Northeastern University

Graduate School of Engineering

Dissertation Title: Objective estimation of loudness growth using tone burst evoked auditory responses

Author: Ikaro Garcia Araujo da Silva

Department: Department of Electrical and Computer Engineering

Approved for Dissertation Requirement for the Doctor of Philosophy Degree

_____

Dissertation Adviser Date

_____

Dissertation Reader Date

_____

Dissertation Reader Date

_____

Department Chair Date

Graduate School Notified of Acceptance:

_____

Director of Graduate School Date

Objective estimation of loudness growth using tone burst evoked auditory responses

A Dissertation presented by

Ikaro Garcia Araujo da Silva

to

The Department of Computer and Electrical Engineering

In partial fulfillment of the requirements
for the degree of

Doctor of Philosophy

in

Electrical and Computer Engineering

Northeastern University
Boston, Massachusetts

June 2009

Abstract:

Human perception of loudness as function of a wide range of stimulus intensity is hypothesized to vary in complex ways that cannot be accounted for simply by the listener's threshold and discomfort level. The ability to estimate loudness growth as function of stimulus intensity using an objective procedure could potentially allow for customized non-linear hearing aid fitting for patients unable to perform standard subjective auditory tests. Some researchers have investigated the feasibility of using evoked auditory brainstem responses (ABR) and otoacoustic emissions (OAE) as physiological predictors of loudness growth in humans. Results from OAE studies were promising, but lacked an investigation into the selection of its parameters and a measure of the methodological performance in hearing impaired listeners. Previous studies using ABR, on the other hand, did not use frequency specific stimuli and were based on subjective segmentation techniques that did not take into account the residual noise levels in the averaged waveform. The work in this thesis aims at investigating and improving the use of simultaneously recorded tone burst OAE and ABR for assessing frequency specific loudness growth in normal-hearing and hearing-impaired listeners. First, an optimal set of parameters for the OAE estimator is selected based on mean square error criteria. Second, a new signal-to-noise-ratio (SNR) estimation procedure is developed that attempts to take into account non-stationary noise sources present in ABR recordings. Third, several objective ABR segmentation and noise control procedures are tested for estimating loudness growth. Additionally, SNR estimates are compared with ABR loudness growth estimation errors in an attempt to examine dependencies between residual noise levels in averaged ABR waveforms and estimation performance. Results show that for normal listeners and with 1 kHz tone bursts, both OAE and ABR loudness growth estimators can perform within range of standard psychoacoustical procedures. With hearing impaired listeners, or with 4 kHz tone bursts, the performance of the OAE estimator produced a higher mean-square error with respect to a psychoacoustical measure of loudness than most of the ABR estimators. The performance of ABR estimators with hearing-impaired listeners is close to, but not as accurate as the standard psychoacoustical procedures.

List of Acronyms

ABR:	Auditory Brainstem Response
ALSE:	Average Least Square Error
AMSE:	Average Mean Square Error
AMLR:	Auditory Middle Latency Response
BM:	Basilar Membrane
CA:	Cochlea Amplifier
CMM:	Cross-Modality-Matching
ECG:	Electrocardiogram
EEG:	Electroencephalogram
FFT :	Fast Fourier Transform
HL:	Hearing Loss
HIL:	Hearing Impaired Listener
IHC:	Inner Hair Cells
INEX:	INflected EXponential
LDL:	Loudness Discomfort Level
nHL:	Normal Hearing Level
OAE:	Otoacoustic Emissions
OHC:	Outer Hair Cells
SL:	Sensation Level
SNR:	Signal to Noise Ratio
SPL:	Sound Pressure Level
TBABR:	Tone-Burst Auditory Brainstem Response
TBOAE:	Tone-Burst Otoacoustic Emissions

List of Figures

List of Tables:

Table of Contents:

I. Introduction

The specific aim of this thesis is to develop an objective and automated procedure for the estimation of *individual* loudness-growth functions[1] through the use of frequency-specific stimulation of evoked auditory brainstem responses (ABRs) and otoacoustic emissions (OAEs). The ABR and OAE are early electrophysiological and acoustic responses to acoustic stimulation. These responses are routinely used in clinical setting to check the integrity of the auditory system since they do not require a co-operating listener. Thus a procedure for estimating loudness growth based on these two responses would be particularly useful for assessing loudness growth in populations that are incapable of participating in psychoacoustical experiments, such as infants and cognitively impaired patients. Because loudness growth is well known to have significant variability for both normal-hearing and hearing-impaired individuals (e.g., Buus and Florentine 2001; Epstein and Florentine 2005; Whilby, Florentine et al. 2006; Marozeau and Florentine (In Press)), this implies that hearing-impaired listeners with similar thresholds can have vastly different loudness-growth functions, particularly different

[1] The term 'loudness growth', commonly used in the literature and in this manuscript, implies specifically loudness growth as function of stimulus level.

compressive regions. An accurate and objective way of estimating these compressive regions across individuals will allow for a better understanding of loudness growth and its implications towards hearing rehabilitation.

Current clinical psychoacoustical measurements of loudness are limited in their approach to estimate individual loudness growth functions. For example, some procedures in use today only acquire three data points to estimate the loudness growth function for the purpose of hearing-aid gain control. These are: threshold, most comfortable loudness, and uncomfortable loudness (for review, see Dillon 2001). Other procedures attempt to obtain loudness growth estimates through categorical scaling, which is known to be confounded by strong biases depending on the instructions given and a large normative variance (Poulton 1989); (Elberling 1999). In addition, all measurements of loudness growth in clinical settings today require a cooperating and focused listener, being sensitive to fatigue and being potentially difficult/unreliable to use with children, elderly, and cognitively impaired listeners (Kiessling, Dyrlund et al. 1995).

On the physiological side, several attempts have been made to estimate loudness growth from evoked responses. While it has recently been shown that OAEs can be used to estimate the compressive region of the loudness growth functions in normal listeners (Epstein, Buus et al. 2004; Epstein and Florentine 2005), further studies are still required in order to select an optimal and robust set of parameters for improving loudness estimates and to determine its accuracy with hearing impaired listeners (HILs). Several investigators have attempted to use ABRs to find correlations between particular features on the recorded waveform and loudness (Pratt and Sohmer 1977; Serpanos, O'Malley et al. 1997) and have met with moderate success. However, those studies have all had

several short-comings: subjective labeling and interpretation of the ABR waveform features were required, no investigation was done with frequency specific stimuli, and no attempt was made to control for non-stationary residual noise levels in the averaged waveform.

With these issues in mind, the work in this dissertation will focus on four tasks pertaining to the development of an objective procedure for the estimation of *individual* loudness growth functions through ABRs and OAEs. First, an attempt to improve the original procedure for estimating loudness growth through OAEs (Epstein, Buus et al. 2004) will be investigated by selecting parameters from the procedure that yield a minimum mean-square error (MSE) fit to psychoacoustically estimated loudness functions from normal listeners. Second, a new method for estimating residual noise levels in an averaged ABR collected under non-stationary noise sources is proposed. Third, an estimate of loudness growth from ABRs is proposed that: controls for non-stationary residual noise levels, is fully objective in the sense that it does not require a trained human to classify relevant waveform features, and is obtained through frequency specific (narrowband) stimulation. The fourth and final part combines (in a statistical sense) the loudness estimates obtained from concurrently recorded OAEs and ABRs in order to generate a single robust loudness estimate from both evoked responses.

II. Background

2.1 General Auditory Anatomy and Physiology

The main purpose of this section is to provide a brief description of the anatomy and physiology of the auditory system and its relevance towards loudness perception and

the recordings of OAEs and ABRs. Emphasis will be given to studies, theories, and hypotheses that attempt to link auditory anatomy and physiology to loudness and auditory evoked responses. This section will focus mostly on the auditory system from the periphery up to the brainstem. For the sake of size and space, this section will not include any discussion of the higher central auditory system, such as the auditory cortex.

2.1.1 Outer and Middle Ear

Figure 1 shows a general schematic overview of the middle, outer, and inner ear (cochlea). If a sound source is presented to a human listener through a free-field source (i.e., loudspeaker), then the outer ear, or pinna, can significantly alter the properties of the acoustic signal in such a way that is commensurate with sound source spatial location (Musiek and Baran 2007). In typical ABR and OAE recordings, the filtering effects of the pinna are not an issue since the stimulus is presented via an insert earphone that is placed in the ear canal.

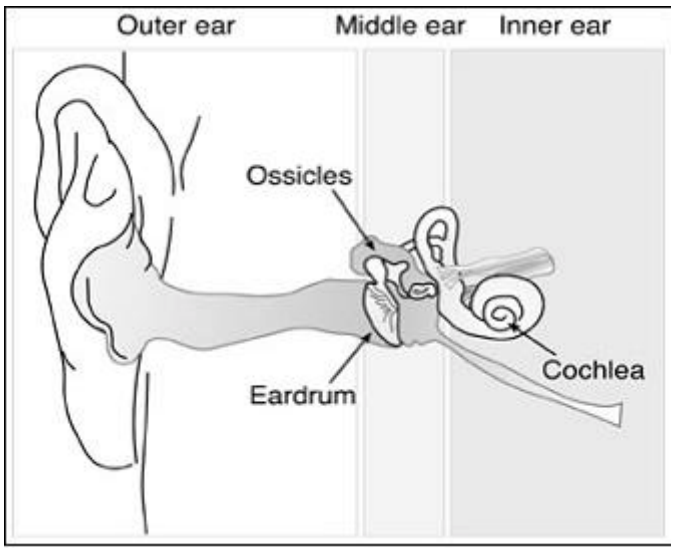

Figure 1- The peripheral auditory system (Baker 1997).

The second peripheral stage of the auditory system involves the ear canal and the middle ear, which consists of the tympanic membrane (the eardrum) and a set of three miniscule bones, the ossicles. The major function of the middle ear is to perform impedance matching between airborne acoustic signals interacting with the tympanic membrane and the fluid-filled cochlear system responsible for the actual mechanical to electro-chemical transduction of sound (Burkard, Don et al. 2006). Both the ear canal and the middle ear have filtering properties that do need to be taken into account when performing evoked response recordings, in particular, OAEs. Figure 2 show the estimated transfer function for the ear canal.

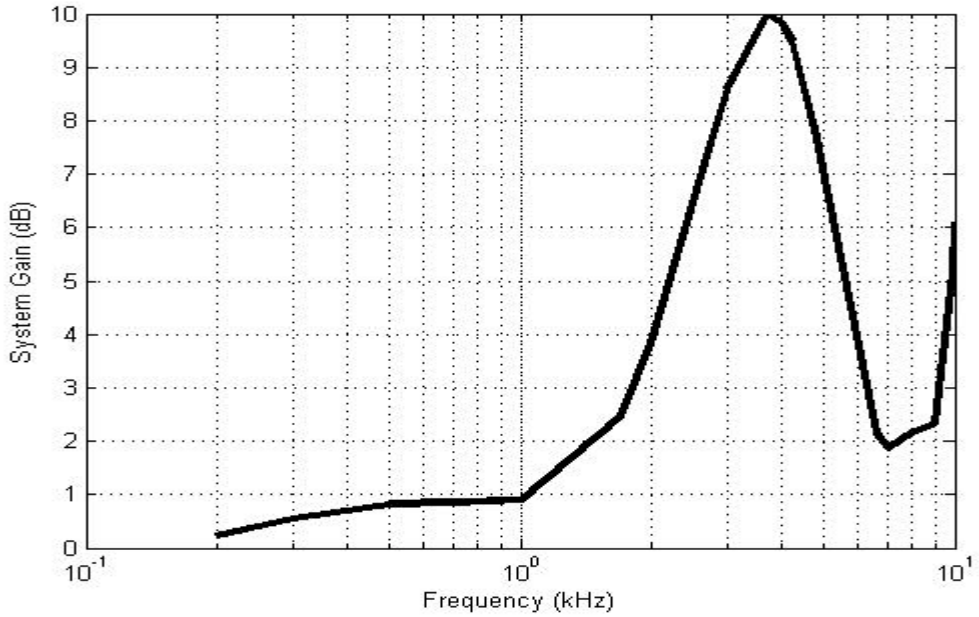

Figure 2- Estimated averaged ear canal transfer function (Keidel and Neff 1974).

Figure 2 shows the change in level (dB) as a function of frequency from a source located outside the ear canal to a source located at the tympanic membrane. From Fig. 2, it is clear that the ear canal has a strong resonance peak between 3 and 5 kHz. The practical implications of this resonance peak, from an OAE recording point of view, is that the sounds coming from the cochlea will be contaminated by a large component of stimulus artifact (ringing) if the stimulus itself has any frequency content around 4 kHz. This artifact also has the potential to contaminate any further response coming from the later stages of the auditory system. Thus the middle ear transfer function presents one of the major difficulties in obtaining artifact-free OAEs, especially those elicited by 4 kHz tone-bursts (Ravazzani and Grandori 1993; Whitehead, Stagner et al. 1994). Furthermore, any damage or degradation of the middle-ear mechanical system can result in absent OAEs for all stimulus frequency, since these emissions originate from the cochlea and require a transmission path to reach the recording microphones in the ear canal (the

transfer function of the middle ear needs to be symmetrical for both incoming and outgoing sounds) (Hall 2006). Hearing losses that only involve the middle ear are classified as *conductive hearing losses* and typically result in less than 40 dB of attenuation particularly at low frequencies.

2.1.2 Cochlea: General Overview

Figure 3 shows a diagram of the Organ of Corti, which is located inside the cochlea and is responsible for the transduction of the mechanical energy coming from the middle ear into electro-chemical energy that is relayed along the central auditory system. This process is mainly driven by fluid vibrations along the basilar membrane (BM) that induce a shearing motion across hair-like projections of the inner hair cells (IHCs), called stereocilia, that result in the opening of ionic gates at the IHC surface. There is a general consensus in the scientific community that the outer hair cells (OHCs) play a key role in a non-linear amplification of the fluid vibrations along the BM (*a.k.a.*, the cochlear amplifier) (Musiek and Baran 2007). This cochlear amplifier mechanism is thought to be the active source of OAEs (Kemp 1978; Musiek and Baran 2007).

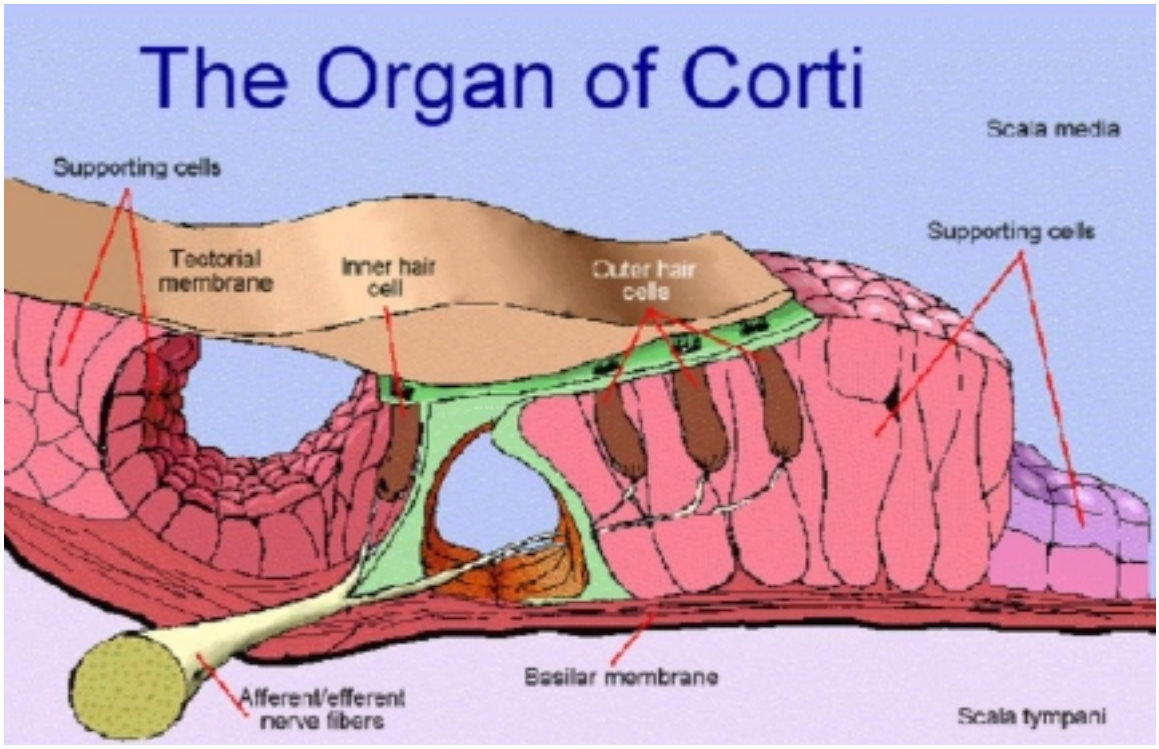

Figure 3- The Organ of Corti, located inside the cochlea and responsible for energy transduction of the auditory system (MEdIC).

Hearing losses resulting from damage to the cochlea, often specifically the Organ of Corti, are classified as sensorineural hearing losses. Due to the crucial role of the non-linear cochlear amplifier and the IHC-transduction system, sensorineural hearing losses can reduce sensitivity tremendously (> 100 dB) and can result in a much more complex spectral pattern of loss than the conductive hearing loss caused by damage or disruption in the middle ear.

2.1.3 Cochlea: Frequency-place Coding as a Function of Level

In addition to performing energy transduction, the cochlea is also responsible for coding the spectral characteristics of the input stimuli by selective activation of IHCs at specific BM locations. The selective activation of IHCs is a direct result of the physical design of the cochlea and the physical properties of the BM. At the basal end of the cochlear, the BM is stiff and thin, and as the BM progress to the apical region of the cochlea, it becomes softer and wider. These physical properties allow the BM to behave as a bank of band-pass filter where the BM's resonant frequency at a given location is determined by the BM's stiffness and width. Thus BM-excitation pattern peaks at locations that resonate with the spectral properties of the stimulus. These excitation peaks provide the IHCs at those locations with the maximum activation.

The frequency-place coding of the BM however, is not fixed as a function of level. In fact, the excitation pattern of the BM changes significantly as a function of level. Figure 4 shows an example of the excitation pattern in the BM of an anesthetized chinchilla to a 10 kHz stimulus (Ruggero, Robles et al. 1992).. It is clear the BM excitation peak location changes and its pattern broadens as a function of level. This broadening of excitation is asymmetrical, in the sense that there is more excitation spreading towards the high frequency region of the BM.

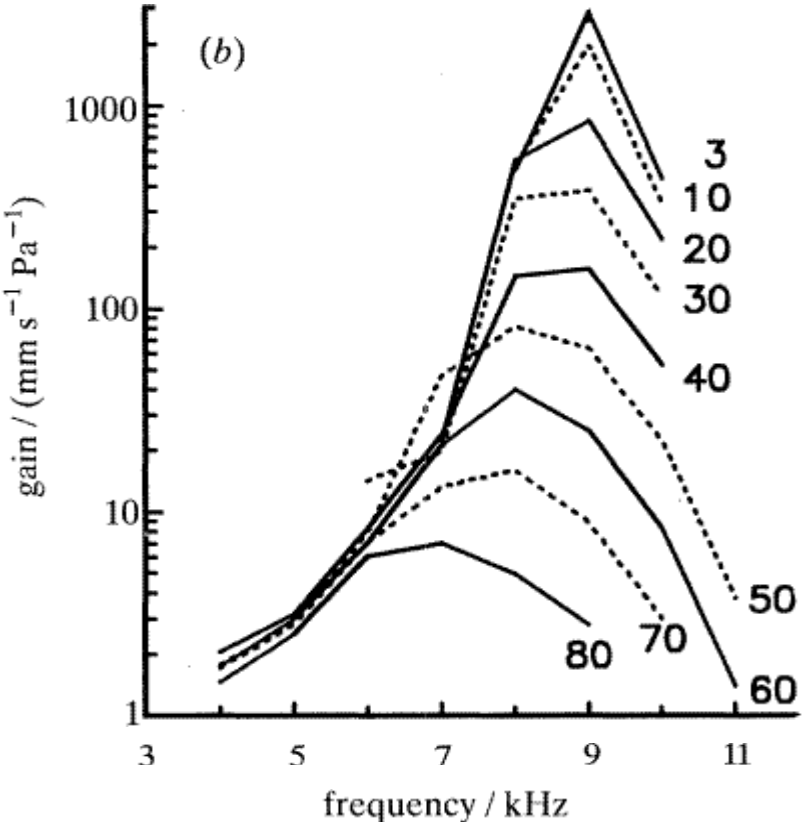

Figure 4- BM gain as a function of level and frequency mapping of the BM (Ruggero, Robles et al. 1992).

2.1.4 Cochlea: The Cochlear Amplifier

The cochlear amplifier (CA) plays a crucial role in maintaining the ~ 120 dB dynamic hearing range in humans. The OHCs are the key component of the CA, and are credited with maintaining a large part of this dynamic range through a mechanism that involves electro-motile expansion of OHC proteins. This non-linear mechanical amplification of the BM vibration is thought to result in improved temporal processing, sharp frequency tuning (i.e.,, frequency selectivity), enhanced sensitivity, and a compressed loudness growth as a function of level (Burkard, Don et al. 2006).

The cochlea amplifier is understood to provide a non-linear amplification of the BM excitation through efferent excitation of the OHC. This has several important implications. Because of this non-linear amplification, the ear is able to emit otoacoustic emissions (OAEs) (Kemp 1978). In addition, this non-linear amplification mechanism can have significant effects if damaged or absent, including an estimated hearing loss between 30 to 50 dB HL (Perlman and Case 1941). In particular, the excitation pattern of the BM at a particular location is linear as a function of sound pressure level (SPL) if no OHC activity is present and/or if the stimulus being presented is outside the resonant frequency range of the specific BM location being measured (Ruggero and Rich 1991). Figure 5 shows a comparison of BM velocity during different stages of the application of a temporary ototoxic agent that disables OHCs.

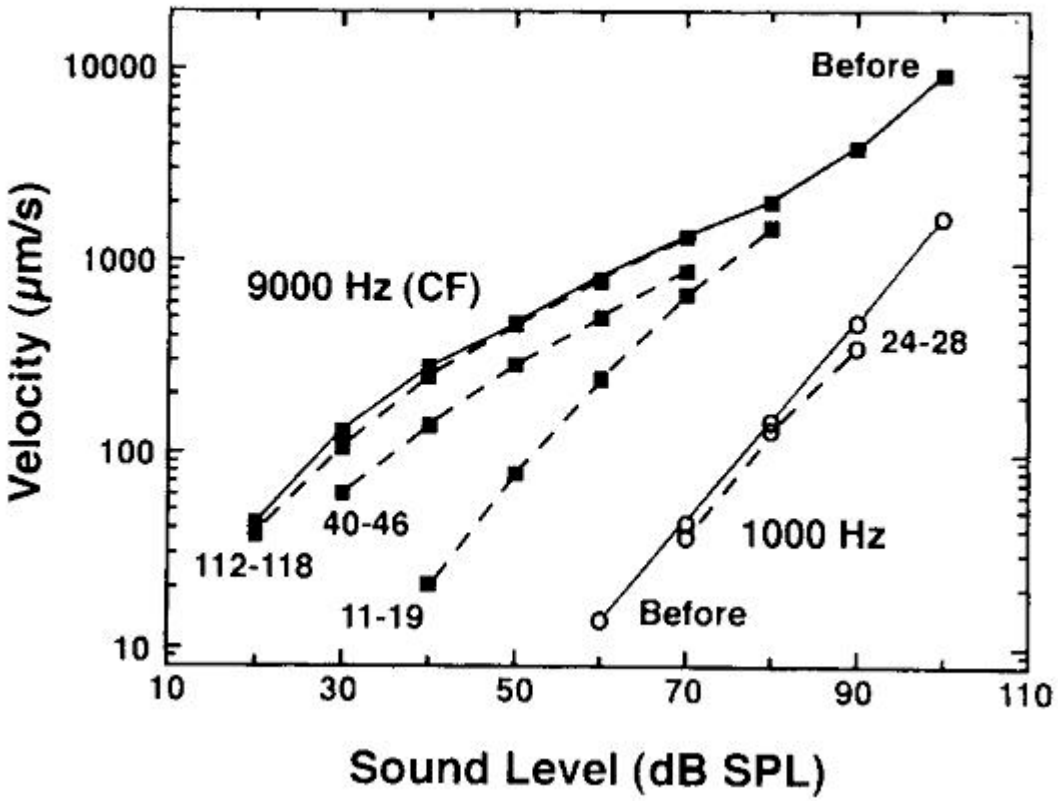

Figure 5 -BM velocity as a function of sound pressure level and OHC activity. BM response is essentially linear as a function of level if no OHC if functioning of if the location of the measured BM velocity is outside the stimulus frequency range (Ruggero and Rich 1991). The number on each curve represents the time in minutes after the administration of a temporary OHC inhibitory drug. The stimulus had center frequency of 9 kHz. Notice the loss of the compressive component where OHC is not active.

A key point from Fig. 5 is that the OHC activity seems to play an important role in determining the compressive characteristic of the BM input-output function. Thus, loss of healthy OHC in the auditory system leads to a loss in the BM dynamic range, which in turn can reduce the dynamic range of the higher-level processing centers in the auditory system (Yates, Winter et al. 1990). Because BM activity is a precursor to any other form of higher auditory processing, it is natural to ask if the BM input-output function characteristics are mirrored in other higher order sensory processes, and if so, what

changes these perceptual processes undergo when the cochlear amplification is no longer present. In fact, some literature studies have indeed established a relationship between BM input-output function and loudness growth (Yates 1990; Yates, Winter et al. 1990; Zhang and Zwislocki 1995; Buus, Florentine et al. 1997; Oxenham and Plack 1997; Schlauch, DiGiovanni et al. 1998; Buus and Florentine 2001; Epstein and Florentine 2005; Johannesen and Lopez-Poveda 2008; Schairer, Messersmith et al. 2008). Figure 8 shows a study (Buus and Florentine 2001) that compared data from BM velocity input-output functions in chinchilla (Ruggero, Rich et al. 1997) with psychoacoustical loudness growth functions in humans (Buus, Florentine et al. 1998; Florentine, Buus et al. 1998). The data seem to have good general agreement at low and moderate levels, particularly given that one might expect across-species differences. At high levels, however, the psychoacoustical and physiological measurements diverge somewhat. The chinchilla cochlear was stimulated by a 10 kHz stimulus, so that BM responses away from the stimulus center frequency have a general linear shape consistent with that from Fig. 5. At high stimulus levels(> 70 dB SPL), however, neighboring lower frequencies start to show a compressive trend. This could be in part explained by the spread of excitation and peak shifting at high levels described in Section 2.1.2.2 (i.e., Fig. 5) (also see: (Zhang and Zwislocki 1995; Buus and Florentine 2001).

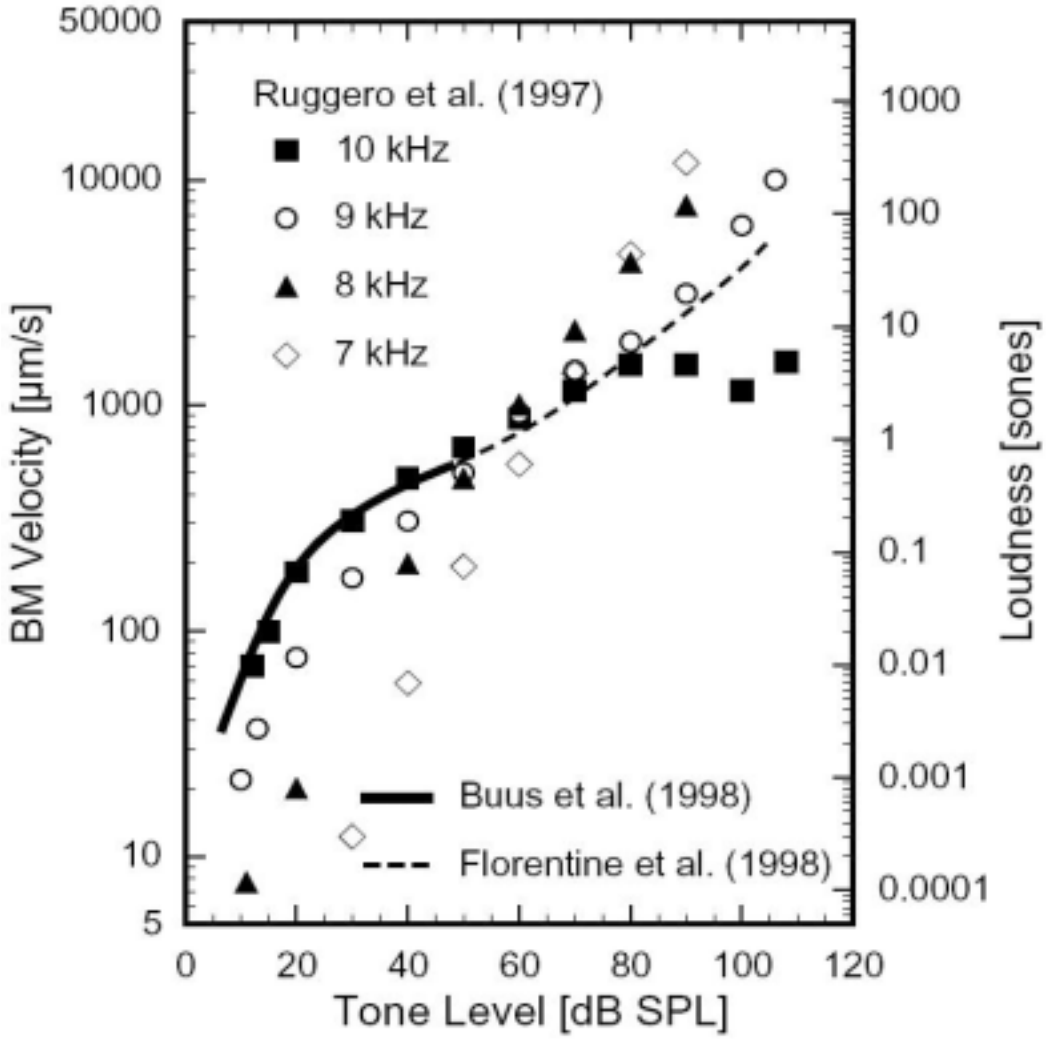

Figure 6- Input-output BM function in chinchilla and loudness growth functions in human (Buus and Florentine 2001). The stimulus used to excite the BM had a center frequency of 10 kHz.

Given that BM input-output function shares some properties with the higher sensory process of loudness growth, it natural to ask how reliable loudness growth can be estimated from BM input-output functions. The usefulness of such relationship and method for estimating loudness growth will become more apparent after a brief analysis of issues involving psychoacoustical estimation of loudness growth functions.

2.2 Psychoacoustical Estimation of Loudness Growth Function

Loudness is typically defined as the quality of a sound that is the primary psychological correlate of sound intensity. Thus it's a subjective trait that is related to the sensation and perception of a stimulus. A couple of methods and procedures have been devised to measure and estimate loudness in humans in order to acquire a better understanding of the processing involved in human auditory systems. However, several issues involving loudness estimation are still a factor in limiting our understanding of loudness. Among such issues is the difficulty of obtaining reliable individual loudness growth measurements. Nevertheless, researchers have heuristically developed reliable models that describe the growth of loudness as a function of level on the *average* human listener.

2.2.1 Loudness Growth Models for Normal Listeners

One of the first attempts to develop a mathematical model for the growth of loudness as a function of stimulus intensity was done by S. S. Stevens (Stevens 1955; Stevens 1957) for review see (Scharf 1978). Stevens modeled loudness growth as a power function of the stimulus sound pressure:

$$L = 0.01 \cdot P^{0.6} \qquad (1)$$

Where L is the estimated loudness in sones and P is the sound pressure (in micropascals). The exponent of 0.6 has for the loudness function is also the reference for the ISO

loudness exponent standard (R131- 1959). Equation (1), however, is not accurate at low levels where loudness is measured close to the hearing threshold. In an attempt to overcome this shortcoming, Hellman and Zwislocki (Hellman and Swislocki 1961) modified equation (1) to account for loudness near threshold:

$$L = 0.01 \cdot (P - P_0)^{0.6} \qquad (2)$$

Where P_0 is a value that approximates the effective threshold (a value of 45 for P_0 is commonly used). Unlike (1), the loudness model in (2) is able to account for a faster growth of loudness at low (threshold) levels, with both models converging at higher SPLs (Fig. 9). Over time, however, it was clear that the models of loudness in (1) and (2) were not able to account for key issues in the growth of loudness. In particular, it was observed that: i) the local exponent of loudness growth close to threshold is bigger than 2, ii) at moderate levels the exponent seems to be lower than 0.6 and grow slower than at higher levels. In order to account for these new findings (i.e., change in slope as a function of level), Buus and Florentine devised a new loudness growth model based on a version of Zwislocki's 1965 model (Zwislocki 1965; Buus and Florentine 2001; Buus and Florentine 2001):

$$L = k[(1 + (snr_{th} \cdot 10^{\frac{SL}{10}})^{s_{-\infty}})^{s_{\infty}/s_{-\infty}} - 1] \qquad (3)$$

Equation (3) estimates the loudness given k, a scale factor, snr_{th}, which is the signal-to-noise ratio (SNR) for a tone at threshold, SL is the sensation level of a tone in dB, s_{∞} is the asymptotic exponent at high levels, and $s_{-\infty}$ is the asymptotic exponent at low levels. Values for these parameters that fit the data for six normal listeners for loudness as a function of stimulus intensity (*not* pressure) are: k=1, snr_{th}=0.25, $s_{-\infty}$=1.4, s_{∞} =0.11.

In an attempt to devise a model of loudness perception that was capable of taking into account hearing loss and binaural stimulation, (Moore and Glasberg 2004)) modified previous models of loudness in literature to create the following level dependent models (for stimulus with frequency higher than 500 Hz):

$$L = k \cdot \left(\frac{2E}{(E + E_{TH})}\right)^{1.5} \cdot [(G \cdot E + A^{0.2})^{0.2}] \qquad 4a$$

$$L = k[(G \cdot E + A^{0.2})^{0.2}] \qquad 4b$$

$$L = k\left(\frac{E}{1.115}\right)^{0.2} \qquad 4c$$

Where E is the excitation in power units, k is a constant, A is a constant that is frequency dependent, G is term that represents the low-level gain of the cochlear amplifier at a specific frequency relative to 500 Hz, and E_{TH} is the excitation produced by a sinusoid at absolute threshold. And where equation (4a) is used for E < E_{TH}, equation (4b) is used for E_{TH} <E< 10^9, and equation (4c) is used for 10^9 <E . An interesting feature of the loudness growth model in (4) is that it attempts to model loudness growth with a component due to the cochlear amplifier through the parameter G in the model. Typical values for the parameters, assuming normal hearing are: E_{TH} =2.31, A=2* E_{TH} =6.62,

More recently, a new loudness growth model, the INflected EXponential (INEX) loudness model has been proposed in order to account for changes in the slope of the

loudness function at low, moderate, and high levels (Buus and Florentine 2001; Florentine and Epstein 2006). It is a simple set of modifications to the classical power function that includes the subtle variations in slope as a function of level observed in a number of studies by a variety of investigators (Robinson 1957; Hellman and Zwislocki 1961; Zwislocki 1965; Stevens 1972; Buus, Florentine et al. 1997; Buus, Florentine et al. 1998; Buus and Florentine 2001). The INEX is computed by the following polynomial:

$$f(L)=1.7058x10^{-9}L^{5} -6.587\ x10^{-7}L^{4} + 9.7515\ x10^{-5}L^{3}-6.6964\ x10^{-3}L^{2+}0.2367L-3.4831 \quad (5)$$

where *f(L)* is the loudness in sones (Stevens 1936) and L is the level in dB SPL. In particular, equation (5) attempts to account for the fact that the loudness growth function has different slopes at low, moderate, and high SPL (other models only account for different slopes in two regions). Figure 7 shows the loudness growth functions based on some of the models for normal hearing mentioned in this section.

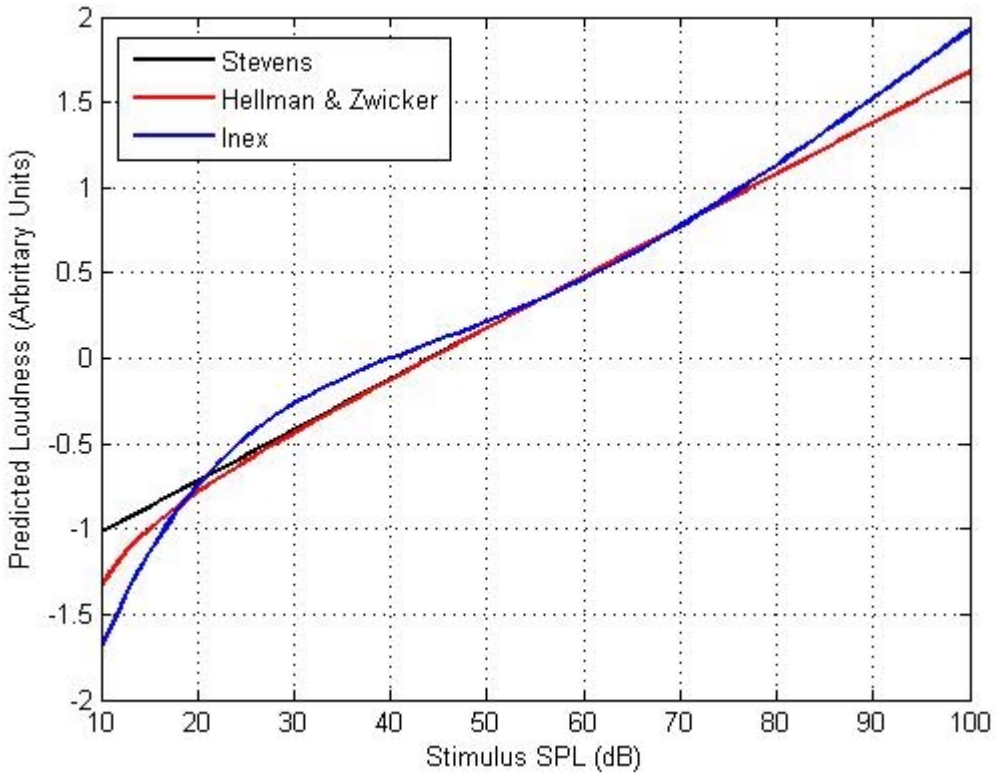

Figure 7- Example of three models for loudness growth in normals. Notice the disagreement of all three models at low levels.

2.2.2 Loudness Growth in Hearing Impaired

Loudness growth functions for HILs with sensorineural hearing loss are thought to behave according to at least two hypotheses: loudness recruitment (i.e.,, rapid-growth) and softness imperception (Fig. 8) (for review see (Marozeau and Florentine 2007). Under the loudness recruitment hypothesis (Fowler 1928; Steinberg and Gardner 1937), the loudness function has a steeper than normal slope close to thresholds (0-30 dB SL) and eventually the loudness function and loudness function slope converges to the normal values at moderate to high SPLs (for review see (Moore 2003) slope estimates assuming the classical model of equation (2) (Hellman and Meiselman 1992). A possible physiological explanation for recruitment seems to be damage to the active feedback of

the cochlear amplifier, since recruitment is observed as peripheral as the OHC level (Zhang and Zwislocki 1995). For HILs exhibiting loudness recruitment, the loudness growth function for both the normal and HILs are significantly different at low levels but identical and high levels.

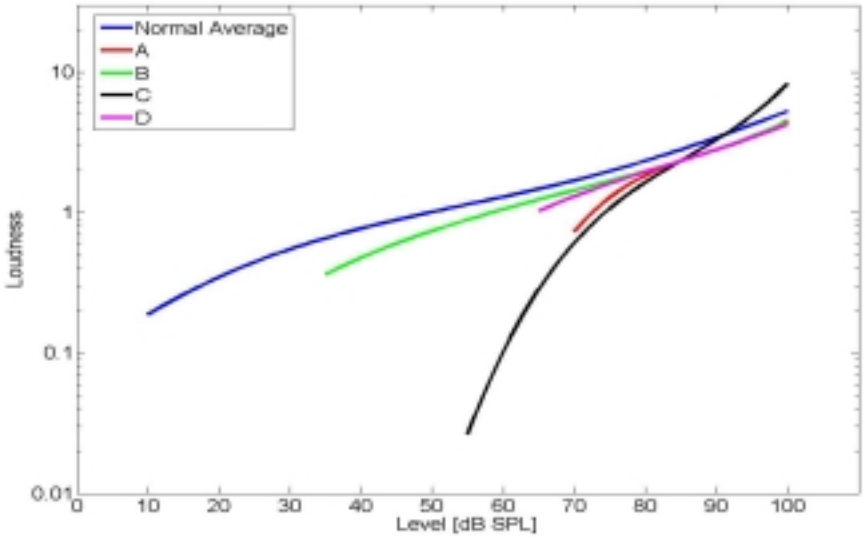

Figure 8- Variability of loudness growth in humans. The blue line represent loudness growth on the average normal listener. The black line is the classical rapid loudness growth called loudness recruitment. The red line is a classical example of softness imperception, where the listener is unable to hear soft sounds. The other lines represent a combination of loudness recruitment and softness imperception.

Two studies done by Hellman and Meiselman looked at loudness growth in 100 HILs (Hellman and Meiselman 1990; Hellman and Meiselman 1992) and found a correlation between degree of hearing loss and the rate of growth of the loudness function and low SL levels. Their findings also indicated an increase in the variance of the estimated slopes as the mean slope value increases. Their investigation found that loudness growth rate ranged from 0.4 to 0.7 with a mean of 0.6 for normals, and a range of 1.04 and 2.92 with a mean close to 2.0 for HILs (three times larger than in normals).

Unlike loudness recruitment, loudness growth for HILs with softness imperception has the same slope as normal listeners, but their perceived dynamic range is much smaller than normal-hearing listeners (NHLs) or HILs with loudness recruitment (Fig. 8) (Buus and Florentine 2001; Florentine 2004). In this study, Buus and Florentine modeled the loudness growth in five HILs using Eq. (3) and found that the loudness growth near threshold is the same for NHLs and HILs. At moderate levels HILs showed an increase in the rate of the loudness growth function, which was attributed to several factors including OHC damage. The models of loudness-recruitment and softness imperception are not exclusive; thus, it is possible that a HIL has both a reduced perceptual range and an abnormalrate of loudness growth near threshold (Marozeau and Florentine 2007). Because of the variability in the shapes of the individual loudness growth function for HILs, it may be crucial to measure loudness growth for proper, personalized hearing-aid fitting.

2.2.3 Psychoacoustical Procedures for Estimating Loudness Growth

There are several psychoacoustical procedures for the estimation of loudness growth over a wide range of SPLs, two of the most common ones are: Absolute Magnitude Estimation (AME or just ME if a reference sound and number is used) and Cross-Modality Matching (CMM). The AME procedure is performed by listeners assigning a numerical value to their perceived loudness of a sound. In similar manner, the general CMM procedure consists of having the listener match the loudness of a sound to another perceptual continua (i.e., the listener can be instructed to cut a string with a length that he/she thinks matches his/her perceived loudness). Both procedures have

been used extensively and in some cases have shown to yield reliable results yielding a loudness exponent of 0.6 using Eq. (1) (Stevens 1955; Hellman and Meiselman 1988; Collins and Gescheider 1989; Hellman and Meiselman 1993).

While these procedures seem reliable, there are, however several issues that might limit their application. Among one of the issues, is a myriad of biases a human can have when ask to make a judgment (for review see (Poulton 1989). Another, albeit obvious, limitation is that the listener is required to be conscious and able to understand and do the task; which might not necessarily be possible for young children, elderly people, and cognitively impaired people. In addition to that, even normal listeners may have their level of concentration reduced by a long experimental task, fatigue, stress, or other factors that can result in inconsistent perceptual judgments. These issues are what, in part, motivate the search for an objective loudness assessment procedure using OAEs and ABRs.

2.3 Estimation of Loudness Growth Through OAEs

This section provides an overview of OAE topics that are relevant to the work done in this thesis. The issues that are covered include: a general background of OAEs and their relation to physiology of the auditory system as described in Section 2.1, issues and limitations involved in measuring and recording OAEs, and a review of studies that have investigated the relationship between loudness growth and a specific type of OAEs from a unified mathematical framework consistent with the notation used throughout this thesis.

2.3.1 General Overview of OAEs

Otoacoustic emissions are sounds emitted, either spontaneously or by stimulation, from the inner ear (typically recorded by a microphone in the ear canal, Fig. 9). These emissions were first discovered by Kemp in 1978 (Kemp 1978; Whitehead, Stagner et al. 1994), but they have been previously speculated to exist due to the non-linear amplification of the inner ear (Gold 1948). Otoacoustic emissions are largely believed to be a result of the non-linear nature of the cochlear amplifier and its response to transient stimulation (for review see Section 2.1.2.3). Evidence for this hypothesis is further corroborated by the fact that in cases of moderate to severe hearing impairment, where the cochlear amplifier mechanism is damaged, no OAEs are observed in clinical tests, though they likely could be observed at a significantly reduced level (Kennedy, Kimm et al. 1991; Whitehead, Stagner et al. 1994).

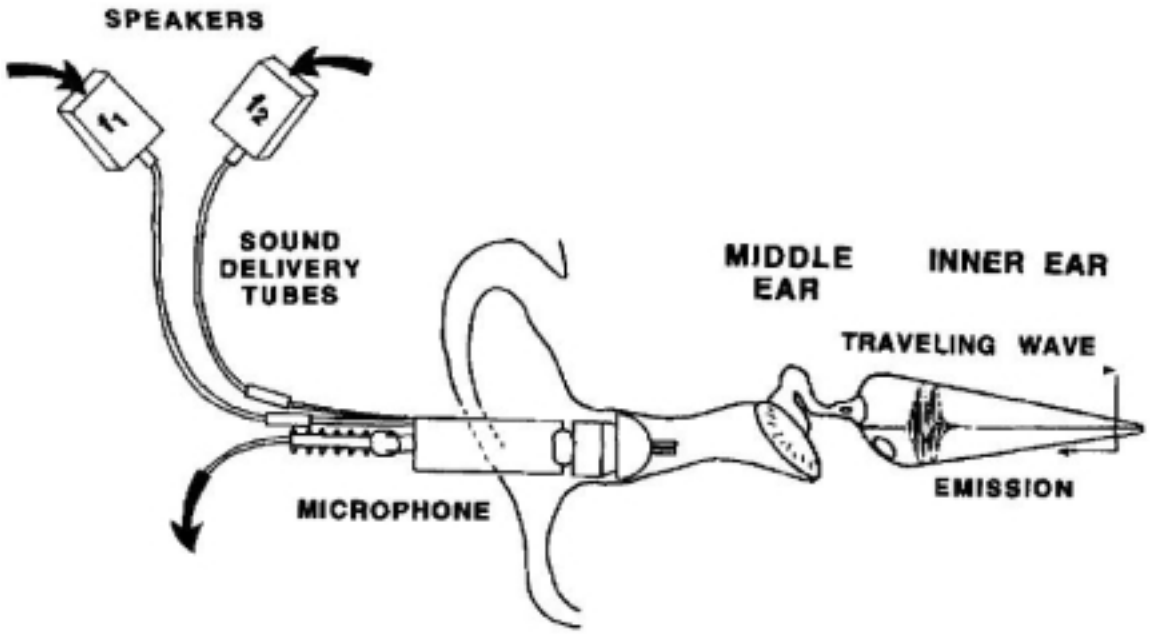

Figure 9- Example of an OAE recording apparatus and the emission being generated in the inner-ear (Whitehead, Stagner et al. 1994).

One common stimulus used for measuring OAEs is the tone burst. Emissions elicited by tone burts are called tone-burst otoacoustics emissions (TBOAEs). Using this methodology, the stimulus is a pure tone, windowed by a short temporal window, such as Hanning window, in order to reduce spectral splatter (Harris 1978). One of the main advantages of TBOAEs is that measurement is relatively fast and the stimulus is narrowband, allowing for a frequency specific assessment of the hearing of the listener. A typical TBOAE trial consists of a windowed tone burst of about a few milliseconds (e.g., 6 ms) followed by the recording of the response for about 20 to 30 ms post-stimulation; these trial are repeated around 300 times and averaged to reduce noise that is not synchronized with the stimulation rate. Because of the low level of the evoked response, the recordings from the microphone located in the ear canal are usually amplified by 40 dB prior to any analog to digital conversion.

Although TBOAEs are fast and reliable, they have some well-known issues that restrict their applications. Because TBOAEs are generated in the inner ear, and travel through the middle ear before reaching the microphone in the ear canal, a healthy middle ear and ear canal are required in order to measure TBOAEs. If the middle-ear-transmission system is not working properly (as in otitis media) the evoked acoustic responses might be severely attenuated, if measurable at all. In addition to that, any material present in the ear canal, such as earwax (cerumen), could alter the response. Thus, it's common practice to check the integrity of the middle ear and ear-canal through tympanometry exams (Burkard, Don et al. 2006). However, even a healthy middle ear and ear canal impose two significant limitations on TBOAEs: limited frequency range and artifact ringing. The first limitation is a simply due to the fact that the middle ear is a band pass system, where frequencies below 500 Hz and above 5 kHz are severely attenuated by the mechanics of the middle ear transmission system (for review see Burkard, Don et al. 2006). The second limitation, artifact ringing, is a direct result of the ear canal shape and size. Because of its size and shape, the ear canal amplifies frequencies close to its resonant peak (Fig. 2) around 3-5 kHz, thus tone-burst stimulation with frequencies in this range are strongly corrupted by stimulus ringing from the ear canal, which can be difficult to separate from the evoked response (Ravazzani and Grandori 1993; Whitehead, Stagner et al. 1994).

Besides the frequency limitation of the middle ear, the physics of the basilar membrane in combination with the nature of the response also imposes a stimulus frequency and duration limitation on TBOAE analysis. In order to record a strong TBOAE, the cochlea needs to be stimulated for a few milliseconds. However, due to the

frequency-place coding of the BM high frequencies will reach the microphone earlier in time than low frequencies (where the site of the response generator is further down towards the apex of the cochlea). The practical implications of frequency-place coding in the BM with respect to TBOAEs recording is that as the frequency of the stimulus increases, the recorded response will move closer to stimulus onset and the artifact ringing. Eventually, if the stimulation is done at a high enough frequency and at a long stimulus duration, the response will overlap with the ear canal stimulus ringing and possibly even the tail end of the stimulus itself, making the use of OAEs for estimation of loudness and hearing thresholds for frequencies of 4 kHz and higher challenging.

2.3.2 Using TBOAEs to Estimate Loudness Growth

As mentioned in Section 2.1.2.3, some researchers have suggested a link between the BM input-output function and loudness growth (Zhang and Zwislocki 1995; Buus, Florentine et al. 1997; Oxenham and Plack 1997; Schlauch, DiGiovanni et al. 1998; Buus and Florentine 2001; Epstein and Florentine 2005; Johannesen and Lopez-Poveda 2008; Schairer, Messersmith et al. 2008). Based on this hypothesis, and the premise that OAEs can give an indirect measure of BM activity, Epstein and Florentine (Epstein, Buus et al. 2004; Epstein and Florentine 2005; Epstein, Marozeau et al. 2006) have devised a method for estimating loudness from TBOAEs with NHLs.). Their method has recently shown that OAEs can be used to estimate the compressive region of the loudness growth functions in NHLs (Epstein, Buus et al. 2004; Epstein and Florentine 2005). The measured signal is assumed to be a combination of the desired response (a combination

of the cochlear amplifier's response to stimulation and acoustic reflection within the cochlea (cite shera)) and noise:

$$x(t,k) = s_{ca}(t) + \eta(t,k) \qquad (6)$$

Where $x(t,k)$ is the recorded response at post-stimulus time *t* and trial (*a.k.a.*, epoch) *k*, $s_{ca}(t)$ is the desired CA response at post-stimulus time *t* (the desired response is assumed to be deterministic and constant across trials), and $\eta(t,k)$ is a mixture of stochastic noise whose properties can vary across trials *k* (i.e., it can be non-stationary) and a deterministic stimulus ringing component. The estimation of the CA response is done in several steps; the first is to discard any trial where the recording in (6) surpasses a predetermined threshold limit *TH*. This common artifact rejection technique is done in order to avoid including trials where a high level of noise is present. After artifact rejection is performed, the trials are broken into two different sets in order to generate two different ensemble averages and truncated with a temporal window of short duration with the aim to temporally filtering the recorded stimulus and stimulus ringing:

$$\hat{s}_{CA1}(t) = \frac{1}{K_1}\sum_j x_1(t,\, j)\cdot w(t) \qquad for \qquad t1 \le t \le t2 \quad (7a)$$

$$\hat{s}_{CA2}(t) = \frac{1}{K_2}\sum_j x_2(t,\, j)\cdot w(t) \qquad for \qquad t1 \le t \le t2 \quad (7b)$$

Where $\hat{s}_{CA1}(t)$ is the estimated OAEs activity, $x_1(t,\, j)$ is the first subset containing half of the trials after artifact rejection has been applied, and $w(t)$ is a temporal window such

as Gaussian, Hanning, or a rectangular window of width equal to *t2-t1*, and the time indices *t1* and *t2* determine the starting and ending times respectively of the estimated OAE activity. In this way, Eq (7) is expected to decrease the stochastic noise component through ensemble averaging, and the deterministic noise component through temporal selection of the recorded waveform (the assumption being that eventually both the stimulus and the stimulus-induced ringing will decay sufficiently while the OAE activity is still present) (Epstein, Buus et al. 2004; Epstein and Florentine 2005; Epstein, Marozeau et al. 2006). The final loudness estimate at the particular level is obtained by summing the positive real component of the cross-spectra of the two averages over a limited frequency range:

$$\hat{\phi}_1(w) = FFT\left(\hat{s}_{CA1}(t) \right) \qquad (8a)$$

$$\hat{\phi}_2(w) = FFT\left(\hat{s}_{CA2}(t) \right) \qquad (8b)$$

$$\hat{\theta}_{LOAE} = \sum_{w=w1}^{w2} (REAL(\hat{\phi}_1(w) \cdot \hat{\phi}_2^{\,*}(w)) > 0) \qquad (8c)$$

Where $\hat{\theta}_{LOAE}$ is the estimated loudness through the TBOAEs, FFT is the Fast Fourier Transform operation, REAL( ) is the real components of a complex variable, $\hat{\phi}_2^{\,*}(w)$ is the complex conjugate of $\hat{\phi}_2(w)$, and *w1* and *w2* are parameters that determine which frequency bins (range) to use of the cross-spectrum for the estimation of

loudness(obtained through FFT cross multiplication). Altogether, this procedure has six parameters: *t1*, *t2*, *w1*, *w2*, *w(t)*, and *TH*.

2.4 Estimation of Loudness Growth Through ABR

This section, like its previous OAE counterpart, provides an overview of ABR topics that are relevant to the work done in this thesis. The issues that are covered include: a general background of ABR, in particular its relation to physiology of the auditory system as described in Section 2.1, issues and limitations involved in measuring and recording ABRs, and a review of studies that have investigated the relationship between loudness growth and ABRs from a unified mathematical framework consistent with the notation throughout this thesis.

2.4.1 General Overview of ABR

The evoked auditory-brainstem response (ABR) is an electrical response generated by the auditory system (from the periphery up to the brainstem) in the first 15-ms of auditory stimulation (for comprehensive review see (Burkard, Don et al. 2006; Hall 2006). The ABR response was first described specifically for humans and animals in 1970 (Jewett 1970; Jewett, Romano et al. 1970). The recorded response contains several morphological features (i.e., peaks and valleys) that are given standard labels named sequentially as waves I, II, III, IV and V (VI and VII also exist). These features of the recorded evoked potential are thought to originate in specific sites in the earlier portions of the central auditory system (Fig. 10).

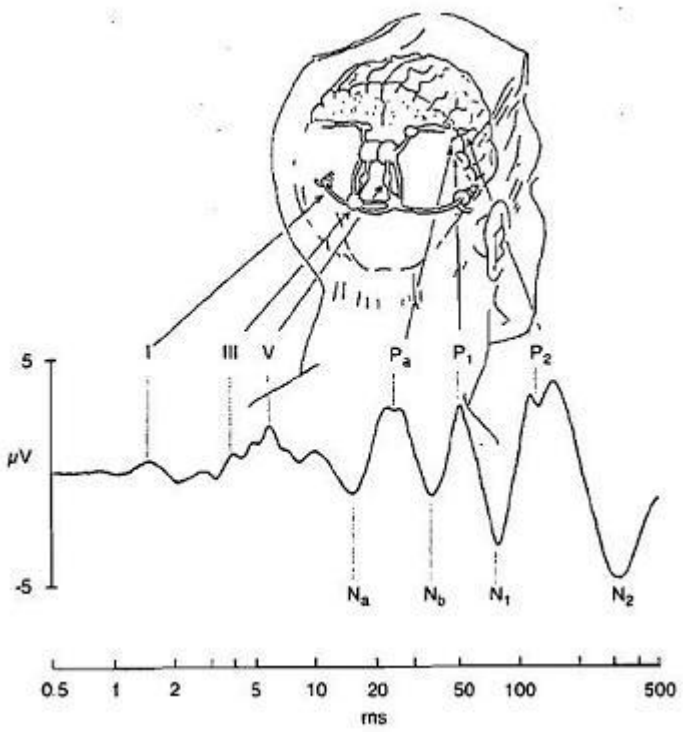

Figure 10- Recorded ABR waveform (first 15 ms of the response) and the speculated site generators of electrical activity (Burkard, Don et al. 2006). The ABR waveforms have a morphology in which peaks and troughs are labeled sequentially. These peaks and troughs are thought to originate at specific locations in the auditory system as the sound transmission progress from the inner ear all the way to the auditory cortex.

The most informative and studied peaks in the ABR are waves I, III, and V. The first component of the ABR, wave I, is thought to originate from the afferent activity of the eight nerve fibers as they leave the cochlea and move toward the brainstem. Wave I typically occurs 1.67 ms after the acoustic stimulation (using clicks). Wave III is believed to originate in the cochlear nucleus, the trapezoid body, and the superior olivary complex; it usually occurs at about 3.8 ms after acoustic stimulation (using clicks). The biggest and most robust landmark, wave V, is thought to be generated by the lateral lemniscus termination in the inferior colliculus and typically occurs at 5.6 ms after acoustical stimulation (using clicks) (Hall 2006). The normative data just described for the three

ABR components are referenced to a brief click stimulus (~ 0.1 ms) at high levels and measured in adult NHLs. The ABR morphology changes in a predictable manner as a function of level. In particular, as stimulus intensity increases, peak locations increase in amplitude and occur sooner afterstimulus onset. A more detailed analysis of the ABR features as a function of level is presented in Section 2.3.3. Figure 11 shows a time-frequency representation of the click ABR waveform for the first four milliseconds (Kiang 1975)

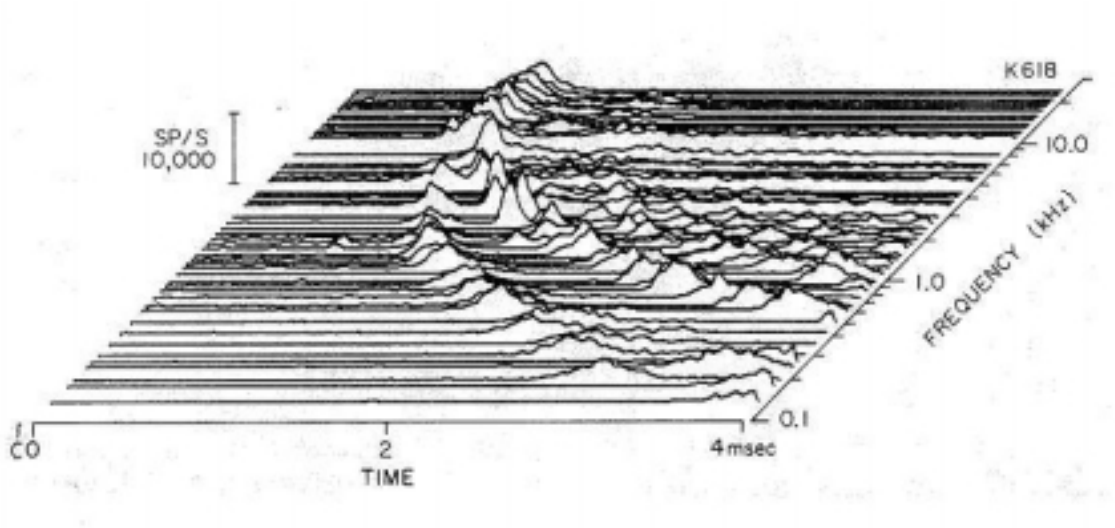

Figure 11- Time-Frequency representation of the click ABR for the first 4 ms (Kiang 1975).

The time-frequency representation of the click ABR as seen on Fig. 11 shows some important characteristics of the response that are thought to be driven in part by the cochlear physical properties: the first components of the response are higher in frequency and in amplitude, and as time progresses, the evoked response systematically decreases in amplitude and frequency, with some frequency bifurcation occurring late in the response. This is in agreement with our understanding of the mechanics of the traveling wave along

the cochlea in conjunction with the assumption that stronger electrical responses are elicited when an ensemble of neurons are activated synchronously. After acoustic stimulation by a broadband signal such as a click, the first components of the cochlear to be activated are the high frequency IHCs located in the basal region of the cochlear. At this region, the peak response is sharp and should elicit an almost synchronous response from the transducer cells. As the wave in the cochlea resulting from the click stimulus travels along towards the apical region, its peak broadens significantly, which results in a less synchronous activation of the IHCs and therefore a smaller electrical response relative to its earlier part. Among the practical implication of this is that ABR responses are typically easier to record/detect with short, high-frequency stimuli, because the neural electrical activity is large due to synchronous discharge of the population of neurons. It has also been suggested that cochlear size can play a significant factor in explaining gender differences in ABR amplitudes, where the smaller size yields better neural synchronization with the stimulus (Don, Ponton et al. 1993; Don, Ponton et al. 1994). In addition to that, because of the dispersion and delay effect along the BM, ABRs recorded from *click* stimuli reflects mostly the combined response of the peripheral auditory system that operates along the frequency range of about 1-4 kHz (Moller and Blegvad 1976; Coates and Martin 1977). Finally, a HIL with a hearing loss at high frequency might show a delayed ABR simply because the stimulus has to transverse the inactive high-frequency region in the cochlear first, prior to eliciting a response in the healthy low-frequency region (Gorga, Reiland et al. 1985; Musiek 1991).

2.4.2 ABR Recording and Signal Processing

The ABR is typically measured by using three electrodes placed at the: forehead (non-inverting), ipsi-lateral earlobe (inverting), and contra-lateral earlobe (ground). The stimuli are usually short tone-bursts (about 6 ms) or clicks (about 0.1 ms) presented via insert earphones similar to those used for OAEs. The SNR for the ABR measurements is very low and often in the range of –10 dB. Thus, averaging can require up to 10,000 trials per level for a reliable estimation. The physiological signal itself is very weak (in the range of nanovolts), so that the signals recorded by the electrodes are amplified by at least by 96 dB and band-pass filtered from 100 to 3,000 Hz. The background noise in the ABR measurements is well known to be non-stationary, in part due to independent muscle, EEG, and ECG activities (for general review see (Burkard, Don et al. 2006; Hall 2006). The ABR signal recording is typically modeled as consisting of two components:

$$x_{ABR}(t,k) = s_{ABR}(t) + \eta(t,k) \qquad (9)$$

Where $x_{abr}(t,k)$ is the ABR recording at post-stimulus time *t* and trial *k*, $s_{abr}(t)$ is a deterministic component related to the acoustic stimulation, and $\eta(t,k)$ is stochastic noise whose properties can vary with trial *k*. Because the noise $\eta(t,k)$ in the ABR signal is non-stationary, it is common practice to use weighted (i.e., Bayesian) averaging in the estimation of the deterministic component of (9) (Elberling and Wahlgreen 1985). Using weighted averaging, the noise source $\eta(t,k)$ is assumed to have a zero-mean Gaussian distribution whose variance is allowed to vary within fixed blocks of trials: $\eta(t,k) : N(0,\sigma_\eta(k))$. The variance within a block of trials is estimated by selecting an

arbitrary fixed post-stimulus time, t_0, and measuring the variance across trials within this selected point (Fig. 12).

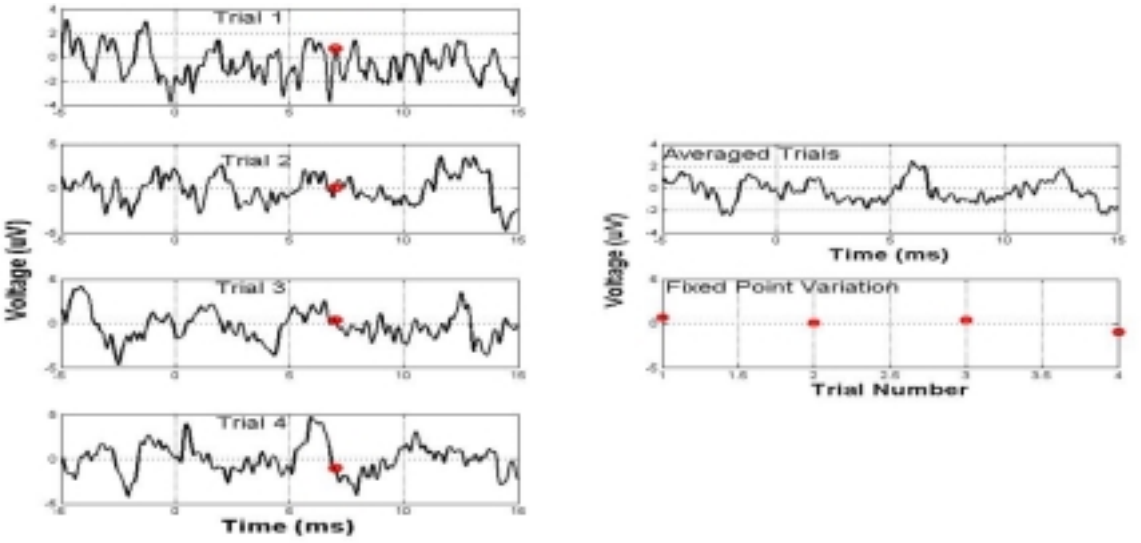

Figure 12- Estimation of noise variance over a block of trials. An arbitrary fixed point is selected with respect to stimulus onset time and the variance of this point across trials (bottom right graph) is the estimated noise variance.

Thus the Single-Point (SP) noise variance estimation described above can be written mathematically as:

$$\hat{\sigma}_{\eta}^{2}(k) = \frac{1}{k_1 - k_0 - 1} \sum_{m=k_0}^{k_1} \left(x_{ABR}(t_0, m) - \overline{x_{ABR}(t_0)}\right)^2 \qquad for \quad k_0 \le k \le k_1 \qquad (10)$$

Where $\overline{x_{ABR}(t_0)}$ is the normal sample mean of the ABR voltage at the selected reference point so that Eq. 10 is an unbiased estimate of the across-trials variance within that block. Once the block variances have been estimated using Eq. 10, the final average is given by the sum of the trials weighted proportionally to the inverse of their estimated variance:

$$\rho(k) = \frac{\frac{1}{\hat{\sigma}_{\eta}^{2}(k)}}{\sum_{m} \frac{1}{\hat{\sigma}_{\eta}^{2}(k)}} \qquad (11\ a)$$

$$\hat{s}_{ABR}^{sp}(t) = \sum_{k} x_{ABR}(t,k) \cdot \rho(k) \qquad (11b)$$

The use of Eq. 11 for estimating the ABR signal has been shown to make the SNR more robust than normal averaging, especially under non-stationary noise. Artifact rejection, similar to that employed on OAEs averaging, can also be applied to ABR normal averaging technique, but the selection of the rejection threshold level itself is not trivial and can have significant effects in both the residual noise levels and the number of trials required to generate an accurate estimate (Elberling and Wahlgreen 1985; Don and Elberling 1994).

A common modification to the single-point noise variance estimation of Eq. 10 is to use multiple fixed points as opposed to a single point. The points selected in the waveform, however, have to be sufficiently far apart as to considered essentially independent. Points that are close together will be inherently correlated due to filtering and the auto-regressive nature of the noise and signal as well as the sampling rate, which can result in overestimation of noise variance (McEwen and Anderson 1975; Hoke, Ross et al. 1984; Elberling and Wahlgreen 1985; Don and Elberling 1994). Modifying Eq. 10 to have L multiple fixed points equally space by Δ samples apart we then have:

$$\hat{\sigma}_{\eta}^{2}(k) = \frac{1}{L}\sum_{n=0}^{L-1} \frac{1}{k_1 - k_0 - 1} \sum_{m=k_0}^{k_1} \left(x(t\cdot\Delta, m) - \overline{x(t\cdot\Delta)}\right)^2 \qquad for \quad k_0 \le k \le k_1 \qquad (12)$$

The inner summation of the Eq. 12 yields L estimates of the noise variance within that block of trials; these L estimates are then averaged to yield the sample mean using the

outer summation. The use of multiple points in (12) allows the block used in the estimation of the noise variance to be smaller and thus better able to detect quick changes in the variance of the noise power as a function of trials. For instance, under the fixed single point scheme, Elberling and Wahlgreen (1985) have suggested the use of 256 trials within a block to yield a stable estimate of the background noise power. Using eight multiple fixed points however, the number of trials within a block can be decreased to only 32 trials while still yielding an accurate noise variance estimate. This allows more accurate detection of the timing of increases in noise power and allows for an appropriate weighting for each trial. It has been speculated however, that the number of points used in Eq. (12) cannot exceed eight samples due the noise characteristics (in particular the bandwidth) of the ABR signal (Elberling and Wahlgreen 1985).

An important application of the ABR is to detect hearing losses based on the SNR as a function of trials used on the averaged waveform. The most common way to estimate the SNR in the ABR is through the Fixed-Single-Point (Fsp) procedure (Don, Elberling et al. 1984; Burkard, Don et al. 2006). This procedure estimates the SNR as the average response of the ratio of the averaged response variance and the estimated background power :

$$\hat{SNR}_{Fsp}(K) = K \cdot \mathrm{var}(\hat{s}_{ABR}(t)) / \hat{\sigma}_n^{\,2} \qquad (13)$$

Where $\hat{SNR}_{Fsp}(K)$ is the estimate SNR of the ensemble average of K trials, $\mathrm{var}(\hat{s}_{ABR}(t))$ is the estimated variance (along time) of the estimated signal, and $\hat{\sigma}_n^{\,2}$ is the average of background noise within a trial (i.e., the average along trials *k* of all estimated noise

variance given by Eq. 10). Notice that Eq. 13 in reality assumes a stationary environment where the background noise is constant across trials, $\hat{\sigma}_n^2$, and where SNR increases proportionally according to the number of trials being added to the average (the classical noise variance reduction of 1/K). A value of $\hat{SNR}_{Fsp}(K)$ =3.1 or higher is commonly used to determine the presence of an ABR signal, and therefore of some auditory functioning (Don, Elberling et al. 1984; Elberling and Don 1984; Elberling and Don 1987; Elberling and Don 1987).

2.4.3 Using ABR to Estimate Loudness Growth

As mentioned briefly on Section 2.3.1, the ABR waveform undergoes systematic changes as function of stimulus level as shown in Fig 13 using tone burst stimulation presented at 0 ms. As the stimulus level decreases the amplitude of the response decreases as well. In addition, the peak location also changes: the peaks are shifted farther away from the stimulus onset as the stimulus intensity is reduced (this delay in peak location is termed "latency" in standard ABR nomenclature); some of the peaks exhibit a broadening effect as well. Clearly, the response has some non-linear components as a function of stimulus level and its natural to wonder if this can bears any correlation with loudness perception. In fact several studies have looked into ABR and loudness or have a assumed a correlation between ABR and loudness as a basis for hearing aid fitting (Bauer, Elmasian et al. 1975; Kiessling 1982; Kileny 1982; Hecox 1983; Kiessling 1983; Suter and Brewer 1983; Howe and Decker 1984; Mahoney 1985; Thornton, Yardley et al. 1987; Thornton, Farrell et al. 1989; Darling and Price 1990;

Davidson, Wall et al. 1990). In particular, some of these studies report even a relative success in hearing-aid fitting based on normalization techniques (for review see Mahoney 1985).

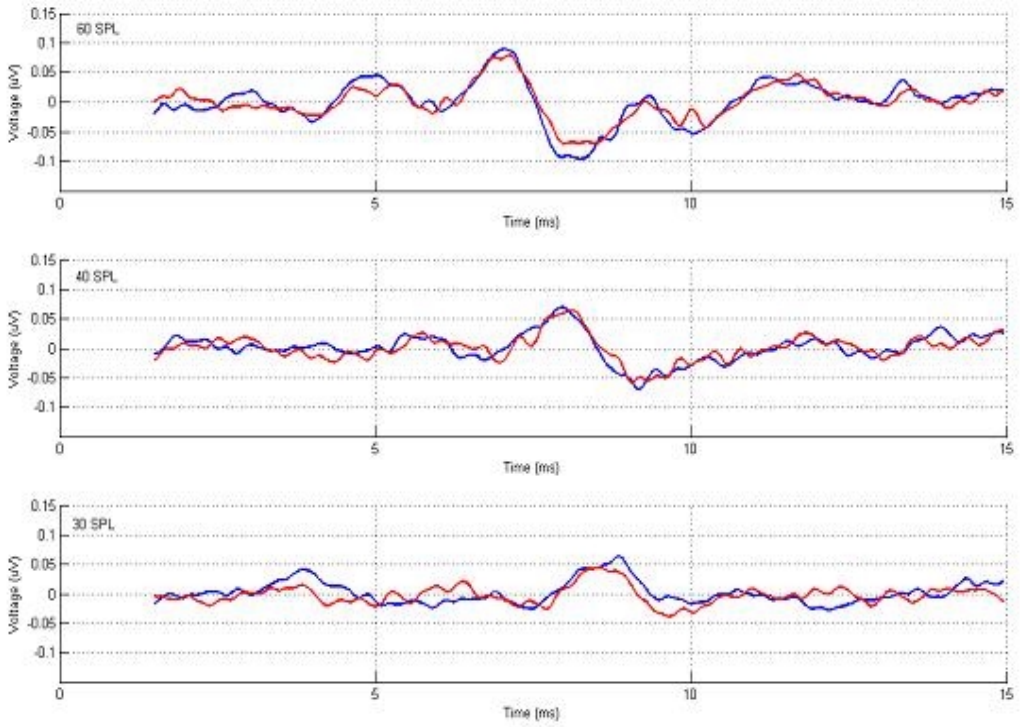

Figure 13- The ABR waveform as a function of stimulus level (using tone-burst stimulus). The systematic changes in the waveform include a decrease in amplitude and an increase in the delay of the ABR peaks as well as a broadening of some peaks. The red and blue signals represent two independent averages at each level.

While the studies previously mentioned examined ABRs and loudness, it is also necessary to investigate the relationship between ABR and the specific shape of the loudness growth function. Several studies have searched for ABR features that correlate directly with loudness growth function (Table I). One of the first studies attempting to estimate loudness growth through ABR was done by Pratt an Sohmer (Pratt and Sohmer 1977). This study used 22 normal young listeners, with both the ABR recording and the

psychoacoustical estimate of loudness done simultaneously. The acoustical stimuli were broadband clicks at 10 different levels (0 SL to 75 SL in steps of 5 dB (0 to 75 in steps of 5 is not 10 levels). During a session, 20-30 random repetitions of each stimulus were presented, with each repetition consisting of a series of 10 short 0.05-ms clicks. The psychoacoustical estimate was obtained using an ME procedure (see Section 2.2.3) with the use of a reference stimulus. The ABR recording for each level consisted of 200-300 trials (10 trials per repetition). The authors looked for correlation between the psychoacoustical loudness growth and three features (latency, amplitude, area) of each of the five waves on the recorded ABR. The experimenter subjectively labeled (or segmented) the ABR waveform into the five different waves. The data analysis was performed in two different ways: i) the individual data for each feature was averaged across subjects and than smoothed by fitting a power function similar to that of Eq. 2; ii) data from each individual listener was smoothed by fitting it to a power function similar to that of Eq.2, and than the average coefficients of the power function for each feature were obtained by averaging the individual coefficients across listeners. No correlation was found for the first (type i) analysis; for the second analysis (type ii) correlations were found between the psychoacoustical estimates and the ABR amplitudes, with a percent of similarity of 99% (wave I), 80% (wave II), 50% (wave III), 10% (wave IV and V), and 20% (wave VI). With the power function exponent fits of wave I being 0.27 and of wave V being 0.14. Despite these high values of similarity, however, the authors concluded that the results were inconclusive (or no correlation existed) because "…a closer analysis proved this similarity superficial, since MEs showed an appreciable intersubject and intersession variability while the auditory nerve and brainstem responses were

approximately constant" (Pratt and Sohmer 1977). They also admitted that the proper set of neural parameters that correlate with loudness might have not been examined with their methodology. Following this exact same study of Pratt and Sohmer (1977), Babkoff et al. (Babkoff, Pratt et al. 1984), attempted a reanalysis of that data by correcting the latency-intensity curves for their asymptotic values (the latency-intensity curves show a saturation effect). These corrected latency-intensity curves yielded exponents in closer agreement with those obtained under the psychoacoustical procedure, but according to the authors, " the closer values of electrophysiological and psychophysical exponents do not necessarily change the conclusions of the original study by Pratt and Sohmer (1977)".

In 1982, Wilson and Stelmack (1982) attempted to reproduce the results of the study by Pratt and Sohmer. Wilson and Stelmack's study compared loudness estimated psychoacoustically using ME and physiologically through ABR intensity-amplitude and intensity-latency function from several wave components of the ABR waveform. The study used 36 student NHLs (mean age of 19.3), with the stimulus being presented from 55 through 90 dB SPL on steps of 5 dB. A major difference between this study and the previous 1977 study was that the 1982 study used 2 minute click-trains as the ME stimuli as opposed to the single clicks used in the 1977 study. The study concluded that power fits to the intensity-amplitude and intensity-latency function of major ABR components did not show any correlation with the ME.

Another study, by Howe and Decker (1984), attempted to look for correlation between wave V latency on the ABR and loudness differences between monaural and binaural listening conditions across a wide range of SPLs. The idea being that if there is an increase in loudness between in the binaural condition relative to the monaural

condition, then the binaural ABR when compared to the monaural ABR, should also reflect this difference. Thus the authors quantified the psychoacoustical loudness difference between binaural and monaural conditions through direct loudness matches; while the physiological ABR difference was quantified in terms of the difference in wave V latency across conditions (based on their discretion that the ABR amplitudes in general are too unreliable to be compared accurately). Their experimental design gathered data from different NHLs for the physiological (12 listeners) and psychoacoustical (10 listeners) conditions. Their stimuli consisted of broadband clicks with 2048 trials collected for each of the levels used (30, 50, 70 and 90 dB SL for ABR and 20, 60, and 80 dB SL for loudness matches). Comparison between the ABR and loudness matching conditions showed no significant correlation. Howe and Decker concluded that because the wave V latency measured from the ABR waveform was equal regardless of listening condition (binaural vs. monaural), this parameter of the ABR waveform was not correlated with loudness.

Despite a history of inconclusive studies, a study by Davidson *et al.* (Davidson, Wall et al. 1990), examined correlations between wave V and loudness-growth functions. Their motivation was in part, due to a preliminary study done by Kiessling (Kiessling 1982; Kiessling 1983) showing the relative success of hearing-aid fitting based on ABR data and a amplitude projection procedure-where the hearing-aid gain was chosen such that the ABR wave V amplitude-intensity is shifted towards a reference normal level across the whole intensity range. A major drawback of Kiessling study, however, was the unverified assumption that the ABR wave V amplitude is correlated with loudness, an assumption that Davidson et al. (1990) sought to investigate. Davidson et al. recorded

click evoked ABRs from 13 listeners (3 HI) in four different session across a four week period (one session per week). During each session, the ABR was averaged across 2048 trials for stimuli presented from 10 to 100 dB SPL in steps of 10 dB. In similar manner to the Kiessling (1982 and 1983) studies, the physiological estimate for loudness was obtained my measuring the negative amplitude of wave V of the ABR response. The psychoacoustical estimate of loudness was obtained through ME on trains of clicks (of about 5 s). Davidson et al. decided to use a rank order Spearman Rho correlation test between the wave V and ME variables to examine if a relationship existed. Based on their analysis, both the NHLs ($r = 0.5 - 0.99$) and HILs ($r = 0.7 - 0.99$) showed significant correlations when all trials were averaged across all four sessions (8192 trials per level). The authors did not, however, find a significant correlation within single sessions (2048 trials). This discrepancies can, in part, be explained by the total number of trials averaged in the comparison, rather than poor ABR "test-retest" characteristics, particularly because no systematic control for residual noise levels was applied.

More recently, a study done by Serpanos et al. (1997), has shown that wave V latency-intensity functions of ABRs correlates with loudness growth for NHLs and for HILS with flat hearing loss (Serpanos, O'Malley et al. 1997). The psychoacoustical loudness measurements were obtained using AME and production in 10 NHLs, 10 HILs with flat frequency loss, and 10 HILs with sloping, high-frequency loss. The ABR analysis was done on two averaged responses consisting of 2,000 trials for each level (20-90 dB nHL in steps of 10 dB). In this particular study, the authors were able to establish a correlation between wave V latency and loudness in normal and flat hearing loss impaired listeners (hearing loss of cochlear origin). No correlation was found in a group

of listener with a sloping configuration of cochlear hearing loss. Of particular interest here, is the fact that the analysis was done on the average of the data and not on individual loudness growth, which is known to be variable even within normal listeners, (Hellman and Meiselman 1990; Marozeau and Florentine 2007).

One of the most recent studies to further look into a link between ABR and loudness growth was done by Gallego et al (Gallego, Garnier et al. 1999). In this study the authors looked for relationship between loudness growth functions in cochlear implantees and electrically evoked ABRs. The experiment was performed on 14 listeners, with the loudness growth functions obtained through categorical scaling (seven levels) and with 2,000 trials collected for each ABR level (analyzed over a 10 ms time window). An analysis of variance (ANOVA) on their data was performed with wave latencies and amplitude as a function of stimulus loudness. The ANOVA showed a significant effect on waves II and V amplitudes and on wave's II latency. The authors also observed a pattern of saturation of wave V amplitude with the "Too Loud" last category of loudness in eight out of the 14 listeners.

Table I shows a summary of all studies- to the best of the authors knowledge- that have investigated a relationship between ABR features as a function of level and loudness-growth functions. While the results from these studies are intriguing and even promising, there are still some major areas for further work and an opening for improvements that could generate significant contributions towards the goal of obtaining objective physiological estimates of loudness-growth functions with possible clinical applications.

Table I. List of previous studies that have looked into relationship between ABR parameters and loudness growth. See text for more detail on each study.

Study	Stimulus	Number of Trials (per level)	Noise Control Mthd	Levels Range (dB)	Psycho. Procedure	# Listeners (Age)	ABR-Loudness Est. Method	Comparison Method	Result
Prattt & Sohmer (1977)	Clicks	200-300	NA	0-75 (steps of 5)	ME per single click	22 Normals (Young)	Power fits to Amp., Latency, and Area	Student's Test on the Coefficients	Correlation with Amp. But Authors claim results are inconclusive
Wilson & Stelmack (1982)	Clicks	4096	NA	55-90 (step of 5)	ME per click train	36 Normals (mean 19.3)	Power fits to Amp, Latency of waves I through VI	Correlation coefficient and comparison of power coefficients	No significant effect found
Babkoff et al (1984) (reanalysis of Pratt & Sohmer 1977)	Clicks	200-300	NA	0-75 (steps of 5)	ME per single click	22 Normals (Young)	Power fits to Latency corrected for asymptote	Comparison of exponents	Latency exponent showed close agreement with Loudness
Davidson et al (1990)	Clicks	8192	NA	20-100 (steps of 10)	ME per click train	10 Normals (22-24) 3 HI	Wave V Amplitude	Rank order correlation	A significant correlation was found.
Serpanos et al (1997)	Clicks	4000	0.025 mV reject	20-90 dB nHL	ME and MP per click train	10 Normals (mean 34.7) 20 HI (mean 50.1)	Wave V Latency	Direct correlation	Correlation was found for Normals and HI with flat HL. No correlation was found for HI with sloping HL.

Gallego et al (1999)	Electric Clicks	2000	NA	20-40 dB	Categorical Scaling (7 levels)	14 Cochlear Implantees	Wave latencies and amplitude	ANOVA with loudness level	Wave II and V amplitudes and Wave II latency were significant

III. Specific Aims

3.1 Introduction

There are four specific aims of this thesis, with the general goal of creating an accurate procedure for the objective estimation of loudness growth. The first aim looks at improving the original method by Epstein and Florentine through a thorough investigation of the effects of various parameters on the TBOAE procedure (Epstein and Florentine 2005; Epstein, Marozeau et al. 2006). The second aim is to devise a method for estimating the post-average SNR of evoked TBABR under non-stationary noise conditions. Briefly, the importance of this aim, in the context of loudness estimation, is underscored by the fact that most investigators who looked into a relation between loudness growth and ABR amplitude-intensity functions, did not control adequately for residual noise levels in the ABR measurements resulting, leading them to concluded that either no correlation was present or that the ABR amplitudes were too unreliable of a

measure (fore review see Table I on Section 2.3.3). The third aim is to attempt to devise a fully objective method for determining loudness growth of frequency specific stimuli from TBABR measurements that rely on the novel residual noise estimation scheme devised from aim 2 without subjective labeling of the ABR waveform morphology. The fourth and final aim of is to attempt to combine the loudness growth estimations from TBOAEs and TBABR to yield an improved general loudness growth estimation for normal and hearing impaired listeners.

3.2 Improving Loudness Growth Estimation through TBOAEs

There is some preliminary evidence that TBOAEs might be useful for estimating loudness (Epstein, Buus et al. 2004; Epstein and Florentine 2005) at frequency-specific sites, as the response for a narrow-band stimulus is likely generated at the characteristic location along the cochlea (Norton and Neely 1987; Shera, Guinan et al. 2002) and OAEs are useful for estimating the peak of the traveling wave (Zweig and Shera 1995). Unfortunately, however, no systematic analysis was done to determine an optimal set of parameters to be used for this estimation procedure (for review see Section 2.3.2.). If TBOAEs are to be used for loudness estimation, it is important and desired that the estimation procedure is insensitive to small variations in parametric choices in order to allow a robust procedure that accounts for anatomical and physiological variability across listeners. In addition, great care must be taken to avoid the linear, non-cochlear portion of the response that results from ear-canal resonance and other trivial acoustic reflections within the auditory system (Kemp 1978; Whitehead, Stagner et al. 1994;

Ravazzani, Tognola et al. 1996) which are particularly prevalent around 4 kHz. For the present study, TBOAEs were recorded in response to 1- and 4-kHz tone bursts and then analyzed using a wide range of parameters to determine which parameter set yielded the minimum average mean-square-error (AMSE) (the average taken across listeners) estimation of loudness growth in reference to loudness growth functions derived using a CMM procedure and the INEX loudness model (Eq. 5). The five parameters studied were parameters: *t1*, *t2*, *w1*, *w2*, *w(t)* where *t1* and *t2* are the starting time and ending time of the temporal analysis window respectively; *w2* and *w2* are the starting and ending frequency of the spectral range used in the analysis; and *w(t)* is the type of temporal window used. In a addition to these five parameters in the original procedure, a sixth parameter, *TH* - which is a rejection threshold for trials above a certain level- was eliminated by using a weighted averaging (a.k.a., Bayesian Averaging) adapted from Elberling and Wahlgreen's procedure on ABR (Elberling and Wahlgreen 1985).

3.3 Estimation of post-average SNR under non-stationary noise

While the main goal of this thesis is to estimate loudness growth from evoked responses, it is fairly well known that in order to get accurate amplitude measurements, specially at the low SNR level of the ABR, it is necessary to control for noise (Elberling and Don 1987; Don and Elberling 1994; Don and Elberling 1996; Gentiletti-Faenze, Yanez-Suarez et al. 2003). Moreover, the methods devised in this section, can a have broader range of application and need not be limited to ABR loudness growth estimation in particular. In any measure of event-related potentials, it is crucially important to be

able to estimate the post-average SNR in order to assess the quality of the measured signals. For instance, the estimated post-average SNR of evoked auditory potentials is important, among several applications, in: detection criterion in identifying infant hearing loss (Don, Elberling et al. 1984), comparing responses obtained across different conditions as well a quality metric in comparing different methods attempting to improve the signal acquisition process (Elberling and Don 1984; Sininger 1995; Don and Elberling 1996; Jacquin, Causevic et al. 2006), and finally, as an important criterion in determining if a sub average is of sufficient quality for use as a reference signal in adaptive noise cancellation in evoked–response measurements (Chan, Lam et al. 1995; Qiu, Chan et al. 1998).

As discussed in Section 2.3.2, the ABR signal is typically modeled consisting of a deterministic component of interest and stochastic zero-mean Gaussian noise (Elberling and Don 1984) (Eq. 9) :

$$x_{ABR}(t,k) = s_{ABR}(t) + \eta(t,k) \qquad (9)$$

It has been observed that for evoked responses, such as evoked auditory brainstem response (ABR), the power of the noise source, $n_k(t)$, is non-stationary, even across several nearby trials (Elberling and Wahlgreen 1985; Don and Elberling 1994). The noise source in (9) is, however, assumed to be *locally* stationary as a function of trial *k*.

Following a matrix notation similar to that of (Davila and Mobin 1992), the ensemble data can be rewritten as:

$$S(t)=\begin{bmatrix} s_1(t) \\ s_2(t) \\ : \\ s_M(t) \end{bmatrix} \qquad N(t)=\begin{bmatrix} \eta_1(t) \\ \eta_2(t) \\ : \\ \eta_M(t) \end{bmatrix} \qquad X(t)=\begin{bmatrix} x_1(t) \\ x_2(t) \\ : \\ x_M(t) \end{bmatrix} = S(t)+N(t) \quad (14)$$

Where *S(t)*, *N(t)* and *X(t)* are all M by T matrices (T being the trial length). The post-average SNR expressed in terms of (14) becomes:

$$SNR=\frac{E\left\{\overline{w}^T S(t)S(t)^T \overline{w}\right\}}{E\left\{\overline{w}^T N(t)N(t)^T \overline{w}\right\}}=\frac{\overline{w}^T R_s \overline{w}}{\overline{w}^T R_\eta \overline{w}} \qquad (15)$$

Where E denotes the expectation operation, $\overline{w}$ is an M by 1 row vector of weights corresponding to the type of averaging being used, the matrices R_s and R_η are M by M covariance matrices for the signal and noise respectively, and σ^2_{res} is the residual noise power in the final average. If the signal is deterministic, all elements of its covariance matrix R_s are constant and equal to the signal's power over the trial, σ_s^2; the signal's covariance matrix R_s can then be written as:

$$R_s=\sigma_s^2\begin{bmatrix} 1 & \dots & 1 \\ : & \ddots & : \\ 1 & \dots & 1 \end{bmatrix} \qquad (16)$$

The assumption that the noise source is locally stationary and independent across trials limits the noise covariance matrix R_η, in (15) to be diagonal. The noise can be modeled as $\hat{P}$ independent noise sources that are stationary across a variable number of trials.

The noise covariance matrix R_η will then have a diagonal structure with P clusters of repeated elements:

$$R_\eta = \begin{bmatrix} R_{\eta 1} & & & 0 \\ & R_{\eta 2} & & \\ & & \ddots & \\ 0 & & & R_{\eta P} \end{bmatrix} \quad where \quad R_{\eta i} = \sigma_i^2 I \qquad (17)$$

Where $R_{\eta i}$ is a variable size diagonal covariance matrix for source *i* (*i.e.*, the number of trials over which a noise source is stationary is variable), *I* is the identity matrix, and the $\hat{P}$ is the total number of distinct, locally stationary noise sources.

3.3.1. Non-stationary SNR Estimation under Normal Averaging

Under normal ensemble averaging conditions the M-by-1 weight vector in (15) is held constant across trials and set to:

$$\overline{w}_{Nave}^T = \begin{bmatrix} \frac{1}{M} & \frac{1}{M} & \ldots & \frac{1}{M} \end{bmatrix} \qquad (18a)$$

$$\hat{s}_{Nave}(t) = \overline{w}_{Nave}^T X(t) \qquad (18b)$$

Under the assumption of a deterministic signal and locally stationary noise sources, we can use equations (16) and (17) to simplify the SNR estimate in (15) to:

$$SNR_{Nave} = \frac{\overline{w}_{Nave}^T R_s \overline{w}_{Nave}}{\overline{w}_{Nave}^T R_\eta \overline{w}_{Nave}} = \frac{\sigma_s^2 (\sum_{i=1}^{M} \frac{1}{M})^2}{\sigma_{Nres}^2} = \frac{\sigma_s^2}{\sigma_{Nres}^2} \qquad (19)$$

Where the residual noise power, σ_{Nres}^2, from $\hat{P}$ locally stationary and independent noise sources is given by (using (17)):

$$\sigma_{Nres}^2 = \overline{w}_{Nave}^T R_\eta \overline{w}_{Nave} = \frac{1}{M^2} tr(R_\eta) = \frac{1}{M^2} \sum_{i=1}^{\hat{P}} \hat{M}_i \hat{\sigma}_{\eta i}^2 \qquad with \quad \sum_{i=1}^{\hat{P}} \hat{M}_i = M \quad (20)$$

Where $\hat{M}_i$ is the number of trials collected under the noise source power $\hat{\sigma}_{\eta i}^2$ (note that $\hat{M}_i$ is random under evoked response experiments and is estimated from a segmentation procedure) and *tr*() is the trace operator. Equation (20) reduces to the classical $\frac{1}{M}$ reduction in the noise power under normal averaging if all P noise sources are equal. The SNR estimation under standard averaging can be done by using the final average and an estimate of the different noise sources:

$$\hat{SNR}_{Nave} = \frac{\hat{\sigma}_s^2}{\hat{\sigma}_{Nres}^2} = \frac{\sigma_{Nave}^2}{\hat{\sigma}_{Nres}^2} - 1 \quad (21)$$

with

$$\hat{\sigma}_{Nave}^2 \equiv \text{var}(\hat{s}_{Nave}(t)) \qquad (22)$$

Where $\hat{\sigma}_{Nave}^2$ is the variance of the averaged trials. The estimation of the residual noise power $\hat{\sigma}_{Nres}^2$ in (20) requires segmentation of trials into stationary regions and the

estimation of the noise variances within these regions. Assuming that the trials have been segmented into $\hat{P}$ sections, the locally stationary noise sources power $\hat{\sigma}_{\eta i}^2$ can be estimated in the same manner as that proposed by (Elberling and Wahlgreen 1985) for weighted (Bayesian) averaging. The noise variance is estimated by selecting L fixed points with respect to the stimulus onset time and measuring the variability of these points across $\hat{M}_i$ trials

$$\hat{\sigma}_{\eta i}^2 = \frac{1}{L}\sum_{n=0}^{L-1}\frac{1}{M_i - 1}\sum_{m=k_i}^{k_i+\hat{M}_i}(x(n\cdot\Delta, m) - \overline{x(n\cdot\Delta)})^2 \qquad (23\,a)$$

$$\hat{\sigma}_{Nres}^2 = \frac{1}{M^2}\sum_{i=1}^{\hat{P}}\hat{M}_i\hat{\sigma}_{\eta i}^2 \quad \textit{with} \quad \sum_{i=1}^{\hat{P}}\hat{M}_i = M \qquad (23b).$$

Where, k_i is number of the first trial collected under the noise source power $\hat{\sigma}_{\eta i}^2$, $\overline{x(n\cdot\Delta)}$ is average value of a fixed point within a set of trials, and Δ is an integer corresponding to the number of samples such that the fixed multiple points are sufficiently far apart to be considered independent and uncorrelated. Points that are close together will be inherently correlated due to filtering and the auto-regressive nature of the noise and signal as well as the sampling rate, which can result in overestimation of noise variance (McEwen and Anderson 1975; Elberling and Wahlgreen 1985; Lutkenhoner, Hoke et al. 1985; Don and Elberling 1994).

3.3.2. Segmentation of Noise Sources

The total number of independent noise sources, $\hat{P}$ in (23), can be estimated recursively by a series of F-tests. The current noise variance, $\hat{\sigma}^2_{\eta i}$, is estimated by using equation (23a) and a parameter that determines the minimum number of trials $M_{\min}$ for all sources. The ratio between the current and previous noise variances are then used in an F-test to determine the likelihood that the current noise power lies within a pre-set confidence interval of the previous noise power

$$F(D_{\eta(i-1)}, D_{\eta i}) = \frac{\hat{\sigma}^2_{\eta(i-1)}}{\hat{\sigma}^2_{\eta i}} \qquad (24)$$

, where the degrees of freedom of the statistic in (24) are related the number of L fixed samples used to estimate the respective noise sources in (23a)

$$D_{\eta i} = L \cdot M_{\min} - 1, \qquad D_{\eta(i-1)} = L \cdot M_{\eta(i-1)} - 1 \qquad (25).$$

If the test statistic in (24) is within the confidence interval of the previous noise source power, the current estimate is updated using recursive averaging

$$\hat{\sigma}^{2\prime}_{\eta i} = \frac{Q\hat{\sigma}^2_{\eta(i-1)} + \hat{\sigma}^2_{\eta i}}{Q+1} \qquad N_{\eta i}' = N_{\eta i} + N_{\eta(i-1)} \qquad (26)$$

, where Q is the total number of variance estimates used to generate the grand average variance estimate of the previous block (*i.e.*, $Q = \hat{M}_{\eta(i-1)} / M_{\min}$). Segmenting the data in this manner allows the noise sources to have a dynamic region of stationarity (but with

minimal segment size of $M_{\min}$). The confidence parameter, *p*, for the statistical test of (24) acts as the smoothing parameter of the segmentation. The extreme value of *p*=1 yields maximum smoothing (all noise source are averaged to a single noise source, similar to current standard methods (Elberling and Don 1984)) and a value of *p*=0 yields the maximum segmentation ($\hat{P}$ = M/ $M_{\min}$).

3.3.3. Non-stationary SNR Estimation under Weighted Averaging

Previous studies have shown that weighted (Bayesian) averaging can significantly improve the SNR of ABR evoked potentials compared to simple artifact rejection (where trials are discarded if a pre-set amplitude limit is reached) (Hoke, Ross et al. 1984; Elberling and Wahlgreen 1985; Don and Elberling 1994). The M-by-1 weighting vector, $\overline{w}_{Wave}$, under the weighted average scheme is given by (Sörnmo and Laguna 2005)

$$\overline{w}_{Wave} = \frac{\hat{R}_\eta^{-1}\overline{1}}{\overline{1}^T \hat{R}_\eta^{-1}\overline{1}} \qquad where \qquad \overline{1}^T = [1 \quad 1 \quad ... \quad 1] \quad (27a)$$

$$\hat{s}_{Wave}(t) = \overline{w}_{Wave}^T X(t) \qquad (27b)$$

, using equation (27) and the covariance for a deterministic signal (16) the estimated SNR from (15) becomes

$$SNR_{Wave} = \frac{\overline{w}_{Wave}^T R_s \overline{w}_{Wave}}{\overline{w}_{Wave}^T \hat{R}_\eta \overline{w}_{Wave}} = \frac{\sigma_s^2 (\sum_{i=1}^{M} w_{Wave}(i))^2}{\sigma_{Wres}^2} = \frac{\sigma_s^2}{\hat{\sigma}_{Wres}^2} \qquad (28a)$$

$$\hat{\sigma}_{Wres}^2 = \overline{w}_{Wave}^T \hat{R}_\eta \overline{w}_{Wave} = \frac{\overline{1}^T \hat{R}_\eta^{-1} \hat{R}_\eta \hat{R}_\eta^{-1} \overline{1}}{(\overline{1}^T \hat{R}_\eta^{-1} \overline{1})^2} \qquad (28b).$$

Where the last equality in (28*a*) follows from the normalization constraint of the weights in (27*a*). From (28*b*) we see that in the case of exact estimation of the noise covariance matrix R_η the SNR under weighted averaging is related to the signal power by

$$SNR_{Wave}^{opt} = \sigma_s^2 \frac{(\bar{1}^T R_\eta^{-1} \bar{1})^2}{\bar{1}^T R_\eta^{-1} R_\eta R_\eta^{-1} \bar{1}} = \sigma_s^2 (\bar{1}^T R_\eta^{-1} \bar{1}) = \sigma_s^2 tr(R_\eta^{-1}) \quad (29)$$

where the last equality follows from equation (17) . Thus the optimal SNR under weighted averaging will increase proportionally to the sum of the inverse of the power of the noise sources (if the environment is stationary, that value converges to a $\frac{1}{M}$ reduction in noise variance) . For a fixed set of known trials (i.e., offline processing), where the trial weights are known ahead of time, the SNR curve in equation (28) is a monotonically increasing function of trial (See Apendix A.1.1). With an exact estimate of the noise covariance matrix, weighted averaging is always superior to normal averaging (with the exception of stationary noise conditions, in which case both techniques are equivalent)(See Apendix A.1.2).

One method to estimate the diagonal elements, $\hat{\sigma}_{\eta i}^2$, of the noise covariance matrix is through the fixed multiple point method from (23*a*). The diagonal elements of $\hat{R}_\eta$ are estimated from (23*a*) and then averaged across statistically similar trials by using the segmentation procedure and equation (26) (this assumes the noise is stationary for at least $M_{\min}$ trials). The estimated SNR under weighted averaging then becomes

$$\hat{SNR}_{Wave} = \frac{\hat{\sigma}^2_{Wave}}{\hat{\sigma}^2_{Wres}} - 1 \qquad (30a)$$

$$\hat{\sigma}^2_{Wave} = \mathrm{var}(\hat{s}_{Wave}) \qquad (30b)$$

$$\hat{\sigma}^2_{Wres} \approx \frac{1}{\overline{1}^T \hat{R}_\eta^{-1} \overline{1}} = (\sum_{i=1}^{\hat{P}} \frac{\hat{M}_i}{\hat{\sigma}^2_{\eta i}})^{-1} \qquad (30c).$$

3.3.4. Predicting Number of Trials Required for Given Residual Noise Level

An issue of practical interest is the ability to predict how many trials are necessary to achieve a minimum post-average residual noise level (this also applies to procedures that use a sub average as a reference signal for adaptive noise cancellation). Under non-stationary noise sources, one way to make this prediction is by assuming that the current noise source will remain stationary for an infinite period ($\hat{M}_{\hat{P}}$ is incremented by the number of trials to predict). Using equation (11) for forecasting, with *M* trials collected

$$BN(\theta) = \sum_{i=1}^{P-1} \frac{M_{i-1}}{(M+\theta)^2} \hat{\sigma}^2_{\eta(i-1)} + \frac{\theta}{(M+\theta)^2} \hat{\sigma}^2_{\eta P} = \frac{c + \theta \cdot \hat{\sigma}^2_{\eta P}}{(M+\theta)^2} \qquad (31a)$$

prior to the current noise source and θ trials under the current noise source with normal averaging, the background noise level, $BN(\theta)$, as a function of trial θ is given by:

Where the constant *c* is given by:

$$c = \sum_{i=1}^{P-1} M_{i-1} \hat{\sigma}^2_{\eta(i-1)} \qquad (31b)$$

The solution to find the number of trials θ to reach a given residual background power level, BN_{TH}, is obtained by applying the quadratic equation to (19):

$$\theta = \frac{\frac{\sigma_P^2}{BN_{TH}} - 2M \pm \frac{1}{BN_{TH}}\sqrt{\hat{\sigma}_P^4 - 4BN_{TH}(M\hat{\sigma}_P^2 - c)}}{2} \qquad (32)$$

A negative or zero value for θ means that the current residual background power is already lower than the desired BN_{TH}. Under a single stationary noise source, $c = M\hat{\sigma}_P^2$ in (20), and the results converge to the classical linear form ($\theta = \frac{\sigma_P^2}{BN_{TH}} - M$).

Determining the number of trials necessary to reach a previous residual power level once an artifact perturbation has occurred might also be of particular interest to clinicians (*i.e.*, $BN_{TH} = \frac{\sigma_1^2}{M}, \hat{\sigma}_P^2 = \sigma_2^2, c = M\sigma_1^2$). Equation (20) with two noise sources then becomes:

$$\theta = M(\frac{\sigma_2^2}{\sigma_1^2} - 2) \qquad (33)$$

If the variance in the new noise source is twice that from the original noise source, the residual noise power in the post-average remains the same. The SNR under normal averaging is also a monotonically increasing function of the number of trials collected as long as the second noise power is less than double the power of the first noise source. From a practical point of view, (21) might be used as a rough estimate of (20) by replacing σ_1^2 with the residual noise power level from (11) after any transients have

decayed. For the case of weighted averaging, a closed form solution is not so straightforward to obtain, however a numerical approach can be implemented by using (18) with the assumption that the power of the current noise source, $\hat{\sigma}_P^2$, remains constant and incrementing $\hat{M}_{\hat{p}}$.

3.4 Loudness Growth Estimation through TBABR

As reviewed in Section 2.3.3, several investigators have studied the relationship between click ABR response and loudness growth (for review see Table I). The work in this thesis plans to expand on those previous studies by taking into consideration the following:

a) The use of frequency specific stimuli (tone bursts) instead of clicks for estimating loudness growth from ABR.

b) The development of a signal-processing scheme that estimates loudness growth from the TBABR waveform in an objective manner without the need for the human labeling (i.e., segmentation).

c) Controlling for residual noise levels in part b through use of a new SNR estimation procedure developed in Section 3.2 that can account for non-stationary background noise activity and use of weighted averaging and collecting a minimum high number of trials (> 2000) per level.

d) The testing of the TBABR and TBOAE loudness procedures with normal listeners as well as HILs.

e) The use of two psycho acoustical procedures (CMM and ME) to serve as reference and a measure of the inherent psychoacoustical variability in estimating individual loudness growth functions.

3.4.1. Use of Frequency Specific Stimuli

Use of frequency specific (tone bursts) for estimating loudness-growth functions from ABR can have strong real-world applications in providing more accurate frequency-specific hearing aid fitting on those not capable of performing psycho acoustical tasks. However there are several challenges present in the use of the tone bursts (specially low frequency) instead of clicks for ABR recordings. Tone bursts ABRs can have a significantly lower SNR (see Section 2.3.1), due in part to a broader peak excitation of the BM, which results in a less synchronized neural response and an overall smaller number of neurons responding. In addition, excitations at different frequencies yield different wave morphologies, which need to be taken into account on the analysis of the response.

3.4.2. Objective Estimation of Loudness Growth from TBABR

The third point towards improving on previous studies was to develop a fully objective procedure from which a loudness growth estimate is obtained from a series of ABR signals collected under different SPLs. Automatic analysis segmentation of the ABR waveform can reduce the cost, analysis, training, and variability of the results. As mentioned in Section 2.3.3 (Fig. 13), there is however an important observation: the

ABR waveform morphology changes non-linearly as function of level, with peak amplitude increasing and latency decreasing with increasing level, in addition to an overall waveform "compression" as the level increases. It is also possible that waveforms obtained at different levels show similarities with each other (i.e., correlation) only at narrow specific regions- an average correlation obtained using the whole signal might not be accurate for certain regions. At low levels its not clear if the wave peaks is simply highly attenuated (a continuous response as function of level around the threshold range) or if there is no response (a binary response as function of level around the threshold range). The estimation of loudness growth from TBABR obtained at different stimulus SPLs consisted of three unique stages: a segmentation stage, a point-estimate stage, and a noise-control stage. The goal of segmentation stage was to select specific regions of the final averaged evoked responses that were to be used in the point-estimate stage. The point-estimate stage is a function that yields single value for a given segmented waveform (i.e., a point estimate can consist of the average power of the waveform or the highest amplitude). The noise-control stage consists of the methods describe in the section above that attempt to correct for non-stationary noise effects on the point estimates by using the NS Fmp statistic. The noise-control methods can operate either on an individual level-by-level basis (Don-Elberling), or across-levels (weighted polynomial fitting).

3.4.2.1 Segmentation Stage

Eight different segmentation techniques were used for segmenting the weighted average evoked response: fullblock, abrblock, amlrblock, waveVamp, amlramp, fullsync,

abrsync, amlrsync (Table II). All of the procedures were applied on the weighted averages of the recorded evoked responses (equation 27, Section 3.3.3). The first three segmentation techniques- fullblock, abrblock, and amrlblock –consisted of simply applying rectangular window to a predefined time region of the evoked response. The fullblock selected the fixed region of the evoked response between 1-ms and 61-ms after the stimulus presentation. The abrblock technique selected a fixed region of the evoked response between 1-ms and 21-ms after the stimulus (the earl component of the evoked response). The amlrblock selected a fixed region of the evoked response between 20 ms and 45-ms after stimulus presentation (late part of the response). The amlrblock response was selected based on some previous findings by Madell and Goldstein (1972), that suggest that this late response, also known as Auditory Middle Latency Response (AMLR), might show some correlation with loudness growth (Madell and Goldstein 1972). The AMLR also has higher amplitude than the ABR, which might make the AMLR more robust to artifact. The drawback of the AMLR, however, is that it might be sensitive to the listener's cognitive state (i.e., deep sleep can significantly change the morphology of the waveform).

The next two segmentation techniques, waveVamp and amlramp, are based on an attempt to select well known specific regions of interest in the evoked response waveform. The waveVamp method is an attempt to select the wave V component of the ABR waveform. It is initiated by selecting the largest amplitude within 4.5 and 10 ms after the stimulus onset on the evoked responses recorded at the highest SPL. The wave V component is then tracked through lower levels by selecting the largest amplitude within 0.5 ms before and 1 ms after the peak location calculated from the previous level. Figure

14 shows an example of how this procedure works with one data set obtained from one of our experiments to be described level (briefly, the stimulus consisted of a 4-ms 1 kHz tone-burst presented at 15 different SPLs). The amlramp segmentation procedure is exactly the same as that of waveVamp procedure except that the initial time region is between 20 and 45 ms after the stimulus and the search is performed over maximum *absolute* peak (as opposed to maximum positive peak in the waveVamp method).

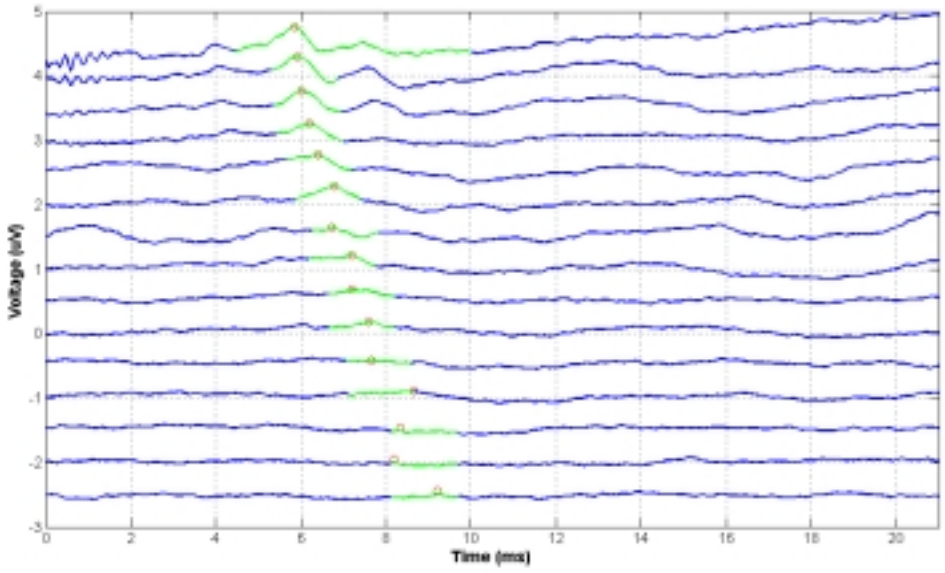

Figure 14- Example of how the wave V segmentation procedures works. The procedure is initiated by selecting the maximum between 4.5 ms and 10 ms after the stimulus onset at the highest recorded level (top). Subsequent wave V locations are determined by selecting the maximum peak within 0.5 ms before and 1 ms after the peak location of the previous level. The green highlight represents the regions over which the maxima were taken. The waveforms in this figure were shifted vertically for ease of comparison.

The last three procedures created (fullsync, abrsync, and amlrsync) segment the evoked responses based on the degree of similarity between the current responses and the responses at the previous higher level. (It is assumed that the overall signal shape between two waveforms is the similar if they were obtained at a close SPL.) The three

procedures are essentially the same with the only modification being the time regions in which they are allowed to operate, with fullsync operating between 0.5 through 45 ms after the stimulus, abrsync operating from 0.5 through 21 ms after the stimulus, and amlrsync operating from 20 through 45 ms after the stimulus. The first step in this procedure is to time-align the previous evoked response (obtained at a higher level) with the current evoked by selecting the time-lag that yields the maximum amount of cross-correlation, $\hat{R}_{\hat{s}_1\hat{s}_2}(\tau)$, between the waveform at the previous highest level, $\hat{s}_1(n)$, and at the current level, $\hat{s}_2(n)$ waveforms (where *n* corresponds to the sample number and N the number of samples),

$$\hat{n} = \arg\max_{\tau} \hat{R}_{\hat{s}_1\hat{s}_2}(\tau) - N/2 \qquad (34\ a)$$

$$\hat{s}_{1b}(n) = \hat{s}_1(n - \hat{n}) \qquad (34\ b)\ .$$

The estimated optimal time lag of the previous level with respect to the current level, $\hat{n}$, is constrained to be within an equivalent 2-ms range (if $\hat{n}/Fs > 2$ ms then $\hat{n}$ is set to 0). The cross-correlation between the two weighted averaged waveforms, $\hat{R}_{\hat{s}_1\hat{s}_2}(\tau)$, has length of N*2-1 and is estimated through (Orfanidis 1996)

$$\hat{R}_{\hat{s}_1\hat{s}_2}(\tau) = \begin{cases} \sum_{n=0}^{N-\tau-1} \hat{s}_1(n+\tau)\hat{s}^*_2(n) & \tau \geq 0 \\ \hat{R}^*_{\hat{s}_1\hat{s}_2}(-\tau) & \tau < 0 \end{cases} \qquad (35)\ .$$

The time-aligned waveform of the previous high level, $\hat{s}_{1b}(n)$, is then used to select regions on the current waveform, $\hat{s}_2(n)$, by performing a two-step process. The first step

divides each waveform into K sections of fixed length *w*, for each of the K sections a dot-product value is calculated between the two waveforms

$$f(k) = \sum_{i=1}^{w} y_1(i,k) \cdot y_2(i,k) \qquad for \quad 1 \le k \le K \quad (36a)$$

$$y_1(i,k) = \hat{s}_{1b}(i + w \cdot (k-1)) \qquad for \quad 1 \le k \le K, \qquad 1 \le i \le w \quad (36b)$$

$$y_2(i,k) = \hat{s}_2(i + w \cdot (k-1)) \qquad for \quad 1 \le k \le K, \qquad 1 \le i \le w \quad (36c)$$

Notice that a high positive value of *f(k)* implies a high degree of similarity between the two segments, a small value of *f(k)* implies a lack of similarity, and a high negative value of *f(k)* implies a high degree of similarity but with one of the signals inverted (multiplied by –1). The second and final step consists in generating a binary gating signal from *f(k)* by applying a threshold and multiplying the gating signal with the original waveform

$$g(k) = \begin{cases} 1 & for \quad f(k) \ge th \\ 0 & for \quad f(k) < th \end{cases} \qquad (37a)$$

$$\hat{s}_2^{seg}(n) = \hat{s}_2(n) \cdot g(\, floor(\frac{n}{w})) \qquad for \quad 1 \le n \le K \cdot w, \qquad (37b)$$

where $\hat{s}_2^{seg}(n)$ is the final segmented signal from $\hat{s}_2(n)$ and floor( ) is a function that rounds a number to the next smallest integer.

The sync segmentation procedure described (equations 34-37) are exactly the same in the fullsync, abrsync, and amlrsync procedures, with the only difference being the time range in which the segmentation procedure operates. The fullsync procedure used the waveform region from 0.5 to 45 ms after the stimulus, the abrsync region is from 0.5 to 21 ms after the stimulus, and the amlrsync region is from 20 to 45 ms after the

stimulus. Figure 15 shows an example of the segmentation procedure, abrsync, on real data (using th=0 and a time window length of 2 ms). The evoked responses have been shifted vertically for ease of comparison. Table II shows a summary of all the segmented procedures mentioned on this section.

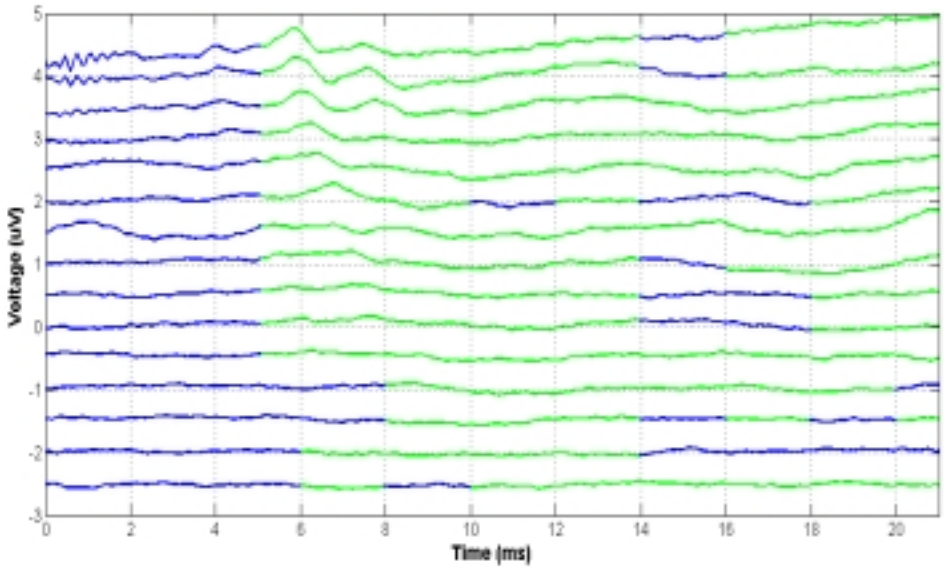

Figure 15- Example of the abrsync segmentation procedure. The green highlight represents the segmented (selected) region that will be used for loudness estimations. The waveforms have been shifted vertically for ease of comparison.

Table II- Summary of the different segmented procedures used.

Name	Summary	Range (ms after stimulus)
fullblock	waveform at the specific range	0.5-45 ms
abrblock	waveform at the specific range	0.5-21 ms
amlrblock	waveform at the specific range	20-45 ms
wave V peak	maximum at specific range	adaptive, starting at 4.5-10 ms
amlr peak	absolute maximum at specific range	adaptive, starting at 20-45 ms
fullsync	selected regions base on dot product	0.5-45 ms
abrsync	selected regions base on dot product	0.5-21 ms
amlrsync	selected regions base on dot product	20 – 45 ms

3.4.3.2 Point Estimate Stage

The purpose of the point estimate stage is to yield a single point estimate given a segmented evoked response waveform. The recorded waveforms were made zero-mean prior to any calculation of the point estimates. Two point estimates were used: the waveform's log of the power and peak amplitude. The peak-amplitude point-estimate was used only for segmentation procedures waveVamp and amlramp.

3.4.3.3 Controlling for Residual Noise Level

The third and final stage of the loudness-growth estimation procedure consisted of attempts to minimize and control for difference in residual noise levels between the point estimates obtained at different SPLs. Two different parametric noise control methods were studied, they were: wpoly, and inex_fitting. These methods used residual noise estimates obtained through the WNS Fmp statistic as described in Section 3.3.3.

The first noise control method investigated, the wpoly, consisted of weighted polynomial fitting to the estimated powers across L different averages (Scharf 1991). The weighted polynomial fitting attempts to minimize the residual noise effects by operating across different L different weighted averages of the waveform. It is a simple polynomial regression fit, where the Px1 polynomial coefficients are given by (Scharf 1991),

$$\vec{\theta}_a = (H'WH)^{-1} H'W\vec{\sigma}_x^2 \qquad (38),$$

where H is a LxP matrix that represents the independent variables in which the Pth order polynomial is to be evaluated (in this case, stimulus SPL), and W is a LxL weighting matrix set to be diagonal and defined in terms of estimated residual noise level for each weighted average

$$W = \begin{bmatrix} \frac{1}{\hat{\sigma}_{\eta 1}^2} & 0 & 0 & 0 \\ 0 & \frac{1}{\hat{\sigma}_{\eta 2}^2} & 0 & 0 \\ 0 & 0 & \ddots & 0 \\ 0 & 0 & 0 & \frac{1}{\hat{\sigma}_{\eta L}^2} \end{bmatrix} \qquad (39)$$

The final noise power for each level is then estimated by evaluating the polynomial at the respective levels L

$$\theta_{wpoly}(L) = \theta_{a5}L^5 + \theta_{a4}L^4 + \theta_{a3}L^3 + \theta_{a2}L^2 + \theta_{a3}L + \theta_{a0} \qquad (40).$$

Where $\hat{\vec{\sigma}}_s^2$ is a vector whose elements correspond to the estimated signal power (or amplitude) as a function of level. For this method the maximum polynomial order was set to 5 (the same order as that of the INEX model described in equation 5) but was allowed

to decrease automatically if the order was bigger than the number of data points or if the matrix $H'WH$ was close to singular (or badly scaled).

The second noise control method, inex fitting, was a parametric approach that involves fitting shifted versions of the INEX function (equation 5) to the point estimates obtained at each level and selecting the best overall fit based on the combination of weighted MSE and residual noise at each level. Thus the estimated loudness growth at level L using the inex fitting is given by

$$\theta_{inex}(L) = a_5(L-\hat{c})^5 + a_4(L-\hat{c})^4 + a_3(L-\hat{c})^3 + a_2(L-\hat{c})^2 + a_1(L-\hat{c}) + a_0 \qquad (41).$$

Where the polynomial coefficients (a_i) are still the same as those in equation (5) and the estimated shift, $\hat{c}$, on the INEX function is give by

$$\hat{c} = \arg\min_x \frac{1}{M}\sum_{i=1}^{M} \frac{(INEX\ (i-x) - \hat{\sigma}_s^2(i))^2}{\hat{\sigma}_\eta^2(i)} \qquad (42),$$

where M is the total number of levels measured, $\hat{\sigma}_s^2(i)$ is the point estimate for level *i*, and $\hat{\sigma}_\eta^2(i)$ is the estimated residual noise level for level i. The general idea behind the parametric loudness estimation based on (42) is based on the concept of softness imperception describe on Section 2.2.2. According to the softness imperception hypothesis, some HILs have normal loudness-growth functions at threshold but their lowest loudness perception level is somewhat elevated beyond normal, thus a shifted the INEX loudness model might yield a good approximation for these types of listeners.

Table III provides a summary of the noise control procedures, and Table IV provides a list of all procedures used throughout the three stages.

Table III- Summary of the different noise control methods described.

Noise Control Name	Summary
wpoly	Operates on full set of point estimates. Weighted polynomial fitting, weights proportional to inverse of residual noise levels.
inex fitting	Operates on full set of point estimates. Best fit determined from weighted combination of MSE and residual noise level.

Table IV- Summary of the three stages for loudness growth estimation and the different methods used.

Stage 1- Segmentation	Stage 2- Point Estimates	Stage 3- Noise Control
Fullblock	Power	wpoly
abrblock	peak (used only with wave V peak and amlr peak)	Inex_fitting
amlrblock		
wave V peak		
amlr peak		
fullsync		

abrsync		
amlrsync		

IV. Experiments

4.1 Introduction

Four major experiments were performed in order to investigate the feasibility of using TBOAEs and TBABRs for the estimation of loudness growth in humans at two different frequencies of interest (1 and 4 kHz). The first experiment consisted of improving loudness estimation through TBOAEs by the selection of a proper set of parameters (as described in Section 3.2) based on the AMSE of a sample set of normal human listeners. The second experiment investigated the accuracy and the selection of a proper set of parameters for the ABR SNR estimation process described on Section 3.3. The third experiment was designed to compare the loudness-growth estimation using TBOAE and TBABRs with that obtained from two other typical psychoacoustical procedures (CMM and ME) for normal-hearing listeners. The fourth and final experiment was to investigate the feasibility of these procedures on a group of HILs.

4.2 Selection of TBOAE Parameters for Loudness Growth Estimation

4.2.1 Listeners

Six listeners with normal hearing (four females, two males), ages 19 to 31, participated in both TBOAE and loudness measurements. No listener had a history of hearing difficulties, and their audiometric thresholds did not exceed 15 dB HL at octave frequencies from 250 Hz to 8 kHz (ANSI 1996). Additionally, all listeners had their middle-ear function evaluated via a clinical exam.

4.2.2. Stimuli

The tone bursts used in all parts of the experiment were two-cycle-up-two-cycle-down pure tones multiplied by Gaussian windows. Two frequencies were tested: 1 kHz and 4 kHz. The 1-kHz tone had a 4-ms duration and the 4-kHz tone had 1-ms duration. This 2-cycle-up-2-cycle-down tone duration ensured energy consistency between the two different frequencies (Hall 2007). The windowed tone was then end-padded with silence to generate a stimulus length of 41.7 ms. The stimulus levels varied from 35 to 100 dB SPL in steps of 5 dB. Levels matched the specifications of the voltage-to-level conversion provided by Etymotic Research for the ER-10c apparatus.

For the CMM procedure, each stimulus presentation consisted of 12 concatenated 41.7-ms intervals in order to generate a train of tone bursts that lasted approximately 0.5 s. This train was presented in place of a single tone-burst in order to minimize any potential temporal integration effects (Florentine, Buus et al. 1996; Buus, Florentine et al.

1997; Zwicker and Fastl 1999). Levels were determined by a pressure-proportional voltage-to-level conversion based on calibration levels measured in a coupler.

4.2.3. Apparatus

The stimuli were generated in MATLAB (2007b running on Windows 2000 for CMM and Ubuntu for TBOAEs) and converted from digital (48-kHz sampling frequency) to analog using a 32-bit Lynx Two Soundcard. The analog signal was then passed through either a Tucker-Davis Technologies (TDT) HB6 (CMM) or a TDT HB7 (TBOAEs) headphone buffer and presented monaurally via Sony MDR-V6 headphones (CMM) or the two transducers of the Etymotic ER-10C (TBOAEs) to a listener inside a double-walled sound-attenuating booth. During the TBOAE measurements, the recordings from the ER-10C were converted from analog to digital (48 kHz sampling frequency) via a Lynx Two soundcard. Routine calibration for each system was performed to test for proper wiring and ER-10C output in a plastic syringe coupler provided by Etymotic. For the TBOAEs, all levels were determined using the RMS of the windowed signal relative to the specifications which were provided by Etymotic and verified by doing an actual in ear measurement for a single listener using a Fonix 6500-CX real-ear system.

4.2.4. TBOAE Recordings

Stimuli were presented in blocks of 1000 trials (about 41.7 ms per trial, at a presentation rate of about 24 Hz). Each block of trials was repeated eight times for each level yielding 8000 recordings per level. For each level, two averages of TBOAE recordings were made. The first average consisted of a weighted mean (Elberling and Wahlgreen 1985) of all the trials in the first half of each of the eight blocks (total of 4000 trials), the second average consisted of a weighted mean of all the trials in the second half of each of the eight blocks (total of 4000 trials). These averages were the basis for the loudness estimation procedure described in Section 4.2.7. Additionally, a separate analysis was done on results recorded in a plastic coupler with the approximate size of a normal ear canal in order to yield a stable reference and an artificial approximation of the acoustic response of the ear canal.

4.2.5. Loudness Growth Estimation through CMM

Listeners were presented with six repetitions of each level in random order and asked to cut a string to be "as long as the sound is loud." After the listener cut each string, they taped it into a notebook and turned the page. Two blocks of trials were run separately, one for each of the two test frequencies. If a particular stimulus was not heard, no string was cut. Levels were omitted if fewer than four out of the six repetitions were heard. All listeners provided at least four strings for levels at 35 dB SPL and higher. The loudness estimate for each level was the transformed geometric mean of the string lengths produced for that level. The transformation was performed in response to the

finding that CMM, though it provides access to the details of the shape of the loudness function for individual listeners, yields functions with shallower slopes than other procedures (Epstein and Florentine 2005). As such, a string-length multiplicative correction factor was determined by using a least squares fit to match the average group data to a power function with an exponent equal to 0.3, widely used as a simple first approximation of the general form of the loudness function (Stevens 1955; Stevens 1957; Stevens 1961; Hellman and Zwislocki 1963; Stevens and Guirao 1964). This correction factor was then applied to the individual data. . The final loudness growth curve was subtracted by an offset in order to yield a zero-mean loudness curve for comparison with loudness curves obtained through other modalities. The comparative parameters of interest and meaning are the slopes of the functions. Thus, the offsetting was performed on individual data sets. Because the scales differ, the vertical offsets between TBOAEs and CMM measurements are arbitrary and essentially meaningless.

4.2.6. INEX Loudness Growth Model

The INEX loudness growth model was used only for comparisons with group average data and not for any individual fitting or parameter optimization. Because this model was designed for longer-duration sounds on the order of around 200-500 ms, the function (Equation 5) was then adjusted by 28.75 dB to account for the difference in sensation level between the very brief tone bursts used in the present study and longer-duration sounds used in most psychoacoustical experiments (Florentine, Buus et al. 1996; Florentine, Epstein et al. 2001; Epstein and Florentine 2005; Epstein and Florentine

2006). This is the equivalent of replacing L in the equation with (L+28.75). This value was determined by calculating a psychometric function from the number of responses to the CMM stimuli and determining the level at which, on average, 50% of trials were detected.

4.2.7. Loudness Growth Estimation from TBOAEs

The procedure used to estimate a loudness growth from the two averaged TBOAE waveforms at each level is similar to that of Epstein and Florentine (2005) (described in detail on Section 3.2). For the present study, four parameters were allowed to vary: window delay from the stimulus onset (from 0 to 39 ms in 1 ms steps), window size (2.5, 10, 20, or 30 ms), window type (rectangular or Hanning), and frequency bandwidth of the analysis region (F-ratio = 1, 1.5, 2, or 3). Combinations in which the sum of window delay and window size exceeded the duration of the recording (41.7 ms) were omitted from analysis. F-Ratio is defined here as the ratio by which the lower and upper bounds of the frequency bandwidth of the analysis region are related to the center frequency. An F-ratio of 1 indicates that only the single FFT bin nearest to the center frequency is used. An F-Ratio of 2 indicates that the lower bound of analysis is one-half of the center frequency and the upper bound of analysis is two times the center frequency. For instance, with a 1-kHz tone-burst, an F-Ratio of two corresponds to an analysis band from 500 Hz to 2 kHz (i.e., an octave wide analysis region on each side of the center frequency).

While varying these parameters, loudness for each level was estimated using three steps. First, each of the two averaged waveforms (weighted, point-by-point means of 4000 trials) was windowed using the window delay, window type, and window size selected. In the second step, FFTs of the two individual averages were calculated and the real components of the cross-spectrum between the two averages determined. Instead of using the FFT absolute magnitude, the real components of cross-spectra were used in order to minimize noise artifacts by including only the portions of the wave that are synchronized in the two averages. In the third and final step, loudness was estimated by summing the positive real components of the cross-spectrum within the frequency region specified by the F-Ratio. The final loudness growth curve was subtracted by an offset in order to yield a zero-mean loudness growth curve for comparison with loudness curves obtained through the other methods utilized.

4.2.8. Peak TBOAE Latency Estimation

The peak latencies of the recorded responses were also estimated as a function of level and stimulus frequency. The recordings at each level were filtered using an octave-wide filter with a center frequency equal to the stimulus frequency. The filter was generated in MATLAB as an 8th-order Butterworth IIR (infinite impulse response) filter. The filtering was performed off-line in conjunction with the function FILTFILT in order to ensure that the filtered response had no phase shifts. The envelopes of these filtered responses were then estimated by passing the magnitude of the Hilbert transform of the filtered signals through a low-pass 8th-order Butterworth filter with cut-off frequency set

to 200 Hz (again in conjunction with the FILTFILT function). The latency of the peak OAE in each condition was the maximum value of the filtered Hilbert transform observed between 2 ms after stimulus offset and 25 ms after the stimulus onset. The initial delay from the stimulus offset was chosen to help reduce the likelihood that the latency observed actually resulted from trivial acoustic artifacts in the ear canal rather than OAEs.

4.3 Estimation of Post-Average SNR under Non-Stationary Noise

The analysis of the quality of the different SNR estimators was done using two experiments. In the first part, an appropriate range for the smoothing parameter *p*, used by the segmentation procedure (the F-test significance level), was determined. The first part, also aimed at comparing the mean-square error of the current standard ABR SNR estimator, the *Fsp* (Burkard, Don et al. 2006; Hall 2006),with the proposed estimator, *NS Fmp* equation (21), and the *WNS Fmp* equation (30); both estimators ignored the offset by negative one in the equations cited. The second experiment compared the receiver operating curves (ROC) for these detectors based on real data (in similar manner to that proposed by (Gentiletti-Faenze, Yanez-Suarez et al. 2003). The choice in deciding to compare the different SNR estimation in terms of detection parameters (ROC curves) was based on the importance that these SNR estimators have in the current hearing-screening procedures. For all experiments, the *Fsp* and the weighted *Fsp* (*WFsp*) noise power estimation, $\hat{\sigma}_n^{\,2}$, was updated using one fixed point on blocks of 256 trials using (23*a*), as originally suggested (Elberling and Wahlgreen 1985). The *Fsp* at any given trial *m* was calculated by

$$Fsp(m) = m \cdot \mathrm{var}(\overline{x(t.m)}) / \hat{\sigma}_n^{\,2}. \qquad (46)$$

The fixed-point location was arbitrarily selected to be the first data point of each waveform. The noise power estimation for *NS Fmp* and *WNS Fmp* was done using (23) with a minimum block length of 32 trials with 8 samples equally spaced per trial and with. For both the *NS Fmp* and *WNS Fmp,* the first point used was arbitrarily selected to be the first data point on each waveform (the same as that of the Fsp procedure), with subsequent points spaced by 50 samples apart (i.e., $\Delta = 50$). These values were chosen based on (Elberling and Wahlgreen 1985) as a result of a trade-off between resolution of the block size, accuracy of estimation, and correlation between samples within a trial. The variances of the individual noise sources were updated accordingly whenever blocks of trials were merged, and the SNR was re-calculated. The noise segmentation for both the *NS Fmp* and *WNS Fmp* was done with the smoothing parameter set to p=0.0005. Notice that the *Fsp* the procedure is equivalent to using the *NS Fmp* procedure with the following parameters: number of points per waveform = 1, number of trials per block=256, and smoothing parameter p = 1.

4.3.1. ABR Measurements

The data from the ABR measurements were from the Gentiletti-Faenze *et al*. (2003) study (courtesy of the original authors). The ABR measurements were done in five normal listeners at 0, 20, 30, 40, 60, and 80 dB SPL. Two records at each level were made (except at 0 dB SPL). The stimuli consisted of monaural rarefaction clicks at a

stimulus rate of 17.5 Hz (with a 5 ms pre-stimulus period followed by a 15 ms recording). A total of 4000 trials were obtained at each level. The signal was sampled at 20 kHz and filtered between 100 and 3000 Hz. The electrodes were place on channels Cz-A2, Cz-A1, and Fz for the ground.

4.3.2. Experiment I: MSE of Residual Noise Estimation

In this experiment, the MSE of the residual noise estimators from (23) and (46) were measured according to the following procedure. For each of the five listeners, an estimate of the noise variance on a trial-by-trial basis was obtained by measuring the variance for each trial under the 0 dB SPL condition. The estimated noise variances of the individual trials were used to generate Monte-Carlo simulations of 50 different sets of experiments for each listener. The Monte-Carlo simulation allowed a measurement of the average decrease in residual noise level in the post-average waveform as a function of the number of trials being added, and under changes in noise variance similar to those observed from real data. The MSE for the *NS Fmp* and the *Fsp* was measured by computing the mean-square-error between the residual noise estimation for each of the 50 simulations and the true simulated value.

4.3.3 Experiment II: ROC Analysis

The ROC analysis for the four procedures (*Fsp, WFsp, NS Fmp,* and *WNS Fmp)* was done in the same way as suggested by (Gentiletti-Faenze, Yanez-Suarez et al. 2003). Briefly, the ABR measurements obtained at 0 SPL were treated as indicative of a hearing loss. The ROC curve for each level was calculated by treating the ABR signal generated

under 0 SPL as the no-signal condition (H0) and the ABR signal generated under a higher level (20, 30, 40, 60, or 80 dB SPL) as the present-signal (H1) condition. For example, if the *Fsp* yielded the following values under H0 (0 dB SPL)*:* 1.93, 1.60, 1.93, and 1.5; with the following SPL values for $H1_{20SPL}$ (20 dB SPL): 2.74, 1.90, 1.75, and 2.41; and the following values for $H1_{30SPL}$ (30 dB SPL): 6.51, 7.09, 6.51, and 6.77; then the ROC areas for the *Fsp* procedure at 20 and 30 dB SPL would be 0.75 and 1.00 respectively.

Hence the ROC area quantifies how well the procedures can discriminate between the ABR measurement being the maximum, and 0.5 equivalent to chance). The median ROC area for all levels (at 0 dB and ABR measurement obtained at higher levels. In order to generate more points for the ROC analysis, the data for each level were broken into three different segments. These three segments were assumed to be independent from each other and were permutated in six different ways in order to simulate six more "experiments" for each level. The performance of each estimator was quantified in terms of the area of the ROC curve (1 20, 30, 40, 60, and 80 dB SPL) was then calculated to yield a final value of performance; the median was used instead of the mean because of possible ceiling effects occurring and skewing the data.

The ROC performance of all four procedures was examined under two parameters: artifact rejection threshold level and a minimum number of accepted trials required to start the SNR estimation procedure. The artifact rejection parameter had two values: $10\ \mu V$ and $\infty\mu V$ (the second value being equivalent to not using artifact rejection and accepting all trials). The minimum number of accepted trials was varied from 256 to 3840 trials in step of 256 trials. Controlling the minimum number of accepted trials prior to any SNR estimation in a procedure allows for minimization of any

initial transients affects, as observed by (Gentiletti-Faenze, Yanez-Suarez et al. 2003). The final SNR estimate value was taken as the maximum value between on the SNR curve between the minimum number of accepted trials and the final accepted trial.

4.4 Estimation of Loudness Growth in Normal Listeners

The purpose of this experiment was to compare estimation of individual loudness growth in eight normal listeners using two psycho-acoustical procedures (CMM and ME) and two procedures based on evoked responses (TBOAEs and TBABRs), with particular emphasis on the TBABR procedures developed on Section 3.4. The qualities of the estimation procedures were quantified in terms of the mean, median, and standard deviation of the MSE with respect to one of the psychoacoustical procedures for each individual loudness growth function (the other psychoacoustical procedure was used a to obtain a reference MSE).

4.4.1 Listeners

Eight listeners with normal hearing (four females, four males), ages 19 to 31 participated in the loudness measurements (some of these listeners were the same that participated on the study mentioned on Section 4.2). No listener had a history of hearing difficulties, and their audiometric thresholds did not exceed 15 dB HL at octave frequencies from 250 Hz to 8 kHz (ANSI 1996). Additionally, all listeners had their middle-ear function evaluated via a clinical exam. All listeners were arbitrarily run on their right ears.

4.4.2. Stimuli, Apparatus, CMM, and TBOAE Recordings

The stimuli, apparatus, CMM, and TBOAE recordings were the same as those described on Section 4.2 through Section 4.2.6. Threshold values estimated from the CMM data were calculated by fitting a polynomial to the percent of non-zero (empty strings) entries as a function of level and interpolating the SPL at which the polynomial value was 0.5. Two other loudness growth estimates were obtained from the CMM procedure by using polynomial and inex fitting as described on Section *3.4.3.3* (with all residual noise levels set to 1). These two smoothed estimates were done in order to provide additional reference for the estimation procedures that utilize the same noise control methods.

Estimates for the slope of the loudness growth function on the raw CMM data were obtained in the following manner: the CMM data points, loudness as a function of SPL, were fitted to a linear least-square best fit. The slope of the fitted line was multiplied by a factor of 10 in order to yield slope values in terms of stimulus intensity rather than SPL.

4.4.3. Loudness Growth Estimation through ME

A second psycho acoustical estimate of the loudness-growth function was obtained by using the magnitude-estimation procedure. Listeners were presented with a series of tones bursts and asked to enter a numerical value that corresponded to their perceived loudness of the stimulus. The listeners were told to use an open set on their numerical answer (with the exception of no negative numbers). The listeners were also told to enter 0 only if no sound was heard and encouraged to use decimals. The stimuli consisted of tone burst from 10 to 100 dB SPL varying in 5 dB steps (19 stimuli) presented in a random order. For each

of the two frequencies tested (1 and 4 kHz) there was a practice test in which each level was presented only once in order to get the listener acquainted with the overall stimuli range. The experimental test occurred right after the training session, with the each of the 19 stimuli being presented eight times and in random order. If a listener entered 0 more than four times for a given level, that level was not used (treated as sub-threshold) during data analysis. The final estimate for a specific level was calculated from the geometric mean of the non-zero numbers. Threshold values estimated from the ME data were calculated by fitting a polynomial to the percent of non-zero (empty strings) entries as a function of level and interpolating the SPL at which the polynomial value was 0.5. Two other loudness growth estimates were obtained from the ME procedure by using polynomial and inex fitting as described on Section *3.4.3.3* (with all residual noise levels set to 1). These two smoothed estimates were done in order to provide additional reference for the estimation procedures that utilize the same noise control methods. Loudness growth function linear slope estimates were obtained in the same manner as that for the CMM data.

4.4.4. TBABR Recordings

The TBABR was recorded simultaneously with the TBOAEs in a sound-attenuating electrically shielded booth. Stimuli were presented in blocks of 1000 trials (about 41.7 ms per trial, at a presentation rate of about 24 Hz). For each frequency the stimulus was presented in ascending order from the listener's threshold to 100 dB SPL on steps of 5 dB. Threshold was determined from the maximum threshold of the CMM or ME procedure. Listeners had three electrodes affixed: the non-inverting electrode was

positioned on the forehead, the inverting electrode was position on the ipsilateral mastoid (behind the ear), and the ground electrode positioned on the contralateral mastoid (listeners were scrubbed with alcohol at the locations prior to the electrode placement). The electrode signal was then sent to a GRASS QP511 Quad AC Amplifier, were it was band-pass filtered from 30-3,000 Hz, amplified by a factor of 50,000, and sent to a 32-bit Lynx Two Soundcard (outside the booth) where it was sampled at 48 kHz. Two different artifact rejection thresholds were applied (one set to 20 uV and one set to 50 uV), these two different sets of averages were stored for later analysis. Throughout the experiment, a computer outside the booth displayed for the current level, two weighted sub-averages of the evoked response, the estimated SNR as function of trial, the estimated noise variance as a function of trial, the estimated acoustical noise as a function of trial, and the estimated power of weighted average as function of blocks of trials. This information was used to monitor the recording.

4.4.5. Estimation of Loudness from TBOAEs and TBABRs

The procedure used to estimate the loudness growth from TBOAEs was exactly the same as that described on Section 4.2.7, with the following parameters held fixed: window delay = 10 ms, window size = 20 ms, window type = hanning, F-ratio = 2. Two other loudness growth estimates were obtained from the TBOAE procedure by using polynomial and inex fitting as described on Section *3.4.3.3* (with all residual noise levels set to 1). These two smoothed estimates were done in order to provide additional reference for the evoked potential estimation procedures that utilize the same noise

control methods. The procedures used to estimate loudness from TBABRs were those described on Section 3.4(summarized on Tables II-IV).

4.5 Estimation of Individual Loudness Growth in Hearing Impaired

The purpose of this experiment was to verify if the methods used for the estimation of loudness growth on normal listeners could be equally applied to HILs. The experimental design was very similar to that of 4.3 with some minor but relevant changes.

4.5.1 Listeners

Eight listeners with a known history of hearing impairment confirmed through clinical test (four females, four males), ages 40 to 85, participated in the loudness measurements. Listeners were tested on the ear that had the best (lowest) thresholds determined through a 2-Interval-2-Alternative- Forced-Choice (2I2-AFC) procedure (Levitt 1971) in which the same stimuli as that of Section 4.2.2 were used. Three listeners were tested on the left ear and five listeners were tested on the right ear). Three listeners could not be tested at 4 kHz because their thresholds were higher than 85 dB SPL for the particular stimulus used.

4.5.2. Stimuli, Apparatus, CMM, ME

The stimuli, apparatus, CMM, and ME details were the same as those described in Section 4.4. With the exception that this time the CMM condition was also run on the

same apparatus of both the ME and evoked response conditions (running on a OS Ubuntu PC).

4.5.4. TBABR and TBOAE Recordings

The TBABR and TBOAE recordings were made in the exact same manner as that described previously on Section 4.4. The only modification was that the recordings were done over three different sessions for each of the frequencies tested (1 and 4 kHz). During each session and for each level the stimulus was presented in four blocks of 1,000 trials each. This yielded a total of 12,000 trials for the whole experiment (3 sessions x 4 blocks x 1,000 trials) per level. The experiment was modified in this way in order to limit the experiment time to no more than two hours per day. Two listeners that provided noisy TBABR recordings were able to come back for further testing and recorded three new additional sessions.

4.5.5. Estimation of Loudness from TBOAEs and TBABRs

The procedure used to estimate a loudness growth from the TBOAEs was exactly the same as that described in Section 4.4.5.

V. Results and Discussion

5.1 Parameter Selection for Loudness Growth Estimation Through TBOAEs

5.1.1 Cross-modality Measurement Correction Factor

The cross-modality matching measurements for 1 and 4 kHz were independently scaled by a fixed correction factor to ensure that the resulting functions were similar to the expected form of the loudness function (i.e., power function exponent equal to 0.3). The same scaling factor, 1.67, was used for all subjects based on the average data. The 1- and 4-kHz scale-factors were calculated separately, but agreed exactly to three decimal places, so the same value was used for both. This is consistent with the expectation that loudness functions for tones at different frequencies are parallel on a log scale within a relatively wide range of frequencies (Hellman 1976).

5.1.2. Loudness Estimation from TBOAEs at 1 kHz

Figure 16 shows the AMSE between loudness measured using CMM (after scaling) and loudness estimated from TBOAEs in response to 1-kHz tone bursts using a rectangular analysis window. The measures are compared directly on an arbitrary normalized logarithmic scale with both functions vertically shifted to zero mean. An increase of one unit on this scale corresponds to the scaled string length multiplied by ten and a change of 10 dB in TBOAE level. The plots show AMSE as a function of window delay, with each curve representing a different window size and each panel representing a

different F-Ratio. Figure 17 is identical to Figure 16 except that the TBOAE loudness estimation was instead performed using a Hanning analysis window. All of the loudness estimations shown in both figures exhibit a similar pattern. Regardless of other parameters, when window delay is very short, the window captures stimulus artifact or the linear reflections within the auditory system. As the delay gets longer, the AMSE improves indicating that the effects of the stimulus artifact diminish and the non-linear cochlear response becomes more pronounced. At very long delays, the window fails to capture the TBOAE at a level strong enough to overcome the background noise.

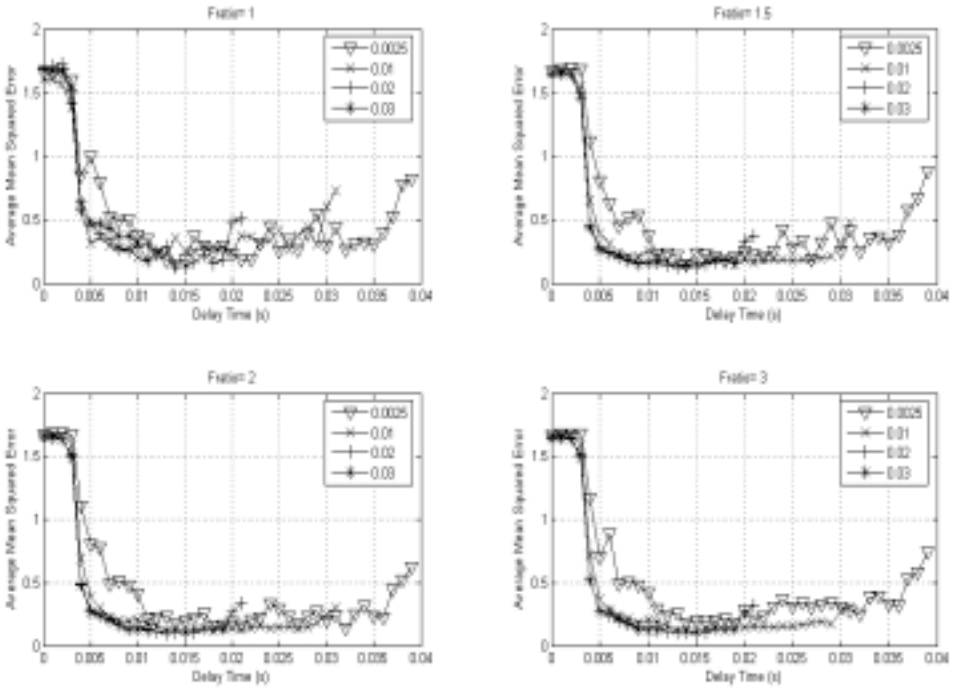

Figure 16- Averaged mean squared error (AMSE) between loudness measured using CMM and loudness estimated from TBOAEs in response to 1-kHz tone bursts using a rectangular analysis window. The plots show AMSE as a function of window delay, with each curve representing different window sizes and each plot representing a different F-Ratio.

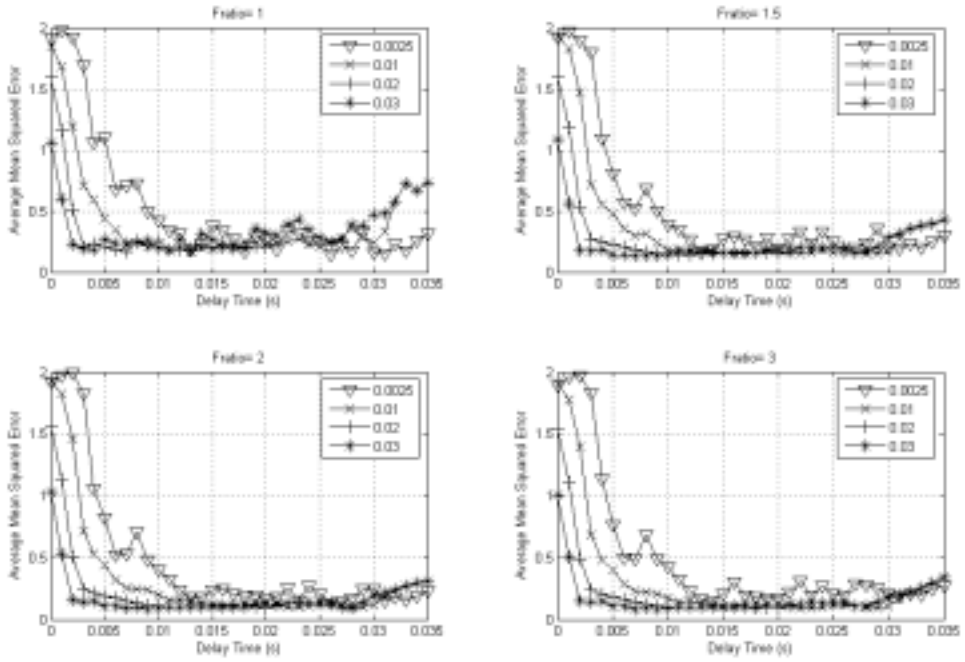

Figure 17- Averaged mean squared error (AMSE) between loudness measured using CMM and loudness estimated from TBOAEs in response to 1-kHz tone bursts using a Hanning analysis window. The plots show AMSE as a function of window delay, with each curve representing different window sizes and each plot representing a different F-Ratio.

The Hanning window (Fig. 17) appears more robust than the rectangular window (Fig. 16) and results in near-optimal AMSE for a larger range of delays. In addition, the Hanning window reaches optimality sooner because it deemphasizes the early portion of the signal more. It is thus important to note that the Hanning window has a total area equal to half of the rectangular window. As such, the Hanning window is more localized in time, resulting in an increased side-lobe attenuation in the spectral domain, which can improve spectral estimation by reducing any spectral leakage (Harris 1978). The AMSE curve is approximately flat for delays of 13 to 20 ms, similar to the findings of Epstein *et al.* (2004) in which no significant effect of window delay was found in that range. It is

noteworthy that the apparatus and the duration of the stimulus differ somewhat between the two experiments and the results are still in agreement.

Each of the curves in Figs. 16 and 17 are for different window sizes. The shortest window size (0.0025 ms) tends to yield high variance estimates as a function of delay. This short window size is very sensitive to local variations in background noise because of the small number of points being collected. In addition, because the window is so short, it has a smaller range of near-optimal delays. At short delays in particular, this window will place a lot of emphasis on the stimulus itself resulting in a poor loudness estimate. The loudness estimate becomes more robust as the window size increases (i.e., a classical trade-off between recording time and susceptibility to noise in measurements). The 20-ms and 30-ms windows perform approximately equally well. As a result, a 20-ms window was arbitrarily chosen for subsequent loudness estimation.

The F-Ratios of 2 and 3 result in lower optimal AMSEs (AMSE = 0.10 for both) than F-Ratios of 1 or 1.5 (AMSE = 0.15, 0.13 respectively). As the AMSEs for F-Ratios of 2 and 3 are approximately equal, an F-Ratio of 2 will be used for subsequent loudness estimation. The combination of parameters that provides the lowest AMSE is: Hanning windows with window delays of 10 ms, window sizes of 20 ms, and an F-Ratio of 2. This exact combination of parameters was chosen for the loudness growth estimation from TBOAEs for the experiments described on Section 3.4 and 3.5

5.1.3. Loudness Estimation from TBOAEs at 4 kHz

It is known that it is difficult to measure TBOAEs at 4 kHz due to ear-canal resonance and signal artifacts (Ravazzani and Grandori 1993; Whitehead, Stagner et al. 1994). Identical sets of analyses were performed for the responses to 4-kHz tone bursts to determine whether the current procedure could be used to estimate loudness of 4-kHz tone bursts using any possible combination of parameters. Although neither window produced particularly good results, both the Hanning and rectangular windows yielded roughly the same results, with the Hanning window providing a slightly more robust response. Therefore, only the Hanning-window results are shown in Figure 18.

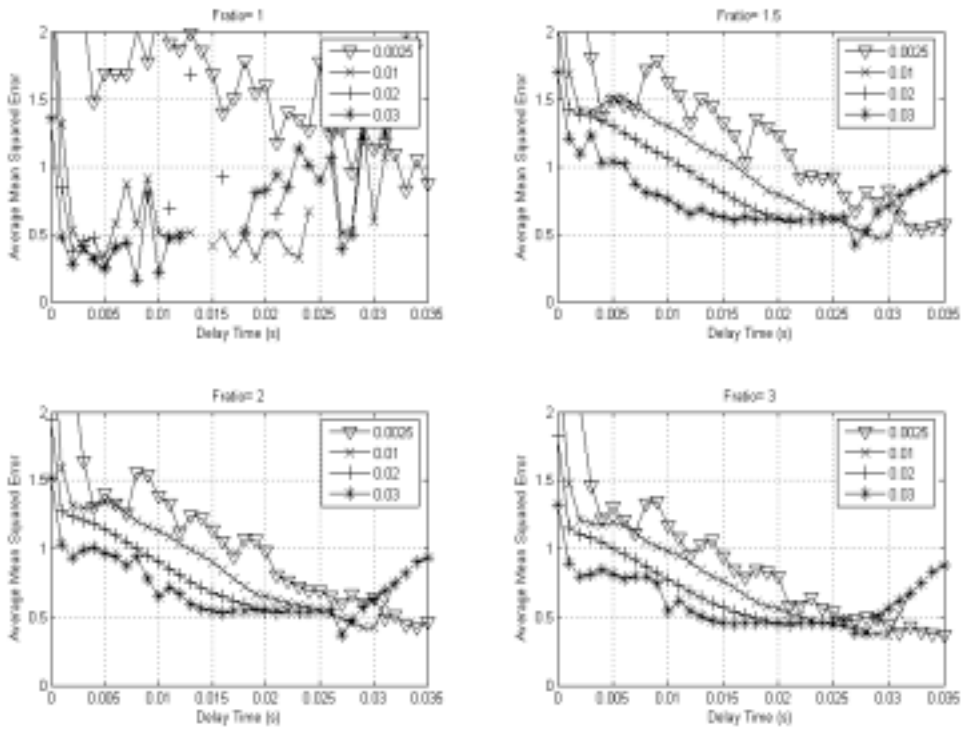

Figure 18- Averaged mean squared error (AMSE) between loudness measured using CMM and loudness estimated from TBOAEs in response to 4-kHz tone bursts using a Hanning analysis

window. The plots show AMSE as a function of window delay, with each curve representing different window sizes and each plot representing a different F-Ratio.

None of the parametric combinations resulted in AMSE approaching the values measured at 1 kHz. As the window delay increases and less of the stimulus resonance is captured, the error decreases. However, it seems like the linear response of the ear-canal resonance overpowers any possible non-linear cochlear response. The results support reports of difficulty making measurements at 4 kHz and support the claim that most of the 1-kHz responses are true non-linear responses and not simple linear reflections. The small number of missing data points resulted from noisy data that did not contain any positive real components in the cross-spectrum.

5.1.4. Coupler Measurements

The goal of all of these analyses was not only to determine the parameters that minimize the error in the loudness estimation, but also to identify a possible spurious relationship with stimulus artifacts. To this end, Fig. 19 shows the MSE comparing the results from the human listeners' average CMM and the loudness–estimation procedure in response to 1-kHz tone bursts compared to measurements in a coupler using a Hanning window as a function of window delay. These results exhibit a relatively clear mean squared error (MSE) minimum despite the fact that the coupler generated no TBOAEs. However, this minimum still has a higher absolute estimation error than any of the minima measured using 1-kHz tones in real ears. Additionally, the lowest values of the MSE result from when the recording is essentially zero. As the CMM function has been shifted to have a zero mean, the MSE is then essentially just a sum of the squares of the

logarithmic values of string length after the shift. This pattern is in great contrast with the results measured in the real ears, which show a relatively wide range of stable delays and lower MSE.

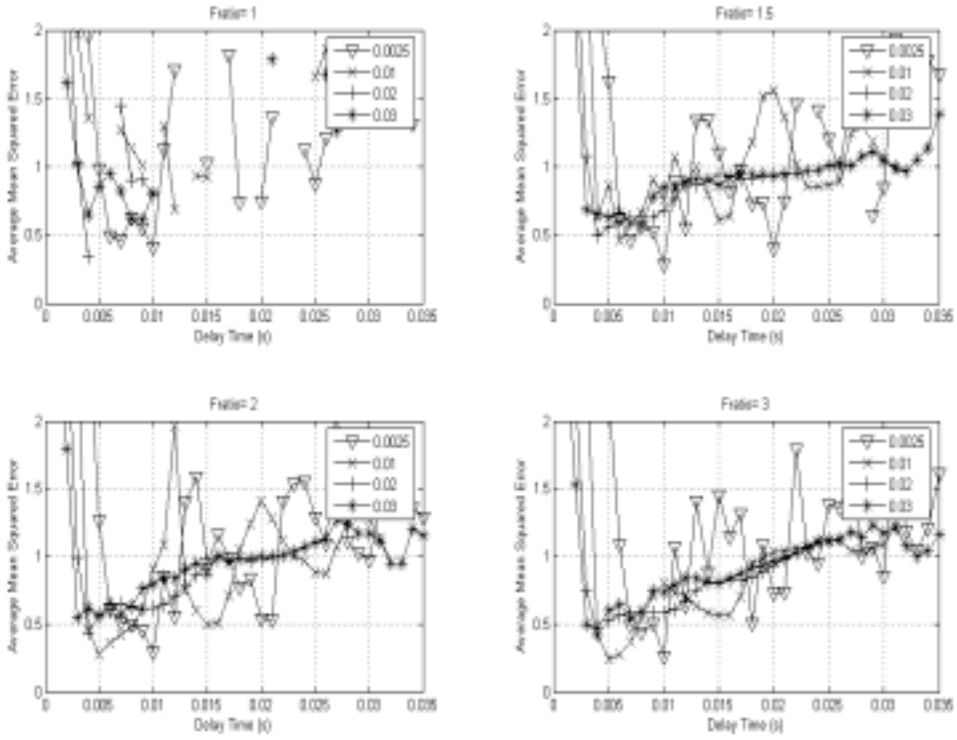

Figure 19- Mean squared error (MSE) between loudness measured using CMM and loudness estimated from the response of the coupler to 1-kHz tone bursts using a Hanning analysis window. The plots show AMSE as a function of window delay, with each curve representing different window sizes and each plot representing a different F-Ratio.

Identical measurements at 4 kHz in the coupler, shown in Figure 20, exhibit a long, but asymptotic, decrease in MSE as the ringing diminishes (clearly, a much longer ringing decay time than for the 1-kHz stimulus). The optimal case results in an MSE much higher than the 1 kHz real-ear response (Figure 16), but results that are not much worse than the 4 kHz real-ear responses. Again, this supports the idea that the 4-kHz

measurements fail to capture the non-linear response, but that the 1-kHz non-linear response cannot be explained by simple resonance or linear reflections from the ear canal.

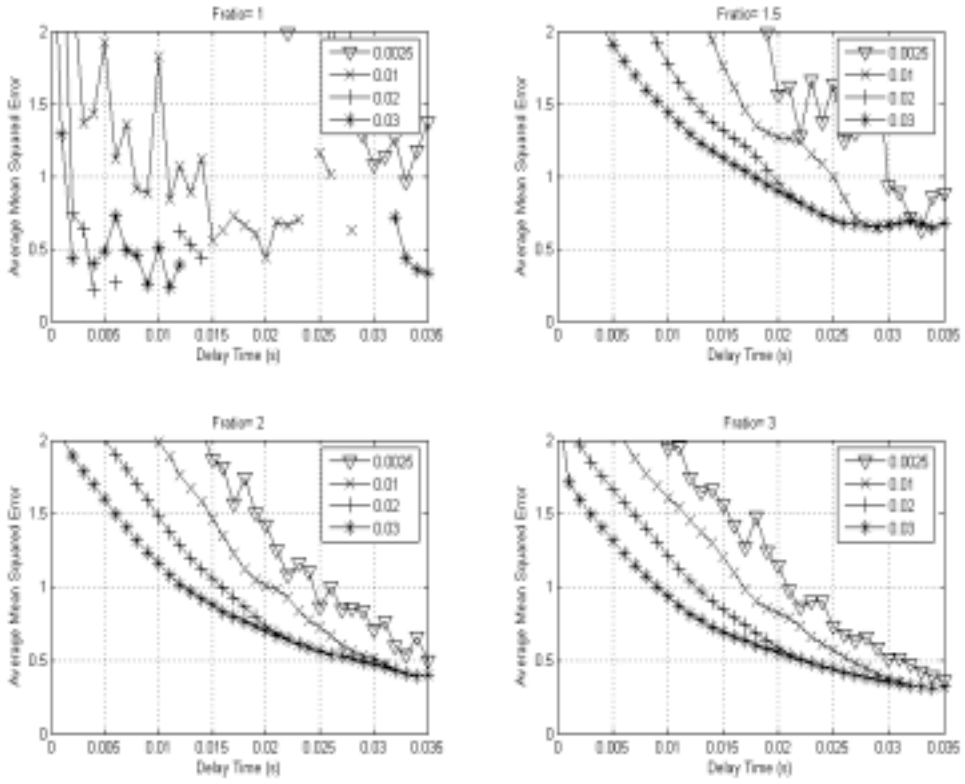

Figure 20- Mean squared error (MSE) between loudness measured using CMM and loudness estimated from the response of the coupler to 4-kHz tone bursts using a Hanning analysis window. The plots show AMSE as a function of window delay, with each curve representing different window sizes and each plot representing a different F-Ratio.

5.1.5. Comparison between TBOAE and CMM Loudness Growth Estimates

In Figs. 21, 22, and 23, functions were allowed to vary by a single parameter in order to adjust the location of curve on the plot to have to zero-mean. Figure 21 shows estimates of loudness in response to 1-kHz tone bursts as a function of level derived from

TBOAEs and CMM. Several loudness estimates are plotted, each for a different window delay to demonstrate the robustness of the estimation procedure and to show the contrast of the zero-delay stimulus response. There is good agreement between the TBOAE loudness estimates and the direct measure of loudness, except at the lowest levels. This is consistent with loss of steepness at low levels resulting from edge effects in the CMM procedure (Hellman and Meiselman 1988). It is noteworthy that the analysis results in a nearly linear response when the window delay is set to zero and the stimulus itself is included in the analysis.

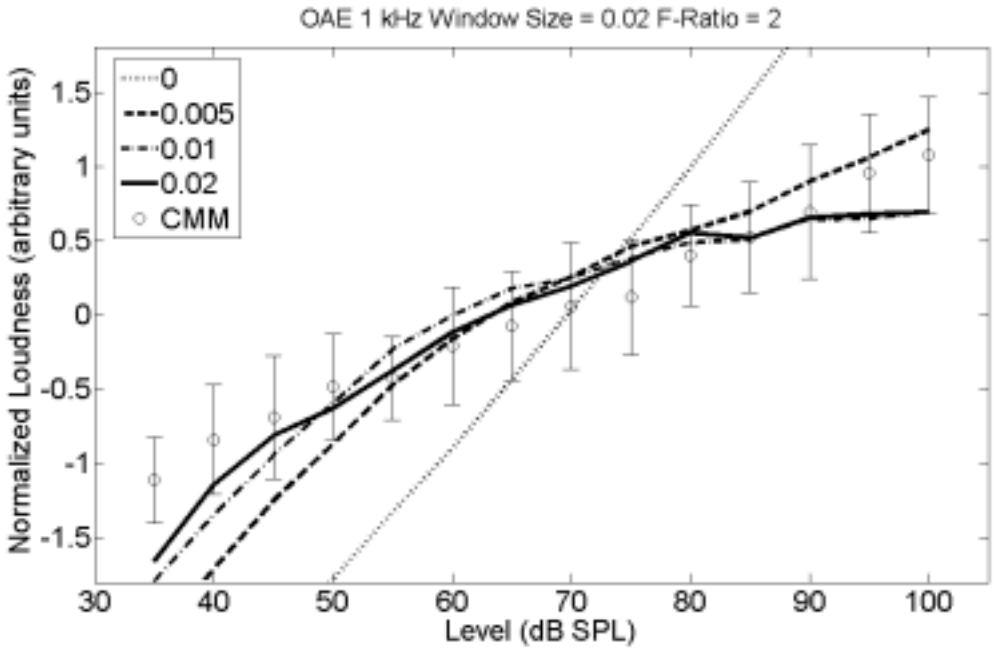

Figure 21- Loudness resulting from 1-kHz tone bursts as a function of level for estimates derived from TBOAEs and for direct loudness measurements using CMM. Several loudness estimates for the TBOAEs are plotted, each for a different window delay. This analysis was performed with a 20-ms Hanning window and an F-Ratio of 2. Error bars show plus and minus 1 standard deviation.

Figure 22 is identical to Figure 21 for the 4-kHz tone bursts. The estimates derived from TBOAEs at 4 kHz are too steep and serve as very poor fits to the CMM

function. In fact, they are more likely to match the linear response obtained when the window delay is set to zero. This is probably due to the difficulty in avoiding ear canal resonances. The resulting OAEs contained too great a linear component to remove by simple windowing and averaging. As these results match those seen in the coupler, they are unlikely to arise from actual cochlear processing. In contrast, the OAEs in response to the 1-kHz tone bursts differed dramatically from the coupler recordings, indicating that these measurements are not results of simple stimulus artifacts.

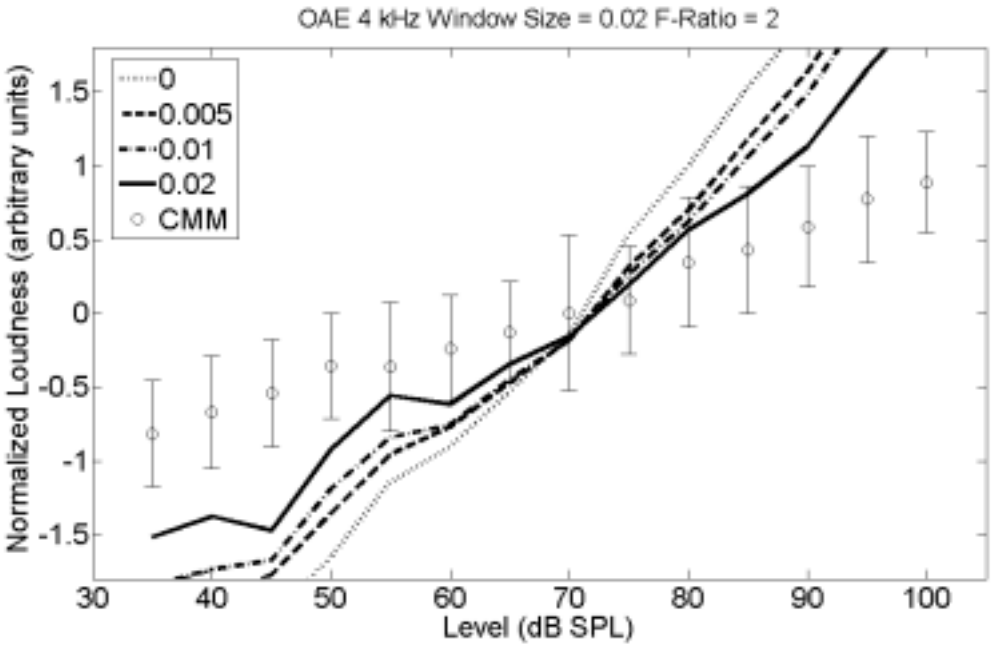

Figure 22- Loudness resulting from 4-kHz tone bursts as a function of level for estimates derived from TBOAEs and for direct loudness measurements using CMM. Several loudness estimates for the TBOAEs are plotted, each for a different window delay. This analysis was performed with a 20-ms Hanning window and an F-Ratio of 2. Error bars show plus and minus 1 standard deviation.

Figure 23 shows a summary of all of the average functions in optimal parametric conditions. It is particularly noteworthy that the CMM functions at the two frequencies correspond quite closely, indicating that the psychacoustical loudness functions are consistent across frequency for normal listeners. The INEX and the loudness function

derived from TBOAEs at 1 kHz show excellent correspondence across the entire range. This indicates that TBOAE functions may be closely related with loudness measured using longer tones for an average of a group of listeners.

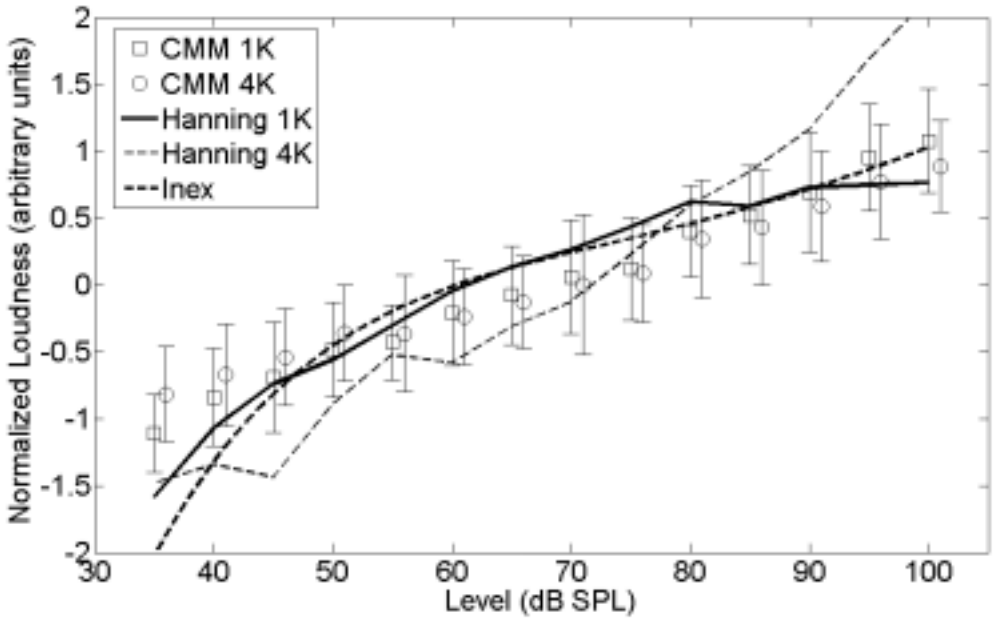

Figure 23- Loudness as a function of level derived from CMM and Hanning-windowed TBOAEs in response to 1- and 4-kHz stimuli (The 4-kHz CMM data points are offset by 1 dB so that the error bars are visible). In addition, the INEX model loudness function is plotted for comparison with loudness measurements in the literature. Error bars show plus and minus 1 standard deviation.

5.1.6. Latency/Level Relationship

Figure 24 shows the estimate of latency as a function of level for both the real ears and coupler at 1 and 4 kHz. For OAE measurements in real ears at 1 kHz, the results show a relatively smooth decrease in latency as a function of level, ranging from 25 ms at 35 dB SPL to 6.3 ms at 100 dB SPL. Although this result differs somewhat from other reports of TBOAE latency (Brass and Kemp 1991; Hoth and Weber 2001; Schairer,

Ellison et al. 2006; Sisto and Moleti 2007), it is not surprising, since the methodologies for estimating OAE latency, the window sizes and shapes, and the definition of latency differ substantially among studies.

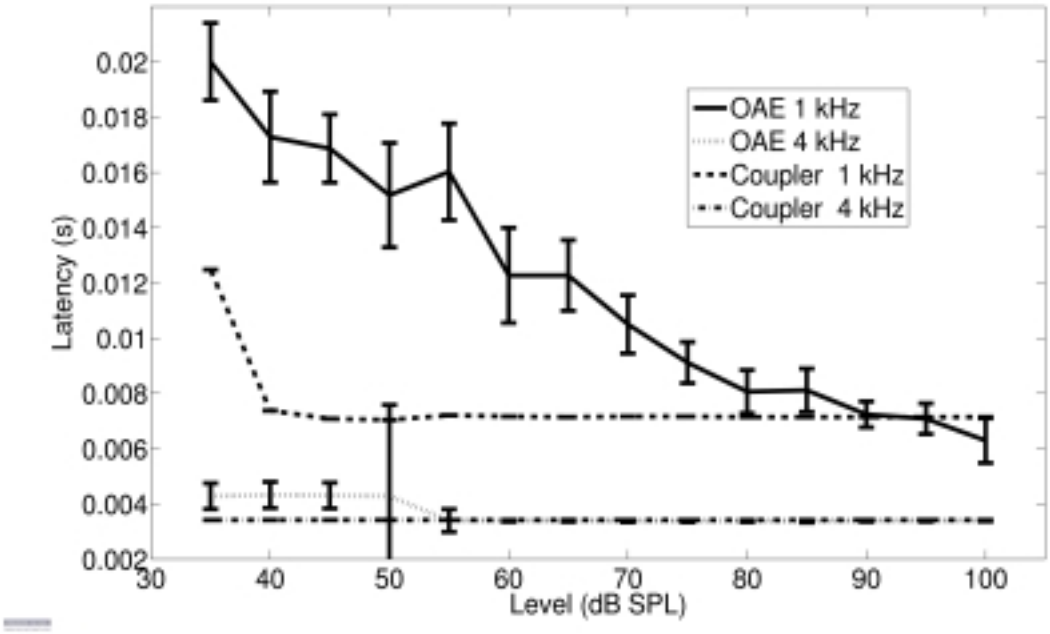

Figure 24- TBOAE latency estimates as a function of level for both the real ears and coupler in response to 1- and 4-kHz stimuli. Error bars show plus and minus 1 standard deviation.

The present real-ear results are in agreement with the fact that the latency measured in the coupler at 1 kHz remains constant as a function of level. This indicates that the coupler recordings show a peak response at the biggest peak of the trivially reflected waveform, observed here to be close to 7 ms. In contrast, the acoustic ringing at 4 kHz was significant enough to render the procedure invariant to the initial delay chosen (i.e., the maximum would always occur at the beginning of the selected time region in both the real ears and the coupler). In this case, the latency was estimated at approximately 3.4 ms for both the real ears and the coupler. Even when the start of the

observation window was delayed by several more ms, the real ears and coupler produced identical, flat latency curves with latencies adjusted to match the start time of the window.

5.1.7. Quantify Accuracy of Estimation

Additional numerical analyses were performed by first assuming a "correct" loudness function and then determining the error (Model – Data) between the model and each of the other functions. Two analyses were performed, in one the "correct" model was the INEX function, in the other the "correct" model was the CMM data at 1 kHz. The variability accounted for by the model, $\sigma^2_{accounted}$, is then quantified by:

$$\sigma^2_{accounted} = 1 - \frac{\mathrm{var}(L_{model} - L_{Data})}{\mathrm{var}(L_{model} - L_{Data}) + \mathrm{var}(L_{model})} \qquad (47)$$

Where L_{model} is the loudness point given by the model, L_{Data} is the loudness pointes estimates through the data, and var( ) is the variance operator. Equation (47) is more useful than a simple correlation analysis because it does not heavily weight trends that are simply in the same linear direction, but rather looks at whether the functions deviate from each other. Table V shows the variability accounted for by each data set given a specific model. It is noteworthy that the INEX resulted from a relatively large quantity of data, but none of the data or listeners in the present study.

Table V- The performance of the estimation procedure was quantified in terms of equation (47) .

Model		**CMM**	**CMM**	**INEX**	**OAE Hann**	**OAE Hann.**	**OAE Rect.**	**OAE Rect.**

		1k	4k		1k	4k	1k	4k
CMM 1k	**Variance Accounted**	--	0.96	0.82	0.89	0.63	0.80	0.45
INEX	**Variance Accounted**	0.88	0.76	--	0.96	0.72	0.95	0.60

When the CMM 1k function was used as the ideal function, the CMM 4k function was found to correspond closely. This indicates that the average loudness functions at these different frequencies are consistent. Other measurements of loudness have shown similar results (e.g., Hellman 1976). Additionally, the Hanning-windowed 1 kHz OAE input/output function also closely matched the CMM 1k function. This supports the idea that TBOAE measurements at 1 kHz are related to loudness. This idea is further supported by the second comparison, performed with the INEX model as the "ideal" function. If the INEX is assumed to be a good average loudness function, CMM is then an acceptable, but not ideal measure of loudness based on the variance accounted for by the CMM functions when using the INEX as the model. In fact, the 1-kHz TBOAE loudness models matched the INEX function better than CMM, despite the fact that the analysis parameters were optimized to fit the CMM function. This indicates that TBOAEs may actually be a better measure of loudness than CMM. As expected, the 4-kHz TBOAE functions did not serve as very good models of either ideal function.

In order to examine potential viability for use as an individual metric for the assessment of loudness growth, the same analysis was performed comparing individual CMM at 1 kHz and loudness estimated using Hanning-window analysis on 1-kHz TBOAEs. The group average variance accounted for in this condition was 0.89 in the earlier analysis. The mean of the individual accounted variances was 0.80 with a

standard deviation of 0.11. Of the six listeners, two estimates matched relatively poorly with variances accounted of 0.62 and 0.68. In both cases, the individuals for whom the variance accounted was lower had particularly shallow CMM slopes indicating that much of the problem may have arisen as a result of unreliable CMM data rather than unreliable OAE data.

The remaining listeners had variances accounted of 0.83, 0.86, 0.89, and 0.90. This indicates that, at least for two-thirds of the listeners, the two methods resulted in roughly the same estimates of loudness.

5.2 Estimation of SNR for TBABRs under Non-Stationary Noise

5.2.1 Experiment I: MSE of Residual Noise Estimation

An example of the simulations based on real data is shown in Fig. 25. The top graph shows the average residual noise level as a function number of trials from the Monte-Carlo simulation and the results from the *Fsp* and the *NS Fmp*. The bottom graph shows the estimated noise variance from each trial that was used to generate the Monte-Carlo simulations. Figure 26 shows the MSE for the residual noise estimation in the *Fsp* and *NS Fmp* as a function of the smoothing segmentation parameter *p*. The MSE of the *Fsp* procedure, which assumes a single noise source, is constant as a function of this parameter. The *Fsp* yielded a higher MSE than the *NS Fmp*.

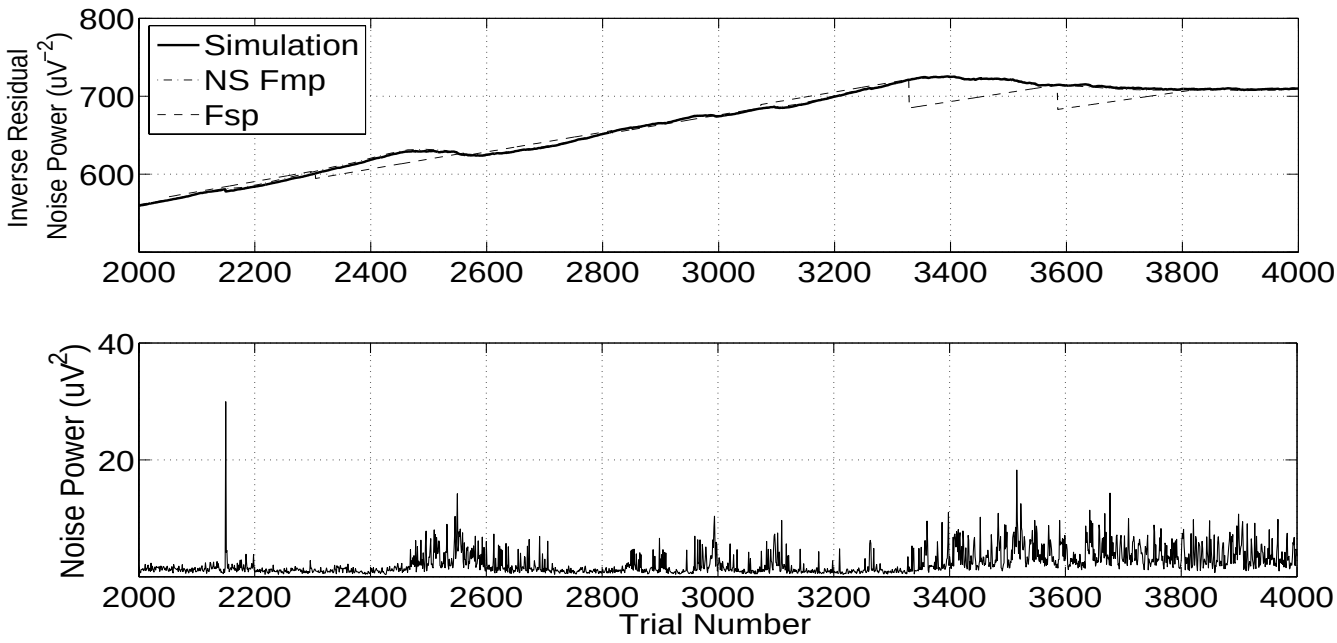

Figure 25- Example of residual noise estimation. The top graph shows the estimated inverse residual noise as a function of number of trials added to the average. There is very close agreement between the residual noise estimation in the *NS Fmp* procedure and the average results from the Monte Carlo simulation, making their distinction difficult. The bottom graph shows the background noise power present in each of the individual trials used for the simulation. For this particular listener, the background power is clearly non- stationary (with changes in noise power of more than one order in magnitude).

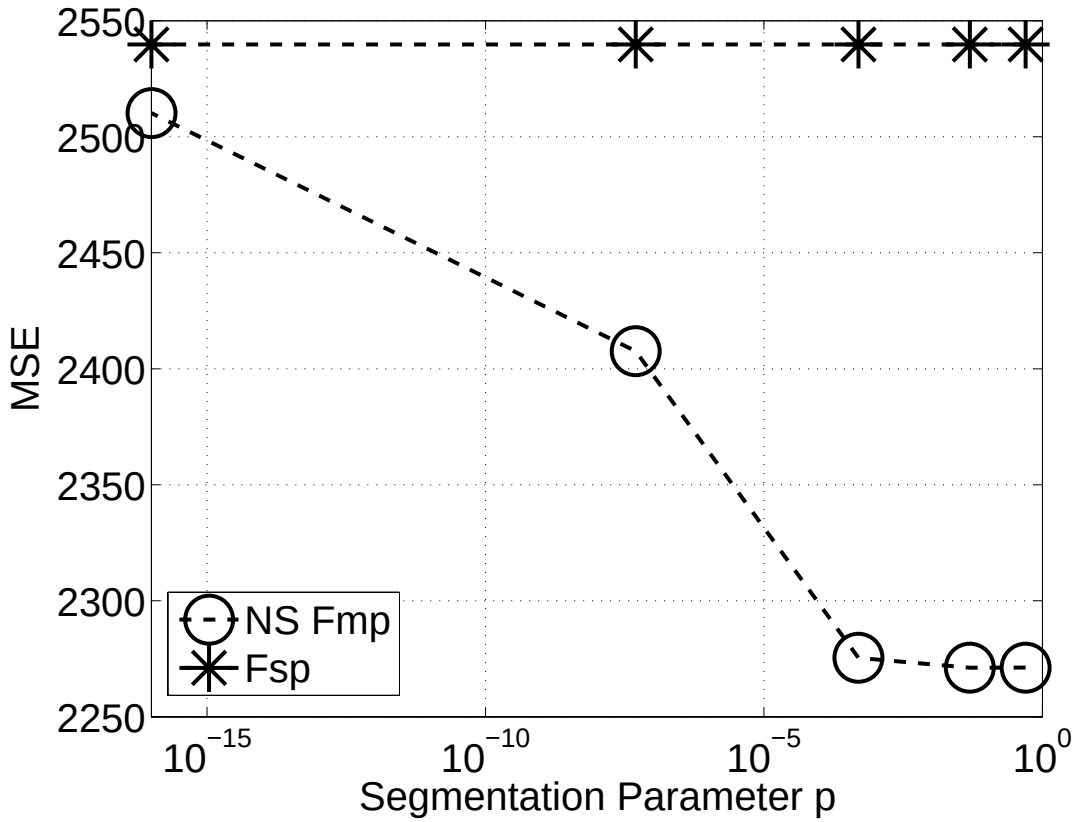

Figure 26- Mean square error between the estimated residual noise and the Monte Carlo simulations as a function of the segmentation parameter p. The MSE of the *Fsp* procedure is plotted for reference. The MSE of the *NS Fmp* procedure is a monotonic function of p. As p approaches 0 the number of noise sources estimated decrease, with *NS Fmp* algorithm being equivalent to the *Fsp* at p=0.

5.2.2 Experiment II: ROC Analysis

The results of the ROC analysis are shown in Figs. 27 and Table VI. Fig. 27 shows the performance of all procedures as a function of minimum number of accepted trials and artifact rejection threshold. In general, the weighted averaging procedures (*WFsp* and *WNS Fmp*) were superior to the normal averaging procedures for both artifact rejection levels. In similar fashion, the *NS Fmp* was superior to the *Fsp* on both artifact rejection levels as well. The performance of all procedures increases as the artifact rejection level was lowered (at the cost of more trials being discarded) and as the

minimum number of accepted trials required for estimation increased. Although the performance for the weighted averaging procedures seem identical, a more careful look at Fig. 27 and Table VI show that the *WNS Fmp* procedure has a small but consistent advantage when the minimum number of accepted trials $\leq$ 1000 for either rejection levels.

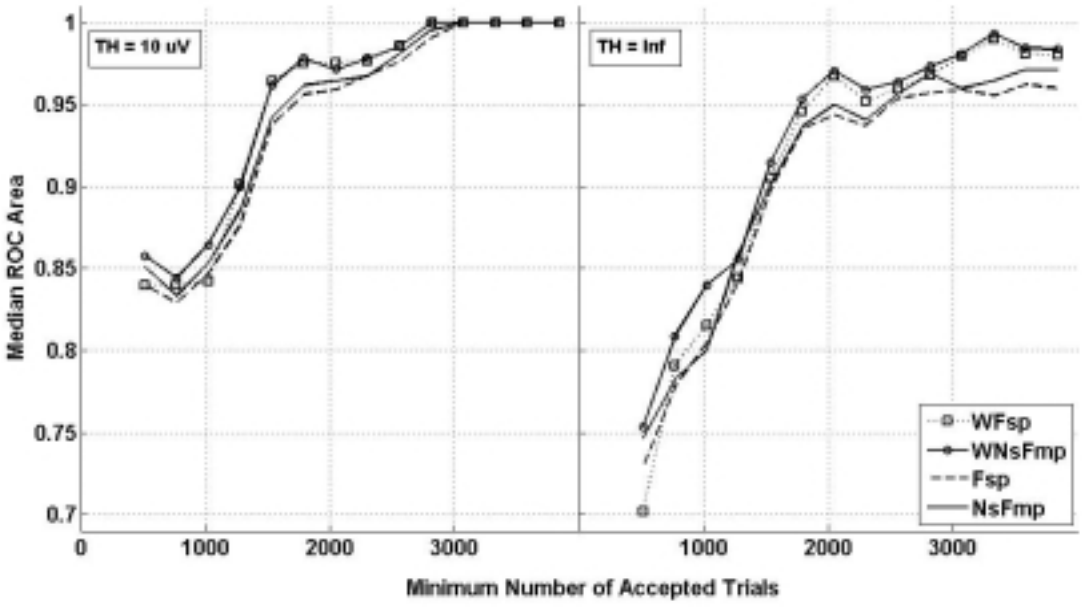

Figure 27- Median ROC area for all four procedures as a function of a minimum number of accepted trials before any estimation begins, and artifact rejection level. The minimum number of accepted trials attempts to control for initial transient effects in the estimation of the SNR curve.

Table VI- Median ROC Area for the four different SNR procedures and artifact rejection.

Procedure	Th=10,M=512	Th=10,M=1024	Th=Inf,M=512	Th=Inf,M=1024
Fsp	84.0%	84.6%	72.8%	80.4%
NS Fmp	85.1%	84.2%	74.6%	80.0%
WFsp	84.0%	85.3%	70.2%	81.5%
WNS Fmp	85.7%	86.4%	75.3%	84.0%

5.2.3 Discussion

The results from Experiment I (Section 5.2.1) show that changes in the ABR background noise power can be of more than an order in magnitude (Fig. 25 bottom graph). This corroborates previous literature on the non-stationary behavior of the ABR background noise and the necessity of accounting for residual noise levels while analyzing the averaged data. Computer simulations of averaging under non-stationary noise sources show that the mean-square-error of the SNR curve obtained under a standard procedure, the *Fsp*, is higher than the *NS Fmp* (which can account for differences in noise power across trials). The *NS Fmp*'s smoothing parameter, *p*, can be adjusted in such a way that the *NS Fmp* performance will asymptote to that of the *Fsp*.

In the second experiment (Section 5.2.2), the performance of all procedures was examined as function of artifact rejection threshold and a minimum number of accepted trials required for estimation. On average, the *WNS Fmp* procedure had a small but consistent advantage, yielding overall best performance. The *WNS Fmp* advantage was greatest at small values for the minimum required number of accepted trials ($\leq$ 1000), which is precisely where transients are more likely to have an effect. In general, as the number of accepted trials increase, the estimation of the signal power component of the SNR (equation 22 or 30b) converges, thus becoming robust to background noise perturbations and improving the overall SNR estimation. Therefore as the minimum number of trials required for estimation increased, all procedures showed a general increase in performance regardless of the artifact rejection level. When the artifact rejection level is taken into account, all procedures showed improvement at lower artifact

rejection level (with all performances converging at high numbers of minimum accepted trials). The biggest differences between performances and across threshold levels occur for small values of minimum required number of accepted trials, although differences at both extremes are clearly visible. Table VI shows a value of 512 for the number of minimum accepted trials the average performance across procedures was 84.7% for a threshold of $10\,\mu V$, this average performance drops to 73.2% when no threshold is applied ($\infty\mu V$). As the number of trials increase to 1024, the estimation becomes more robust to perturbations and the drop is less substantial (from 85.1% at $10\ \mu V$ to 81.4% at $\infty\mu V$).

5.3 Estimation of Loudness Growth in Normal Listeners

5.3.1. Psychoacoustical Results

Figures 28 and 29 show the individual results for the psychoacoustical loudness growth data obtained for the eight normal listeners with 1 and 4 kHz tone-bursts (all functions were arbitrarily shifted to have zero mean). Most of the listeners yielded consistent estimates in the sense that their CMM data was in reasonable agreement with the ME data (Tables VII and VIII). Some listeners, however, showed a clear discrepancy between their CMM and ME data: N1 for 1 and 4 kHz, N3 at 1kHz, and N4 at 1 kHz. The discrepancy had a consistent pattern in that the ME was lower than CMM at near threshold levels. This could be the result of a small cognitive bias of the ME procedure at

low levels and/or edge effects on the CMM at low levels (mechanical limitations in the ability to cut a smaller and smaller strings, this bias effect can also be seen in (Hellman and Meiselman 1988)). The effects of these outliers are quantified in more detail on Tables VII and VIII).

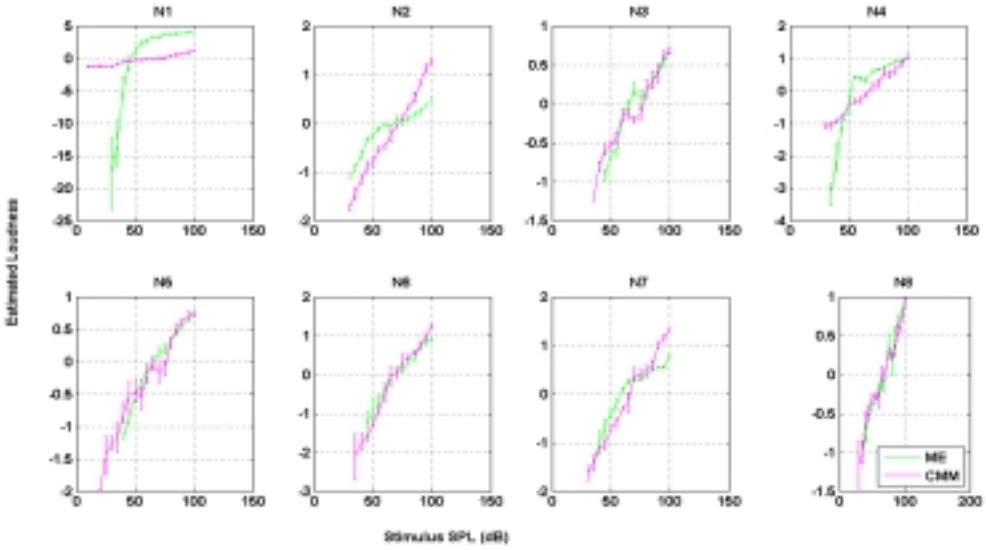

Figure 28- Individual loudness growth functions estimated through CMM and ME with 1 kHz tone-burst. The legend for listener N8 applies to all other plots. Error bars represent plus or minus one standard error. At low level the ME loudness growth was steeper than CMM for some listeners.

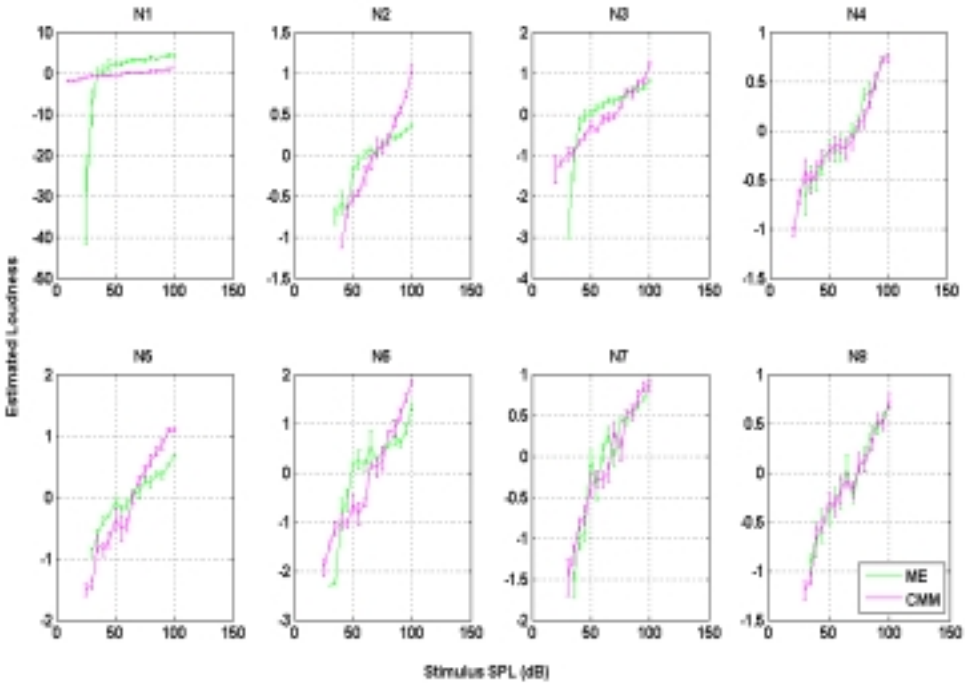

Figure 29- Individual loudness growth function estimated through CMM and ME for 4 kHz tone-burst. The legend for listener N8 applies to all other plots. Error bars represent plus or minus one standard error. At low level the ME loudness growth was steeper than CMM for some listeners.

Table VII- Statistics on the Psychoacoustical measurement of loudness growth with respect to sound pressure levels for 1 kHz tone-burst.

Lst	MSE (CMM vs ME)	CMM TH	ME TH	CMM Slope	ME Slope	CMM Stderr	ME Stderr
N1	35.84717063	10	30	0.279	2.336	0.0895	0.9109
N2	0.240234349	30	30	0.41	0.193	0.0741	0.0447
N3	0.031672125	35	45	0.258	0.269	0.0661	0.0683
N4	0.546711415	30	35	0.315	0.507	0.0853	0.1264
N5	0.039236718	20	40	0.31	0.295	0.1255	0.0562
N6	0.042633455	35	45	0.515	0.388	0.1514	0.121
N7	0.115033661	30	35	0.424	0.299	0.0962	0.0643
N8	0.012967618	30	35	0.259	0.278	0.0964	0.0915
Mean	4.609457496	27.5	36.875	0.34625	0.570625	0.0980625	0.1854125
Median	0.078833558	30	35	0.3125	0.297	0.09285	0.0799
St.dev.	12.62320551	8.451542547	5.93867	0.0931968	0.719415	0.02786687	0.29463357

Table VIII- Statistics on the Psychoacoustical measurement of loudness growth with respect to sound pressure level for 4 kHz tone-burst.

Lst	MSE (CMM vs ME)	CMM TH	ME TH	CMM Slope	ME Slope	CMM Stderr	ME Stderr
N1	66.0356	10	25	0.295	2.289	0.1029	1.3171
N2	0.1101	40	35	0.308	0.164	0.0646	0.0488
N3	0.0127	20	30	0.181	0.197	0.0748	0.0834
N4	0.3899	20	30	0.295	0.336	0.0957	0.0881
N5	0.135	25	30	0.351	0.188	0.099	0.0596
N6	0.3913	25	30	0.468	0.417	0.1255	0.1305
N7	0.0673	30	35	0.323	0.322	0.1033	0.0834
N8	0.0082	30	35	0.243	0.231	0.0794	0.1025
Mean	8.3937625	25	31.25	0.308	0.518	0.09315	0.239175
Median	0.12255	25	30	0.3015	0.2765	0.09735	0.08575
St.dev.	23.2913188	8.864053	3.5355339	0.083122973	0.7208931	0.01936942	0.43625972

The first column of the Tables VII and VIII show the MSE between the CMM and ME estimates. It is clear that for both frequencies, listener N1 was an outlier with the MSE being an order of magnitude higher than the average (it is highly likely that the subject was not doing the task appropriately), therefore the median was also used as a measure of central tendency of the data. The thresholds were slightly higher for ME than for CMM for both frequencies, about 5 dB higher on the median (although there is significant overlap from the standard deviation), this minor difference was most likely due to different apparatus setup between CMM and ME. As expected, the estimated thresholds for the 1 kHz were on average higher than for 4 kHz tone-bursts because the 1 kHz stimulus were shorter by 3-ms, thus having a smaller amount of temporal integration (e.g., (Moore 2003). The mean (median) estimated linear slope at 1 kHz was 0.34 (0.31) for CMM and 0.56 (0.29) for ME. For 4 kHz the mean (median) estimated linear slope was 0.30 (0.30) for CMM and 0.51 (0.27) for ME. Both methods yielded consistent slope estimates across frequencies but the CMM method was more reliable in the sense that the

standard deviation for the slope estimates (last row) was an order of magnitude smaller than that of ME (0.09 vs 0.71 for 1 kHz, and 0.08 vs 0.72 for 4 kHz), this observation is consistent with similar studies that use CMM and ME (e.g., (Collins and Gescheider 1989; Serpanos, O'Malley et al. 1998). The average estimated slope within the range of that reported on some literature using different stimuli: (Epstein and Florentine 2006) reported a mid-to-high-level slope of about 0.18 (using uncorrected CMM), (Hellman 1991) reported a slope of 0.3, (Serpanos, O'Malley et al. 1998) reported a value of 0.32, (Collins and Gescheider 1989) reported a value of 0.292, (McFadden 1975) observed individual values between 0.14-0.24, and (Stevens 1966) reported value of 0.32.

The average loudness growth functions across all subjects for CMM and ME at 4 kHz and 1 kHz are plotted on Figure 30. All functions were vertically shifted to be zero-mean for comparison. It is interesting to note that in general there is more consistency, or less variability, within a modality (CMM or ME) than within frequency of stimuli (1 or 4 kHz). An interesting observation is that although the stimulus at 1 kHz was shorter than the 4 kHz (1ms vs 4 ms), their loudness growths seem to eventually converge at high enough levels (>70 dB SPL) The MSE error (Table VIII) from these two standard psycho-acoustical procedures will serve as benchmark data form which the accuracy of the evoked potentials estimates will be assessed.

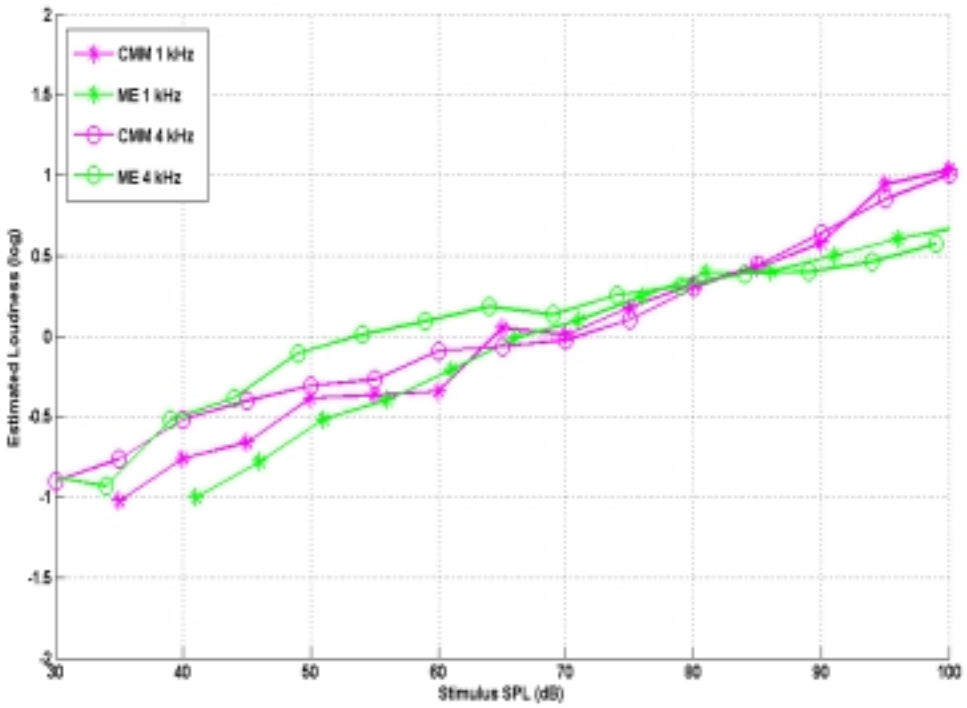

Figure 30- Average psychoacoustical data of loudness growth for tone-burst at 1 kHz (stars) and at 4 kHz (circles) for both CMM (magenta) and ME (green). All functions were shifted vertically in order to have zero-mean. There is a general agreement across frequency within modalities.

5.3.2. TBABR Recordings

A representative sample of the TBABR recording for 1 and 4 kHz are shown on figures 31 and 32. As expected (see Section 2.3.1), evoked responses generated by the 4 kHz tone-burst had more robust ABR components (visible at lower levels) in the range of 4-10 ms after the stimulus presentation. For this particular subject, the AMLR component (21 – 42 ms) is more visible with the 1 kHz than with the 4 kHz stimuli.

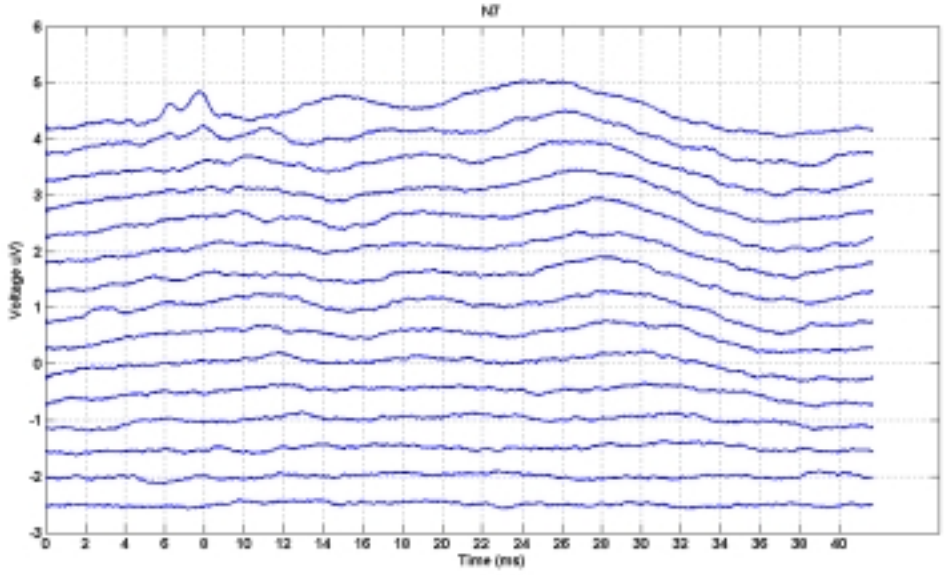

Figure 31- Weighted ABR averages for subject N7 and 1 kHz tone-bursts of 4ms duration. The levels are shifted for comparison. From top to bottom the levels were varied from 100 to 30 dB SPL in steps of 5 dB.

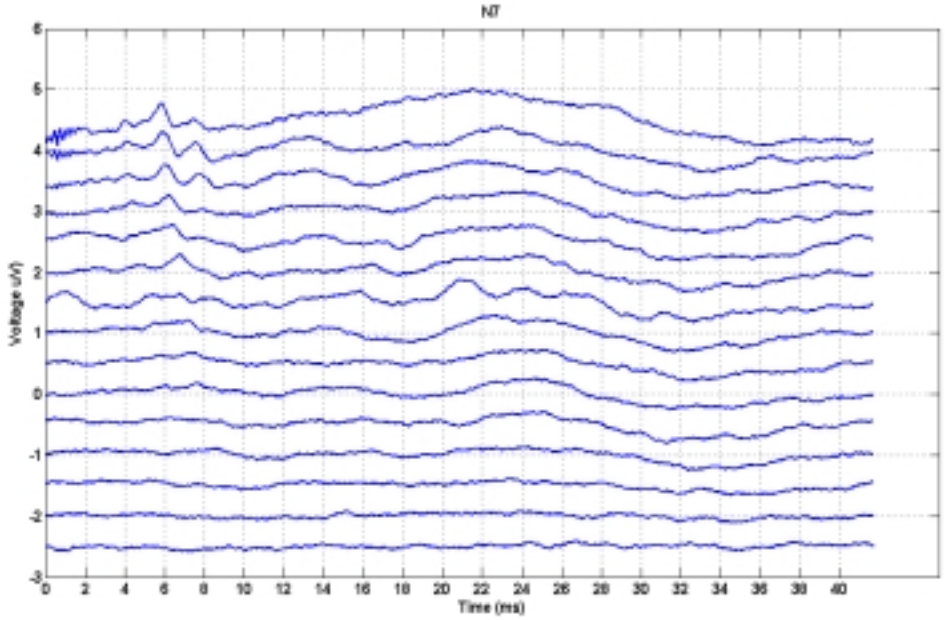

Figure 32- Weighted ABR averages for subject N7 and 4 kHz tone-bursts of 1 ms duration. The levels are shifted for comparison. From top to bottom the levels were varied from 100 to 30 dB SPL in steps of 5 dB.

Table IX shows some statistics of the evoked response recording. The Wave V TBABR latency was estimated for the highest level only using the wavVamp procedure described on Section 3.2 (location of maximum peak within 4.5-10 ms). The measurements are in qualitative agreement with literature (Hall 2006) in that the average latency for the 1 kHz (7.3 ms) was greater than that for the 4 kHz tone (6.4 ms). Columns three and five describe the average (across level) residual noise levels for the responses using the WNS Fmp estimate (in μV^2. There is a general consistency across subjects in that the residual noise power levels were within a 0.001 μV^2 magnitude range. While the standard deviation of the mean residual noise level for 1 kHz (last row third column) was an order of magnitude lower than the mean (third to last row, third column) the standard deviation of the group mean for the 4 kHz measure was within the same magnitude of the group mean (the mean at 4 kHz was also twice as large as the mean at 1 kHz). The same pattern is seen on the distribution of standard deviations of the average residual noise levels (fourth and sixth column). Its not clear why there is such a pattern, but a possible reason could be outlier effects since the medians are in much closer agreement (N2 and N5 had really high noise residual noise levels at 4 kHz).

Table IX- Statistics for the TBABR recordings on eight normal listeners (see text for details). The names represent: V Lat – wave V latency, Ave Noise- average residual noise power (μV^2) across SPL, Std Noise-standard deviation of the residual noise power (μV^2) across SPL (see text for detail).

Lst	V Lat (1 kHz)	V Lat (4 kHz)	Ave Noise (1 kHz)	Std Noise (1 kHz)	Ave Noise (4 kHz)	Std Noise (4 kHz)
N1	7.3	6.1	0.0013	0.0012	0.0023	0.0009
N2	6.5	6.4	0.0015	0.0004	0.0046	0.0024
N3	7.9	6.2	0.0028	0.0016	0.0008	0.0003
N4	7.3	4.5	0.001	0.0003	0.0011	0.0004
N5	7.8	6.1	0.0018	0.0004	0.0031	0.0024
N6	6.9	9.9	0.0009	0.0001	0.0024	0.0033
N7	7.8	5.9	0.0013	0.0003	0.0016	0.0006
N8	7.5	6.3	0.0008	0.0001	0.0006	0.0002

Ave	7.375	6.425	0.001425	0.00055	0.0020625	0.0013125
Median	7.4	6.15	0.0013	0.00035	0.00195	0.00075
Std	0.486239212	1.527603165	0.000645313	0.000547723	0.001339443	0.001200521

5.3.3. TBABR and TBOAE Loudness Growth Estimation with no Noise Control

Figure 33 shows the results in terms of the median MSE calculated across the eight normal listeners for loudness growth estimation obtained through psychoacoustical procedures, TBOAE, and the ten different types of loudness estimation through evoked potentials described on Section 3.4.3.1 and 3.4.3.2 (with no noise control) . For the 1 kHz stimulus, the TBOAE procedures as well as the AMLR Sync and Full Sync, operate within the same range as the “optimal” psychoacoustical procedures. For the 4 kHz tone-burst stimulus, however, the TBOAE yields the highest median MSE (as expected from similar results from Section 5.1), but the AMLR Sync and Full Sync still operate within the same range as the “optimal” psychoacoustical procedures.

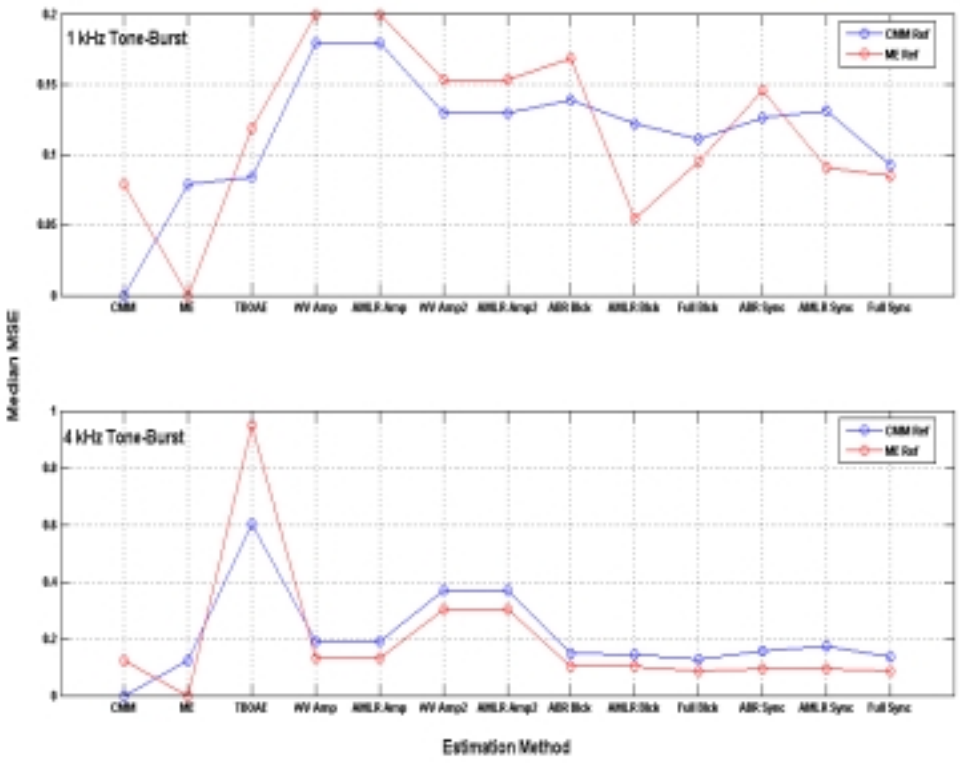

Figure 33- Median MSE across eight normal listeners. The legend on the bottom graph also applies to the top graph (axis between both plots are not in the same range).

Figure 34 shows the exact same information as Figure 33, but the psychoacoustical references are separately in order to visualize the effect of stimulus frequency on the estimation procedure. Figure 34 shows that there is little difference in the MSE with respect to stimulus frequency between the two psychoacoustical procedures (first point on the series at each graph). As expected, however, the estimation of loudness growth through TBOAEs does show a difference with respect to frequency regardless of which psychoacoustical procedure is used to estimate loudness growth. This observation and MSE results are in agreement with the previous study described in Section 5.1. In fact, the TBOAE estimates at 4 kHz might be used as a upper bound on

the quality of the estimation, since it has been suggested (Section 5.1) that TBOAE estimations of loudness growth at 4 kHz, using the previously describe procedure, is essentially linear with respect to the stimulus. Interesting, the estimation through evoked responses also seem to show a bias effect with CMM as reference (top graph). This result is unexpected, because intuitively, the evoked responses should perform better with a 4 kHz stimulus than with a 1 kHz tone-burst (due to neural synchronization with the stimulus). This bias is not observed (except for the ABR Amp2 and AMLR Amp2) when ME is used as the reference loudness growth. In general it seems like the 'sync' segmentation procedures (fullsync, abrsync, and amlrsync) are the ones that yield the best performances of the evoked potential procedures as well as showing some robustness across stimulus frequency. Their performance are within the same range as the psychoacoustical references (and with the TBOAE at 1 kHz). Figures 35 and 36 show the individual loudness growth estimates obtained through CMM, ME, TBOAEs and using the fullsync procedures on the evoked response. The smallest MSE for the fullsync procedure at 1 kHz was given by listener N8 for CMM (0.0212) and N7 for ME (0.0140). The smallest MSE for the fullsync procedure at 4 kHz was given by listener N3 for CMM (0.0191) and listener N2 (0.0255).

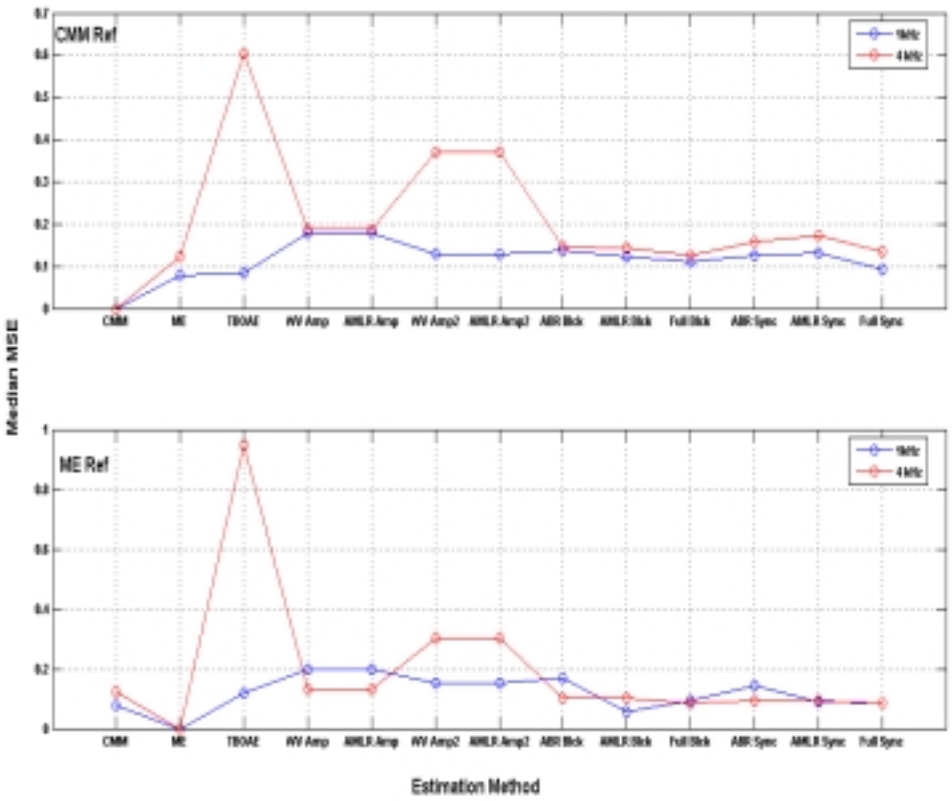

Figure 34- Effect of stimulus frequency on procedure. This is plot has the same information as Figure 33, but the psycho acoustical references are separated from each other and plotted with respect to stimulus frequencies.

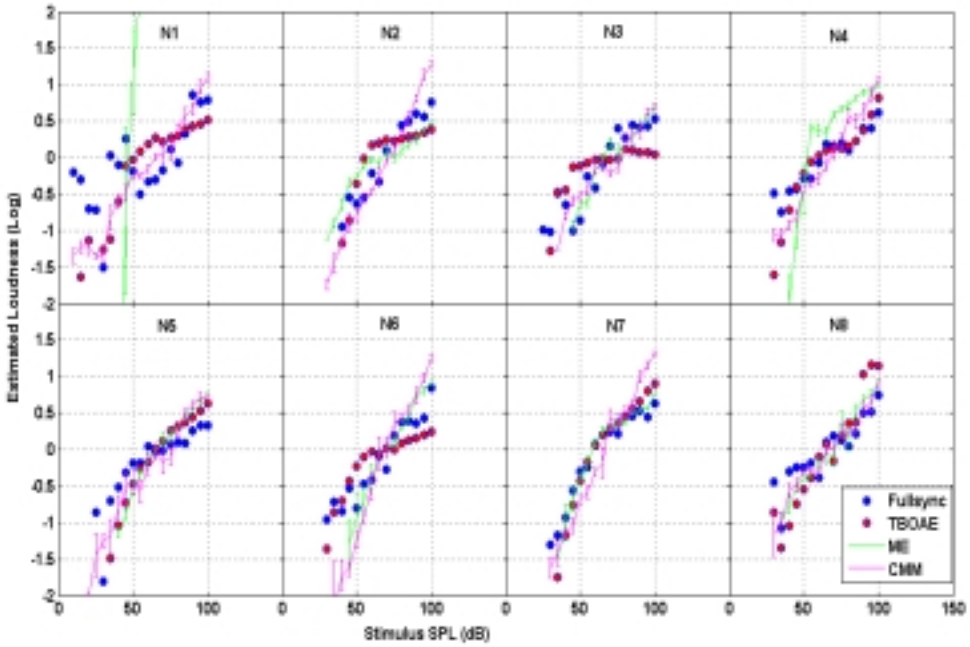

Figure 35- Individual loudness growth estimates for 1 kHz tone-burst.

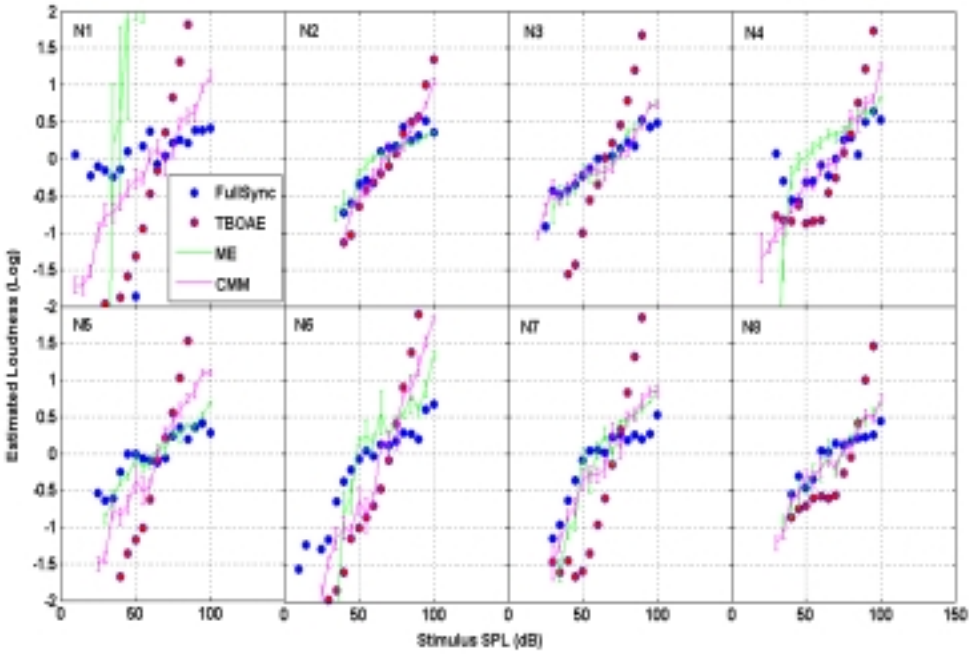

Figure 36- Individual loudness growth estimates for 4 kHz tone-Burst.

5.3.4. TBABR Loudness Growth Estimation Noise Control and Parametric Fitting

Figure 37 compares the results in terms of the median MSE calculated across the eight normal listeners for loudness growth estimation obtained used the two different noise control methods. For the psychoacoustical procedures and TBOAE, the noise control methods were applied with the residual weights all set to 1 (uniform) across all SPLs. In general, the parametric noise control using the INEX function yielded the lowest median MSE. The weighted polynomial fitting in some cases (especially at 4 kHz) yields a small improvement in the MSE. It interesting to observe that the noise control methods also seem to have a very small advantage when used on the ME estimation data with the

CMM treated as the reference loudness growth (first column). The noise control methods showed greatest benefit under the 1 kHz tone-burst, with the wave V amp procedures showing the greatest benefit.

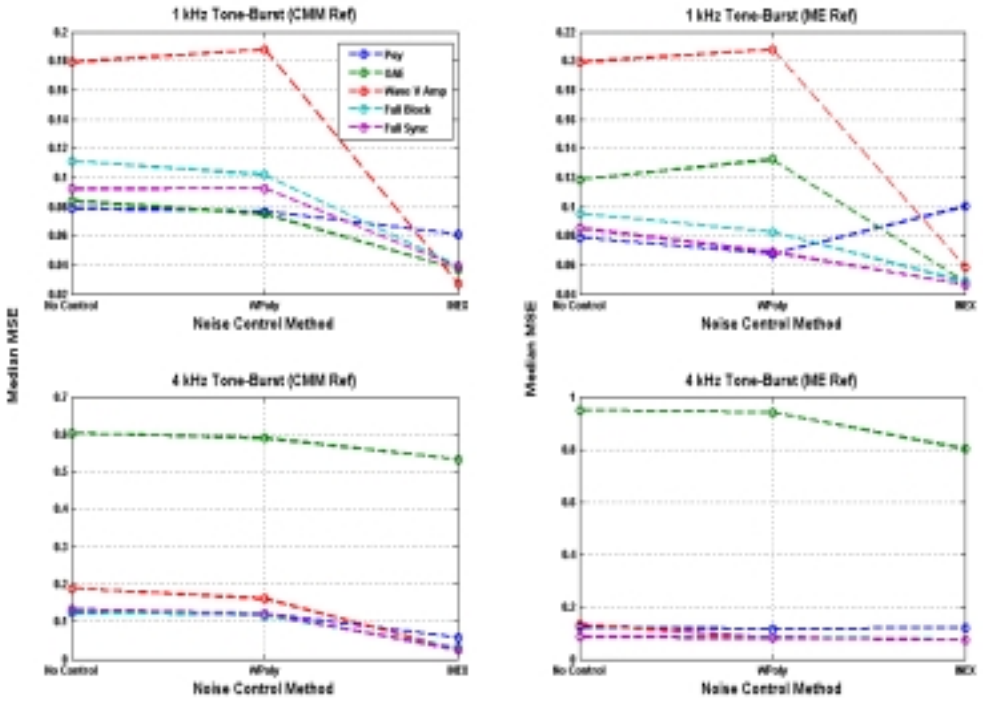

Figure 37- Median MSE four five different loudness growth procedures according to the noise control method applied. For the psychoacoustical and TBOAE procedures, the noise control was used with all residual weights being equal (uniform) across SPL (the scales are not the same for each graph).

Figures 38 through 41 show the individual loudness growth estimation for the fullsync procedure using the wpoly and inex noise control methods respectively along with the CMM and ME raw data for comparison (Figure 35 and 36 show the same data but with no noise control applied). In general subject N3 showed good agreement

between the physiological estimates and the psychoacoustical estimates for both noise control methods.

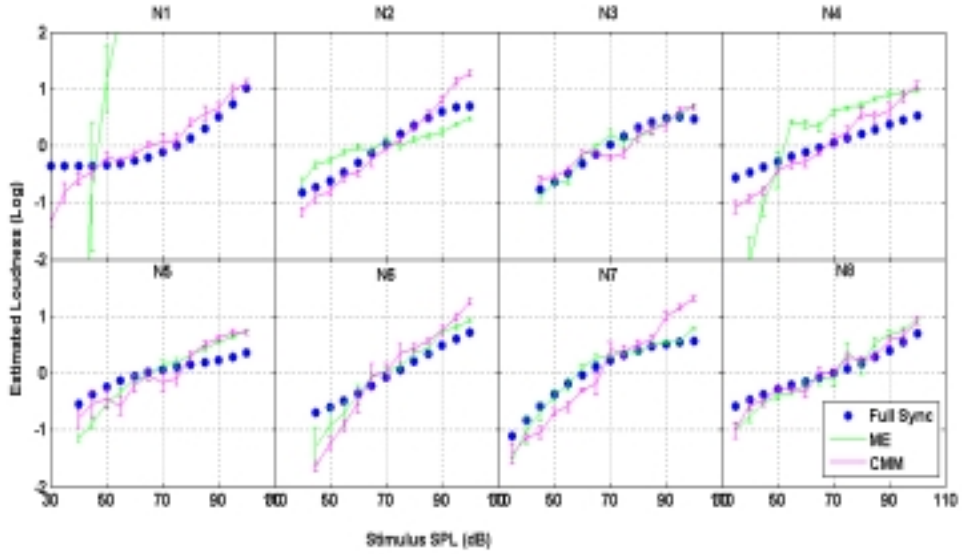

Figure 38- Individual loudness growth estimation for 1 kHz tone-burst using CMM, ME and fullsync procedures with weighted polynomial noise control applied to fullsync.

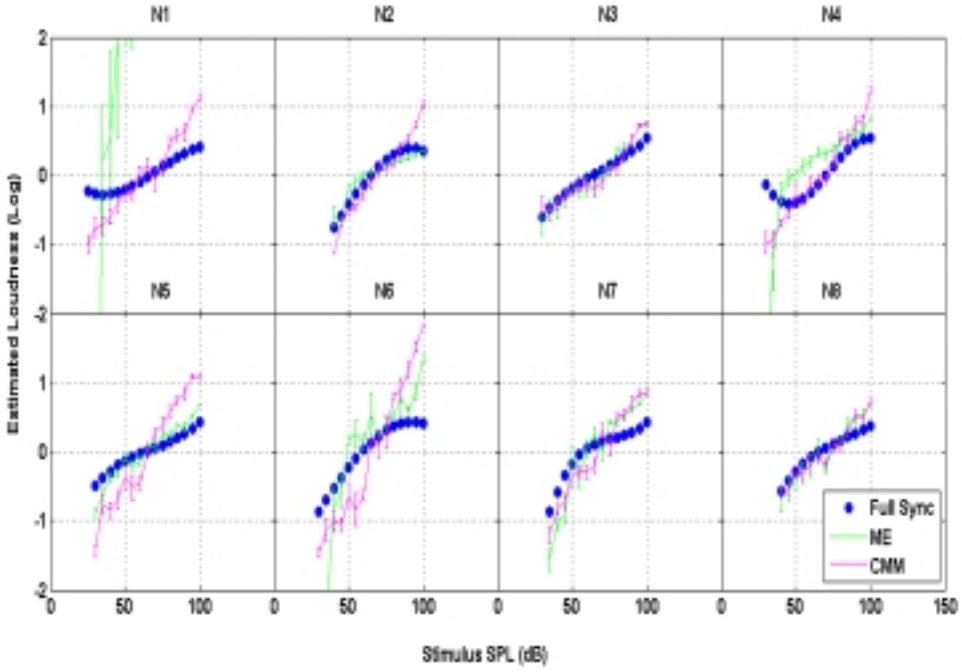

Figure 39- Individual loudness growth estimation for 4 kHz tone-burst using CMM, ME and fullsync procedures with weighted polynomial noise control applied to fullsync.

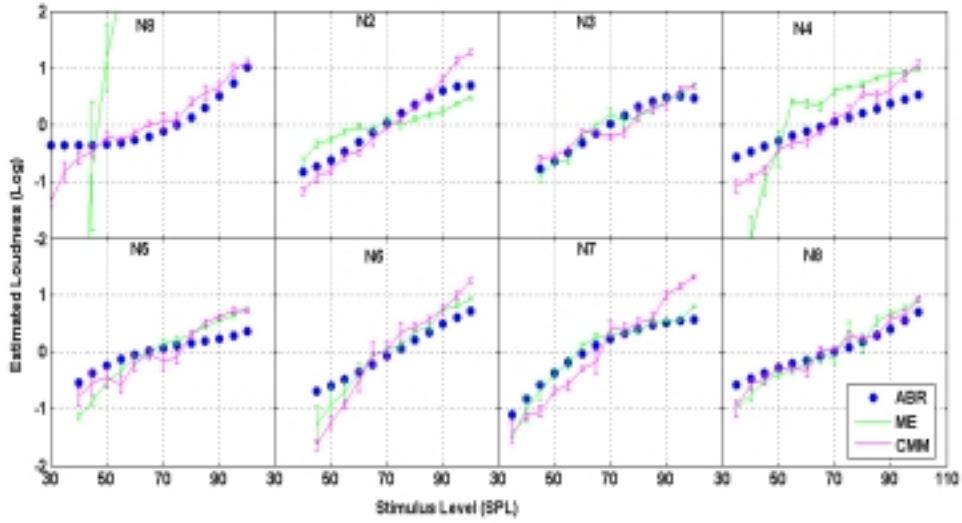

Figure 40- Individual loudness growth estimation for 1 kHz tone-burst using CMM, ME and fullsync procedures with inex noise control applied to fullsync.

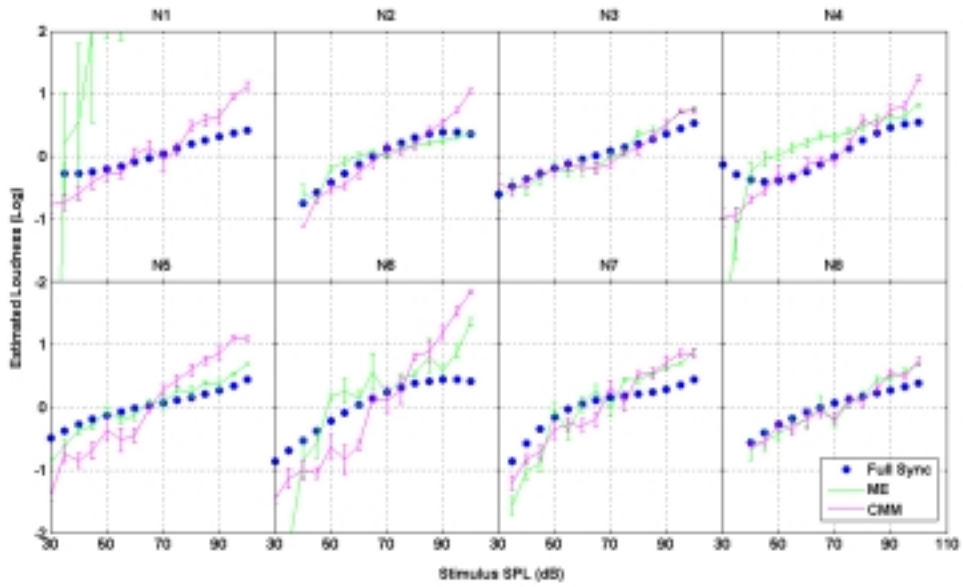

Figure 41- Individual loudness growth estimation for 1 kHz tone-burst using CMM, ME and fullsync procedures with weighted polynomial noise control applied to fullsync.

5.4 Estimation of Loudness Growth in Hearing Impaired Listeners

5.3.1. Psychoacoustical Results

Figures 42 and 43 show the individual results for the psychoacoustical loudness growth data obtained for eight HILs using 1 kHz tone-bursts and four HILs using 4 kHz tone-bursts respectively (all functions were arbitrarily shifted to have zero mean). The HILs yielded more consistent estimates between the CMM data and the ME data (Tables X and XI). This could be due to the outlier (N1) on the normal data, which had a MSE of 35.84 compared to 0.258 (H3) highest value of the HIL dataset at 1 kHz and 66.03 (N1) and 0.04 (H3) for highest values at 4 kHz; in addition to the fact that the HIL data had

fewer levels. Unlike the normal data, the HILs' loudness growth function does show much discrepancy between CMM and ME close to threshold values, only at 1 kHz, listener's H3 show the same behavior as normals, with the ME data having a higher slope at threshold than the CMM data; on the other-hand, listener H5 shows the exact opposite behavior at 1 kHz as well. It is possible that the listener's perceptual range is one of the factors influencing this edge effect. The median estimated threshold of the CMM (ME) data for HILs was 60 (62) dB SPL, which was about twice that of the normals 30 (35) dB SPL for 1 kHz. The threshold values estimated from CMM and ME on HILs were in closer agreement than that with normal listeners (in this case CMM and ME were run on the exact same apparatus). Two HILs, had threshold values within a 5-15 dB range of the normal average at 1 kHz: 45 (45) for H5 and 40 (40) for H7. A similar analysis for the 4 kHz data shows that the median CMM (ME) estimated threshold for HILs were 75 (70) dB SPL versus 25 (30) dB SPL for normal listeners. Listener H7 had the lowest threshold 40 (45) of the all HILs at a 15 dB range of normal median. The median slope values obtained were within similar range to that of normals (0.19-0.26).

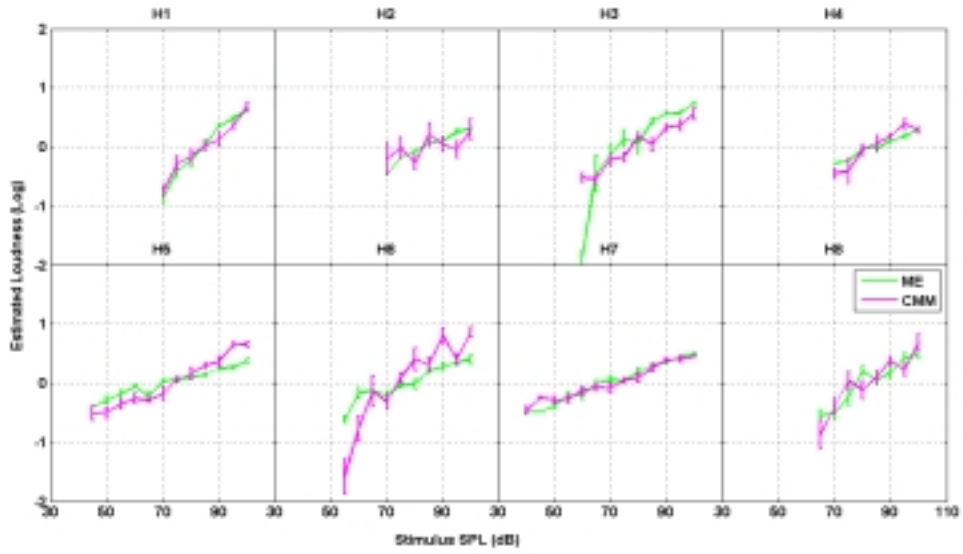

Figure 42- Individual loudness growth functions estimated through CMM and ME with 1 kHz tone-burst for HIL. The legend for listener N8 applies to all other plots. Error bars represent plus or minus one standard error.

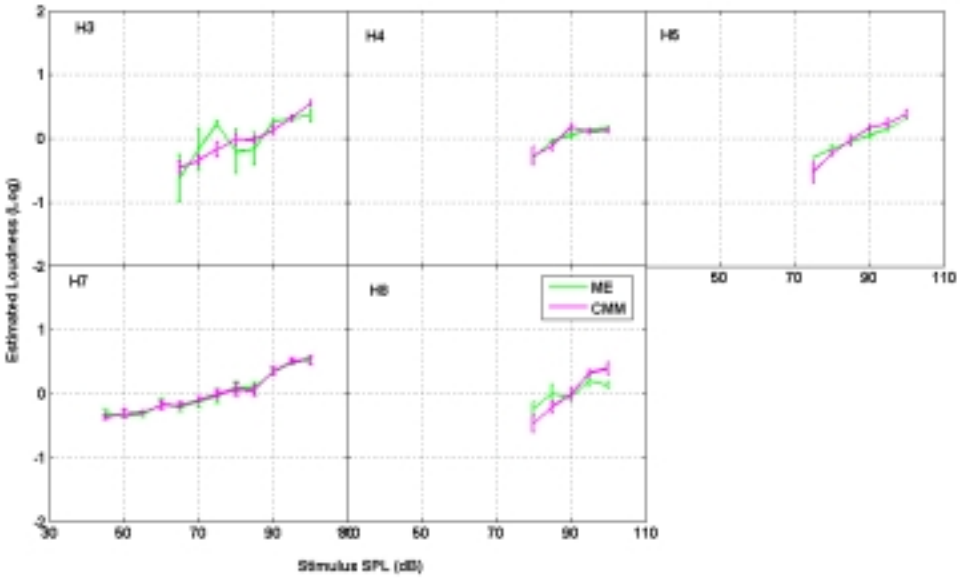

Figure 43- Individual loudness growth function estimated through CMM and ME for 4 kHz tone-burst for HIL (same listeners of Figure 42). Three listeners had thresholds too high to collect any data at 4 kHz.

Table X- Statistics on the Psychoacoustical measurement of loudness growth with respect to sound pressure for 1 kHz tone-burst on HILs.

Lst	MSE (CMM vs ME)	CMM TH	ME TH	CMM Slope	ME Slope	CMM Stderr	ME Stderr
H1	0.014941087	70	70	0.41922483	0.488131	0.09578871	0.06762218
H2	0.03433365	70	70	0.12429242	0.243576	0.17262108	0.03502164
H3	0.258217135	60	60	0.27901569	0.512245	0.07947442	0.12278234
H4	0.015887301	65	70	0.22586745	0.192619	0.08749846	0.03666038
H5	0.035343357	45	45	0.22657338	0.13106	0.07176658	0.0345628
H6	0.198609432	55	55	0.44827004	0.19405	0.15544111	0.06497822
H7	0.008708099	40	40	0.14945356	0.175145	0.05063174	0.03729003
H8	0.053565636	60	65	0.45778569	0.308257	0.11760376	0.09239573
Ave	0.077450712	58.125	59.375	0.29131038	0.280635	0.10385323	0.06141416
Median	0.034838504	60	62.5	0.25279454	0.218813	0.09164359	0.05113412
Std	0.09560454	10.99918828	11.783	0.13382059	0.145199	0.04203455	0.03249052

Table XI- Statistics on the Psychoacoustical measurement of loudness growth with respect to sound pressure for 4 kHz tone-burst on HILs.

Lst	MSE (CMM vs ME)	CMM TH	ME TH	CMM Slope	ME Slope	CMM Stderr	ME Stderr
H3	0.041164367	65	65	0.26744836	0.228137	0.06612763	0.184391
H4	0.005867375	80	75	0.20148097	0.14092	0.0618658	0.02412537
H5	0.01374505	75	70	0.35234594	0.211179	0.06919339	0.03334591
H7	0.00139163	40	45	0.16994428	0.16757	0.05260055	0.0548393
H8	0.03620979	75	80	0.55014164	0.190147	0.08674621	0.0890867
Ave	0.019675642	67	67	0.30827224	0.18759	0.06730671	0.07715766
Median	0.01374505	75	70	0.26744836	0.190147	0.06612763	0.0548393
Std	0.017995059	16.04680654	13.5093	0.1521914	0.03459	0.01253713	0.06494477

Figure 44 shows a plot of the threshold value versus the log of the perceptual range of all listeners. The perceptual range was calculated by subtracting the minimum from the maximum of the estimated loudness growth function on a log scale. The final range was the average range between the CMM and ME procedures. Although Figure 44 shows an interesting pattern, it has to be interpreted with some caution, because its possible that the dynamic ranges of the stimuli presented does not fully encompass the listeners' perceptual ranges. Having said that, it is still noteworthy to observe that

listeners with the same threshold can have drastically different perceptual ranges. In particular, thresholds around 60 dB SPL seem to result in large individual differences in perceptual ranges, with some listeners within two standard deviations of the normal-hearing listeners and some listeners have perceptual ranges more than two standard deviations below the normal listeners (dotted line). A significant ($p<0.01$) linear correlation, $r=-0.697$, was found between log perceptual range and threshold. This analysis is in agreement with the current understanding of loudness-growth models, where loudness growth in HILs seems to behave anywhere between two model types: loudness recruitment and softness imperception. Under loudness recruitment, the HILs have the same perceptual range as normals, but a higher threshold, so that the loudness growth function is steeper than normal. Under softness imperception, on the other hand, HILs have not only a reduced physical dynamic range of hearing, but also a reduced perceptual range. This variability is also consistent with the variability of estimated loudness slope function in HIL (e.g., (Hellman and Meiselman 1992).

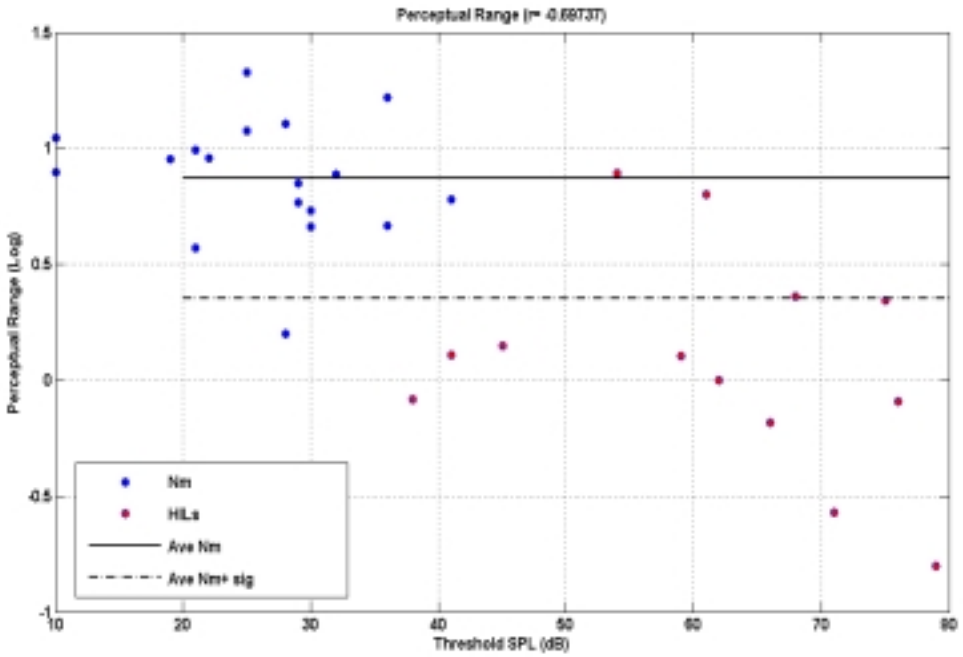

Figure 44- Listener's threshold vs log Perceptual Range. A significant linear correlation (r=-0.69) was found between perceptual range and threshold. The solid black line reprensent the normal average and the dotted black line represents minus two standard deviations from the average. Nevertheless significant variability can still be found for perceptual ranges given a fixed threshold (i.e., see data points around 60 dB SPL).

5.3.2. TBABR Recordings

Table XII shows some descriptive statistics from the evoked response recordings in a similar manner to Table IX. The residual noise levels were within the same range as that for normal-hearing listeners at 1 kHz; for the 4 kHz stimuli, the residual noise levels of the HILs were about two orders of magnitude higher than normal listeners. The listener with highest residual noise levels for both frequencies was listener H5.

Table XII- Statistics for the TBABR recordings on HIL (see text for details). The names represent: V Lat – wave V latency, Ave Noise- average residual noise power (μV^2) across SPL, Std Noise- standard deviation of the residual noise power (μV^2) across SPL (see text for detail).

Lst	V Lat (1 kHz)	V Lat (4 kHz)	Ave Noise (1 kHz)	Std Noise (1 kHz)	Ave Noise (4 kHz)	Std Noise (4 kHz)
H1	7.1		0.0002	0.0001		
H2	9.4		0.0011	0.001		
H3	7.5	4.7	0.0011	0.0009	0.001	0.00372
H4	7.6	6.9	0.0008	0.0002	0.0006	0.001164
H5	7.7	4.6	0.0014	0.0002	0.002	0.004642
H6	9.4		0.0012	0.0001		
H7	7.4	4.5	0.002	0.0011	0.0006	0.00166
H8	7.4	6.3	0.0006	0.0003	0.0008	0.004031
Ave	7.9375	5.4	0.00105	0.0004875	0.001	0.0030434
Median	7.55	4.7	0.0011	0.00025	0.0008	0.00372
Std	0.91953016	1.118033989	0.000539841	0.000432394	0.000583095	0.00153579

5.3.3. TBABR and TBOAE Loudness Growth Estimation with no Noise Control

Figures 45 and 46 show the results in terms of the median MSE calculated across the HILs for loudness growth estimation obtained through the psychoacoustical procedures, TBOAE, and the ten different types of loudness estimation through evoked potentials described on Section 3.4.3.1 and 3.4.3.2 (with no noise control). In general the estimation done by the TBOAE yields a higher median MSE compared to the estimations done with evoked potentials (e.g., wave V amp, and fullsync). There is however, as expected a consistent improvement in the TBOAE estimation when the stimulus is 1 kHz rather than 4 kHz (Figure 46). The performances of the estimators are not as close as the psychoacoustical estimators (especially at 4 kHz), but it is still well below the upper bound reference represented by the 4 kHz TBOAE performance.

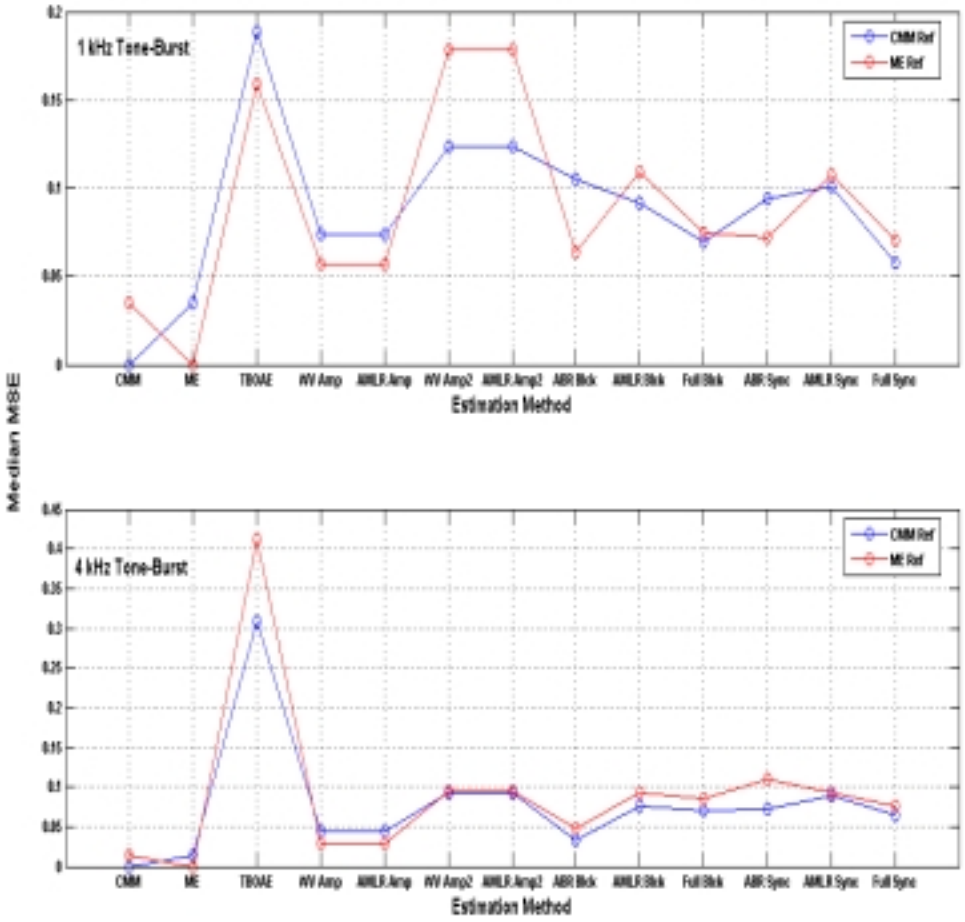

Figure 45- Median MSE across the HIL. Axis between both plots are not in the same range. In general the performance of the waveV amp and fullsync are robust across frequency seem to perform the close to optimal within the procedures using evoked potentials. The TBOAE procedure for some reason does not yield as good MSE (the ranges between both plots are not the same).

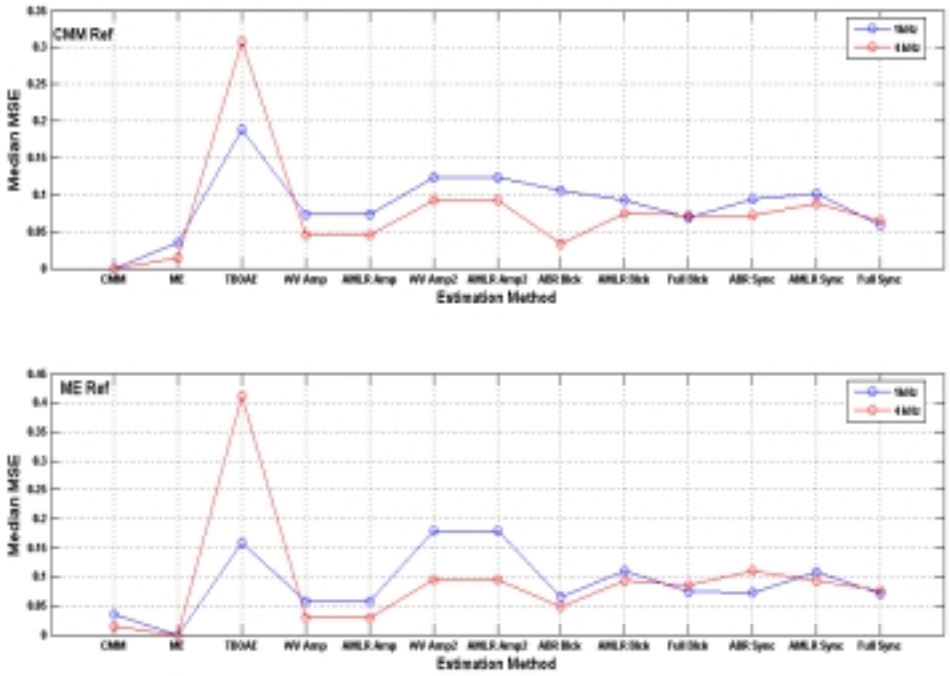

Figure 46- MSE as a function of estimation procedure. This plot represents the same information as that of Figure 45 but MSE is grouped with respect to the psychoacoustical reference used. From this graph it is clear that there is an overall improvement in the TBOAE procedure from 4 kHz to 1 kHz as expected.

Figures 47 and 48 show the individual loudness estimates obtained with the full sync procedure along with the TBOAE, CMM, and ME data. The data from 1 kHz TBOAE yields somewhat similar results to the full sync on some subjects (H1, H2, and H8), but at 4 kHz the TBOAE essentially estimates a linear growth (probably a result of artifact contamination).

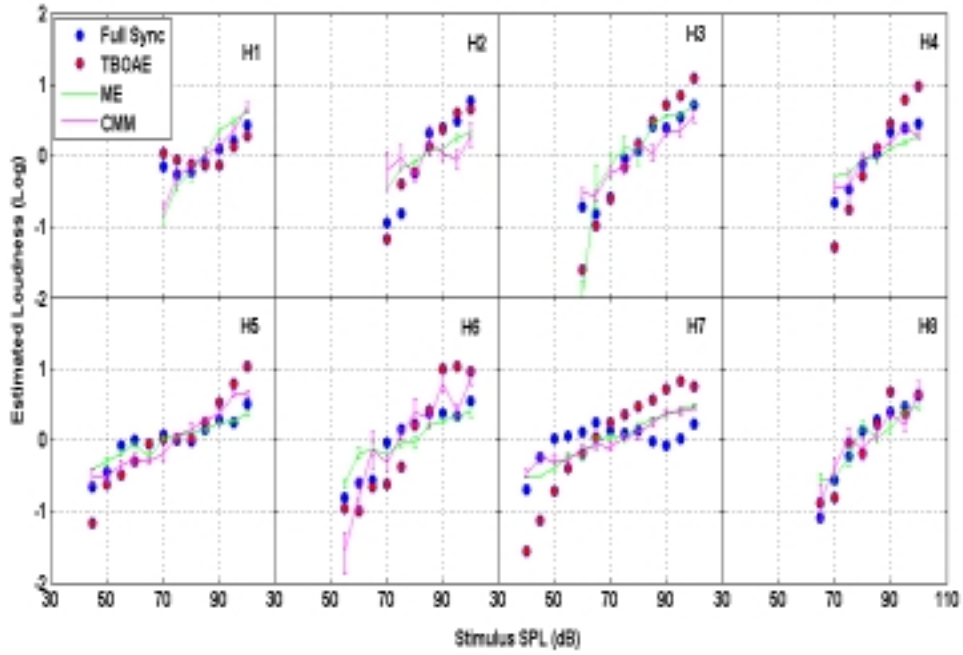

Figure 47- Individual loudness growth estimates for HIL using 1 kHz tone-burst.

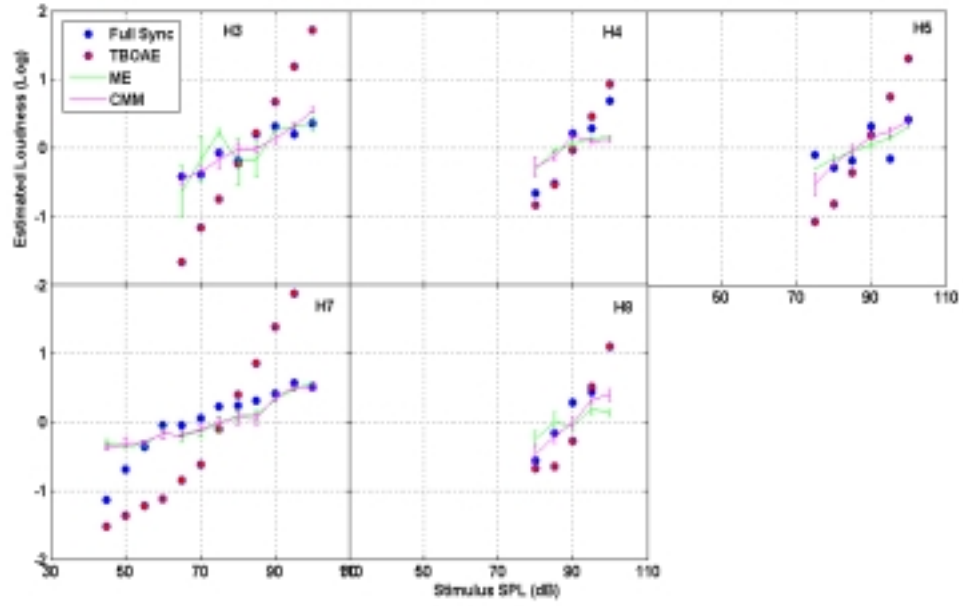

Figure 48- Individual loudness growth estimates for HIL using 4 kHz tone-Burst.

5.3.4. TBABR Loudness Growth Estimation with Noise Control with Parametric Fitting

Figure 49 compares the results in terms of the median MSE calculated across the HILs for loudness growth estimation obtained used the two different noise control methods. For the psychoacoustical procedures and TBOAE estimators, the noise control methods were applied with the residual weights all set to 1 (uniform) across all SPLs. Unlike the normal counterpart (Figure 37) the TBOAE estimation yields higher MSE compared to the other procedures. The wave V amp procedure also does not show a drastic drop in noise control as seen with normal listeners. In general the evoked potential estimators almost show a monotonic improvement on performance with respect to the three different conditions, achieving equivalent performance with the psychoacoustical reference (except perhaps at the 1 kHz ME Ref condition).

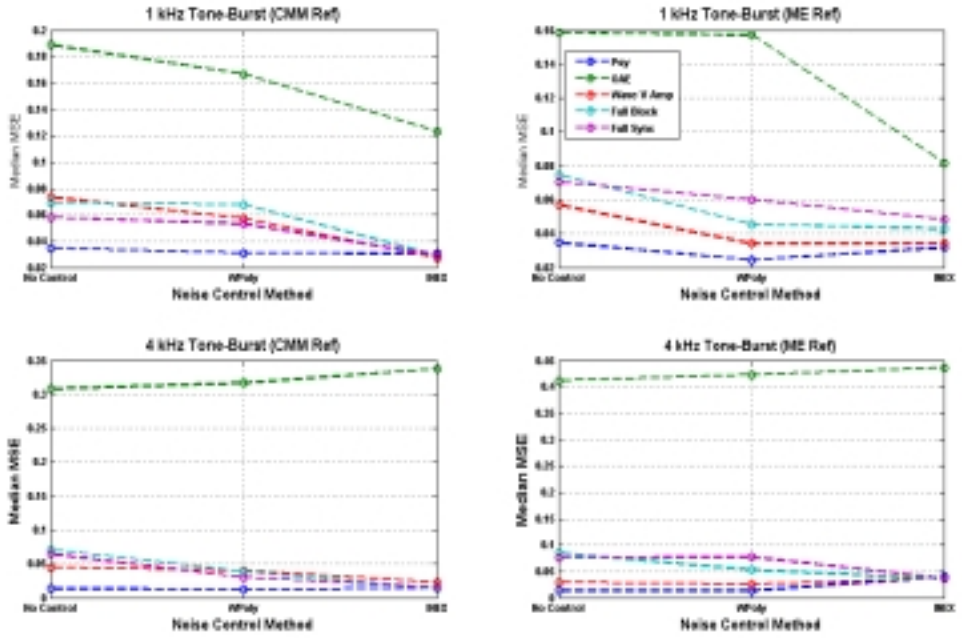

Figure 49- Median MSE four five different loudness growth procedures according to the noise control method applied. For the psychoacoustical and TBOAE procedures, the noise control was used with all residual weights being equal (uniform) across SPL (the scales are not the same for each graph).

Figures 50 through 53 show the individual loudness growth estimation for the fullsync procedure using the wpoly and inex noise control methods respectively along with the CMM and ME raw data for comparison (Figures 47 and 48 show the same data but with no noise control applied). In general it seems like listeners H3, H5, and H8 yielded the best results across the different conditions.

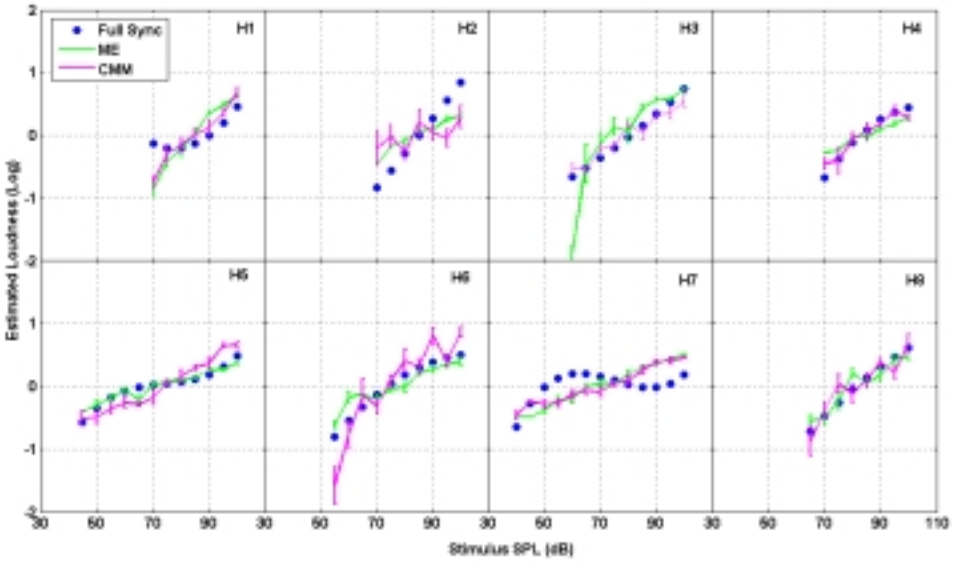

Figure 50- Individual loudness growth estimation for HIL using 1 kHz tone-burst and the CMM, ME and fullsync procedures with weighted polynomial noise control applied to fullsync.

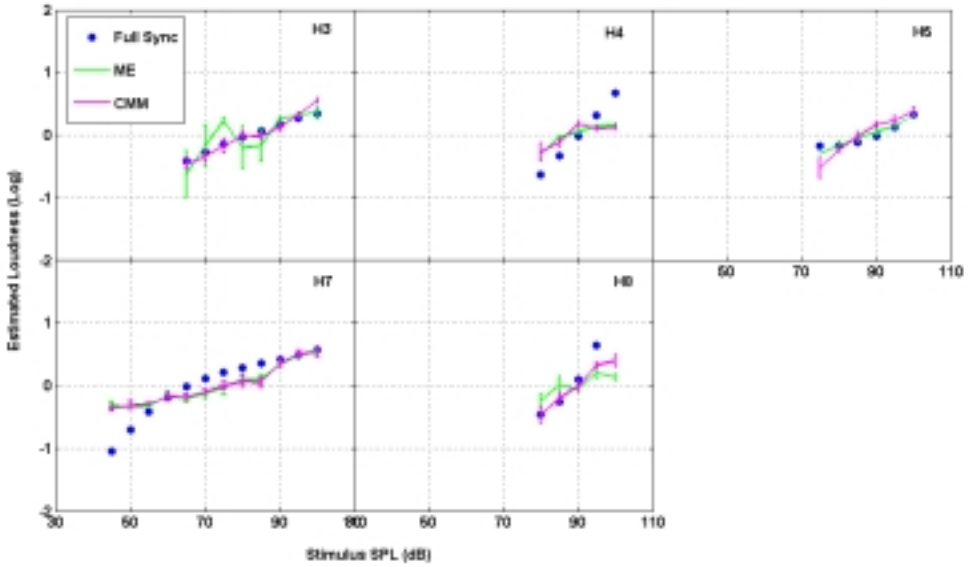

Figure 51- Individual loudness growth estimation for HIL using 4 kHz tone-burst and the CMM, ME and fullsync procedures with weighted polynomial noise control applied to fullsync.

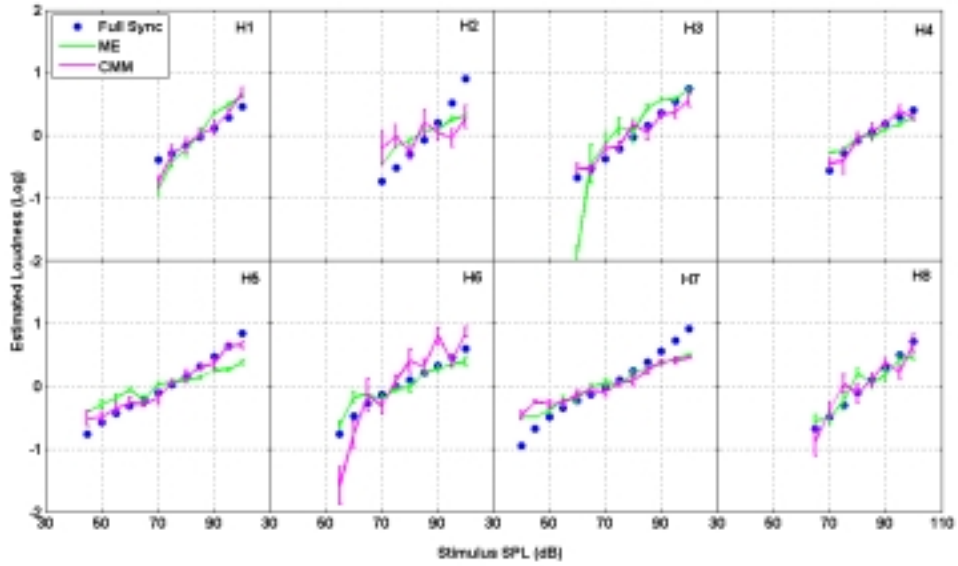

Figure 52- Individual loudness growth estimation for HIL using 1 kHz tone-burst and the CMM, ME and fullsync procedures with inex noise control applied to fullsync.

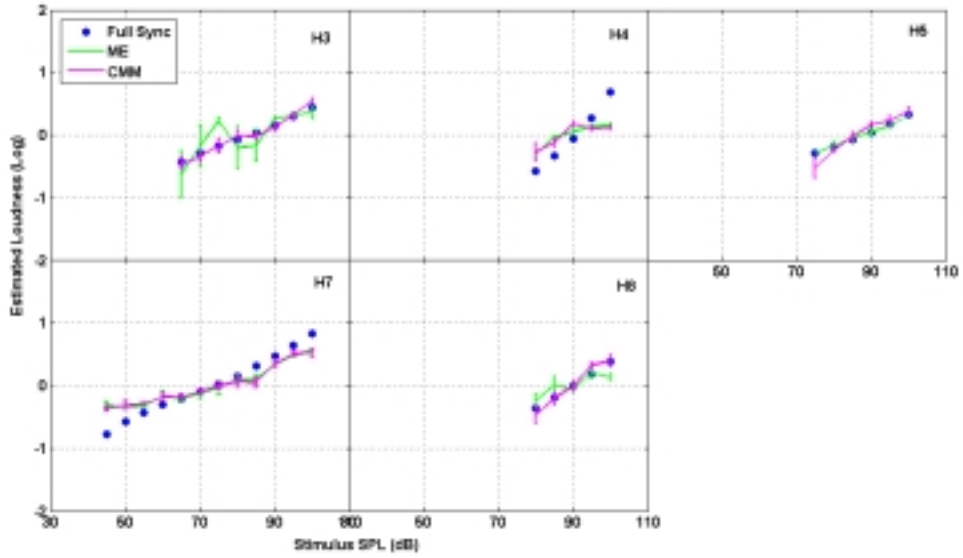

Figure 53- Individual loudness growth estimation for HIL using 4 kHz tone-burst and the CMM, ME and fullsync procedures with inex noise control applied to fullsync.

5.3.5 TBABR Loudness Growth Estimation and Residual Noise Levels

In order to investigate the effects of TBABR post-average residual noise levels on the performance the estimation of loudness growth function two analyses of the data were performed. One analysis looked into the averaged evoked response residual noise level (estimated through the WNS Fmp) in all listeners versus the MSE and linear correlation coefficient between the TBABR wave V amp estimator and the CMM psychoacoustical estimator. The second investigation looked into how the MSE and linear correlation coefficient between CMM and two of the TBABR loudness growth estimators varied as function of residual noise level (estimated through the WNS Fmp) when the number of trials averaged varied from 1,000 to 12,0000 in steps of 1,000 epochs. In order to obtained a wide range of MSE values this analysis used the data from one of the subjects that yielded the lowest MSE (H5 at 1 kHz).

Figure 54 shows the results for the first analysis. The top graph shows the estimated averaged evoked response residual noise level vs. the MSE between wave V amp estimator and the CMM psychoacoustical estimator (normals in blue, and HILs in red). An estimate of the linear correlation was significant only at the $p< 0.01$ significance level ($r = 0.3544$, $p = 0.059$). The bottom graph of Fig. 54 show a plot of the estimated averaged evoked response residual noise level and the linear correlation coefficient between the wave V amp estimator and the CMM estimator. In this case there is a significant negative linear relationship between averaged evoked response residual noise level and a correlation between the two modalities at $p< 0.05$ ($r = -0.3837$, $p = 0.039$). In general the HILs had a better performance compared to normals even within similar

WNS Fmp ranges, this could be due to differences in the experimental setup or to fewer data points. If a MSE value of 0.15 is the cutoff for a reasonable fit, based for example, on the performance seen from Figure 45, then it seems that setting an average value of WNS Fmp to be between 0.0001 and 0.00015 might, for the most part, yield MSE estimation using wave V amp around the 0.15 range. The linear correlation coefficient plot shows a similar behavior, but with wider scatter; in this case setting the WNS Fmp to less than or equal to 0.0001 will, in general, yield a correlation coefficient between the two loudness growth estimators of 0.7 or higher between the wave V amp procedure and CMM.

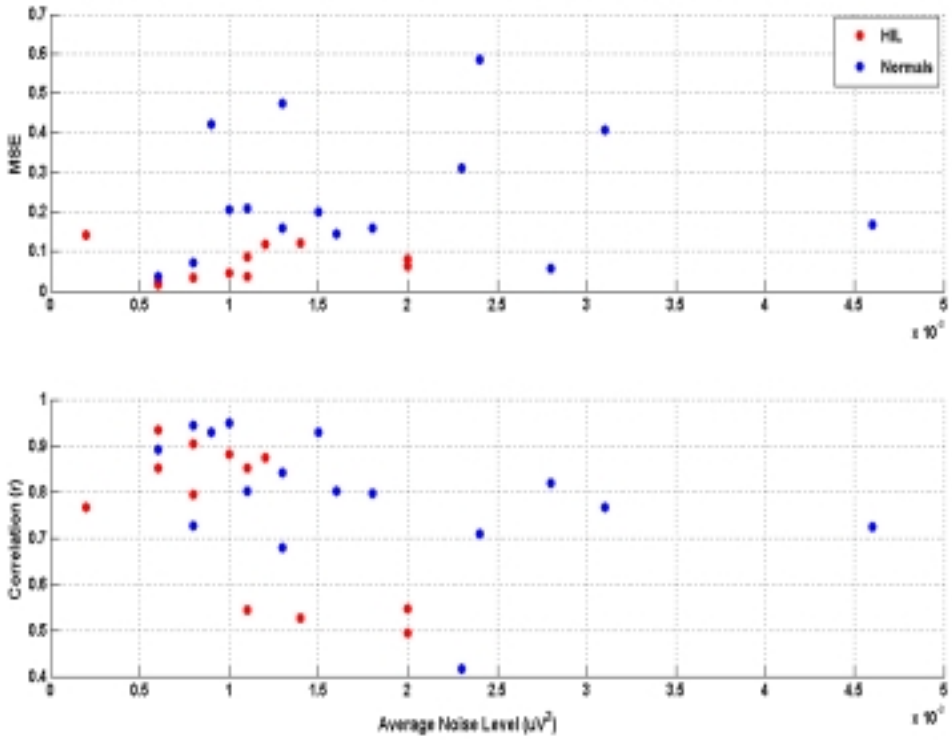

Figure 54- Performance of the estimators as function of WNS Fmp. Top graph shows the MSE and bottom graph the linear correlation between the wave V amp procedure and CMM.

Figure 55 show the results for the second type of analysis. For this analysis, TBABR data from one of the subject that yielded the best MSE result (H5 at 1 kHz) was weighted averaged in a cumulative fashion from 1,000 to 12,000 trials to yield 12 data points with different WNS Fmp values. No noise control method was applied to the evoked response loudness growth estimators. The mean residual noise level of the weighted averaged evoked response waveform was monotonic with respect to the number of trials used (i.e., the point to far left corresponds to using all 12,000 trials and the point to the far right corresponds to using the first 1,000 trials). The top graph shows the estimated averaged evoked response residual noise level vs. the MSE between wave two of the evoked response estimators and the CMM psyhchoacoustical estimator (wave V amp in blue, and full sync in red). The estimated linear correlation for the top graph was $r = 0.9612$ ($p<0.001$) for the full sync procedure and $r= 0.7748$ ($p<0.001$) for the wave V amp procedure. The bottom graph of Fig. 54 shows a plot of the estimated averaged evoked response residual noise level and the linear correlation coefficient between two evoked response loudness growth estimators and the CMM loudness growth estimator (wave V amp in blue, and full sync in red). The estimated linear correlation for the bottom graph of Fig. 54 was $r = -0.9697$ ($p<0.001$) for the full sync procedure and $r= -0.9844$ ($p<0.001$) fro the wave V amp procedure. Overall, Fig. 55 shows that both the MSE and correlation coefficients of the wave V amp procedure are consistently worst than the full sync procedure. The graph that describes the correlation between the psychoacoustical estimate and the evoked response estimates (bottom graph) seems to show ceiling effects for the wave V amp at about 0.8 (y-axis) being reached at about a an average evoked response residual noise level of 0.0002 (and a limiting performance

slightly higher than 0.95 for the full sync procedure). This ceiling behavior can somewhat also be seen on the MSE data as a floor effect, with the wave V amp and full sync procedure converging on MSE values of around 0.04 and 0.01 for WNS lower than 0.0002. This suggests that the performance of the evoked response estimators for this particular subject using 12,000 could have also been can be achieved with half the number of trials used (the number of trials used was inversely related to averaged residual noise level).

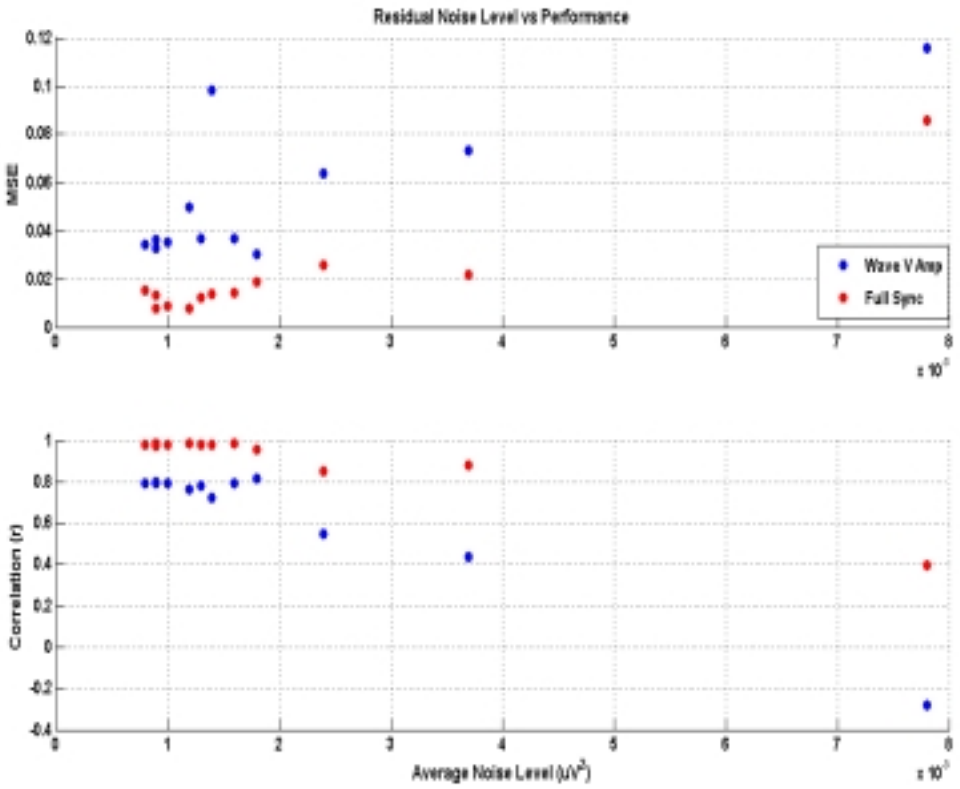

Figure 55- WNS Fmp level vs Performance for two TBABR procedures, wave V amp, and full sync, measured against the CMM psychoacoustical procedure for listener H5.

VI. Conclusion

Four experiments were performed with the overall aim of improving methods for the estimation of loudness growth through objective means. The first experiment looked into finding an optimal set of parameters for the (Epstein and Florentine 2005) TBOAE loudness growth estimation procedure at 1 kHz and how the ear canal resonance affects the performance for the 4 kHz stimulus. The results form this experiment indicate that there is significant potential for using tone-burst otoacoustic emissions for estimating loudness in response to 1-kHz stimuli across a wide range of levels in normal-hearing listeners. However, at frequencies near to ear-canal resonance, a simple reflection analysis procedure may not be sufficient to eliminate acoustic artifacts for loudness estimation.

The second experiment focused on developing a new procedure for the estimation of residual noise levels from the normal and weighted averaging of evoked responses under non-stationary conditions. Because the TBABRs waveforms are notorious for high amplitude variability, it important that loudness growth estimation procedures obtained from TBABRs take into account residual noise levels in their analysis. This experiment proposed two new methods (*NS Fmp* and *WNS Fmp*) that deals with non-stationary noise sources based on a modification to a widely used procedure (*Fsp*) for estimating SNR. The new method attempts to account for the non-stationary background by assuming $\hat{P}$ discrete locally stationary noise sources. A new noise segmentation procedure was also introduced that statistically partitions the background noise into dynamic discrete segments of different noise power based on a series of F-tests. The combined *NS Fmp* and segmentation parameters values can be set to result in the standard *Fsp* estimation.

Modifications to the *NS Fmp* were also done to account for weighted averaging, yielding the *WNS Fmp*. The results in general showed a better performance than the current standard, the *Fsp*.

The third experiment investigated the estimation of loudness growth from CMM, ME, TBAOEs, and TBABRs for normal listener at 1 and 4 kHz tone-burst stimuli. Several different fully objective procedures for estimating loudness from TBABR were developed along with two noise control methods that utilized the *WNS Fmp* (a weighted polynomial fitting and a parametric fitting based on the INEX mode). The full sync procedure in general yielded the best estimates for loudness growth of the evoked potential responses, performing around the same range as the TBOAE estimate for 1 kHz tone-burst stimulus and as best of the physiological estimators for the 4 kHz tone-burst stimuli. Most of the evoked response procedures, benefited from the noise control methods, with the inex parametric fitting giving the best overall MSE results. However, further work is still needed to determine when it is appropriate to use the INEX parametric modeling for a particular listener. While the weighted polynomial noise control method did not perform as good as the INEX method, it nevertheless provide a close performance without the need to make specific assumptions about the form of the loudness growth function.

The fourth and final study looked the loudness growth estimation with HILs. An analysis of the perceptual range of the all listeners (including normals from the third study) as a function of their threshold showed that there could be large variability in perceptual range between two listeners with the same threshold (Figure 44). This supports the current knowledge of two different extremes of loudness growth models

(softness imperception and recruitment) and in qualitative agreement with the observation that the standard deviation of estimated slopes of the loudness growth function for a group of people increase as function of the average hearing threshold of the group (Hellman and Meiselman 1992) . From a more practical point of view, this variability underscores the general idea that simply knowing a listener's threshold is not sufficient to determine the person's perception of loudness as a function of stimulus intensity. Procedures that use categorical scaling of loudness, which typically use only seven unique categories, and measurements of discomfort level, might mask this perceptual range deficit because a perceptual range scale is already being imposed on the listener in addition to being poorer estimates of loudness growth (Elberling 1999). In general, the loudness growth estimation results are in similar agreement with the results obtained from normal listeners. Overall, the full sync procedure was the best estimate from the evoked response, and its performance was also superior to that of the TBOAE response under all conditions (i.e., Figure 46). It is possible that part of decrease in performance of the TBOAE procedure could be due to the parameter selection based only on normal young adults, and/or on the types of HILs tested. A parameter selection controlling for age might have improved the TBOAE results with the HIL, since on average they constituted an older group (TBAOEs are known to yield different responses according to age, for review see Hall 2006). The performance of the full sync procedure was better for the HIL group than for the normals listeners. It is not clear why this is so, particularly because an analysis of the WNS Fmp (Figure 54) does not seem to indicate any specific patterns across the two groups (this could be a result of using fewer data points with the HILs). The WNS Fmp analysis does show, as expected, a mild correlation between

residual noise levels and the performance of the estimators. This corroborates the importance of controlling for residual noise levels when attempting to compare the amplitude of evoked responses across levels. Perhaps part of the controversy in the scientific literature regarding the relationship between ABR amplitude and loudness (e.g., see Table I) can be explained in terms of the relatively few number of trials collected on those studies that no relationship was found and lack of adequate assessment and reporting of the confounding residual noise levels in all studies. In fact, even submitting all listeners to the exact same number of trials is not sufficient to guarantee equal residual noise level across subjects or experiments with the same number of trials

A preliminary attempt has been made at understanding the relationship between the performance of the evoked response loudness growth estimator and the estimated averaged evoked response residual noise levels. Using data from our best listener, a liberal strategy suggests that WNS Fmp values around 0.0002 generates appropriate conditions for loudness growth estimation from the auditory evoked responses tested. On the other hand, a conservative strategy based on data from all the listeners, suggest an average lower bound of the WNS Fmp between 0.0001 and 0.00015. The data does not suggest, however, that a low average WNS Fmp is a necessary condition for accurate loudness growth estimation, since on a very small number of occasions good performance was obtained at relatively high WNS Fmp values (which in a way may help explain why some of the studies on Table I found a relationship with ABR and loudness using so few trials).

VII. Acknowledgements

The author would like to thank his thesis advisor, Dr. Michael Epstein, for the valuable support and comments throughout the work. The author is also grateful for Dr. Jeremy Marozeau for contributing with the initial development of the software for the recording of the evoked responses, and Dr. Ying-Yee Kong's help in calibration of the apparatus. Finally, the investigation of the SNR estimation procedure (Experiment II) would not have been possible without the generosity of Gentiletti-Faenze and Yanez-Suarez in providing us with their own data.

VIII. References

Babkoff, H., H. Pratt, et al. (1984). "Auditory brainstem evoked potential latency-intensity functions: a corrective algorithm." Hearing Research 16(3): 243-249.

Baker, D. E. (1997). "Noise: The Invisible Hazard." from http://extension.missouri.edu/explore/agguides/agengin/g01962.htm.

Bauer, J. W., R. O. Elmasian, et al. (1975). "Loudness enhancement in man. I. Brainstem-evoked response correlates." J. Acoust. Soc. Am. 57(1): 165-171.

Brass, D. and D. T. Kemp (1991). "Time-domain observation of otoacoustic emissions during constant tone stimulation." J. Acoust. Soc. Am. 90(5): 2415-27.

Burkard, R. F., M. Don, et al. (2006). Auditory Evoked Potentials: Basic Principles and Clinical Application, Lippincott Williams & Wilkins.

Buus, S. and M. Florentine (2001). "Growth of loudness in listeners with cochlear hearing losses: Recruitment reconsidered." J. Assoc. Res. Otolaryngol. 3(2): 120-139.

Buus, S. and M. Florentine (2001). Modifications to the power function for loudness. Fechner Day 2001, Pabst, Berlin.

Buus, S., M. Florentine, et al. (1998). Loudness function for tones at low levels derived from loudness summation. Psychophysical and Physiological Advances in Hearing. A. R. Palmer, A. Rees, A. Q. Summerfield and R. Meddis. London, Whurr: 449-457.

Buus, S., M. Florentine, et al. (1998). "On loudness at threshold." J. Acoust. Soc. Am. 104: 399-410.

Buus, S., M. Florentine, et al. (1997). "Temporal integration of loudness, loudness discrimination, and the form of the loudness function." J. Acoust. Soc. Am. 101: 669-680.

Chan, F. H. Y., F. K. Lam, et al. (1995). "Detection of brainstem auditory evoked potential by adaptive filtering." Medical and Biological Engineering and Computing 33(1): 69-75.

Coates, A. C. and J. L. Martin (1977). "Human auditory nerve action potentials and brainstem evoked responses." Archives of Otolaryngology 103: 605-622.

Collins, A. A. and G. A. Gescheider (1989). "The measurement of loudness in individual children and adults by absolute magnitude estimation and cross-modality matching." J. Acoust. Soc. Am. 85(5): 2012-2021.

Darling, R. M. and L. Price (1990). "Loudness and auditory brain stem evoked response." Ear and Hearing 11(4): 289-295.

Davidson, S. A., L. G. Wall, et al. (1990). "Preliminary studies on the use of an ABR amplitude projection procedure for hearing aid selection." Ear and Hearing 11(5): 332-339.

Davila, C. E. and M. S. Mobin (1992). "Weighted averaging of evoked potentials." IEEE Transactions on Biomedical Engineering 39(4): 338-345.

Dillon, H. (2001). Hearing Aids. New York, Thieme.

Don, M. and C. Elberling (1994). "Evaluating residual background noise in human auditory brain-stem responses." Journal of the Acoustical Society of America 96(5): 2746-2757.

Don, M. and C. Elberling (1996). "Use of quantitative measures of auditory brain-stem response peak amplitude and residual background noise in the decision to stop averaging." Journal of the Acoustical Society of America 99(1): 491-499.

Don, M., C. Elberling, et al. (1984). "Objective detection of averaged brainstem responses." Scandinavian Audiology 13: 219-228.

Don, M., C. W. Ponton, et al. (1993). "Gender differences in cochlear response time: An explanation for gender amplitude differences in the unmasked auditory brain-stem response." J. Acoust. Soc. Am. 93: 940-951.

Don, M., C. W. Ponton, et al. (1994). "Auditory brainstem response (ABR) peak amplitude variability reflects individual differences in cochlear response time." J. Acoust. Soc. Am. 96(6): 3476-3491.

Elberling, C. (1999). "Loudness scaling revisited." J Am Acad Audiol 10(5): 248-60.

Elberling, C. and M. Don (1987). "Threshold charactersitics of the human auditory brain stem response." Journal of the Acoustical Society of America 81(1): 115-121.

Elberling, C. and M. Don (1984). "Quality estimation of averaged auditory brainstem responses." Scandinavian Audiology 13: 187-197.

Elberling, C. and M. Don (1987). "Detection function for the human auditory brainstem response." Scandinavian Audiology 16: 89-92.

Elberling, C. and O. Wahlgreen (1985). "Estimation of auditory brainstem response, ABR, by means of bayesian inference." Scandinavian Audiology 14: 89-96.

Epstein, M., S. Buus, et al. (2004). "The effects of window delay, delinearization, and frequency on tone-burst otoacoustic emission input/output measurements." J. Acoust. Soc. Am. 116(2): 1160-7.

ANSI (1996). American National Standard Specification for Audiometers. New York, American National Standards Institute. ANSI S3.6-1996.

Babkoff, H., H. Pratt, et al. (1984). "Auditory brainstem evoked potential latency-intensity functions: a corrective algorithm." Hearing Research 16(3): 243-249.

Baker, D. E. (1997). "Noise: The Invisible Hazard." from http://extension.missouri.edu/explore/agguides/agengin/g01962.htm.

Bauer, J. W., R. O. Elmasian, et al. (1975). "Loudness enhancement in man. I. Brainstem-evoked response correlates." J. Acoust. Soc. Am. 57(1): 165-171.

Brass, D. and D. T. Kemp (1991). "Time-domain observation of otoacoustic emissions during constant tone stimulation." J. Acoust. Soc. Am. 90(5): 2415-27.

Burkard, R. F., M. Don, et al. (2006). Auditory Evoked Potentials: Basic Principles and Clinical Application, Lippincott Williams & Wilkins.

Buus, S. and M. Florentine (2001). "Growth of loudness in listeners with cochlear hearing losses: Recruitment reconsidered." J. Assoc. Res. Otolaryngol. 3(2): 120-139.

Buus, S. and M. Florentine (2001). Modifications to the power function for loudness. Fechner Day 2001, Pabst, Berlin.

Buus, S., M. Florentine, et al. (1998). Loudness function for tones at low levels derived from loudness summation. Psychophysical and Physiological Advances in Hearing. A. R. Palmer, A. Rees, A. Q. Summerfield and R. Meddis. London, Whurr: 449-457.

Buus, S., M. Florentine, et al. (1998). "On loudness at threshold." J. Acoust. Soc. Am. 104: 399-410.

Buus, S., M. Florentine, et al. (1997). "Temporal integration of loudness, loudness discrimination, and the form of the loudness function." J. Acoust. Soc. Am. 101: 669-680.

Chan, F. H. Y., F. K. Lam, et al. (1995). "Detection of brainstem auditory evoked potential by adaptive filtering." Medical and Biological Engineering and Computing 33(1): 69-75.

Coates, A. C. and J. L. Martin (1977). "Human auditory nerve action potentials and brainstem evoked responses." Archives of Otolaryngology 103: 605-622.

Collins, A. A. and G. A. Gescheider (1989). "The measurement of loudness in individual children and adults by absolute magnitude estimation and cross-modality matching." J. Acoust. Soc. Am. 85(5): 2012-2021.

Darling, R. M. and L. Price (1990). "Loudness and auditory brain stem evoked response." Ear and Hearing 11(4): 289-295.

Davidson, S. A., L. G. Wall, et al. (1990). "Preliminary studies on the use of an ABR amplitude projection procedure for hearing aid selection." Ear and Hearing 11(5): 332-339.

Davila, C. E. and M. S. Mobin (1992). "Weighted averaging of evoked potentials." IEEE Transactions on Biomedical Engineering 39(4): 338-345.

Dillon, H. (2001). Hearing Aids. New York, Thieme.

Don, M. and C. Elberling (1994). "Evaluating residual background noise in human auditory brain-stem responses." Journal of the Acoustical Society of America 96(5): 2746-2757.

Don, M. and C. Elberling (1996). "Use of quantitative measures of auditory brain-stem response peak amplitude and residual background noise in the decision to stop averaging." Journal of the Acoustical Society of America 99(1): 491-499.

Don, M., C. Elberling, et al. (1984). "Objective detection of averaged brainstem responses." Scandinavian Audiology 13: 219-228.

Don, M., C. W. Ponton, et al. (1993). "Gender differences in cochlear response time: An explanation for gender amplitude differences in the unmasked auditory brain-stem response." J. Acoust. Soc. Am. 93: 940-951.

Don, M., C. W. Ponton, et al. (1994). "Auditory brainstem response (ABR) peak amplitude variability reflects individual differences in cochlear response time." J. Acoust. Soc. Am. 96(6): 3476-3491.

Elberling, C. (1999). "Loudness scaling revisited." J Am Acad Audiol 10(5): 248-60.

Elberling, C. and M. Don (1984). "Quality estimation of averaged auditory brainstem responses." Scandinavian Audiology 13: 187-197.

Elberling, C. and M. Don (1987). "Detection function for the humand auditory brainstem response." Scandinavian Audiology 16: 89-92.

Elberling, C. and M. Don (1987). "Threshold charactersitics of the human auditory brain stem response." Journal of the Acoustical Society of America 81(1): 115-121.

Elberling, C. and O. Wahlgreen (1985). "Estimation of auditory brainstem response, ABR, by means of Bayesian inference." Scand. Audiol. 14(2): 89-96.

Elberling, C. and O. Wahlgreen (1985). "Estimation of auditory brainstem response, ABR, by means of bayesian inference." Scandinavian Audiology 14: 89-96.

Epstein, M., S. Buus, et al. (2004). "The effects of window delay, delinearization, and frequency on tone-burst otoacoustic emission input/output measurements." J. Acoust. Soc. Am. 116(2): 1160-7.

Epstein, M. and M. Florentine (2005). "Inferring basilar-membrane motion from tone-burst otoacoustic emissions and psychoacoustic measurements." J. Acoust. Soc. Am. 117(1): 263-74.

Epstein, M. and M. Florentine (2005). "A test of the Equal-Loudness-Ratio hypothesis using cross-modality matching functions." J. Acoust. Soc. Am. 118(2): 907 - 913.

Epstein, M. and M. Florentine (2006). "Loudness of brief tones measured by magnitude estimation and loudness matching." J. Acoust. Soc. Am. 119(4): 1943-1945.

Epstein, M., J. Marozeau, et al. (2006). Basilar-membrane activity and loudness. Fechner Day.

Florentine, M. (2004). "Softness imperception- Defining a puzzling problem." Hearing Health 20(1): 31-34.

Florentine, M., S. Buus, et al. (1996). "Temporal integration of loudness as a function of level." J. Acoust. Soc. Am. 99: 1633-1644.

Florentine, M., S. Buus, et al. (1998). "Temporal integration of loudness under partial masking." J. Acoust. Soc. Am. 104: 999-1007.

Florentine, M. and M. Epstein (2006). To honor Stevens and repeal his law (for the auditory system). Fechner Day, St. Albans, England.

Florentine, M., M. Epstein, et al. (2001). Loudness functions for long and short tones. Fechner Day 2001. E. Sommerfeld, R. Kompass and T. Lachmann. Berlin, Pabst: 361-366.

Fowler, E. P. (1928). "Marked deafened areas in normal ears." Arch. Otolaryngol. 8: 151-155.

Gallego, S., S. Garnier, et al. (1999). "Loudness Growth Functions and EABR Characteristics in Digisonic Cochlear Implantees." Acta Otorolaryngol 119: 234-238.

Gentiletti-Faenze, G. G., O. Yanez-Suarez, et al. (2003). Evaluation of automatic identification algorithm for auditory brainstem response used in universal hearing loss screening. Proceedings of the 25th Annual International Conference of the IEEE EMBS, Cancun, Mexico.

Gold, T. (1948). "Hearing. II. The physical basis of the action of the cochlea." Proc. Roy. Soc. B135: 492-498.

Gorga, M. P., J. K. Reiland, et al. (1985). "Auditory brainstem responses in a case of high-frequency conductive hearing loss." Journal of Speech and Hearing Disorders 50: 346-350.

Hall, J. W. (2006). New Handbook for Auditory Evoked Responses Allyn & Bacon.

Hall, J. W. (2007). Handbook of Auditory Evoked Responses. Boston, MA, Allyn & Bacon.

Harris, F. J. (1978). "On the use of Windows for Harmonic Analysis with the Discrete Fourier Transform." Proceedings of the IEEE 61: 51-84.

Harris, F. J. (1978). "On the Use of Windows for Harmonic Analysis with the Discrete Fourier Transform." Proceedings of the IEEE 66(1): 51-84.

Hecox, K. E. (1983). "Role of auditory brain stem response in the selection of hearing aids." Ear and Hearing 4(1): 51-55.

Hellman, R. P. (1976). "Growth of loudness at 1000 and 3000 Hz." J. Acoust. Soc. Am. 60: 672-679.

Hellman, R. P. (1991). Loudness scaling by magnitude scaling: Implications for intensity coding. Ratio Scaling of Psychological Magnitude: In Honor of the Memory of S. S. Stevens. G. A. Gescheider and S. J. Bolanowski. Hillsdale, NJ, Erlbaum.

Hellman, R. P. and C. H. Meiselman (1988). "Prediction of individual loudness exponents from cross-modality matching." Journal of Speech and Hearing Research 31: 605-615.

Hellman, R. P. and C. H. Meiselman (1988). "Prediction of individual loudness exponents from cross-modality matching." J. Speech Hear. Res. 31: 605-615.

Hellman, R. P. and C. H. Meiselman (1990). "Loudness relations for individuals and groups in normal and impaired hearing." J. Acoust. Soc. Am. 88(6): 2596-2606.

Hellman, R. P. and C. H. Meiselman (1992). "Rate of loudness growth for pure tones in normal and impaired hearing." J. Acoust. Soc. Am. 93(2): 966-975.

Hellman, R. P. and C. H. Meiselman (1993). "Rate of loudness growth for pure tones in normal and impaired hearing." J. Acoust. Soc. Am. 93(2): 966-975.

Hellman, R. P. and J. J. Swislocki (1961). "Some factors affecting the estimation of loudness." J. Acoust. Soc. Am. 33: 687-694.

Hellman, R. P. and J. J. Zwislocki (1961). "Some factors affecting the estimation of loudness." J. Acoust. Soc. Am. 33: 687-694.

Hellman, R. P. and J. J. Zwislocki (1963). "Monaural loudness function at 1000 cps and interaural summation." J. Acoust. Soc. Am. 35: 856-865.

Hoke, M., B. Ross, et al. (1984). "Weighted averaging- theory and application to electric response audiometry." Electroencephalography and Clinical Neurophysiology 57: 484-489.

Hoth, S. and F. N. Weber (2001). "The latency of evoked otoacoustic emissions: its relation to hearing loss and auditory evoked potentials." Scand. Audiol. 30(3): 173-83.

Howe, S. W. and T. N. Decker (1984). "Monaural and binaural auditory brainstem responses in relationto the psychophysical loudness growth function." J. Acoust. Soc. Am. 76(3): 787-793.

Jacquin, A., E. Causevic, et al. (2006). Optimal denoising of brainstem auditory evoked response (BAER) for automatic peak identification and brainstem assessment. Proceedings of the 28th IEEE EMBS Annual International Conference, New York City, USA.

Jewett, D. L. (1970). "Volume-conducted potentials in response to auditory stimuli as detected by averaging in the cat." Electroencephalogr Clin Neurophysiol 28: 609-618.

Jewett, D. L., M. N. Romano, et al. (1970). "Human audiotry evoked potentials: Possible brainstem components detected on the scalp." Science 167: 1517-1518.

Johannesen, P. T. and E. A. Lopez-Poveda (2008). "Cochlear nonlinearity in normal-hearing subjects as inferred psychophysically and from distortion-product otoacoustic emissions." J. Acoust. Soc. Am. 124(4): 2149-2163.

Keidel, W. D. and W. D. Neff (1974). "Handbook of Sensory Physiology, Springer-Verlag, New York.

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

Kennedy, C. R., L. Kimm, et al. (1991). "Otoacoustic emissions and auditory brainstem responses in newborn." Archives of Disease in Childhood 66: 1124-1129.

Kiang, N. Y.-S. (1975). Stimulus representation in the discharge patterns of auditory neurons. The Nervous System. Human Communication and Its Disorders. D. B. Tower. New York, Raven Press. 3: 81-96.

Kiessling, J. (1982). "Hearing aid selection by brainstem audiometry." Scand Audiol 11: 269-275.

Kiessling, J. (1983). "Clinical experience in hearing aid adjustment by means of BER amplitudes." Archives of Otorhinolaryngology 238: 233-240.

Kiessling, J., O. Dyrlund, et al. (1995). Loudness scaling - towards a generally accepted clinical method. European Conference on Audiology, Noordwijkerhout, The Netherlands.

Kileny, P. (1982). "Auditory brainstem response as indicators of hearing aid performance." Ann. Otol. Rhinol. Laryngol. 91: 61-64.

Levitt, H. (1971). "Transformed up-down methods in psychoacoustics." J. Acoust. Soc. Am. 49: 467-477.

Lutkenhoner, B., M. Hoke, et al. (1985). "Possibilities and limitations of weighted averaging." Biological Cybernetics 52: 409-416.

Madell, J. R. and R. Goldstein (1972). "Relation between loudness and the amplitude of the early components of the averaged electroencephalic response." Journal of Speech and Hearing Research 15: 134-141.

Mahoney, T. M. (1985). Auditory brainstem response hearing aid applications. The Auditory Brainstem Response. J. T. Jacobson. San Diego, CA, College-Hill Press: 349-370.

Marozeau, J. and M. Florentine (2007). "Loudness growth in individual listeners with hearing losses: A review." J. Acoust. Soc. Am. 122(3): EL81- EL87.

Marozeau, J. and M. Florentine ((In Press)). "Loudness growth in individual listeners with hearing losses: A review." J. Acoust. Soc. Am.

McEwen, J. A. and G. B. Anderson (1975). "Modeling the stationarity and gaussianity of spontaneous electroencephalographic activity." IEEE Transactions on Biomedical Engineering 22(5): 361-369.

McFadden, D. (1975). "Duration-intensity reciprocity for equal loudness." J. Acoust. Soc. Am. 57: 702-704.

MEdIC. from http://medic.med.uth.tmc.edu/Lecture/Main/flap2.gif.

Moller, K. and B. Blegvad (1976). "Brainstem potentials in subjects with sensorineural hearing loss." Scand Audiol 5: 115-127.

Moore, B. C. J. (2003). An Introduction to the Psychology of Hearing. San Diego, Academic.

Moore, B. C. J. and B. R. Glasberg (2004). "A revised model of loudness perception applied to cochlear hearing loss." Hearing Research 188: 70-88.

Musiek, F. E. (1991). Auditory evoked responses in site-of-lesion assessment. Hearing assessment. W. F. Rintelmann. Austin, TX, Pro-ed: 383-428.

Musiek, F. E. and J. A. Baran (2007). The Auditory System: Anatomy, Physiology, and Clinical Correlates, Pearson Educatoin Inc.

Norton, S. J. and S. T. Neely (1987). "Tone-burst-evoked otoacoustic emissions from normal-hearing subjects." J. Acoust. Soc. Am. 81(6): 1860-72.

Orfanidis, S. J. (1996). Optimum Signal Processing. An Introduction Englewood Cliffs- NJ, Prentice-Hall.

Oxenham, A. J. and C. J. Plack (1997). "A behavioral measure of basilar-membrane non-linearity in listeners with normal and impaired hearing." J. Acoust. Soc. Am. 101(6): 3666-3675.

Perlman, H. B. and T. J. Case (1941). "Electrical phenomena of the cochlea in man." Arch. Otolaryngol. 34: 710-718.

Poulton, E. C. (1989). Bias in Quantifying Judgments. Hillsdale, NJ, Erlbaum.

Pratt, H. and H. Sohmer (1977). "Correlations between psychophysical magnitude estimates and simultaneously obtained auditory nerve, brain stem and cortical responses to click stimuli in man." Electroencephalogr Clin Neurophysiol 43: 802-812.

Qiu, W., F. H. Y. Chan, et al. (1998). An adaptive approach for processing evoked potentials from the auditory cortex of man. Proceedings of the 20th Annual International Conference of the IEEE Engineering in Medicine and Biology Society.

R131-, I. (1959). Expression of the physcial and subjective magnitudes of sound [ISO/R-131-1959(E)]. Geneva: International Organization for Standardization.

Ravazzani, P. and F. Grandori (1993). "Evoked Otoacoustic Emissions: Nonlinearities and Response Interpretation." IEEE Transactions on Biomedical Engineering 40(5): 500-504.

Ravazzani, P. and F. Grandori (1993). "Evoked Otoacoustic Emissions: Nonlinearities and Response Interpretation." IEEE Transactions on Biomedical Engineering 40(5).

Ravazzani, P., G. Tognola, et al. (1996). "Derived 'nonlinear' versus 'linear' click-evoked otoacoustic emissions." Audiology 35: 73-86.

Robinson, D. W. (1957). "The subjective loudness scale." Acustica 7: 217-233.

Ruggero, M. A. and N. C. Rich (1991). "Furosemide alters organ of Corti mechanics: evidence for feedback of outer hair cells upon the basilar membrane." J. Neurosci. 11: 1057-1067.

Ruggero, M. A., N. C. Rich, et al. (1997). "Basilar-membrane responses to tones at the base of the chinchilla cochlea." J. Acoust. Soc. Am. 101: 2151-2163.

Ruggero, M. A., L. Robles, et al. (1992). "Basilar membrane responses to two-tone and broadband stimuli." Philos. Trans. R. Soc. London 336: 307-3115.

Schairer, K. S., J. C. Ellison, et al. (2006). "Use of stimulus-frequency otoacoustic emission latency and level to investigate cochlear mechanics in human ears." J. Acoust. Soc. Am. 120(2): 901-14.

Schairer, K. S., J. Messersmith, et al. (2008). "Use of psychometric-function slopes for forward-masked tones to investigate cochlear nonlinearity." J. Acoust. Soc. Am. 124(4): 2196-2215.

Scharf, B. (1978). Loudness. Handbook of Perception New York, Academic Press. IV: Hearing: 187-242.

Scharf, L. (1991). Statistical Signal Processing: detection, estimation, and time series analysis. Reading, Addison- Wesley Publishing Company Inc.

Schlauch, R. S., J. J. DiGiovanni, et al. (1998). "Basilar membrane nonlinearity and loudness." J. Acoust. Soc. Am. 103: 2010-2020.

Serpanos, Y. C., H. O'Malley, et al. (1997). "The relationship between loudness intensity functions and the click-ABR wave V latency." Ear and Hearing 18: 409-419.

Serpanos, Y. C., H. O'Malley, et al. (1998). "Cross-modality matching and the loudness growth function for click stimuli." J. Acoust. Soc. Am. 103(2): 1022-1032.

Shera, C. A., J. J. Guinan, Jr., et al. (2002). "Revised estimates of human cochlear tuning from otoacoustic and behavioral measurements." Proc Natl Acad Sci U S A 99(5): 3318-23.

Sininger, Y. S. (1995). "Filtering and spectral characteristics of averaged auditory brain-stem response and background noise in infants." Journal of the Acoustical Society of America 98(4): 2048-2055.

Sisto, R. and A. Moleti (2007). "Transient evoked otoacoustic emission latency and cochlear tuning at different stimulus levels." J. Acoust. Soc. Am. 122(4): 2183-90.

Sörnmo, L. and P. Laguna (2005). Bioelectrical Signal Processing in Cardiac and Neurological Applications Academic Press.

Steinberg, J. C. and M. B. Gardner (1937). "The dependence of hearing impairment on sound intensity." J. Acoust. Soc. Am. 9: 11-23.

Stevens, J. C. and M. Guirao (1964). "Individual loudness functions." J. Acoust. Soc. Am. 36: 2210-2213.

Stevens, S. S. (1936). "A scale for the measurement of a psychological magnitude: loudness." Psych. Rev. 43: 405-416.

Stevens, S. S. (1955). "The measurement of loudness." J. Acoust. Soc. Am. 25: 815-829.

Stevens, S. S. (1955). "The measurement of loudness." J. Acoust. Soc. Am. 27: 815-827.

Stevens, S. S. (1957). "Concerning the form of the loudness function." J. Acoust. Soc. Am. 29: 603-606.

Stevens, S. S. (1961). "To honor Fechner and repeal his law." Science 13: 80-86.

Stevens, S. S. (1966). "Brightness and loudness as a function of stimulus duration." Percept. Psychophys. 1(319-327).

Stevens, S. S. (1972). "Perceived level of noise by Mark VII and decibels (E)." J. Acoust. Soc. Am. 51: 575-601.

Suter, C. M. and C. C. Brewer (1983). "Electrophysiologic techniques in audiology and otology." Ear and Hearing 4(4): 212-219.

Thornton, R. D., G. Farrell, et al. (1989). "Clinical methods for the objective estimation of loudness discomfort level (LDL) using auditory brainstem response in patients." Scand Audiol 18: 225-230.

Thornton, R. D., L. Yardley, et al. (1987). "The objective estimation of loudness discomfort level using auditory brainstem evoked responses." Scand Audiol 16: 219-225.

Whilby, S., M. Florentine, et al. (2006). "Monaural and binaural loudness of 5- and 200-ms tones in normal and impaired hearing." J. Acoust. Soc. Am. 119(6): 3931-9.

Whitehead, M. L., B. B. Stagner, et al. (1994). "Measurement of Otoacoustic Emissions For Hearing Assessment." IEEE Engineering in Medicine and Biology 13(2): 210-226.

Whitehead, M. L., B. B. Stagner, et al. (1994). "Measurement of Otoacoustic Emissions For Hearing Assessment." IEEE Engineering in Medicine and Biology(13): 210-226.

Yates, G. K. (1990). "Basilar membrane nonlinearity and its influence on auditory nerve rate-intensity functions." Hearing Research 50: 145-162.

Yates, G. K., I. M. Winter, et al. (1990). "Basilar membrane nonlinearity determines auditory nerver rate-intensity functions and cochlear dynamic range." Hearing Research 45(3): 203-219.

Zhang, M. and J. J. Zwislocki (1995). "OHC response recruitment and its correlation with loudness recruitment." Hearing Research 85(1-2): 1-10.

Zweig, G. and C. A. Shera (1995). "The origin of periodicity in the spectrum of evoked otoacoustic emissions." J. Acoust. Soc. Am. 98(4): 2018-47.

Zwicker, E. and H. Fastl (1999). Psychoacoustics: facts and models. New York, Springer.

Zwislocki, J. J. (1965). Analysis of some auditory characteristics. Handbook of Mathematical Psychology. R. D. Luce, R. R. Bush and E. Galanter. New York, Wiley.

Epstein, M. and M. Florentine (2005). "A test of the Equal-Loudness-Ratio hypothesis using cross-modality matching functions." J. Acoust. Soc. Am. 118(2): 907 - 913.

Epstein, M., J. Marozeau, et al. (2006). Basilar-membrane activity and loudness. Fechner Day.

Florentine, M. (2004). "Softness imperception- Defining a puzzling problem." Hearing Health 20(1): 31-34.

Florentine, M., S. Buus, et al. (1998). "Temporal integration of loudness under partial masking." J. Acoust. Soc. Am. 104: 999-1007.

Florentine, M., S. Buus, et al. (1996). "Temporal integration of loudness as a function of level." J. Acoust. Soc. Am. 99: 1633-1644.

Florentine, M., M. Epstein, et al. (2001). Loudness functions for long and short tones. Fechner Day 2001. E. Sommerfeld, R. Kompass and T. Lachmann. Berlin, Pabst: 361-366.

Florentine, M. and M. Epstein (2006). To honor Stevens and repeal his law (for the auditory system). Fechner Day, St. Albans, England.

Fowler, E. P. (1928). "Marked deafened areas in normal ears." Arch. Otolaryngol. 8: 151-155.

Gallego, S., S. Garnier, et al. (1999). "Loudness Growth Functions and EABR Characteristics in Digisonic Cochlear Implantees." Acta Otorolaryngol 119: 234-238.

Gentiletti-Faenze, G. G., O. Yanez-Suarez, et al. (2003). Evaluation of automatic identification algorithm for auditory brainstem response used in universal hearing loss screening. Proceedings of the 25th Annual International Conference of the IEEE EMBS, Cancun, Mexico.

Gold, T. (1948). "Hearing. II. The physical basis of the action of the cochlea." Proc. Roy. Soc. B135: 492-498.

Gorga, M. P., J. K. Reiland, et al. (1985). "Auditory brainstem responses in a case of high-frequency conductive hearing loss." Journal of Speech and Hearing Disorders 50: 346-350.

Hall, J. W. (2006). New Handbook for Auditory Evoked Responses Allyn & Bacon.

Harris, F. J. (1978). "On the Use of Windows for Harmonic Analysis with the Discrete Fourier Transform." Proceedings of the IEEE 66(1): 51-84.

Hecox, K. E. (1983). "Role of auditory brain stem response in the selection of hearing aids." Ear and Hearing 4(1): 51-55.

Hellman, R. P. and C. H. Meiselman (1988). "Prediction of individual loudness exponents from cross-modality matching." Journal of Speech and Hearing Research 31: 605-615.

Hellman, R. P. and C. H. Meiselman (1990). "Loudness relations for individuals and groups in normal and impaired hearing." J. Acoust. Soc. Am. 88(6): 2596-2606.

Hellman, R. P. and C. H. Meiselman (1992). "Rate of loudness growth for pure tones in normal and impaired hearing." J. Acoust. Soc. Am. 93(2): 966-975.

Hellman, R. P. and J. J. Zwislocki (1961). "Some factors affecting the estimation of loudness." J. Acoust. Soc. Am. 33: 687-694.

Hellman, R. P. (1976). "Growth of loudness at 1000 and 3000 Hz." J. Acoust. Soc. Am. 60: 672-679.

Hellman, R. P. (1991). Loudness scaling by magnitude scaling: Implications for intensity coding. Ratio Scaling of Psychological Magnitude: In Honor of the Memory of S. S. Stevens. G. A. Gescheider and S. J. Bolanowski. Hillsdale, NJ, Erlbaum.

Hellman, R. P. and J. J. Zwislocki (1963). "Monaural loudness function at 1000 cps and interaural summation." J. Acoust. Soc. Am. 35: 856-865.
Hellman, R. P. and C. H. Meiselman (1988). "Prediction of individual loudness exponents from cross-modality matching." J. Speech Hear. Res. 31: 605-615.

Hoke, M., B. Ross, et al. (1984). "Weighted averaging- theory and application to electric response audiometry." Electroencephalography and Clinical Neurophysiology 57: 484-489.

Hoth, S. and F. N. Weber (2001). "The latency of evoked otoacoustic emissions: its relation to hearing loss and auditory evoked potentials." Scand. Audiol. 30(3): 173-83.

Howe, S. W. and T. N. Decker (1984). "Monaural and binaural auditory brainstem responses in relationto the psychophysical loudness growth function." J. Acoust. Soc. Am. 76(3): 787-793.

Qiu, W., F. H. Y. Chan, et al. (1998). An adaptive approach for processing evoked potentials from the auditory cortex of man. Proceedings of the 20th Annual International Conference of the IEEE Engineering in Medicine and Biology Society.

Jacquin, A., E. Causevic, et al. (2006). Optimal denoising of brainstem auditory evoked response (BAER) for automatic peak identification and brainstem assessment. Proceedings of the 28th IEEE EMBS Annual International Conference, New York City, USA.

Jewett, D. L. (1970). "Volume-conducted potentials in response to auditory stimuli as detected by averaging in the cat." Electroencephalogr Clin Neurophysiol 28: 609-618.

Jewett, D. L., M. N. Romano, et al. (1970). "Human audiotry evoked potentials: Possible brainstem components detected on the scalp." Science 167: 1517-1518.

Johannesen, P. T. and E. A. Lopez-Poveda (2008). "Cochlear nonlinearity in normal-hearing subjects as inferred psychophysically and from distortion-product otoacoustic emissions." J. Acoust. Soc. Am. 124(4): 2149-2163.

Keidel, W. D. and W. D. Neff (1974). "Handbook of Sensory Physiology, Springer-Verlag, New York.

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

Kennedy, C. R., L. Kimm, et al. (1991). "Otoacoustic emissions and auditory brainstem responses in newborn." Archives of Disease in Childhood 66: 1124-1129.

Kiang, N. Y.-S. (1975). Stimulus representation in the discharge patterns of auditory neurons. The Nervous System. Human Communication and Its Disorders. D. B. Tower. New York, Raven Press. 3: 81-96.

Kiessling, J. (1982). "Hearing aid selection by brainstem audiometry." Scand Audiol 11: 269-275.

Kiessling, J. (1983). "Clinical experience in hearing aid adjustment by means of BER amplitudes." Archives of Otorhinolaryngology 238: 233-240.

Kiessling, J., O. Dyrlund, et al. (1995). Loudness scaling - towards a generally accepted clinical method. European Conference on Audiology, Noordwijkerhout, The Netherlands.

Kileny, P. (1982). "Auditory brainstem response as indicators of hearing aid performance." Ann. Otol. Rhinol. Laryngol. 91: 61-64.

Levitt, H. (1971). "Transformed up-down methods in psychoacoustics." J. Acoust. Soc. Am. 49: 467-477.

Lutkenhoner, B., M. Hoke, et al. (1985). "Possibilities and limitations of weighted averaging." Biological Cybernetics 52: 409-416.

Madell, J. R. and R. Goldstein (1972). "Relation between loudness and the amplitude of the early components of the averaged electroencephalic response." Journal of Speech and Hearing Research 15: 134-141.

Mahoney, T. M. (1985). Auditory brainstem response hearing aid applications. The Auditory Brainstem Response. J. T. Jacobson. San Diego, CA, College-Hill Press: 349-370.

Marozeau, J. and M. Florentine (2007). "Loudness growth in individual listeners with hearing losses: A review." J. Acoust. Soc. Am. 122(3): EL81- EL87.

McEwen, J. A. and G. B. Anderson (1975). "Modeling the stationarity and gaussianity of spontaneous electroencephalographic activity." IEEE Transactions on Biomedical Engineering 22(5): 361-369.

McFadden, D. (1975). "Duration-intensity reciprocity for equal loudness." J. Acoust. Soc. Am. 57: 702-704.

MEdIC. from http://medic.med.uth.tmc.edu/Lecture/Main/flap2.gif.

Moller, K. and B. Blegvad (1976). "Brainstem potentials in subjects with sensorineural hearing loss." Scand Audiol 5: 115-127.

Moore, B. C. J. (2003). An Introduction to the Psychology of Hearing. San Diego, Academic.

Moore, B. C. J. and B. R. Glasberg (2004). "A revised model of loudness perception applied to cochlear hearing loss." Hearing Research 188: 70-88.

Musiek, F. E. (1991). Auditory evoked responses in site-of-lesion assessment. Hearing assessment. W. F. Rintelmann. Austin, TX, Pro-ed: 383-428.

Musiek, F. E. and J. A. Baran (2007). The Auditory System: Anatomy, Physiology, and Clinical Correlates, Pearson Educatoin Inc.

Norton, S. J. and S. T. Neely (1987). "Tone-burst-evoked otoacoustic emissions from normal-hearing subjects." J. Acoust. Soc. Am. 81(6): 1860-72.

Orfanidis, S. J. (1996). Optimum Signal Processing. An Introduction Englewood Cliffs- NJ, Prentice-Hall.

Oxenham, A. J. and C. J. Plack (1997). "A behavioral measure of basilar-membrane non-linearity in listeners with normal and impaired hearing." J. Acoust. Soc. Am. 101(6): 3666-3675.

Perlman, H. B. and T. J. Case (1941). "Electrical phenomena of the cochlea in man." Arch. Otolaryngol. 34: 710-718.

Poulton, E. C. (1989). Bias in Quantifying Judgments. Hillsdale, NJ, Erlbaum.

Pratt, H. and H. Sohmer (1977). "Correlations between psychophysical magnitude estimates and simultaneously obtained auditory nerve, brain stem and cortical responses to click stimuli in man." Electroencephalogr Clin Neurophysiol 43: 802-812.

Qiu, W., F. H. Y. Chan, et al. (1998). An adaptive approach for processing evoked potentials from the auditory cortex of man. Proceedings of the 20th Annual International Conference of the IEEE Engineering in Medicine and Biology Society.

R131-, I. (1959). Expression of the physcial and subjective magnitudes of sound [ISO/R-131-1959(E)]. Geneva: International Organization for Standardization.

Ravazzani, P. and F. Grandori (1993). "Evoked Otoacoustic Emissions: Nonlinearities and Response Interpretation." IEEE Transactions on Biomedical Engineering 40(5): 500-504.

Ravazzani, P., G. Tognola, et al. (1996). "Derived 'nonlinear' versus 'linear' click-evoked otoacoustic emissions." Audiology 35: 73-86.

Robinson, D. W. (1957). "The subjective loudness scale." Acustica 7: 217-233.

Ruggero, M. A. and N. C. Rich (1991). "Furosemide alters organ of Corti mechanics: evidence for feedback of outer hair cells upon the basilar membrane." J. Neurosci. 11: 1057-1067.

Ruggero, M. A., N. C. Rich, et al. (1997). "Basilar-membrane responses to tones at the base of the chinchilla cochlea." J. Acoust. Soc. Am. 101: 2151-2163.

Ruggero, M. A., L. Robles, et al. (1992). "Basilar membrane responses to two-tone and broadband stimuli." Philos. Trans. R. Soc. London 336: 307-3115.

Schairer, K. S., J. C. Ellison, et al. (2006). "Use of stimulus-frequency otoacoustic emission latency and level to investigate cochlear mechanics in human ears." J. Acoust. Soc. Am. 120(2): 901-14.

Schairer, K. S., J. Messersmith, et al. (2008). "Use of psychometric-function slopes for forward-masked tones to investigate cochlear nonlinearity." J. Acoust. Soc. Am. 124(4): 2196-2215.

Scharf, B. (1978). Loudness. Handbook of Perception New York, Academic Press. IV: Hearing: 187-242.

Scharf, L. (1991). Statistical Signal Processing: detection, estimation, and time series analysis. Reading, Addison- Wesley Publishing Company Inc.

Schlauch, R. S., J. J. DiGiovanni, et al. (1998). "Basilar membrane nonlinearity and loudness." J. Acoust. Soc. Am. 103: 2010-2020.

Serpanos, Y. C., H. O'Malley, et al. (1997). "The relationship between loudness intensity functions and the click-ABR wave V latency." Ear and Hearing 18: 409-419.

Serpanos, Y. C., H. O'Malley, et al. (1998). "Cross-modality matching and the loudness growth function for click stimuli." J. Acoust. Soc. Am. 103(2): 1022-1032.

Shera, C. A., J. J. Guinan, Jr., et al. (2002). "Revised estimates of human cochlear tuning from otoacoustic and behavioral measurements." Proc Natl Acad Sci U S A 99(5): 3318-23.

Sininger, Y. S. (1995). "Filtering and spectral characteristics of averaged auditory brain-stem response and background noise in infants." Journal of the Acoustical Society of America 98(4): 2048-2055.

Sisto, R. and A. Moleti (2007). "Transient evoked otoacoustic emission latency and cochlear tuning at different stimulus levels." J. Acoust. Soc. Am. 122(4): 2183-90.

Sörnmo, L. and P. Laguna (2005). Bioelectrical Signal Processing in Cardiac and Neurological Applications Academic Press.

Steinberg, J. C. and M. B. Gardner (1937). "The dependence of hearing impairment on sound intensity." J. Acoust. Soc. Am. 9: 11-23.

Stevens, S. S. (1936). "A scale for the measurement of a psychological magnitude: loudness." Psych. Rev. 43: 405-416.

Stevens, S. S. (1972). "Perceived level of noise by Mark VII and decibels (E)." J. Acoust. Soc. Am. 51: 575-601.

Stevens, S. S. (1955). "The measurement of loudness." J. Acoust. Soc. Am. 27: 815-827.

Stevens, S. S. (1957). "Concerning the form of the loudness function." J. Acoust. Soc. Am. 29: 603-606.

Stevens, S. S. (1961). "To honor Fechner and repeal his law." Science 13: 80-86.

Stevens, S. S. (1966). "Brightness and loudness as a function of stimulus duration." Percept. Psychophys. 1(319-327).

Stevens, J. C. and M. Guirao (1964). "Individual loudness functions." J. Acoust. Soc. Am. 36: 2210-2213.

Suter, C. M. and C. C. Brewer (1983). "Electrophysiologic techniques in audiology and tology." Ear and Hearing 4(4): 212-219.

Thornton, R. D., G. Farrell, et al. (1989). "Clinical methods for the objective estimation of loudness discomfort level (LDL) using auditory brainstem response in patients." Scand Audiol 18: 225-230.

Thornton, R. D., L. Yardley, et al. (1987). "The objective estimation of loudness discomfort level using auditory brainstem evoked responses." Scand Audiol 16: 219-225.

Whitehead, M. L., B. B. Stagner, et al. (1994). "Measurement of Otoacoustic Emissions For Hearing Assessment." IEEE Engineering in Medicine and Biology(13): 210-226.

Whilby, S., M. Florentine, et al. (2006). "Monaural and binaural loudness of 5- and 200-ms tones in normal and impaired hearing." J. Acoust. Soc. Am. 119(6): 3931-9.
Yates, G. K. (1990). "Basilar membrane nonlinearity and its influence on auditory nerve rate-intensity functions." Hearing Research 50: 145-162.

Yates, G. K., I. M. Winter, et al. (1990). "Basilar membrane nonlinearity determines auditory nerver rate-intensity functions and cochlear dynamic range." Hearing Research 45(3): 203-219.

Zhang, M. and J. J. Zwislocki (1995). "OHC response recruitment and its correlation with loudness recruitment." Hearing Research 85(1-2): 1-10.

Zwicker, E. and H. Fastl (1999). Psychoacoustics: facts and models. New York, Springer.

Zweig, G. and C. A. Shera (1995). "The origin of periodicity in the spectrum of evoked otoacoustic emissions." J. Acoust. Soc. Am. 98(4): 2018-47.

Zwislocki, J. J. (1965). Analysis of some auditory characteristics. Handbook of Mathematical Psychology. R. D. Luce, R. R. Bush and E. Galanter. New York, Wiley.

IX. Appendix

A.1.1

The final (i.e., offline) SNR curve as a function of trial under weighted averaging is monotonically increasing. Using (29) we rewrite the SNR as a function of trial in piece-wise syntax

$$SNR_{Wave}^{opt}(\theta) = \begin{cases} \dfrac{\sigma_s^2 \theta}{\sigma_{1\eta}^2} & for \quad 1 \le \theta \le \hat{M}_1 \\ \sigma_s^2 \left(\dfrac{\hat{M}_1}{\sigma_{1\eta}^2} + \dfrac{\theta - \hat{M}_1}{\sigma_{2\eta}^2} \right) & for \quad \hat{M}_1 < \theta \le \hat{M}_1 + \hat{M}_2 \\ \sigma_s^2 \left(\sum_i \dfrac{\hat{M}_i}{\sigma_{i\eta}^2} + \dfrac{\theta - \sum_{i=1}^{M-1} \hat{M}_i}{\sigma_{\hat{P}\eta}^2} \right) & for \quad \sum_{i=1}^{M-1} \hat{M}_i < \theta \le \sum_{i=1}^{\hat{P}} \hat{M}_i \end{cases}. \quad (I)$$

From (V) it is clear that

$$SNR_{Wave}^{opt}(\theta) = SNR_{Wave}^{opt}(\theta - 1) + \frac{\sigma_s^2 (\theta - \sum_{i=1}^{M-1} \hat{M}_i)}{\sigma_{i\eta}^2} \quad > \quad SNR_{Wave}^{opt}(\theta - 1). \qquad (II)$$

A.1.2

If we define the ratio between the optimal SNR under weighted averaging (29) and the SNR under normal averaging (19) we see that:

$$\delta = \frac{SNR_{Wave}^{opt}}{SNR_{Nave}} = (\frac{tr(R_\eta)}{\sigma_s^2 M^2})\sigma_s^2 tr(R_\eta^{-1}) = \frac{tr(R_\eta)\cdot tr(R_\eta^{-1})}{M^2} \geq 1 \qquad (III)$$

where the inequality in III is derived as following. From the assumption that the noise covariance matrix is diagonal with positive elements, the trace operator is related to the 2-norm of a vector whose elements are the square root of the diagonal elements of the covariance matrix

$$tr(R_\eta) = \bar{\sigma}_\eta^T \bar{\sigma}_\eta \qquad (IV)$$

where: $\bar{\sigma}^T{}_\eta = [\sigma_{\eta 1}, \sigma_{\eta 2}, ..., \sigma_{\eta M}]$ and $\bar{\sigma}_\eta^T \bar{\sigma}_\eta$ is the vector dot product. From Cauchy-Schwarz's Theorem we can then derive the lower bound to (III)

$$(\bar{\sigma}_\eta{}^T \bar{\sigma}_\eta)\cdot((\bar{\sigma}_\eta^{-1})^T \bar{\sigma}_\eta^{-1}) \geq \left|\bar{\sigma}_\eta^T \bar{\sigma}_\eta^{-1}\right|^2 \qquad (V)$$

so that from IV and V we have

$$tr(R)\cdot tr(R^{-1}) \geq M^2. \qquad (VI)$$

Using VI in the numerator of (III) and canceling the *M* terms we get the final inequality.