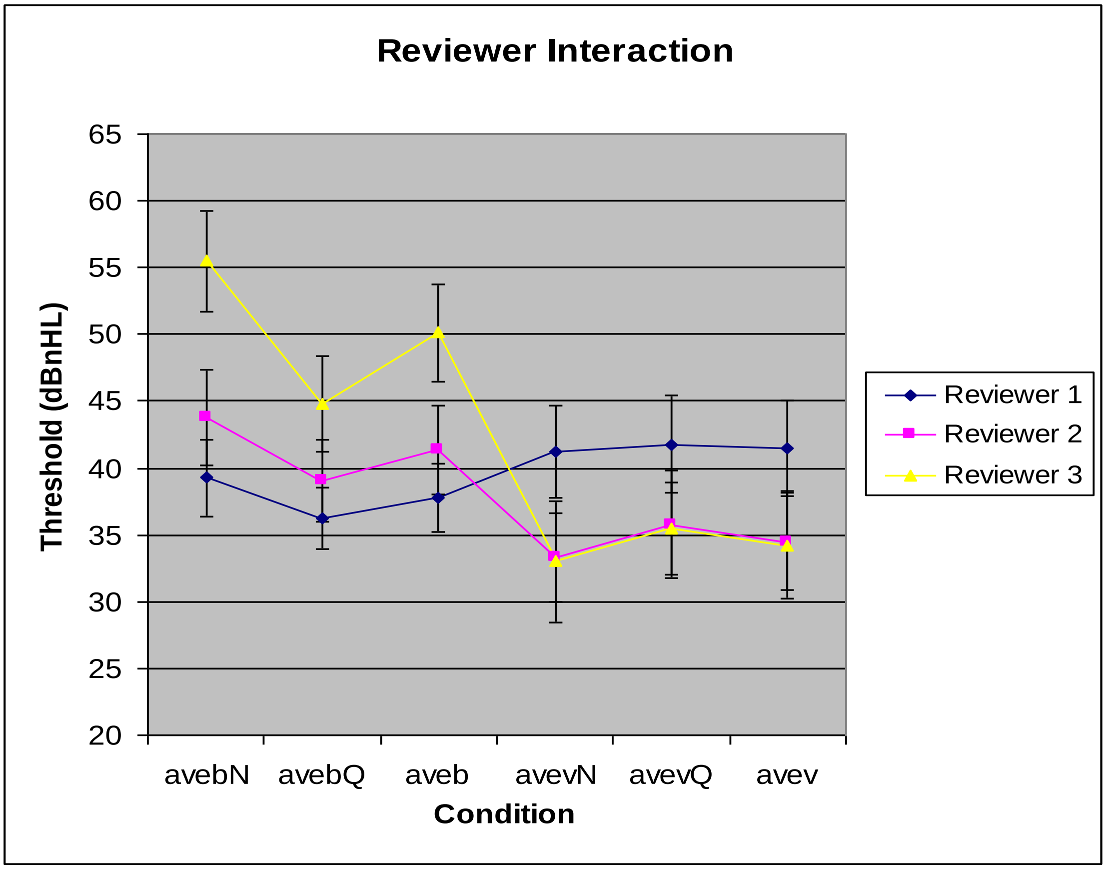

Figure 12. Reviewer by Condition Interaction. Figure indicates that there is a significant interaction between Reviewers with Reviewer 1 yielding an opposite trend than Reviewers 2 and 3 with lower thresholds on the Bio-logic equipment and higher thresholds on the Vivosonic equipment.

ABR Thresholds Accuracy and Efficacy (excluding "No Response" data and accounting for known frequency effects)

Frequency.

Given the above findings of the equipment by Reviewer interaction, the weak correlations between threshold and time when all data was used for calculations, and the large differences between the number of "No Responses" by Reviewer, the decision was made to exclude the "No Response" data from analysis. The following analysis was run with the “No Response” data removed *and* was divided into 500 and 4000 Hz analyses. Results were divided by frequency based on research which shows that 4000 Hz typically generates a more reliable ABR waveform with clearer morphology than at 500 Hz; therefore, it would be expected that different results would be obtained for each frequency in this study. Thus, for the following results, the data sets were divided by frequency.

Repeated measures ANOVA analyses were used to assess the significance of differences pertaining only to threshold measurements between the different test conditions.

4000 Hz- accuracy.

There was a main effect for machine at 4000 Hz, *F*(1, 45)=7.547, *p*=.009. The Vivosonic Integrity ($\bar{x}$ =29.83 dBnHL) measured significantly lower thresholds than the Biologic Navigator ($\bar{x}$ =37.17 dBnHL) (Figure 13). With the “No Responses” and the data for 500 Hz removed, there is no longer a significant interaction noted for equipment by Reviewer, *F*(2, 43)=1.529, *p*=.228). No other significant effects were noted i.e. there was no significant main effect of activity and there were no interactions for activity by Reviewer or equipment by activity.

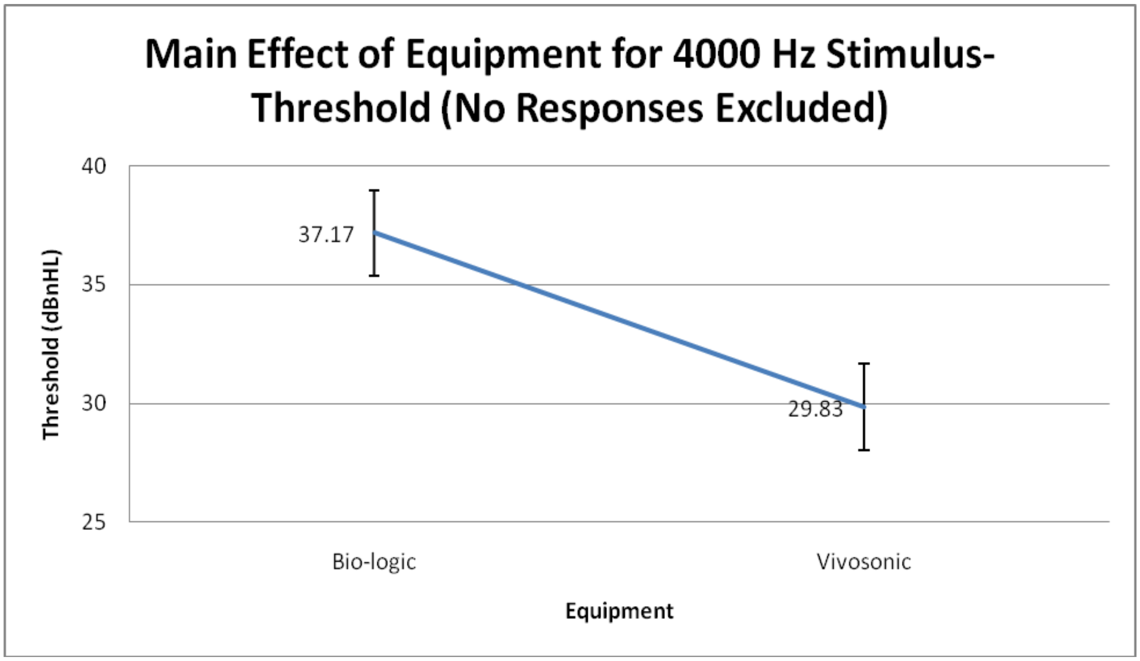

Figure 13. Main Effect of Equipment at 4000 Hz. Figure shows that the Vivosonic measures significantly lower thresholds than the Bio-logic equipment.

Thresholds obtained at 4000 Hz were analyzed both by equipment and by activity level. Results can be found in Figure 14 below. Multiple pairwise comparisons were run to examine the significance of differences between the conditions. No significant differences were found between thresholds obtained with the Bio-logic in quiet ($\bar{x}$ = 34.83, *SEM*=2.01) and the Bio-logic in noise ($\bar{x}$ =38.4, *SEM*=1.99) (*t*=1.095, *p*=.279). No significant differences were found between thresholds obtained with the Vivosonic in quiet ($\bar{x}$ =27.89, *SEM*=1.93) and the Vivosonic in noise ($\bar{x}$ =31.33, *SEM*=2.39) (*t*=1.186, *p*=.241). There was no significant difference found between thresholds obtained with the Vivosonic in noise and the Bio-logic in noise (*t*=1.878, *p*=.066). The only significant finding with a Bonferroni correction was between the Vivosonic in Quiet and the Bio-logic in Quiet (*t*=3.370, *p*=.001).

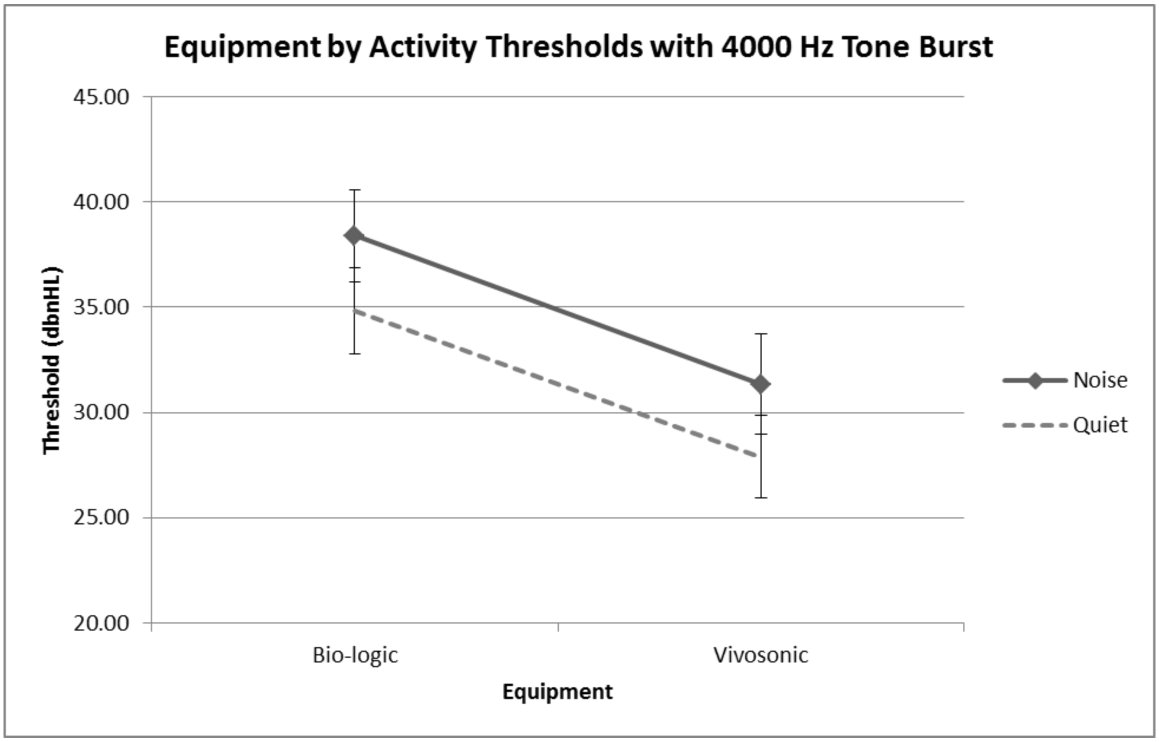

Figure 14. Equipment by Activity Thresholds. Figure indicates a significant different between thresholds obtained with the Vivosonic in Quiet and thresholds obtained with the Bio-logic in quiet with a 4000 Hz tone burst.

4000 Hz- efficiency.

Four different scenarios were possible for efficiency (acquisition time). These include a high threshold and a short time, a high threshold and a long time, a low threshold and a short time, and, finally, a low threshold and a long time (see Table 5 below). In order to include both time and threshold in a single measurement, time (in seconds) was multiplied by threshold (in dBnHL). This method was used because threshold and time were expected to be highly correlated (as was found previously). In other words, ideally, it is expected that there would be a strong negative correlation between time and threshold (low threshold, long time or high threshold, short time). Multiplying time and threshold together effectively "corrected" for threshold differences. Otherwise, if time had been analyzed apart from threshold, a short time may have been misconstrued as a good result when the reason for the short time was actually due to the selection of a suprathreshold response i.e. 70dBnHL. In order to ensure a short time to obtain threshold was not misconstrued as being either “good” or “bad,” each time selection was multiplied by its threshold. As it is not correct to call this unit “Time” the term “Efficiency” was used to describe the variable dBnHL*sec.

Table 5

Possible Efficiency Combinations

Low Threshold*Short Time	Smallest Value	Best
High Threshold*Short Time	↓	↓
Low Threshold*Long Time		
High Threshold*Long Time	Largest Value	Worst

Table 5. Possible Efficiency Combinations.

A repeated measures ANOVA analysis was performed to examine the differences between the conditions. A significant main effect for equipment was found for efficiency, *F*(1, 43)=59.372, *p*<.001, with the mean efficiency at 4000 Hz for the Bio-logic Navigator ($\bar{x}$ =5496.47 s*dBnHL, *SEM*=302.55) being significantly longer than the Vivosonic Integrity ($\bar{x}$ =4227.99 s*dBnHL, *SEM*=249.54) (Figure 15). Further, both sets of equipment were noted to have approximately the same degree of variability.

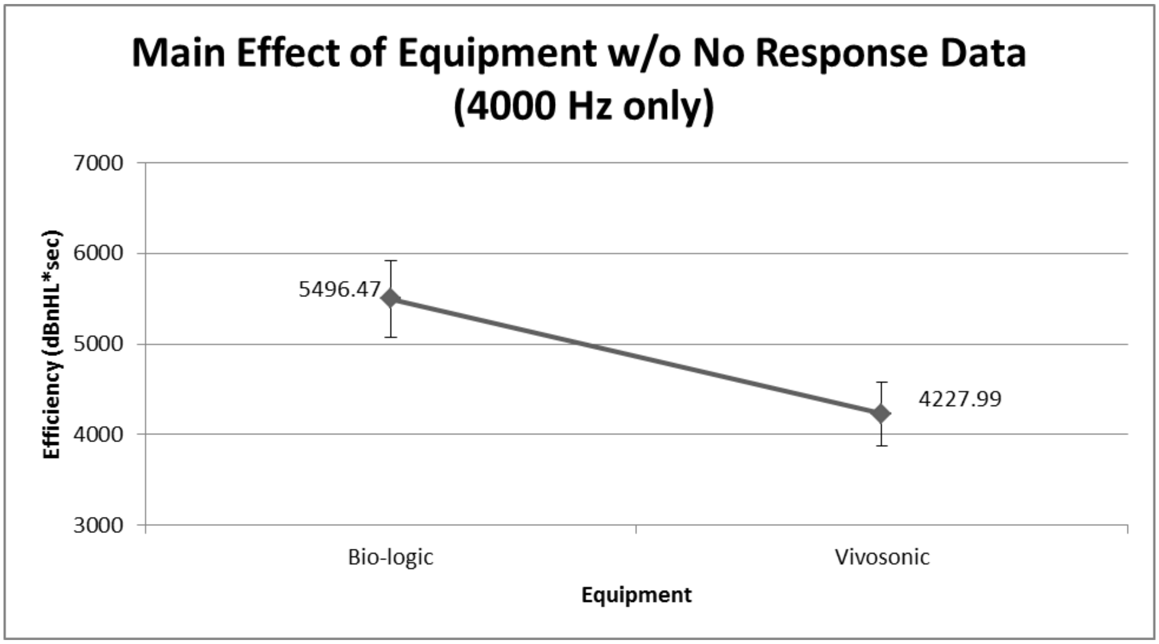

Figure 15. Main effect of equipment with No Response Data Excluded. Figure shows that the Vivosonic is significantly more efficient than the Bio-logic Navigator when using a 4000 Hz tone burst.

Not surprisingly, a significant main effect was also found for activity level, *F*(1,43)=167.53, *p*<.001, with the quiet conditions being significantly more efficient ($\bar{x}$ =3252.88, *SEM*=86.95) than the noisy conditions ($\bar{x}$ =6471.58, *SEM*=189.098) (Figure 16). As expected, both sets of equipment were less efficient in noise and the noisy activity level had a much higher degree of variability than the quiet activity level.

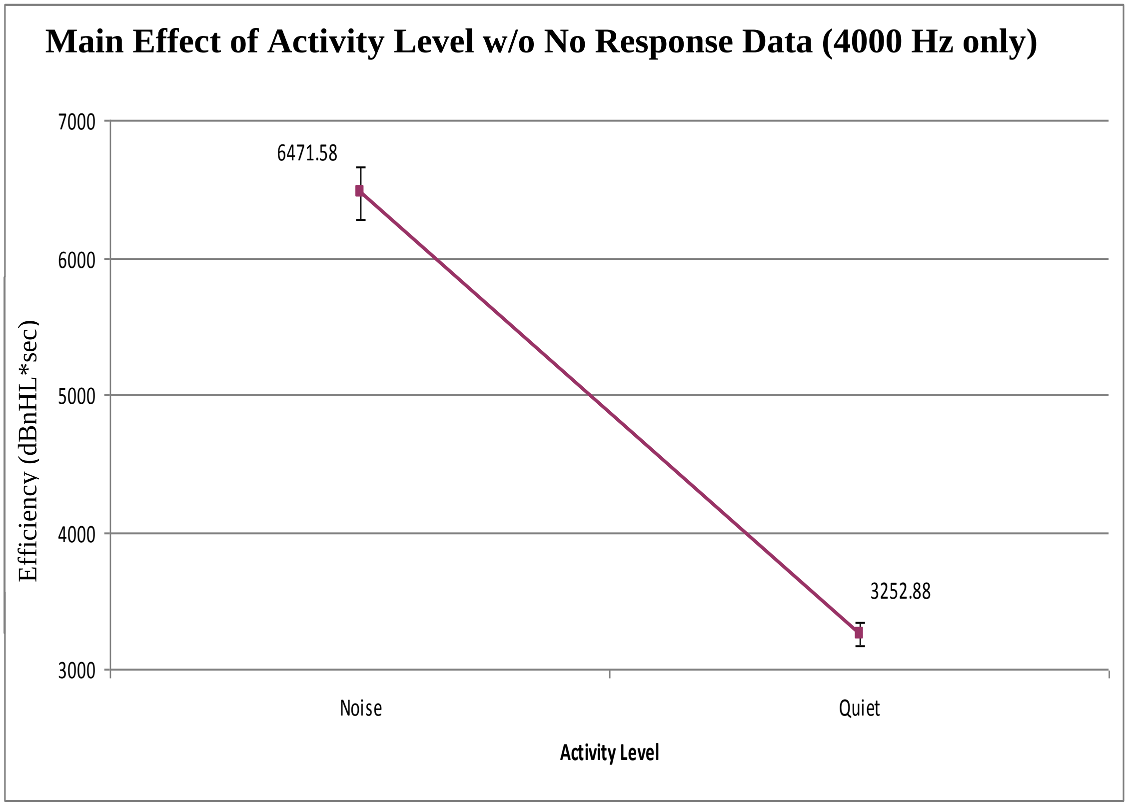

Figure 16. Main Effect of activity level without No Response data obtained using a 4000 Hz tone burst. Figure shows that threshold acquisition is significantly more efficient for the quiet condition than the noisy activity level.

A significant interaction was found for equipment as a function of activity level $F(1,43)=6.681$, $p=.01$, indicating that the effect of activity level on the efficiency of obtaining threshold varies depending on the equipment. While both systems revealed increased efficiency when obtaining thresholds in quiet versus noise, the difference was much smaller for the Vivosonic system. Post hoc analysis revealed a significant difference between the efficiency of obtaining threshold for the Bio-logic Navigator in quiet ($\bar{x}$ =3644.91 dBnHL*sec, $SEM=134.87$) and in noise ($\bar{x}$ =7348.03 dBnHL*sec, $SEM=293.0$) ($t=11.6765$, $p<.001$) with the Bio-logic in quiet found to be significantly more efficient. There was also a significant difference found between the Vivosonic Integrity in quiet ($\bar{x}$ =2860.85 dBnHL*sec, $SEM=111.05$) and in noise ($\bar{x}$ =5595.13 dBnHL*sec, $SEM=241.84$) ($t=10.39$, $p<.001$) albeit a smaller difference than seen with the Bio-logic Navigator with the quiet condition being significantly more efficient. There was a significant difference between the Bio-logic Navigator in noise and the Vivosonic Integrity in noise ($t=4.7550$, $p<.001$) with the Vivosonic being significantly more efficient. Finally, there was a significant difference found between the efficiency of the Bio-logic Navigator and Vivosonic Integrity to acquire thresholds in quiet ($t=4.6102$, $p<.001$) with the Vivosonic Integrity being significantly more efficient (Figure 17). Of particular interest is the lack of Reviewer interaction noted with the “No Response” data removed at 4000 Hz, $F(2,43)=2.12$, $p=.146$; therefore, it can be said that for the 4000 Hz data, all Reviewers varied in the selection of efficiency by equipment in the same manner.

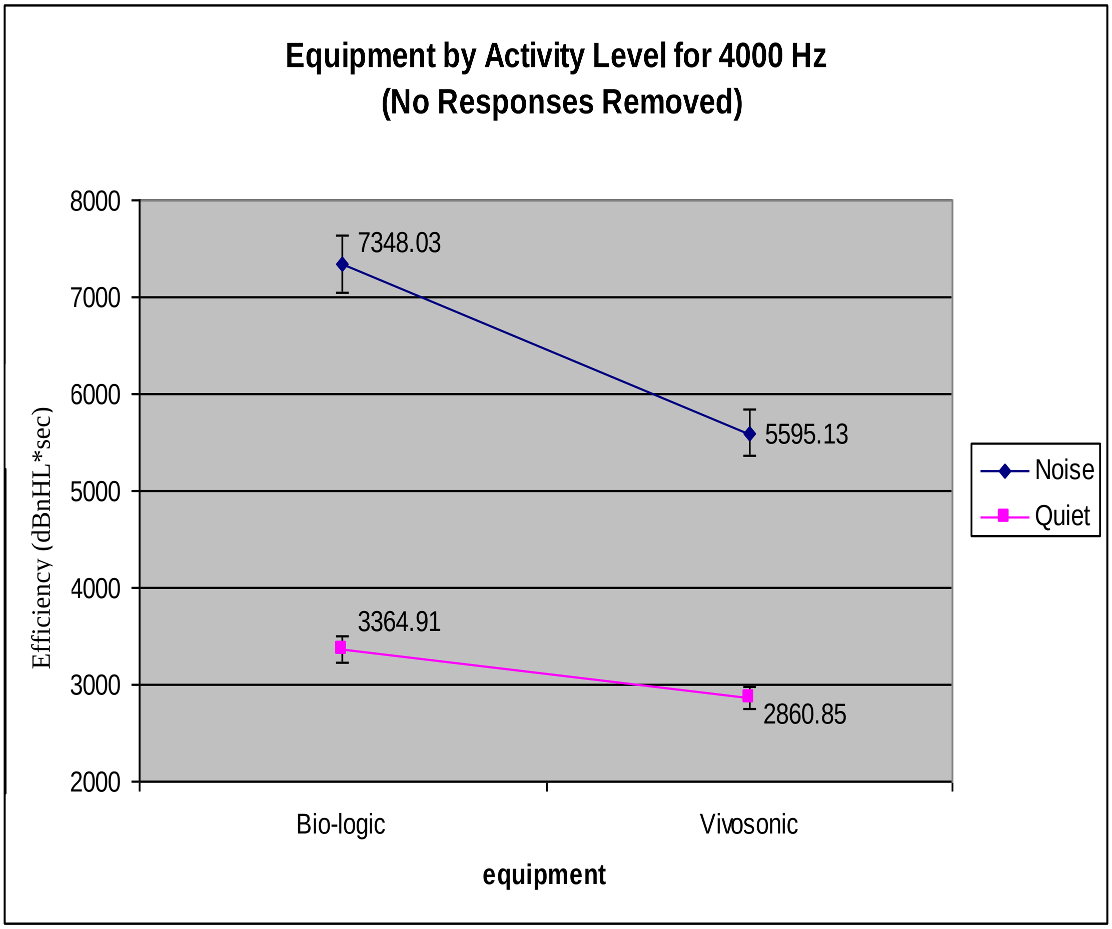

Figure 17. Equipment by Activity Level interaction obtained with the No Responses excluded. There is a significant interaction noted between equipment and activity level in terms of efficiency. There is also a significant difference between the Bio-logic in Quiet and Bio-logic in Noise, the Bio-logic in Quiet and the Vivosonic in Quiet, the Vivosonic in Quiet and the Vivosonic in Noise, and the Vivosonic in Noise and the Bio-logic in Noise.

500 Hz- accuracy.

With the data for 4000 Hz removed, there were no significant findings for 500 Hz. The main effect of equipment at 500 Hz was insignificant, *F*(1,41)=.003, *p*=.956. The statistics indicate that both sets of equipment were equally accurate or, rather, inaccurate when using a 500 Hz tone burst. Although results were not statistically significant, the *trend* shows that the Vivosonic Integrity ($\bar{x}$ =38.84 dBnHL) measured lower thresholds than the Biologic Navigator ($\bar{x}$ =41.78 dBnHL) (Figure 18 Below).

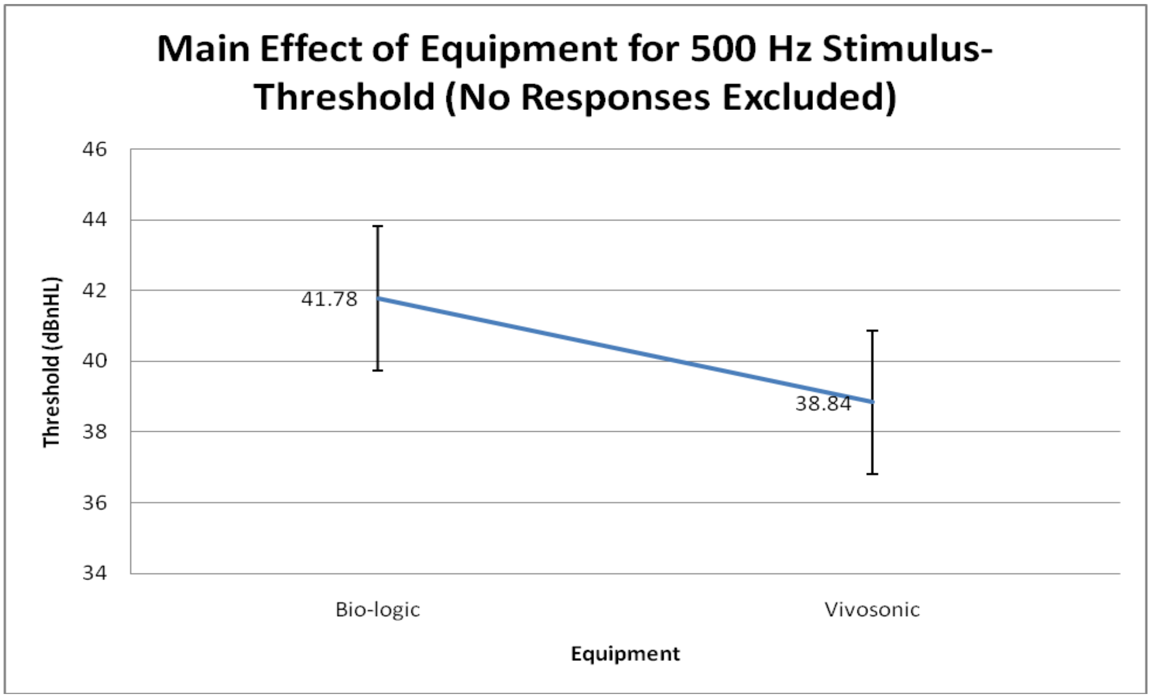

Figure 18. Main effect of equipment for 500 Hz (threshold only). There was no main effect of equipment; however, the trend shows that the Vivosonic measured slightly lower thresholds than the Bio-logic.

Further, there is a significant interaction, *F*(2, 41)=3.942, *p*=.027, noted for equipment by Reviewer when exclusively examining data obtained with the 500 Hz tone burst. This means that for this data set, threshold varies as a function of Reviewer. There is no longer a significant main effect of activity nor is there an interaction for activity by Reviewer or equipment by activity. Multiple pairwise comparisons were calculated to examine the differences in thresholds for the Vivosonic in quiet vs. noise, *t*=1.020, *p*=.314, the Vivosonic in noise and the Bio-logic in noise, *t*=.066, *p*=.948, the Bio-logic in quiet and the Vivosonic in quiet, *t*=.724, *p*=.473, and the Bio-logic in quiet vs. noise, *t*=.317, *p*=.753. This indicates that no significant differences were found between any of the conditions at 500 Hz.

500 Hz- efficiency.

A repeated measures ANOVA was run with all efficiency data for 500 Hz. A marginally significant main effect for equipment was noted, *F*(1,57)=4.330, *p*=.044, with the Vivosonic Integrity ($\bar{x}$ =4384.41 dBnHL*sec, *SEM*= 189.38) acquiring threshold slightly more efficiently than the Bio-logic Navigator ($\bar{x}$ =4827.06 dBnHL*sec, *SEM*=173.0) (Figure 19).

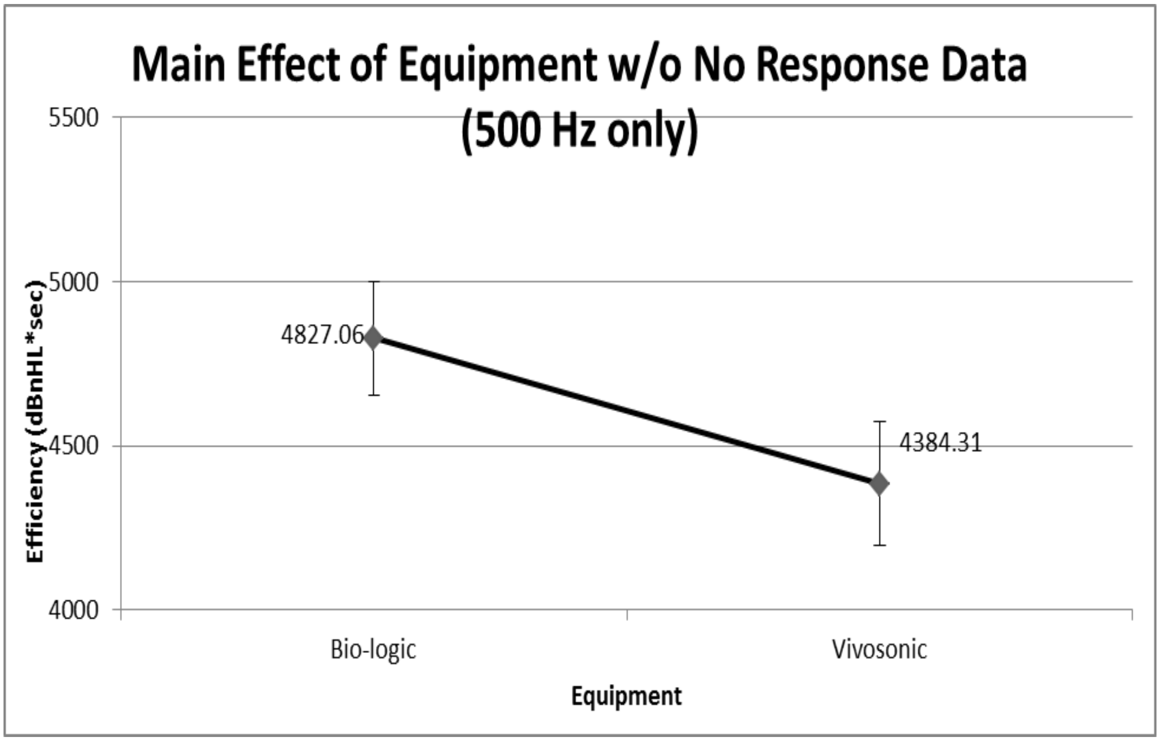

Figure 19. Main effect of equipment with No Response data excluded obtained using a 500 Hz tone burst showing that the Vivosonic Integrity is more efficient at acquiring threshold than the Bio-logic Navigator when a 500 Hz tone burst is used.

A significant main effect was also found for activity level, *F*(1,57)=146.58, p<.001, with the quiet conditions being significantly more efficient overall ($\bar{x}$ =3060.0 dBnHL*sec, *SEM*=98.57) than the noisy conditions ($\bar{x}$ =6151.47 dBnHL*sec, *SEM*=257.59) (Figure 20). The noisy activity level also had a much higher degree of variability than the quiet activity level.

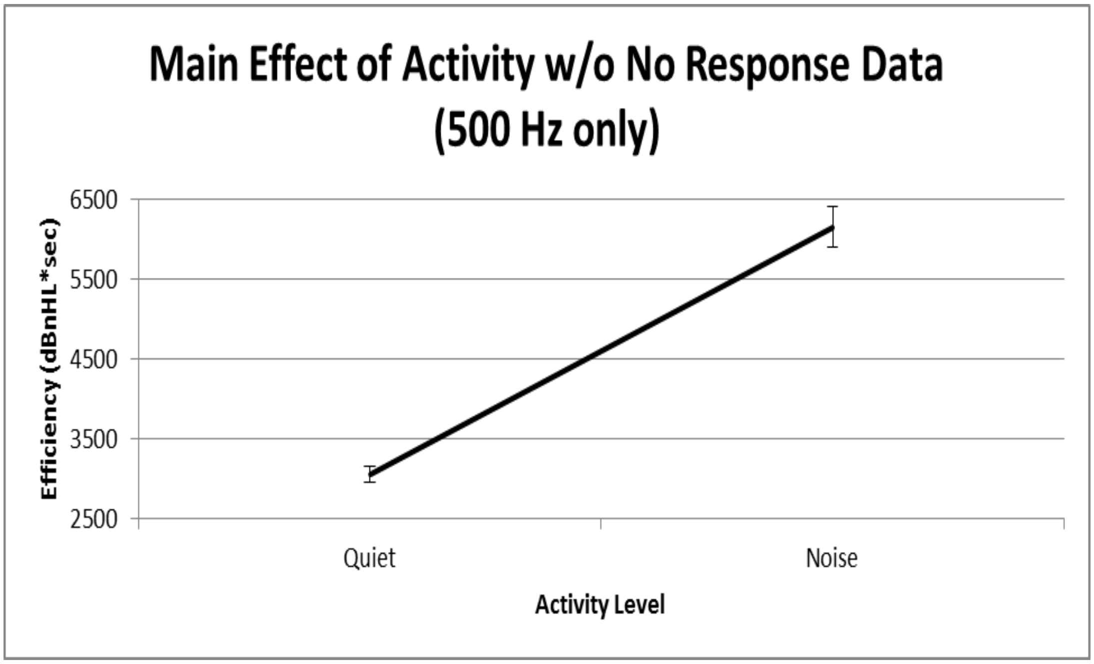

Figure 20. Main effect of activity level with No Response data excluded obtained using a 500 Hz tone burst. Figure shows that thresholds are more efficiently acquired in quiet than in noise.

While both systems revealed increased efficiency to obtain thresholds in quiet versus thresholds obtained in noise, the difference was larger for the Vivosonic system. Post hoc analysis revealed a significant difference between the efficiency of obtaining threshold for the Bio-logic Navigator in quiet ($\bar{x}$ =3372.35 dBnHL*sec, SEM=140.11) and in noise ($\bar{x}$ =6281.76 dBnHL*sec, SEM=274.93) (*t*=11.461, *p*<.001). There was also a significant difference found between the Vivosonic Integrity in quiet ($\bar{x}$ =2747.65 dBnHL*sec, SEM=139.81) and in noise ($\bar{x}$ =6021.18 dBnHL*sec, SEM=357.9) (*t*=10.66, *p*<.001). Finally, there was a significant difference found for efficiency between the Bio-logic Navigator and Vivosonic Integrity to acquire thresholds in quiet (*t*=3.243, *p*<.002). There was no significant difference between the Bio-logic Navigator in noise and the Vivosonic Integrity in noise (*t*=.494, *p*=.624).

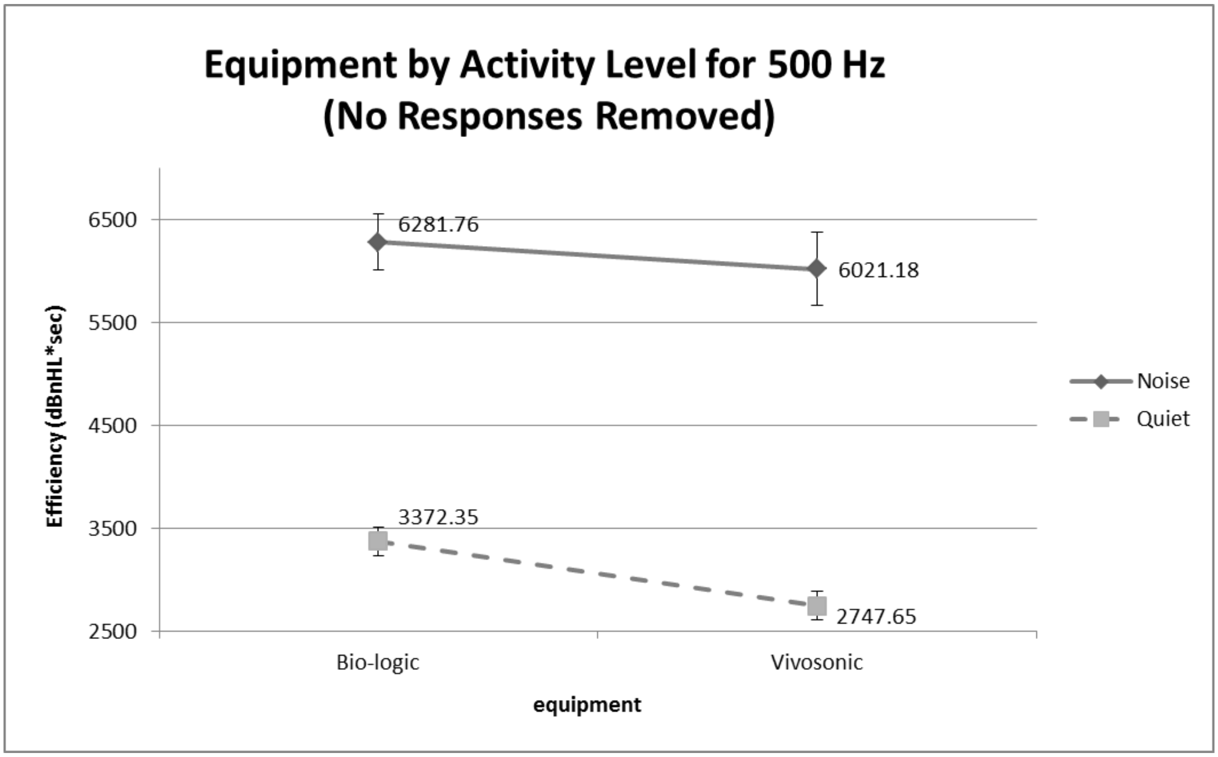

Figure 21. Equipment by Activity Level for 500 Hz with No Response data excluded. Figure shows a significant difference between the efficiency of the Bio-logic in Quiet and in Noise, the Vivosonic in Quiet and in Noise, the Bio-logic in Quiet and the Vivosonic in Quiet, and the Bio-logic in Noise and the Vivosonic in Noise.

There was a significant interaction noted for equipment by Reviewer, *F*(2,41)=4.019, *p*=.025. This means that Reviewers differed in the way that they chose the efficiency ratings by equipment. Because of this interaction, the results at 500 Hz must be cautiously interpreted. More participants and more Reviewers are needed to fully explore the data at 500 Hz. No other significant interactions were found for the efficiency data at 500 Hz i.e. equipment by activity, activity by Reviewer, equipment by activity by Reviewer.

Chapter V

DISCUSSION

The Vivosonic Integrity was used in this study as a means of assessing the effects of Kalman weighted filtering and in-situ pre-amplification on acquisition of ABR thresholds in the presence of physiologic noise. It was compared to thresholds obtained using the Bio-logic Navigator which employs conventional analog filtering and pre-amplification. The purpose of the study was not to assess a particular brand of ABR equipment, rather it was to assess the underlying effects of Kalman weighted filtering and in-situ pre-amplification on ABR threshold acquisition in physiologic noise.

Qualitative Evaluation of Reliability

The inclusion of “No Response” data biased results when considering threshold accuracy and efficiency.   However, the “No Response” data is worthy of consideration because a “No Response” would be a false positive for hearing loss if recorded from normal hearing patients. Statistical analysis revealed that Reviewers indicated that, as a whole, the Integrity system yielded significantly fewer “No Responses” than the Navigator system indicating that Reviewers believed the Integrity was more reliable in threshold estimation. In general, regardless of frequency or equipment, thresholds measured in noise were more often judged as “No Response” than those measured in quiet. Of particular interest is the finding that thresholds in noise at 4000 Hz measured using the Vivosonic Integrity were significantly more reliable- fewer No Responses- than under the same conditions using the Biologic Navigator. The trends indicated that Reviewers found the Integrity to be less reliable in the 500 Hz condition than the 4000 Hz condition regardless of activity level and found the Navigator to be less reliable in the noisy condition than the quiet condition regardless of frequency. Further, there was a statistically significant difference between the Reviewers with Reviewer 3 being far more conservative (i.e. indicated more "No Responses") than Reviewers 1 and 2.

A major component of this study turned out to be the variability of the Reviewers. As noted by Stapells (1998), ABR assessment is highly variable from audiologist to audiologist. It is a highly subjective measure that requires the examiner to have a great deal of expertise and experience. Reviewer 3 was more conservative when choosing thresholds for the Bio-logic system than the other two Reviewers and, likewise, Reviewer 1 was more conservative when choosing thresholds for the Vivosonic system. Although each Reviewer was given normative data indicating the latency of wave V, each had a separate set of internal guidelines that they had developed through their own past experiences. The high degree of inter-rater variability was closely examined in this study as it became apparent that removing one Reviewer would drastically change the outcome of the study. For instance, if Reviewer 3, who chose higher than average thresholds for the Bio-logic and the lowest thresholds for the Vivosonic, was removed then all significant findings would be eradicated. Similarly, if Reviewer 1, who chose the lowest thresholds out of each of the Reviewers for the Bio-logic and the highest threshold out of each of the Reviewers for the Vivosonic, was eliminated then the significance of the findings would have been augmented. After careful consideration, it was decided to include each of the Reviewers in the data analysis as it is a perfect depiction of the subjectivity and variability inherent in ABR measurements. Although Reviewers were variable, there were some consistent trends noted. Interestingly, ABR thresholds were equally variable regardless of which frequency was being assessed- 4000 Hz or 500 Hz- although, as one would expect, the efficiency of obtaining a 500 Hz ABR threshold was significantly poorer and more variable than the time required to obtain a threshold at 4000 Hz. This indicates that Reviewers as a group were less sure of the 500 Hz thresholds and required more time to definitively decide that a wave V was present. When examining differences between the Reviewers in terms of each system as a whole, there was a higher degree of variability for the Bio-logic equipment than the Vivosonic equipment for both absolute threshold measurements (accuracy) and efficiency. Further, within each system variability is similar within each frequency and between frequencies. That is to say that the Reviewers had only a small amount of variability when choosing threshold for each system at each frequency. Not surprisingly, the variability for each Reviewer was found to be much lower for ABR thresholds obtained in quiet versus those obtained in noise. This trend holds true for both absolute threshold as well as efficiency indicating that Reviewers are more confident in thresholds obtained in quiet than in noise.

Additionally, Reviewers were each asked to rate the threshold they chose as reliable or unreliable. If they could not identify a wave V at even the highest intensity, this was considered a No Response and was included as an unreliable response. A statistically significant difference was found between the number of unreliable responses for the Bio-logic in quiet and the Bio-logic in noise with Bio-logic in noise yielding far fewer reliable ratings. There was no significant difference found for the Vivosonic in quiet and the Vivosonic in noise indicating that the Reviewers found the Vivosonic to be equally as reliable in quiet and in noise. The Vivosonic Quiet and the Vivosonic Noise were rated almost equally reliable with 82 reliable votes for the quiet condition and 78 reliable votes for the noise condition. Interestingly, the Vivosonic was rated as more reliable in both the quiet and noisy conditions than the Bio-logic quiet condition. This indicates that Reviewers deemed the Vivosonic waveforms to be more reliable both in quiet and in noise than those obtained under the best of conditions by the Bio-logic. There was a significant difference between the two sets of equipment in noise indicating that Reviewers rated the Vivosonic waveforms significantly more reliable than waveforms obtained using the Bio-logic. Reviewers indicated that both systems were equally reliable in quiet. The reliable vs. unreliable results were further subdivided by Reviewer which revealed that Reviewer 3 indicated far more unreliable thresholds than Reviewers 1 and 2. As above, where Reviewer 3 was noted to have more “No Responses” than Reviewers 1 and 2, Reviewer 3 continued to be the most conservative Reviewer.

Behavioral Threshold vs. ABR Threshold

Average ABR thresholds in this study are likely supra-threshold as the softest intensity stimulus was 20dBnHL; therefore, even if a response was observed at 20dBnHL the Reviewer was not provided the option to observe 10dBnHL. This means that when the average was taken, there are no thresholds below 20dBnHL to make the average better. The Reviewers were also not given the opportunity to see replications, with the exception of 20 dBnHL. It is likely that this caused the Reviewers to be more conservative than they might otherwise have been, leading to supra-threshold measurements. It was expected based on the work of Beattie et al. (2005) that 85% of ABR thresholds obtained at 500 Hz will be within 16 dB of behavioral thresholds.

Similarly, Beattie et al. (2005) found that 85% of ABR thresholds obtained at 4000 Hz were within 9 dB of behavioral thresholds. Furthermore, in 2000, Stapells conducted a meta-analysis of 32 different studies examining ABR threshold estimation using tone burst stimuli. Stapells indicated that results were consistent across the studies with tone burst ABR thresholds typically measured at 10-20 dBnHL in normal hearing participants and, in adult participants with sensorineural hearing loss, ABR thresholds are measured approximately 5-15 dB higher than behavioral thresholds. In the current study, the percentage of ABR thresholds that fall within 20dB of behavioral thresholds was calculated indicating the specificity of each set of equipment. It revealed that the Vivosonic had a much higher degree of specificity than the Bio-logic system within each condition- 500 Hz Noise, 500 Hz Quiet, 4000 Hz Noise, 4000 Hz Quiet (Table 4). In other words, the Bio-logic system indicated hearing loss in participants known to have normal hearing more often than the Vivosonic Integrity system.

Stapells (2000) also indicated that the studies included in the meta-analysis agreed that, as a whole, ABRs evoked by a 500 Hz tone burst are not as reliable as those evoked by a 4000 Hz tone burst. He indicates that, on average, ABR thresholds obtained using a 500 Hz tone burst are 7 dB higher than those obtained using a 4000 Hz tone burst. Each of these studies indicate a strong relationship between behavioral and tone burst evoked auditory brainstem response when measurements are obtained under ideal conditions (i.e. relaxed patient, quiet, electrically shielded environment). In the current study, at 500 Hz in quiet, the Navigator and Integrity systems yielded average thresholds which were 35-40 dBnHL higher than the corresponding behavioral threshold averages. At 4000 Hz in quiet, the Navigator and Integrity systems yielded average thresholds which were 22-29 dBnHL higher than the corresponding behavioral threshold averages. It is likely that these averages would be more in line with the findings in Beattie et al's. study and Stapells' meta-analysis if the stimulus intensity in this study had gone below 20 dBnHL; however, it is noted that the *trends* found in this study were similar to those found by Beattie et al. (2005) and by Stapells (2000) in that the ABR thresholds obtained at 4000 Hz were much more closely aligned with behavioral thresholds than the ABR thresholds obtained at 500 Hz. This study found that in all conditions, thresholds obtained with the 4000 Hz tone burst were more accurate that those obtained with the 500 Hz tone burst. Further, the differences in ABR thresholds obtained within each system at a given frequency in quiet and noise are quite telling. For the Bio-logic, thresholds were consistently more accurate (lower) when obtained in quiet while thresholds obtained by the Vivosonic in noise were fairly consistent with (or even lower than) thresholds obtained by the Vivosonic in quiet.

According to Sininger and Cone-Wesson (2002), electrophysiologic thresholds are highly correlated with behavioral hearing thresholds. The pure tone average and thresholds obtained using click stimuli in an ideal recording (i.e. relaxed patient, quiet environment) have a .979 correlation coefficient (Sininger & Cone-Wesson, 2002). Stapells and Oates (1997) conducted a study examining the correlations between ABR thresholds and behavioral thresholds at 500, 2000, and 4000 Hz for normal hearing participants and participants with sensorineural hearing which revealed a .94 (73 ears), .95 (96 ears), and .97 (51 ears) correlation coefficient, respectively, indicating that behavioral thresholds can accurately be predicted by ABR thresholds under ideal recording conditions. Pearson-product moment correlations were calculated for the participants' behavioral thresholds as a function of their corresponding average ABR threshold in each condition- Bio-logic Noise, Bio-logic Quiet, Vivosonic Noise, and Vivosonic Quiet. It was expected that there would be a strong positive correlation between behavioral threshold and ABR threshold with stronger correlations for quiet conditions versus noisy conditions; however, all calculations revealed weak correlations between behavioral and ABR thresholds. This is possibly a product of the population which includes only normal hearing participants who all had behavioral thresholds better than 20 dB HL. This led to little variance in the behavioral thresholds coupled with potentially larger variations in the ABR thresholds. The correlations between the behavioral threshold and each system at 4000 Hz were slightly higher than the correlations between the behavioral threshold and 500 Hz. Correlations were similar for behavioral thresholds and ABR thresholds obtained in noisy vs. quiet conditions and neither ABR system produced stronger correlations than the other system. Further, behavioral thresholds are determined based on a specific procedure- Modified Hughson Westlake- which eliminates subjective error on the part of the tester. ABR thresholds, however, are a highly subjective measure that require a great deal of expertise and experience making them a more variable measure prone to human error. ABR thresholds obtained in the study did not correlate well with behavioral thresholds, further evidence of the variability associated with subjective ABR threshold estimation.

ABR Threshold: Accuracy and Efficiency

Statistical significance was found for a number of the comparisons made throughout this study, with the exception of the data obtained at 500 Hz. As predicted, the Vivosonic Integrity which uses in-situ pre-amplification and Kalman weighted filtering, returned more accurate thresholds (lower thresholds) than the Bio-logic Navigator at 4000 Hz regardless of whether or not the “No Response” data was included. There was no significant main effect for activity level at 4000 Hz indicating that thresholds obtained at 4000 Hz are equally reliable in quiet and in noise (when data for both systems are combined together). The Vivosonic Integrity was found to measure significantly lower thresholds in quiet than the Bio-logic Navigator; however no differences between ABR systems were noted when thresholds were obtained in noise. Both ABR systems were noted to have similar variability among Reviewers and no significant interactions were observed. Of particular note is the fact that with the “No Response” data and the 500 Hz data removed, there was no longer an equipment by Reviewer interaction which indicates that at 4000 Hz the Reviewers varied in their selection of threshold in a similar manner regardless of which system they were analyzing. Further, at 4000 Hz with the efficiency data (dBnHL*sec), the Vivosonic Integrity was found to be significantly more efficient at acquiring threshold than the Bio-logic Navigator regardless of activity level. As expected, at 4000 Hz, the quiet condition was significantly more efficient than the noisy condition. The inter-rater reliability was similar between systems. Results for the 4000 Hz efficiency data must be cautiously interpreted, however, due to the fact that there was an equipment by activity interaction. This indicates that the effect of activity level on acquisition time to obtain threshold varies by equipment. While both systems require longer acquisition times in noise than in quiet, the Bio-logic unit reveals a much greater discrepancy between the quiet and noisy conditions than is seen with the Vivosonic system. This suggests that that the

Vivosonic is significantly more efficient in noise than the Bio-logic system in noise, but it is also evident that the Vivosonic is significantly more efficient in quiet than the Bio-logic system is in quiet (at 4000 Hz). There was no Reviewer interaction for the efficiency data at 4000 Hz when the No Response data was removed indicating that each of the Reviewers varied in selecting efficiency by equipment in the same manner.

With the No Response data removed, statistically significant findings were not observed for threshold estimation ABRs obtained with the 500 Hz tone burst. These findings indicate that, statistically, both systems are equally reliable (but inaccurate); however, the *trend* shows that the Vivosonic measured lower thresholds on average than the Bio-logic. This trend must be cautiously interpreted, however, due to the fact that there is a significant Reviewer interaction indicating that Reviewers varied in the way that they chose threshold as a function of the equipment. This indicates that Reviewers were not in agreement when choosing thresholds at 500 Hz. This is not surprising as 500 Hz waveforms are typically morphologically unclear, particularly in comparison with 4000 Hz waveforms. There was no significant main effect for activity level, indicating that the Vivosonic and Bio-logic systems were equally reliable (but inaccurate) in determining threshold in quiet and in the presence of physiologic noise at 500 Hz. When examining the efficiency data at 500 Hz, a marginally significant main effect was noted for equipment with the Vivosonic being significantly more efficient at attaining threshold than the Bio-logic system. As expected, a significant main effect was noted for activity level with the quiet condition being significantly more efficient than the noisy condition for the 500 Hz condition. Of note, there was a higher degree of variability in the data for the noisy condition versus the quiet condition.

Summary

- Hypothesis 1: In quiet, there will be no differences between the ABR thresholds obtained using the Vivosonic Integrity system and the Bio-logic Navigator system.

  There was a significant main effect for equipment at 4000 Hz indicating that the Vivosonic Integrity yielded significantly more accurate thresholds overall than the Bio-logic Navigator. Further, there was a significant difference between thresholds measured by the Bio-logic Navigator in quiet ($\bar{x}$ = 34.83dBnHL) versus the Vivosonic Integrity in quiet ($\bar{x}$ =27.89dBnHL). At 500 Hz, there was no significant difference between thresholds measured by the Bio-logic Navigator in quiet versus the Vivosonic Integrity in quiet indicating that the two systems are equally reliable (and inaccurate) at 500 Hz for threshold estimation in quiet. Although results were not statistically significant, the trend shows that the Vivosonic Integrity ($\bar{x}$ =38.84 dBnHL) measured lower thresholds than the Biologic Navigator ($\bar{x}$ =41.78 dBnHL) in quiet.

- Hypothesis 2: In noise, ABR thresholds will be estimated more accurately using the Vivosonic Integrity system than the Biologic Navigator system.

  At 4000 Hz, there was no significant difference between thresholds measured by the Bio-logic Navigator in noise ($\bar{x}$ =38.4dBnHL) versus the Vivosonic Integrity in noise ($\bar{x}$ =31.33dBnHL); however, it should be noted that this finding just missed significance, p=.066. There were no significant differences at 500 Hz.

- Hypothesis 3: There will be no difference in ABR thresholds in quiet versus noise as measured by the Vivosonic Integrity system.

  The Vivosonic Integrity recorded no difference in ABR thresholds in quiet and in noise for thresholds obtained with both the 4000 Hz tone burst as well as the 500 Hz tone burst indicating that that the Vivosonic Integrity was equally accurate in quiet as it was in noise. Although not specifically addressed in the original hypotheses, the above findings were also true for the thresholds obtained with the Bio-logic Navigator.

- Hypothesis 4: The Vivosonic Integrity will be significantly more efficient at obtaining threshold in noise than the Biologic Navigator system.

  At 4000 Hz, the Vivosonic Integrity ($\bar{x}$ =7348.03) was significantly more efficient at obtaining thresholds in noise than the conventional ABR Bio-logic Navigator system ($\bar{x}$ =5595.13). At 500 Hz, the Vivosonic Integrity ($\bar{x}$ =6021.18) and the Bio-logic Navigator ($\bar{x}$ =6281.76) were equally efficient.

Conclusions

Auditory brainstem response testing has long been used to objectively determine peripheral hearing thresholds for children and difficult-to-test populations. With the advent of mandatory newborn hearing screenings, a means of attaining hearing thresholds in infants is imperative and it must typically be achieved in a hospital suite on the maternity ward with an active infant. Because ABR thresholds are known to be highly correlated with behavioral thresholds, ABRs are the current method of objective assessment of peripheral hearing. Current methods of ABR acquisition using a conventional artifact rejection paradigm have proven to be unreliable in the presence of physiologic artifact particularly in cases where the patient is awake and active. In the past, this has meant that children and difficult-to-test populations have had to undergo these procedures while sedated. Sedation is both a costly and risky procedure making an ABR system that is resistant to physiologic artifact imperative. The findings of this study indicate that a system which uses Kalman weighted filtering and in-situ pre-amplification measures significantly better thresholds at 4000 Hz in quiet than a system which uses a standard artifact rejection; however, the system did not return significantly better thresholds at 4000 Hz in noise (although it just barely missed significance) or at 500 Hz in noise. Within the Vivosonic system data, there were no differences found between thresholds obtained in the presence of physiologic noise vs. quiet when using a 4000 Hz or 500 Hz tone burst. An interesting finding of the present study indicated that although there were no differences between each ABR system in quiet and in noise, the Vivosonic Integrity was significantly more efficient at obtaining threshold than the Bio-logic Navigator (at 4000 Hz only). The interactions and lack of significant findings with the 500 Hz tone burst reinforces the fact that 500 Hz is more difficult to interpret for many audiologists as the waveforms are typically not as morphologically clear as those obtained with a 4000 Hz tone burst or click stimuli.

While the Kalman weighted filtering and in-situ pre-amplification did not provide statistically significant differences from the conventional ABR system with the standard artifact rejection paradigm in all conditions, the finding that it was able to measure at least as accurately (if not better) as a conventional system in a shorter period of time should not be discounted. The Vivosonic Integrity was rated as significantly more reliable by the Reviewers in noise and had significantly fewer “No Responses.”

Limitations of the Study

There were several limitations of this study that, had they been addressed from the advent of the study, may have changed the findings. First, the only waveforms that were repeated were those obtained at 20dBnHL. The Reviewers indicated on multiple occasions that they could not see a definitive response at 30dBnHL, but that when they saw the tracings at 20dBnHL where there was a replication they were able to indicate that there was a definitive response. The drawback to this study was that in an instance such as this, the Reviewers were required to indicate threshold at 40dBnHL because they could not identify a wave V at 30dBnHL. If all waveforms were replicated, it is likely that average thresholds would have been considerably better (lower). Further, the present study did not test below 20dBnHL and, therefore, there was a floor effect. If a response was observed at 20dBnHL that response was considered threshold whereas they may have had a response much lower given that all participants had normal hearing acuity. Likewise, the fact that acquisition time was a maximum of two minutes in quiet or four minutes in noise, may have prevented the Reviewers from seeing a response where there would have been one given a longer acquisition window. All of these factors could have caused the thresholds obtained in this study to be over-estimated.

Although every attempt was made to ensure that the two systems were set-up to be equivalent, the acquisition process and physical hardware could not be made identical. For instance, ABR waveforms could not be added together post hoc using the Vivosonic equipment; therefore, the waveforms had to be divided out in 15 or 30 second intervals by Vivosonic Incorporated while those waveforms obtained using the Bio-logic system were able to be obtained in 15 second intervals and added together post hoc. Also, the electrodes used by the Vivosonic equipment are adhesive electrodes while those used by the Bio-logic system are disc electrodes. Variance was controlled by ensuring that impedance values were similar, however, the ability to use the same type of electrode would have been advantageous. Furthermore, muscular movement was not directly measured in this study although participants were given guidelines for how fast and how strong to chew. Because the muscular movement was not directly measured, this may have caused variance between the participants. The Bio-logic system allows the clinician to view the number of rejected sweeps in real-time as acquisition is taking place. In an instance where a large degree of physiologic artifact was present, the clinician would likely stop testing to assess the source. This would increase the accuracy of testing while drastically reducing the efficiency. The Vivosonic system, on the other hand, does not allow the clinician to view physiologic artifact in real-time. Because each sweep is used with Kalman weighted filtering, at least to some extent, a clinician has no basis on which to determine if a large degree of artifact is present. This means that, although the system that uses Kalman weighted filtering was found to be significantly more efficient, it would be difficult for a clinician to know how much physiologic artifact is inherent in the waveforms.

Finally, it was assumed in this study that the physical act of chewing gum mimicked the same motion as that of a baby sucking on a bottle or pacifier; however, as there was no direct assessment of the muscles involved in either chewing gum or a baby sucking on a bottle or pacifier, the relationship was only assumed in this study.

Future Research

Future research is needed to further investigate the accuracy and efficacy of the in-situ pre-amplification and Kalman weighted filter in different populations beyond normal hearing young adults such as the hearing impaired and pediatrics. Although Stapells (2000) indicated in his meta-analysis that ABRs obtained with a conventional system are as reliable in normal hearing participants as in hearing impaired participants, the findings of this study *may* have yielded entirely different results had the population had hearing impairment. Given the floor effect in this study with normal hearing participants, it is possible that results would have been more accurate with a hearing impaired population. Further, the present study was only able to assess the specificity of the Vivosonic system for accurate ABR threshold estimation as all participants were normal hearing. A future study of hearing impaired participants would allow for assessment of the sensitivity of the Vivosonic system.

To further control the study, the data obtained from the Vivosonic system could be reprocessed such that instead of using Kalman weighted filtering, a standard artifact rejection level employed with a conventional ABR system could be used. This would eliminate the in-situ pre-amplification variable and allow for direct comparison between Kalman weighted filtering and a standard artifact rejection paradigm using the same system.

Furthermore, in conjunction with corrections for the limitations listed above, future research should include a larger panel of Reviewers as the variability in reviewers was a major finding of this study which led to a large degree of variability in the data. Although each Reviewer was considered to be highly trained at ABR waveform assessment, more in-depth training and a template for picking waveforms should be given to Reviewers prior to beginning threshold determination.

Finally, now that this study shows that an ABR system which uses Kalman weighted filtering and in-situ pre-amplification is at least as accurate as a conventional system in physiologic noise, future research should examine the effects of other noise sources such as electromagnetic noise.

More research is needed to investigate the possible benefits of Kalman weighted filtering in ABR threshold estimation, but the results (both significant findings and trends) of the current study indicate that Kalman weighted filtering and in-situ pre-amplification have the potential to make ABR threshold estimation in the presence of physiologic noise more accurate and more efficient.

Appendix A

Behavioral and ABR Thresholds

Reviewer	Participant	Behavioral (500)	Behavioral (4k)	b5N	b5Q	b4N	b4Q	v5N	v5Q	v4N	v4Q
1	1	5	5	40	40	20	30	70	70	20	20
1	2	0	5	30	30	40	20	40	30	20	20
1	3	0	0	20	40	20	20	30	70	40	20
1	4	0	10	70	40	30	70	70	20	70	20
1	5	0	10	70	30	70	70	80	80	70	70
1	6	5	0	70	30	30	20	40	40	70	20
1	7	5	5	20	40	80	30	40	30	20	20
1	8	5	0	70	40	20	30	40	20	30	20
1	9	10	10	40	30	30	30	40	70	40	70
1	10	5	0	30	70	40	20	20	80	30	40
1	11	5	0	30	30	40	30	70	40	20	20
1	12	15	10	30	40	30	20	40	70	20	30
1	13	10	5	30	70	30	20	20	70	40	20
1	14	0	0	20	20	20	70	70	70	20	40
1	15	10	10	40	20	30	30	30	70	20	20
1	16	5	0	30	70	30	30	40	30	70	40
1	17	0	15	30	30	40	30	20	40	20	20
1	18	10	10	30	40	80	30	80	80	20	30
1	19	5	5	70	20	30	20	40	70	20	20
1	20	0	10	20	30	70	70	70	40	40	20
2	1	5	5	30	40	30	20	70	30	30	20
2	2	0	5	30	30	30	20	20	20	30	20
2	3	0	0	30	40	40	40	20	20	20	20
2	4	0	10	30	30	30	40	20	20	70	20
2	5	0	10	40	40	70	30	40	70	20	80
2	6	5	0	30	30	40	30	20	40	20	20
2	7	5	5	30	80	40	30	20	70	20	80
2	8	5	0	80	70	30	30	70	70	40	20
2	9	10	10	30	30	80	20	20	20	20	20
2	10	5	0	80	40	40	30	20	20	20	20
2	11	5	0	70	40	80	30	40	20	70	40
2	12	15	10	40	40	30	20	20	40	20	30
2	13	10	5	30	30	30	30	20	40	30	20
2	14	0	0	20	20	40	30	20	70	20	40
2	15	10	10	40	70	40	40	70	40	20	20

Behavioral and ABR Thresholds Continued

Reviewer	Participant	Behavioral (500)	Behavioral (4k)	b5N	b5Q	b4N	b4Q	v5N	v5Q	v4N	v4Q
2	16	5	0	30	30	20	40	40	30	70	20
2	17	0	15	30	30	40	30	40	20	20	20
2	18	10	10	30	30	80	70	20	40	20	40
2	19	5	5	70	70	80	40	40	70	30	20
2	20	0	10	40	70	70	80	70	40	40	70
3	1	5	5	70	70	40	40	80	80	20	20
3	2	0	5	40	30	30	40	40	70	20	20
3	3	0	0	80	70	80	70	20	40	20	20
3	4	0	10	80	80	70	20	80	20	70	20
3	5	0	10	80	70	70	80	80	80	70	80
3	6	5	0	80	20	40	70	70	20	70	70
3	7	5	5	80	70	30	30	20	40	20	20
3	8	5	0	80	30	30	30	20	20	20	20
3	9	10	10	40	30	80	30	20	70	20	20
3	10	5	0	80	70	40	30	20	70	20	20
3	11	5	0	40	70	40	30	70	20	20	70
3	12	15	10	40	40	30	20	20	20	20	30
3	13	10	5	40	40	70	40	20	20	20	20
3	14	0	0	20	20	20	40	20	40	20	20
3	15	10	10	40	70	70	30	40	20	20	20
3	16	5	0	80	40	20	30	20	20	20	40
3	17	0	15	30	20	40	30	20	30	20	20
3	18	10	10	80	20	80	20	20	20	20	20
3	19	5	5	70	70	80	40	70	70	20	20
3	20	0	10	70	70	40	70	20	40	20	40

Efficiency Data

Reviewer	Participant	sb5N	sb5Q	sb4N	sb4Q	sv5N	sv5Q	sv4N	sv4Q
1	1	150	90	300	135	60	15	270	120
1	2	210	150	150	195	150	150	420	135
1	3	360	75	420	195	300	30	150	105
1	4	120	90	240	45	90	195	60	195
1	5	120	180	120	60	480	240	60	60
1	6	60	150	300	210	210	60	60	150
1	7	450	105	480	135	180	135	240	195
1	8	90	90	420	150	180	165	270	105
1	9	150	120	360	135	180	45	180	45
1	10	480	45	180	225	360	240	210	45
1	11	180	105	210	120	90	75	360	165
1	12	210	90	360	180	180	30	270	75
1	13	210	60	360	135	360	30	150	195
1	14	420	195	450	30	90	30	390	105
1	15	210	180	330	120	180	30	360	180
1	16	270	30	240	165	210	120	30	90
1	17	330	135	210	105	420	120	300	165
1	18	270	105	480	165	480	240	300	105
1	19	90	195	330	195	180	45	300	150
1	20	390	150	60	60	90	75	180	105
2	1	120	60	300	150	60	60	90	120
2	2	180	105	180	120	270	165	180	90
2	3	270	90	150	90	300	135	270	150
2	4	210	105	120	60	120	135	90	150
2	5	90	30	60	120	150	45	300	240
2	6	120	90	210	135	270	75	240	135
2	7	330	240	150	75	330	45	270	240
2	8	480	45	210	60	120	60	150	60
2	9	180	75	480	135	210	210	240	90
2	10	480	105	150	135	270	150	180	180
2	11	60	105	480	60	180	105	90	90
2	12	120	75	270	150	330	60	330	90
2	13	90	60	270	60	120	90	180	135
2	14	300	120	180	120	270	30	210	75
2	15	210	30	210	60	210	60	390	195
2	16	330	180	180	90	210	75	60	135
2	17	90	135	120	120	120	210	270	165

Efficiency Data Continued

Reviewer	Participant	sb5N	sb5Q	sb4N	sb4Q	sv5N	sv5Q	sv4N	sv4Q
2	18	180	120	480	30	210	105	180	120
2	19	120	30	480	60	240	60	150	135
2	20	150	45	60	240	90	105	150	15
3	1	120	60	210	90	480	480	300	165
3	2	180	165	330	105	180	30	390	165
3	3	480	60	480	60	300	75	240	105
3	4	480	240	60	210	480	150	30	135
3	5	480	60	90	240	480	240	30	240
3	6	480	180	120	105	90	195	30	60
3	7	480	45	300	135	390	75	270	150
3	8	480	120	300	120	180	165	390	120
3	9	120	165	480	150	270	30	360	120
3	10	480	45	150	120	210	15	210	105
3	11	90	45	150	120	150	105	300	30
3	12	180	75	300	165	240	120	390	120
3	13	210	75	60	105	150	90	270	150
3	14	300	195	420	105	330	75	240	165
3	15	150	30	120	135	60	90	390	90
3	16	480	120	420	150	270	150	300	60
3	17	210	195	180	105	270	105	210	165
3	18	480	180	480	210	330	135	240	60
3	19	90	30	480	75	30	15	270	150
3	20	90	45	120	60	150	60	270	30

Appendix B

SPSS Output

All Data- Accuracy

Effect		Value	F	Hypothesis df	Error df	Sig.	Observed Power[b]
equipment	Pillai's Trace	.164	11.550[a]	1.000	59.000	.001	.917
	Wilks' Lambda	.836	11.550[a]	1.000	59.000	.001	.917
	Hotelling's Trace	.196	11.550[a]	1.000	59.000	.001	.917
	Roy's Largest Root	.196	11.550[a]	1.000	59.000	.001	.917
activity	Pillai's Trace	.025	1.540[a]	1.000	59.000	.220	.231
	Wilks' Lambda	.975	1.540[a]	1.000	59.000	.220	.231
	Hotelling's Trace	.026	1.540[a]	1.000	59.000	.220	.231
	Roy's Largest Root	.026	1.540[a]	1.000	59.000	.220	.231
frequency	Pillai's Trace	.253	19.975[a]	1.000	59.000	.000	.993
	Wilks' Lambda	.747	19.975[a]	1.000	59.000	.000	.993
	Hotelling's Trace	.339	19.975[a]	1.000	59.000	.000	.993
	Roy's Largest Root	.339	19.975[a]	1.000	59.000	.000	.993
equipment * activity	Pillai's Trace	.096	6.265[a]	1.000	59.000	.015	.692
	Wilks' Lambda	.904	6.265[a]	1.000	59.000	.015	.692
	Hotelling's Trace	.106	6.265[a]	1.000	59.000	.015	.692
	Roy's Largest Root	.106	6.265[a]	1.000	59.000	.015	.692
equipment * frequency	Pillai's Trace	.066	4.168[a]	1.000	59.000	.046	.519
	Wilks' Lambda	.934	4.168[a]	1.000	59.000	.046	.519
	Hotelling's Trace	.071	4.168[a]	1.000	59.000	.046	.519
	Roy's Largest Root	.071	4.168[a]	1.000	59.000	.046	.519
activity * frequency	Pillai's Trace	.054	3.390[a]	1.000	59.000	.071	.441
	Wilks' Lambda	.946	3.390[a]	1.000	59.000	.071	.441
	Hotelling's Trace	.057	3.390[a]	1.000	59.000	.071	.441
	Roy's Largest Root	.057	3.390[a]	1.000	59.000	.071	.441
equipment * activity * frequency	Pillai's Trace	.000	.003[a]	1.000	59.000	.956	.050
	Wilks' Lambda	1.000	.003[a]	1.000	59.000	.956	.050
	Hotelling's Trace	.000	.003[a]	1.000	59.000	.956	.050
	Roy's Largest Root	.000	.003[a]	1.000	59.000	.956	.050

4000 Hz Only- Accuracy

Tests of Within-Subjects Effects

Measure:MEASURE_1

Source		Type III Sum of Squares	df	Mean Square	F	Sig.
equipment	Sphericity Assumed	1517.610	1	1517.610	7.547	.009
	Greenhouse-Geisser	1517.610	1.000	1517.610	7.547	.009
	Huynh-Feldt	1517.610	1.000	1517.610	7.547	.009
	Lower-bound	1517.610	1.000	1517.610	7.547	.009
equipment * Reviewer	Sphericity Assumed	615.003	2	307.501	1.529	.228
	Greenhouse-Geisser	615.003	2.000	307.501	1.529	.228
	Huynh-Feldt	615.003	2.000	307.501	1.529	.228
	Lower-bound	615.003	2.000	307.501	1.529	.228
activity	Sphericity Assumed	484.282	1	484.282	2.485	.122
	Greenhouse-Geisser	484.282	1.000	484.282	2.485	.122
	Huynh-Feldt	484.282	1.000	484.282	2.485	.122
	Lower-bound	484.282	1.000	484.282	2.485	.122
activity * Reviewer	Sphericity Assumed	236.909	2	118.455	.608	.549
	Greenhouse-Geisser	236.909	2.000	118.455	.608	.549
	Huynh-Feldt	236.909	2.000	118.455	.608	.549
	Lower-bound	236.909	2.000	118.455	.608	.549
equipment * activity	Sphericity Assumed	54.655	1	54.655	.291	.592
	Greenhouse-Geisser	54.655	1.000	54.655	.291	.592
	Huynh-Feldt	54.655	1.000	54.655	.291	.592
	Lower-bound	54.655	1.000	54.655	.291	.592
equipment * activity * Reviewer	Sphericity Assumed	531.205	2	265.602	1.414	.254
	Greenhouse-Geisser	531.205	2.000	265.602	1.414	.254
	Huynh-Feldt	531.205	2.000	265.602	1.414	.254
	Lower-bound	531.205	2.000	265.602	1.414	.254

4000 Hz Only- Efficiency

Multivariate Tests[b]

Effect		Value	F	Hypothesis df	Error df	Sig.
equipment	Pillai's Trace	.580	59.372[a]	1.000	43.000	.000
	Wilks' Lambda	.420	59.372[a]	1.000	43.000	.000
	Hotelling's Trace	1.381	59.372[a]	1.000	43.000	.000
	Roy's Largest Root	1.381	59.372[a]	1.000	43.000	.000
equipment * Reviewer	Pillai's Trace	.086	2.012[a]	2.000	43.000	.146
	Wilks' Lambda	.914	2.012[a]	2.000	43.000	.146
	Hotelling's Trace	.094	2.012[a]	2.000	43.000	.146
	Roy's Largest Root	.094	2.012[a]	2.000	43.000	.146
activity	Pillai's Trace	.795	167.153[a]	1.000	43.000	.000
	Wilks' Lambda	.205	167.153[a]	1.000	43.000	.000
	Hotelling's Trace	3.887	167.153[a]	1.000	43.000	.000
	Roy's Largest Root	3.887	167.153[a]	1.000	43.000	.000
activity * Reviewer	Pillai's Trace	.061	1.399[a]	2.000	43.000	.258
	Wilks' Lambda	.939	1.399[a]	2.000	43.000	.258
	Hotelling's Trace	.065	1.399[a]	2.000	43.000	.258
	Roy's Largest Root	.065	1.399[a]	2.000	43.000	.258
equipment * activity	Pillai's Trace	.134	6.681[a]	1.000	43.000	.013
	Wilks' Lambda	.866	6.681[a]	1.000	43.000	.013
	Hotelling's Trace	.155	6.681[a]	1.000	43.000	.013
	Roy's Largest Root	.155	6.681[a]	1.000	43.000	.013
equipment * activity * Reviewer	Pillai's Trace	.042	.941[a]	2.000	43.000	.398
	Wilks' Lambda	.958	.941[a]	2.000	43.000	.398
	Hotelling's Trace	.044	.941[a]	2.000	43.000	.398
	Roy's Largest Root	.044	.941[a]	2.000	43.000	.398

a. Exact statistic

b. Design: Intercept + Reviewer
Within Subjects Design: equipment + activity + equipment * activity

500 Hz Only- Accuracy

Multivariate Tests[b]

Effect		Value	F	Hypothesis df	Error df	Sig.
equipment	Pillai's Trace	.000	.003[a]	1.000	41.000	.956
	Wilks' Lambda	1.000	.003[a]	1.000	41.000	.956
	Hotelling's Trace	.000	.003[a]	1.000	41.000	.956
	Roy's Largest Root	.000	.003[a]	1.000	41.000	.956
equipment * Reviewer	Pillai's Trace	.161	3.942[a]	2.000	41.000	.027
	Wilks' Lambda	.839	3.942[a]	2.000	41.000	.027
	Hotelling's Trace	.192	3.942[a]	2.000	41.000	.027
	Roy's Largest Root	.192	3.942[a]	2.000	41.000	.027
activity	Pillai's Trace	.031	1.318[a]	1.000	41.000	.258
	Wilks' Lambda	.969	1.318[a]	1.000	41.000	.258
	Hotelling's Trace	.032	1.318[a]	1.000	41.000	.258
	Roy's Largest Root	.032	1.318[a]	1.000	41.000	.258
activity * Reviewer	Pillai's Trace	.007	.145[a]	2.000	41.000	.865
	Wilks' Lambda	.993	.145[a]	2.000	41.000	.865
	Hotelling's Trace	.007	.145[a]	2.000	41.000	.865
	Roy's Largest Root	.007	.145[a]	2.000	41.000	.865
equipment * activity	Pillai's Trace	.008	.327[a]	1.000	41.000	.571
	Wilks' Lambda	.992	.327[a]	1.000	41.000	.571
	Hotelling's Trace	.008	.327[a]	1.000	41.000	.571
	Roy's Largest Root	.008	.327[a]	1.000	41.000	.571
equipment * activity * Reviewer	Pillai's Trace	.010	.197[a]	2.000	41.000	.822
	Wilks' Lambda	.990	.197[a]	2.000	41.000	.822
	Hotelling's Trace	.010	.197[a]	2.000	41.000	.822
	Roy's Largest Root	.010	.197[a]	2.000	41.000	.822

500 Hz Only- Efficiency

Multivariate Tests[b]

Effect		Value	F	Hypothesis df	Error df	Sig.
equipment	Pillai's Trace	.096	4.330[a]	1.000	41.000	.044
	Wilks' Lambda	.904	4.330[a]	1.000	41.000	.044
	Hotelling's Trace	.106	4.330[a]	1.000	41.000	.044
	Roy's Largest Root	.106	4.330[a]	1.000	41.000	.044
equipment * Reviewer	Pillai's Trace	.164	4.019[a]	2.000	41.000	.025
	Wilks' Lambda	.836	4.019[a]	2.000	41.000	.025
	Hotelling's Trace	.196	4.019[a]	2.000	41.000	.025
	Roy's Largest Root	.196	4.019[a]	2.000	41.000	.025
activity	Pillai's Trace	.781	146.580[a]	1.000	41.000	.000
	Wilks' Lambda	.219	146.580[a]	1.000	41.000	.000
	Hotelling's Trace	3.575	146.580[a]	1.000	41.000	.000
	Roy's Largest Root	3.575	146.580[a]	1.000	41.000	.000
activity * Reviewer	Pillai's Trace	.065	1.428[a]	2.000	41.000	.251
	Wilks' Lambda	.935	1.428[a]	2.000	41.000	.251
	Hotelling's Trace	.070	1.428[a]	2.000	41.000	.251
	Roy's Largest Root	.070	1.428[a]	2.000	41.000	.251
equipment * activity	Pillai's Trace	.017	.698[a]	1.000	41.000	.408
	Wilks' Lambda	.983	.698[a]	1.000	41.000	.408
	Hotelling's Trace	.017	.698[a]	1.000	41.000	.408
	Roy's Largest Root	.017	.698[a]	1.000	41.000	.408
equipment * activity * Reviewer	Pillai's Trace	.007	.143[a]	2.000	41.000	.867
	Wilks' Lambda	.993	.143[a]	2.000	41.000	.867
	Hotelling's Trace	.007	.143[a]	2.000	41.000	.867
	Roy's Largest Root	.007	.143[a]	2.000	41.000	.867

a. Exact statistic

b. Design: Intercept + Reviewer
Within Subjects Design: equipment + activity + equipment * activity

Works Cited

American Speech-Language-Hearing Association. 1991. Guidelines for the audiologic assessment of children from birth through 36 months of age. *Asha Reports,* 33, 37-43.

Beattie, R.C., Kenworthy, O.T., & Vanides, E.L. (2005). Comparison of ABR thresholds using linear versus blackman gating functions for predicting pure tone thresholds in hearing-impaired subjects. *The Australian and New Zealand Journal of Audiology*, 27(1), 1-9.

Burkard, R.F., Don, M., & Eggermont, J.J. (2007). *Auditory evoked potentials: Basic principles and clinical application.* Philadelphia: Lippincott Williams & Wilkins.

Cutmore, T.R., & James, D. (1999). Identifying and reducing noise in psychophysiological recordings. *International Journal of Psychophysiology*, 32(2), 129-150.

Hall III, J.W. (2007). *New handbook of auditory evoked responses*. Boston: Pearson Education, Inc.

Kurtz, I., & Sokolov, Y. (2004). A new method for the collection of AEPs. *The Hearing Review*, 11(3), 64-66.

Kurtz, I., & Sokolov, Y. (2005, February). Reducing the effect of electric and magnetic fields on auditory evoked potentials. Paper presented at the 28th Midwinter Meeting of the Association for Research in Otolaryngology, New Orleans, LA.

Kurtz, I., & Steinman, A. (2005, February). Kalman filtering in recording auditory evoked potentials. Paper presented at the 28th Midwinter Meeting of the Association for Research in Otolaryngology, New Orleans, LA.

Markovsky, I., Amann, A., & Van Huffel, S. (2008, August). Application of filtering methods for removal of resuscitation artifacts from human ECG signals. In the proceedings of the 30th Annual International Converence of the IEEE and Biology Society (EMBS), Vancouver, Canada.

McCall, S., & Ferraro, J.A. (1991). Pediatric ABR screening: Pass-fail rates in awake versus asleep neonates. *Journal of the American Academy of Audiology,* 2(1), 18-23.

Reich, D.S., & Wiatrak, B.J. (1996). Methods of sedation for auditory brainstem response testing. *International Journal of Pediatric Otorhinolaryngology,* 38(2), 131-141.

Sininger, Y. (1993). Auditory brain stem response for objective measures of hearing. *Ear and Hearing*, 14(1), 23-30.

Sininger, Y., & Cone-Wesson, B. (2002). Threshold prediction using auditory brainstem response and steady-state evoked potentials. In J. Katz (Ed.), *Handbook of clinical audiology* (pp. 298-322). Philadelphia: Lippincott Williams, & Wilkins.

Sokolov, Y. (2007). ABR testing in children made easy. *The Hearing Review*, 2007 International edition.

Sokolov, Y., Kurtz, I., Steinman, A., Long, G., & Sokolova, O. (2006). *Integrity technology: Enabling practical ABR*. Retrieved October 15, 2008, from http://www.vivosonic.com

Stapells, D. R. (1998, October). Frequency specific evoked potential audiometry in infants. Phonak: A Sound Foundation Through Early Amplification.

Stapells, D.R. (2000). Threshold estimation by the tone-evoked auditory brainstem response: A literature meta-analysis. *Journal of Speech-Language Pathology & Audiology,* 24 (2), 74-83.

Stapells, D.R., & Oates, P. (1997). Estimation of the puretone audiogram by the auditory brainstem response: A review. *Audiology and Neuro Otology,* 2 (5), 257-280.

Tannenbaum, S. (2005). Clinical survey: ABRs and ASSRs in post newborn screening applications. *The Hearing Review*, 12(1), 50-51.

Van Looij, M.A., Liem, S.S., Van derBurg, H., Van derWees, J., De Zeeuw, C.I., & Van Zanten, B.G. (2004). Impact of conventional anesthesia on auditory brainstem responses in mice. *Hearing Research*, 193 (1-2), 75-82.