2010

Comparison of loudness calculation procedure results to equal loudness contours

Jeremy Charbonneau
University of Windsor

Comparison of Loudness Calculation Procedure Results to Equal Loudness Contours

By

Jeremy Charbonneau

A Thesis

Submitted to the Faculty of Graduate Studies and Research
through Mechanical Engineering
in Partial Fulfillment of the Requirements for
the Degree of Master of Applied Science at the University of Windsor

Windsor, Ontario, Canada

2010

Your file Votre référence
ISBN: 978-0-494-62716-7
Our file Notre référence
ISBN: 978-0-494-62716-7

AUTHOR'S DECLARATION OF ORIGINALITY

I hereby certify that I am the sole author of this thesis and that no part of this thesis has been published or submitted for publication.

I certify that, to the best of my knowledge, my thesis does not infringe upon anyone's copyright nor violate any proprietary rights and that any ideas, techniques, quotations, or any other material from the work of other people included in my thesis, published or otherwise, are fully acknowledged in accordance with the standard referencing practices. Furthermore, to the extent that I have included copyrighted material that surpasses the bounds of fair dealing within the meaning of the Canada Copyright Act, I certify that I have obtained a written permission from the copyright owner(s) to include such material(s) in my thesis and have included copies of such copyright clearances to my appendix.

I declare that this is a true copy of my thesis, including any final revisions, as approved by my thesis committee and the Graduate Studies office, and that this thesis has not been submitted for a higher degree to any other University or Institution.

ABSTRACT

Advances in the field of psychoacoustics have resulted in the development of more accurate models for the calculation of loudness as well as improved contours representing loudness perception. This study was undertaken to experimentally determine a "best use" stationary loudness model among the standardized methods available. To accomplish this, an investigative study was performed using pure tones at varying frequencies to identify the strengths and weaknesses of these loudness algorithms. The results of the investigation showed that with the recent update to the reference equal loudness contours, several of the models have become outdated in their performance. The recently revised ANSI S3.4:2007 model was shown to have the best correlation to the reference curves based on experimental measurements and was also the easiest to implement. It is recommended that the ANSI S3.4:2007 loudness model be used as the present day standard for calculation of stationary loudness.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

NOMENCLATURE

σ	sample standard deviation
A	amplitude
A1	Amendment 1
atm	atmospheres (pressure unit)
ANSI	American National Standards Institute
Bark	unit of critical bandwidth in Zwicker loudness
°C	degrees Celsius
dB	decibels
dBA	A-weighted decibels
DC	direct current
DS	digital signal value
DIN	Deutsches Institut für Normung
ERB	equivalent rectangular bandwidth
f	frequency (Hz)
F_c	centre frequency (kHz)
FFT	Fast Fourier Transform
Hz	Hertz (1 cycle per second)
ISO	International Organisation for Standardisation
kHz	kilohertz (1000 cycles per second)
kPa	kilopascal
LN	Loudness level (phon)
L_P	sound pressure level (dB)
ΔL_P	change in pressure level (dB)
$L_{P,}$	full-scale sound pressure level (dB)
L_{Pi}	SPL in a band (dB)
m	number of frequency bands
m	metre
MAF	minimum audible field
ΔN	change in loudness (sone)
N	Loudness (sone)
N'	specific loudness (sone/Bark or sone/number of ERBs)
NVH	noise, vibration, and harshness
P	sound pressure (Pa)
Pa	Pascal
PC	personal computer
P_{FS}	full-scale sound pressure (Pa)
phon	unit of loudness level
P_i	sound pressure in a band (Pa)
ΔP_i	change in band pressure (Pa)
p_R	random uncertainty of a result
P_{ref}	reference sound pressure, 20 μPa
R	Peirce's constant
RMS	root-mean-squared
s	second
SLM	sound level meter
S_i	standard deviation of input variables
S_N	standard deviation of loudness (sone)
sone	unit of loudness

s_{Pi}	standard deviation of a band pressure (Pa)
SPL	sound pressure level
s_R	standard deviation of a result
t_s	Student's t-statistic
TVL	Time-Varying Loudness model
u_R	systematic uncertainty of a result
u_{DS}	uncertainty in the digital signal value
u_v	uncertainty in the analog voltage signal (V)
u_P	uncertainty in sound pressure (Pa)
$u_{FR,mp}$	uncertainty due to the frequency response of the preamplifier (Pa)
$u_{LP,}$	uncertainty in the SPL measured by a microphone (dB)
$u_{L,,eamp}$	uncertainty in the SPL due to the frequency response of the preamplifier (dB)
,	quantisation error
u_{LN}	uncertainty in loudness level (phon)
u_N	uncertainty in loudness (sone)
u_{DSi}	uncertainty in the digital signal value of the band pressure
u_{Pi}	uncertainty in band pressure (Pa)
v	analog voltage signal (V)
v_{FS}	full-scale analog voltage (V)
vf_N	degrees of freedom for loudness
vf_i	degrees of freedom of an input variable
vf	degrees of freedom
V	volt
vs.	versus
W	Watts
w_R	overall uncertainty of a result
w_{xi}	uncertainty of each variable value x_i
x	mean value
x_i	data point value
x_i	input variable value

I. INTRODUCTION

Efforts by industry to continuously develop and improve the quality of their products have produced many important findings which have influenced our everyday life. In order to predict customer satisfaction, certain intrinsic quantities, including noise, have been identified to indicate either the desirable or undesirable aspects of a product. In other words, the quietest product is not necessarily always the best product. For this, the use of sound quality metrics can play a major role in determining which acoustic cues are the most desirable to a consumer. For example, when a consumer uses a familiar device, there is an expectation of some feedback when an action is performed such as the sound made from the closing of a car door or the response of an automobile when the accelerator pedal is depressed. The appropriate sound can have the effect of portraying the quality of the product while reassuring the customer that it is functioning properly. This feedback is essential to the product image and can therefore influence the purchase decision of a potential buyer. Various sound characteristics have been identified which are used to predict customer approval. Identifiers, known collectively as the sound quality indicators, are derived primarily through the research and observations within an area referred to as psychoacoustics.

The definition of a pleasant sound has changed slightly over the years. As the listening conditions and expectation levels change, the requirements for product developers change as well. Analyzing this phenomenon, psychoacoustics is the branch of science dedicated to understanding the human response to sound. In other words, it studies how well human beings perceive sound and what characteristics or trends influence this perception. Subjective tests are conducted to identify pleasant or unpleasant aspects of a particular noise source, where jury testing may be used as a way to determine which attributes consumers find more acceptable. Here, people may be placed into an anechoic environment and asked to classify noise samples while identifying the specific 'pleasant' or 'unpleasant' acoustic cues. Such studies have identified several characteristics that are subconsciously used to rate and compare individual differences between sounds. Mathematical models or metrics have since been generated to approximate this response, the most important of which is the perception of loudness; a quantity used as an input to most other sound quality models. Due to its importance, loudness is the target focus of this investigation.

The complex relationship between the intensity of a sound level and its frequency content is the result of the non-linear response of the human ear. This makes the modelling of acoustical characteristics difficult; often resulting in years of research devoted to a single descriptor. Loudness is a psychoacoustic model relating to the perceived intensity of a source. As a subjective quantity, the determination of loudness has been an important research topic in acoustics since the 1930's. For two tones with the same sound pressure level, the perceived loudness can vary markedly depending on the spectral content. In some instances time dependence also plays an important role as the loudness calculation procedure varies depending on the temporal characteristics of the source.

Stationary sound sources can be categorized as having signals that do not vary with respect to time; signals such as pure tones and random noise sources fall into this category, remaining essentially constant. Alternatively, samples of speech or music are classified as time-varying or non-stationary signals, they are essentially unpredictable. As a result, time-varying signals are generally much more difficult to analyze as other acoustic phenomena come into effect which also need to be considered. In regard to the complexity of this loudness modelling, fundamental concepts relating to the calculations of loudness are identified in Chapter 3.

To date, several stationary loudness models have been developed and accepted for different levels of standardization. The most commonly used models include standards developed by organizations around the world including the International Organization for Standardization (ISO), the German Deutsches Institut für Normung e.V. (DIN), and the American National Standards Institute (ANSI). A potential problem exists though where these various standardizing agencies have each accepted a different method for calculating this same acoustic metric. The models vary not only in age of acceptance, but also in their calculation approaches and assumptions. If one were to calculate loudness using one model, the levels recorded cannot be adequately compared against those of another, even though the resultant value would have the same units and meaning. The multiple standardized programs available may be a result of reluctance to change or perhaps due to political disputes. A bias appears to be influencing the selection of a loudness model based on where the model was created and the nationality of the developers; regardless of the models performance. To correct this dilemma, it is suggested that one model should be identified as being the best-practice metric to be used in place of all others. This would eliminate confusion and permit exact loudness comparisons for a variety of products from all industries; thus making sound quality concerns easier to solve.

The objective of this study is to investigate and critically compare the various stationary loudness metrics that are presently available. From this comparison one calculation method will be identified as the best model for use in industry. The comparison will include considerations as to each models ease of use, experimental performance and any apparent limitations. By comparing the models this way it is the intention of the author to provide an unbiased opinion as to which model is most appropriate. In order to examine the performance, each model will be directly compared to a set of reference curves as defined by the ISO 226:2003 Equal Loudness Contours. The ISO 226 standard will serve as a benchmark set of data as it is based on a vast amount of auditory experimental research related to the perception of stationary loudness. It will be assumed that, as the ISO 226:2003 document is based on recent experimental data; any calculation model for the perception of loudness should closely correlate to this set of data as it serves as the target results for performance. This investigation will use a wide variety of experimental data including the collection of pure tones using both a direct feed approach and samples collected using a semi-anechoic room.

Once meaningful results are obtained regarding the stationary models, the next goal of the project is to perform a comparison of the non-stationary loudness metrics on the same stationary signals investigated above. As the non-stationary loudness metrics relate back to concepts from the stationary models, the calculated results determined from this investigation should theoretically correlate well with the stationary loudness performance.

Realistically, it is expected to have some discrepancies between the stationary and non-stationary models developed by the same authors. This will most likely be due to differences associated with the model creation date and the complexities of temporal signal analysis. This investigation will study the degree of such variances and comment as to whether the performance changes were improvements on the stationary model procedures or otherwise.

Given the approach outlined above, it is the primary purpose of this study to provide a meaningful comparison to the acoustics community. By conducting a thorough experimental analysis of the various loudness models, as well as a complete literature review, this project will ensure originality and provide significant insight into the available methods for the analysis of stationary noise signals. This will be done using both the stationary and non-stationary loudness metrics available through experimental testing.

II. LITERATURE SURVEY

Prior to the experimental investigation, a review of the available literature was conducted to ensure that no previous study had attempted a loudness model comparison of this magnitude. No existing studies were found which compared all of the existing loudness models; neither against each other nor against the newly updated equal loudness contours of ISO 226:2003. Therefore, no decisions were found in the literature which concluded on a best use loudness model.

A great deal of research does exist for the study of loudness and the resulting equal loudness contours. This research is important to this study as knowledge relating to an understanding of psychoacoustics and its fundamentals is necessary prior to comparisons of different loudness approaches. The characteristics and trends of loudness will be compared through the results of several published papers on the subject. In order to understand the calculations and procedures for loudness, a brief introduction to what loudness is will first be included with an in-depth description of the underlying theories in the following chapter.

2.1 Definition of Loudness

Loudness is a psychoacoustic descriptor relating to the perceived intensity of a sound source. While it is a subjective quantity, a great deal of research going back to as early as the 1920's has been devoted to quantifying this important characteristic of sound. As a result, one of the first documented breakthroughs for the analysis of loudness was in the work of Fletcher and Munson's "Loudness, Its Definition, and Measurement." **[15]** This work performed at the Bell Telephone laboratories revolutionized the measurement of noise using telephone receivers and a variety of subjective tests. The end result was a detailed description of loudness and the trends present in the human hearing spectrum.

The actual sensation describing the magnitude of a sound is related to the density and location of nerve endings excited within the ear at one time. However, this sensation can vary from person to person and depend greatly on the conditions associated with the excitation. As such, it is important to both define the intensity of the perception, and to take into consideration other factors including the physical composition of the sound and the conditions surrounding the listener. **[15]** The perception of loudness depends not only on the level of the intensity (relative to a reference value of 10^{-16} Watts per square centimetre), but also on the frequency content of the signal and the manner in which the signal was presented. The human ear is more sensitive to higher frequency ranges around 1 kHz than to low frequency content below 100 Hz. This is thought to most likely be an evolutionary trait as the majority of speech signals lie in the higher frequency areas of the hearing spectrum. While low frequency noise is still perceptible down to approximately 20 Hz, pure tones in this range must have very high amplitudes in order to be just audible.

In the application of experimental acoustics, a variety of testing environments can be used for the presentation of the source signal to a listener. The most common controlled environments include free-field, diffuse-field or the presentation of the signal through headphones. A free-field application refers to an environment free of any obstructions within the sound field which may influence the sound propagation from the source to the receptor. This environment essentially has zero reflections associated with the signal and is therefore an ideal testing environment for directivity analysis. A diffuse-field on the other hand is an environment in which sound energy is incident from all directions with equal intensities. **[48]** Thus a measurement may be made anywhere in a diffuse-field environment and would result in the same measured sound pressure level; a useful tool for determining the sound power level of a source. Both listening conditions serve unique purposes in acoustical experimentation and are often used in the research and development industry. Listening via headphones is a commonly used method for jury testing. Although the product source is not usually present for the jury experiment, the use of headphones allows the listener to quickly switch between varying product sounds and removes the unwanted effects associated with poor acoustic memory. For loudness measurements, the most common setting is the free-field with frontal incidence. In this case, the source is placed directly in front of the receiver (or listener), which is directly facing the source.

In order to quantify loudness over the frequency spectrum, Fletcher and Munson chose a reference tone of 1 kHz. **[15]** They chose this frequency based on the several considerations including the observation that a 1 kHz was easily defined and allows for easier mathematical computations, reducing computational time. At 1 kHz, the audible spectrum also has a larger audible range than other frequencies, measured from the threshold of hearing up to the threshold of pain. **[15]** Based on this selection, the 1 kHz tone has subsequently remained the reference frequency value for loudness since. As a result, the loudness level (unit phons) of a signal is numerically equivalent to the sound pressure level (dB) of an equally loud reference tone at 1 kHz. This equal loudness definition was the basis for the development of the equal loudness contours. **[15]**

2.2 Development of the Equal Loudness Contours

As an important tool for the understanding of the limits for the human auditory system, the equal loudness contours represent an important descriptor for the perception of loudness.

In 1933 Fletcher and Munson developed one of the first studies to map a set of contours relating to the sensation of equal loudness in a free-field. Continuing on the work started by Kingsbury in 1927, Fletcher and Munson conducted experiments deriving loudness levels over the complete practical auditory range. **[15, 28]** Resulting from this work, **Figure 2.1** is the first combined set of the contours developed, the trends of which provide extensive insight into the strengths and weaknesses of auditory perception. A contour of equal loudness can be described as a group of equally loud data points which vary in both frequency and sound pressure level. Each individual contour line is referred to by the corresponding sound pressure level value at the corresponding 1 kHz center frequency tone. From the definition of loudness above, it is at this point where the loudness level (in phons) and sound pressure level (in dB) are said to be equal. For the experiments, the authors used telephone receivers to introduce the various intensity levels to the subjects. As this was not an ideal free-field environment, calibration factors were obtained at each frequency to correct for the receiver playback. These corrections values were combined to form a calibration curve or transfer function which was used for adjusting the results. The added correction could have led to a potential error source in the experiment; had the experimenters had access to free-field conditions in which to present the pure tones, the correction factors would not have been necessary. Unfortunately the technology was not available at the time of this experiment, but the results obtained were nevertheless an important foundation for the research to follow.

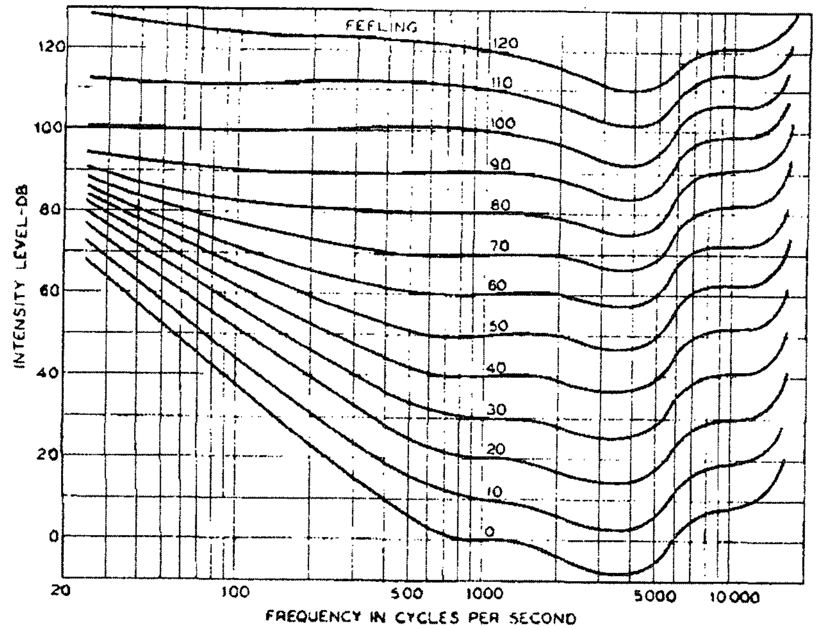


Figure 2.1 - Fletcher's 1933 Equal Loudness Contours [15]

The work by Fletcher and Munson was followed by several others, including Churcher and King in 1937 and soon after by Zwicker and Feldtkeller in 1955. **[24, 57]** Although each of these data sets portrayed experimental contours of equal loudness, Robinson and Dadson identified the fact that the previous investigations displayed considerable discrepancies when compared against each other. As a result, a more extensive investigation was carried out in 1956 by Robinson and Dadson at the National Physical Laboratory which would later be adopted as the first international standard for equal loudness contours. **[37]** The primary target of the project was to provide a comprehensive set of equal loudness contours which would produce consistent results correcting the previous discrepancies. The new study included a threshold for loudness and loudness values for sound pressure levels up to 130 dB. For completeness the frequency range for the experiment extended from 25 Hz up to 15 kHz. **[37]** As a result of the extensive nature of this document, significant portions were used directly in the formulation of the first standardized set of equal loudness contours given in ISO/R 226:1961.

The ISO standardized equal loudness contours originally accepted as ISO 226 have undergone several revisions to permit newer findings and corrections. The first revision in 1987 did not contain any records as to what changed between the 1961 version and the latter. As a copy of the original document was not available, the specific differences cannot be discussed here. The 1987 revision of the standard (ISO 226:1987) provides the equal loudness contours for an ontologically normal person between the ages of 18 and 30 and is intended for free-field listening conditions with binaural perception. **[23]** Within the standard, equations were derived to calculate the loudness level of an independent sound pressure level for each of the preferred third-octave frequencies from 20 Hz up to and including 12.5 kHz. In order to describe the contours graphically, the standard included a table of parameters as well as **Equation (1)** to generate the respective loudness levels. To use the equation, a sound pressure level (L_f) given at a particular frequency (f) is inserted into the formula, while the variables from the built-in table, a_f, b_f, and T_f, are taken corresponding to the desired frequency value. Given these coefficients, a loudness level (L_N) for any desired SPL can be calculated. **[23]** This model is only applicable up to 120 dB for frequencies below 1 kHz and 100 dB below 12.5 kHz. While not known for certain, this limit is most likely due to the physical limitations of both the pain threshold and the hazards present when dealing with SPLs above this amplitude; preventing such information from being collected.

$$L_N = 4.2 + \frac{a_f(L_f - T_f)}{1 + b_f(L_f - T_f)} \qquad (1)$$

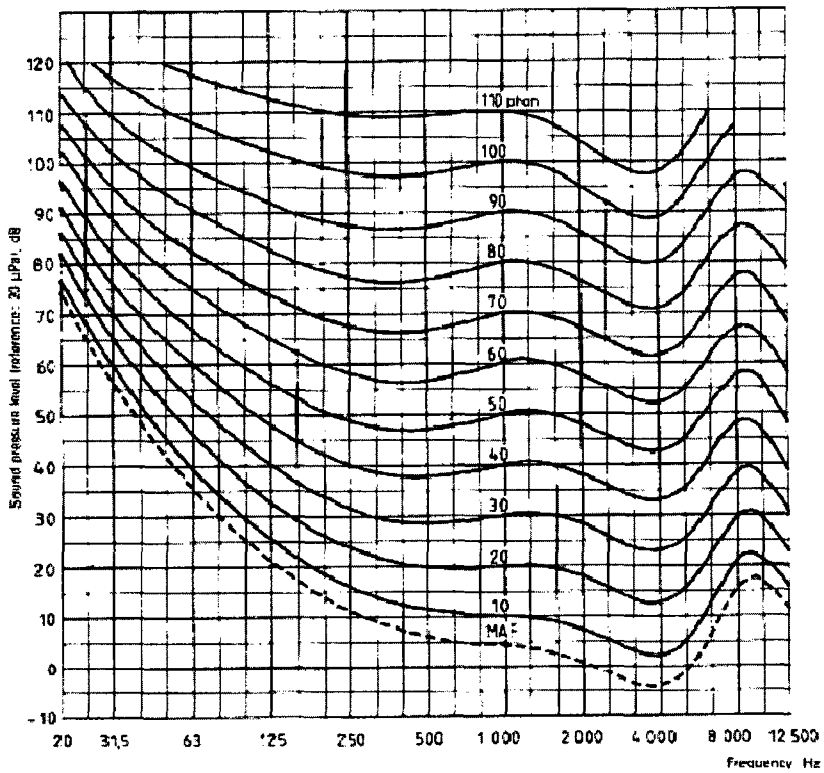


Figure 2.2 - Normal equal-loudness contours for pure tones. [23]

The resulting plots from this equation is given in **Figure 2.2** as the equal loudness contours for binaural free-field listening and frontal incidence; reproduced from ISO 226:1987. **[23]** If compared to the contours from Fletcher and Munson (**Figure 2.1**), one can immediately see the differences between the two. Trends in the newer contours vary smoothly across the frequency spectrum; consistently maintaining the shape of the Minimum Audible Field (MAF) curve indicated by the dashed line. While the contours of Fletcher and Munson's rendition appear to bunch tighter together in the lower loudness levels, indicating an extreme sensitivity to loudness at low frequencies.

The contours derived by Robinson and Dadson do not apply directly to all types of listening conditions. A diffuse-field measurement for example would have contours exhibiting slightly different trends from those described in **Figure 2.2**. Therefore, the 1987 version of the ISO 226 document included the considerations for conversion to a diffuse-field approximation as presented in the since-withdrawn ISO 454 standard. This addition, given as Annex C, gives the document a wider range of applicability as a useful reference for the user.

As knowledge in the acoustic community progressed, an update to the ISO 226:1987 contours was considered to be necessary. Shortly after the release of the 1987 contours, Fastl and Zwicker noted discrepancies between the contours of the standard and their own findings. These results were confirmed in a compilation study produced by Suzuki and Takeshima indicating the research to-date concerning the equal loudness contours. Looking at work from various investigations as well as their own, Suzuki and Takeshima's study confirmed that different trends were in-fact present in the frequencies below 800 Hz. **[47]** The new investigations showed that the values of Robinson and Dadson's 1956 contours were lower than the present results indicated; differing as much as eight decibels at specific frequencies. Suzuki and Takeshima's study clearly illustrates this separation (reproduced in **Figure 2.3)**, where Robinson and Dadson's standardized 40 phon contour (solid line) is plotted against the more recent investigations (see legend in Figure 2.3). The separation below 800 Hz is quite large indicating the need for a revision. Based on their findings, Suzuki and Takeshima used the more recent collection of data to help derive a new set of equal loudness contours. The authors began creating their own by first analyzing the threshold values from each study and generating a best-fit threshold function. As was done in Fletcher and Munson's work, they hypothesized that the equal loudness contours should be smooth and parallel to the threshold function. Likewise this served as starting curve which Suzuki and Takeshima based their new equal loudness contours from. **[15, 47]**

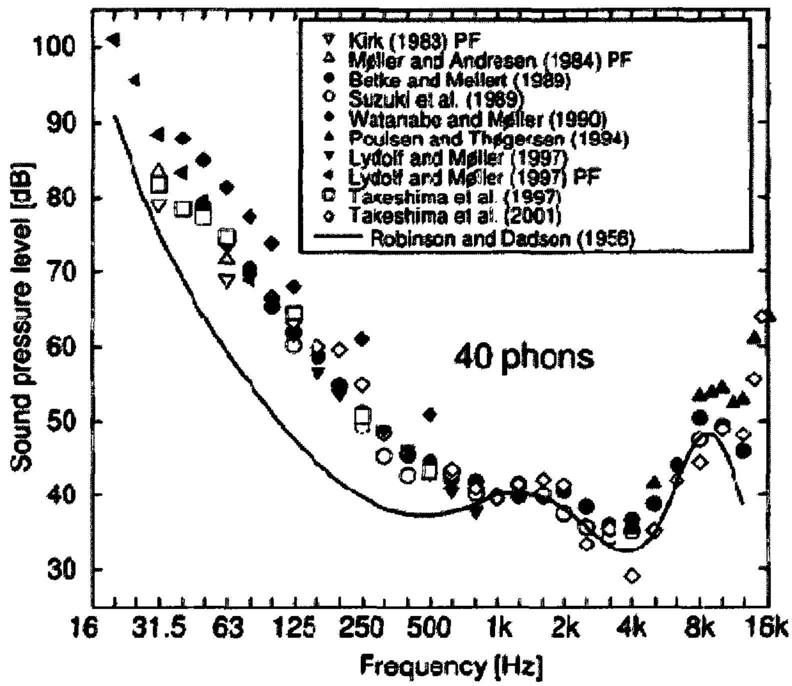


Figure 2.3 - 40 Phon Comparison [47]

From **Equation (2)**, the equal loudness curves can once again be derived using the 1 kHz reference value as a contour identifier and a given reference frequency's sound pressure level (p_r). The equation produces the sound pressure level (p_f) in dB at each centre frequency using the respective frequency dependant coefficients 'α_f' and 'U_f' derived by the authors.

$$p^2{}_f = \frac{1}{U^2{}_f}\left\{(p_r{}^{2\alpha_f} - p_{rt}{}^{2\alpha_f}) + \left(U_f p_{ft}\right)^{2\alpha_f}\right\}^{1/\alpha_f} \quad [47] \ldots \qquad \textbf{(2)}$$

To ensure that the contours extend smoothly with adjacent shapes, the coefficient values were generated at each centre frequency and smoothed under the assumption that the coefficients “do not change abruptly as a function of frequency.” **[47]** This created a loudness function yielding excellent results when compared to the recent loudness studies. The results derived by these authors performed so well that they were used directly to derive the standardized equations stated in the updated ISO 226:2003, (see **Figure 2.4**). **[24]**

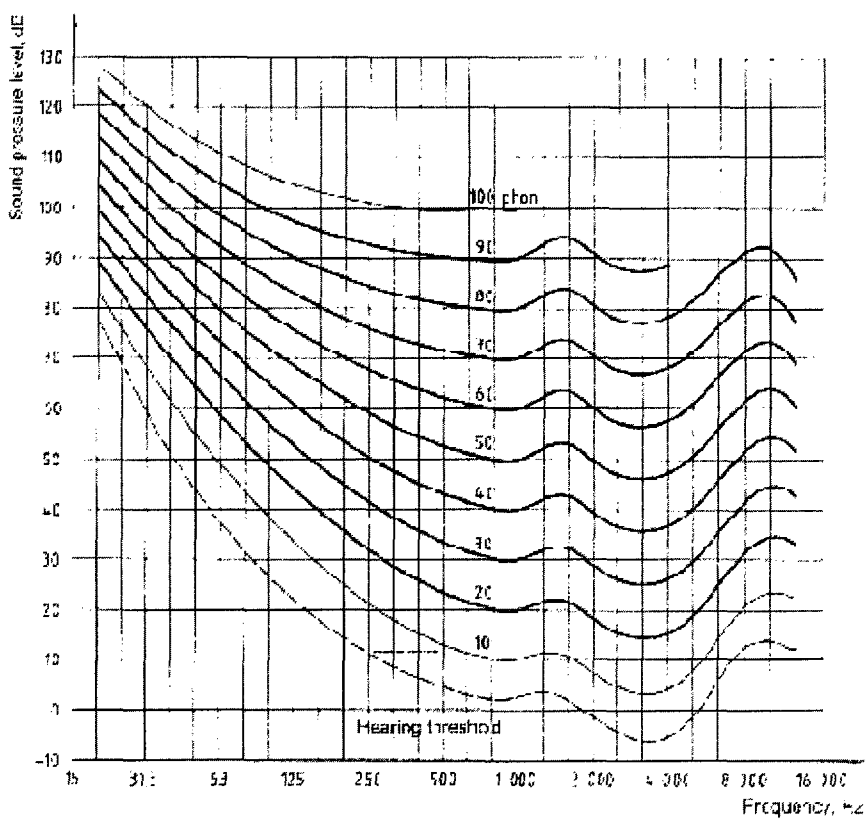


Figure 2.4 - ISO 226:2003 Equal Loudness Contours [24]

The internationally accepted ISO 226:2003 is the most recent update to the standard entitled “Acoustics – Normal equal-loudness level contours.” **[24]** With improvements to the calculation process, the update introduces a set of two equations for deriving the normal equal-loudness contours as reproduced in **Equation (3)** and **Equation (4)**. From the formulae, a sound pressure level (L_p) at a given centre frequency (f) may be determined for any desired loudness level (L_N). As in the previous update to the standard, frequency specific coefficients can be taken from an included table which can be inserted directly into the equations below. The three coefficients used in this case are: the exponent for loudness perception (α_f), the threshold of hearing (T_f) and the magnitude of the linear transfer function (L_U); normalized at 1 kHz.

$$L_p = \left(\frac{10}{\alpha_f}\log A_f\right) dB - L_U + 94\, dB \quad [24] \ldots \tag{3}$$

$$A_f = 4.47 \times 10^{-3}(10^{0.025 L_N} - 1.14) + \left[0.4 \times 10^{\left(\frac{T_f + L_U}{10} - 9\right)}\right]^{\alpha_f} \quad [24] \ldots \tag{4}$$

The new set of equal loudness contours may then be plotted as in **Figure 2.4**. This data set shows a steeper slope when compared to the previous standard, and as a result, matches appropriately to the research compilation of Suzuki and Takeshima seen in **Figure 2.3**. **[47]** It is important to point out that in this new variation, there is a more pronounced 'bump' around 1 kHz which the previous standards did not possess. Also, unlike the previous document, there is no mention of the equal loudness contours for a diffuse sound field; most likely due to the fact that all of the new studies mentioned above were focused on free-field perception. While this area of psychoacoustics has generated a lot of scientific findings, diffuse-field investigations are not as numerous.

The maximum levels available have also been left out. The new contours do not extend to loudness levels higher than 100 phons while the previous 1987 standard contained equal loudness contours up to and including 110 phons. Sound pressure levels exceeding 100 dB are nearing the boundary for discomfort and damage risk. This coupled with a greater emphasis on what is ethical for jury testing and experimentation, it is no surprise that recent studies did not include such elevated levels. Given that the ISO 226:2003 standard was the most recent update to the equal loudness contours at the time this research was undertaken, it serves as the best known reference against which to compare any newly developed models. Therefore, it will remain the focus of all experimental comparisons and acting as a target set of measured or 'real' values to achieve.

2.3 Loudness Metrics

There have been many prediction methods for the calculation of loudness developed over the course of the last few decades. With the experimental data of the ISO 226 describing the perception of stationary loudness exactly, several methods have been developed to predict this phenomenon for everyday signals through calculations; however few of the models developed have reach the level of a standardized calculation document. For this comparative study, three stationary loudness prediction models have been selected. These include the international standard ISO 532B and the German DIN 45631; two stationary loudness models based on the work of Eberhard Zwicker, plus a third, the ANSI S3.4:2007 an American standardized loudness metric developed by Brian Glasberg and Brian Moore.

One of first loudness models accepted by a standards organization was the ISO 532 "Acoustics – Method for calculating loudness level." **[27]** Accepted in 1975, the ISO 532 document contains two separate loudness metrics which have been identified individually as Method A and Method B. The first model Method A, is the lesser well-known of the two and is based on the research of S. Stevens. In 1936, Stevens proposed a new scale for describing loudness based on the use of the unit sones, a set of units on which the majority of recent loudness metrics now base their values. **[42]** Over the course of a few decades, Stevens developed some of the fundamental concepts used in the prediction of loudness; including the power law and eventually the development of the Mark VI loudness model in 1961, now known as Method A. **[44]** Recommended for use with 1/1 octave bandwidth data, this method calculates loudness through the use of given equations and corresponding coefficient look-up charts. Unfortunately, the version included in the ISO 532 document was only applicable for a diffuse-field environment, further limiting its applicability. As a result of various performance comparisons, Method A has often been disregarded due to its poor resolution and known limitations as opposed to the accompanying Method B model.

More commonly used in industry, Method B of the standard is the often preferred method for calculating loudness; commonly referred to as simply ISO 532B. Developed from the loudness model by acousticians Paulus and Zwicker, this model played an important role in the development of the loudness metrics still in use today. **[34]** Using concepts from Zwicker's earlier work, (see **[50, 51, 54, 57 and 58]**) the authors compiled a loudness model which made use of the fundamental concepts of loudness including: critical bandwidths, the various listening conditions, and the effects of simultaneous masking.

By approximating the filtering process of the human auditory system with the use of critical bands, Zwicker's method attempts to better approximate the sensation of loudness, (a detailed description of critical bandwidths is given in Chapter 3). To make the application of the model easier, it was developed such that input values can be either entered as critical bandwidth values or the more commonly found 1/3 octave data sets. During the calculation process, if the input signal is specified using the 1/3 octave values, the procedure simply combines the lower frequency bands into three larger sets to form approximate critical bandwidths.

The ISO 532B version of Zwicker's model is based on a complete graphical approach. One would plot the recorded 1/3 octave data from 25 Hz to 12.5 kHz on an included set of charts where separate sets of stencils were dedicated to either free-field or diffuse-field measurements. From the stencils, a resulting plot of frequency versus specific loudness level was generated providing the specific loudness spectrum for the stationary signal presented. On the horizontal axis, the plot is generated using a Barks scale; one Bark represents one critical bandwidth, resulting in 28 barks across the spectrum. The specific loudness values are therefore provided in the unit sone/Bark, or loudness level per critical band. In order to connect adjacent bands, data points progress from left to right where increasing specific loudness levels are represented by vertical lines and drops and portrayed by decreasing slopes. It is the sloping plots which create the simultaneous masking effect, an important component of loudness mentioned again in the theory of Chapter 3. Once completed, the entire area under the resultant shape is summed to give a total loudness level for the signal in sones (or phons) using the appropriate side scale.

Alternatively, the 1972 Zwicker paper also included a set of FORTRAN-VI computer programs to ease the calculation process; one program was available for each the critical bandwidth and 1/3 octave inputs. However, during the transfer process to the

ISO 532B standard, the FORTRAN-VI programs were neglected and only the 1/3 octave stencils were included - greatly reducing the application of the original model. As the procedure for ISO 532B above was very time-consuming and tedious, Zwicker and his colleagues later reproduced their original program using the modern programming language (BASIC code), to implement the model electronically. **[59]** The resulting loudness metric was easier to use and more popular than Method A of the document, and has been known to produce more accurate results. For this reason, only Method B will be considered in the comparisons that follow.

The second stationary loudness model addressed in this study is the German DIN 45631 standard. **[10]** Accepted in 1991 by the Deutsches Institut für Normung (DIN - translating to the German Institute for Standardization), this loudness model was also originally based on Zwicker's work above for the ISO 532B. As an improvement on the previous model, the standard has come to be known in industry as the Modified Zwicker Method.

The procedure of the DIN model is essentially the same as that of the ISO standard, only this time it included a revised version of Zwicker's program code. Various data files in the program have been adjusted slightly from the BASIC code which is an improvement as the values are a better representation of the original coefficient plots. The same year that the DIN 45631 model was released, the updated code was re-published in English by Zwicker. **[60]** The general consensus of the changes is that the DIN 45631 model does improve on the performance of the model below 300 Hz. The present study will show that it is within this frequency range that the ISO 532B procedure performed poorly.

The last model investigated was the American National Standards Institute's (ANSI) metric entitled ANSI S3.4:2007 "American National Standard Procedure for the Computation of Loudness of Steady Sound." **[2]** Originally produced as ANSI S3.4:1980, the loudness model was based on the work of S. Stevens as in Method A of ISO 532:1975. However, in 1996 Glasberg and Moore developed a new method which would eventually replace the 1980 ANSI standard as an improved estimation of loudness. **[30]**

Glasberg and Moore's approach was another extension of Zwicker's 1972 model. Retaining the main elements of the original, the basic ideas for process remained the same but the data and manner in which the steps are carried out differed markedly. Data is inserted into the model using 1/3 octave bands, which is then altered using functions imitating the effects of the outer and middle ear. In Moore and Glasberg's model, the effects are modelled using transfer function contours based on their earlier work. **[17]** The transfer functions allow for smooth modifications to the signals with no jumps in coefficient values.

The filter shapes used by Glasberg and Moore also differ from Zwicker's approach. The shapes are based on Equivalent Rectangular Bandwidths (ERBs) which are used to calculate the excitation patterns required for loudness analysis. (For reference purposes more information on ERBs will be presented in Chapter 3 with a comparison to the critical bandwidths). Unlike the ISO 532B method, the Glasberg and Moore model is based on a computational approach only, relying heavily on tabulated values and formulae. As Defoe (2007) commented during his synopsis of the model, the relationship between specific loudness and excitation is no longer calculated using the plot results on given charts; it is now primarily based on theoretical ideas and derived constants. [9] This makes a graphical application very complex for models of this magnitude.

Shortly after releasing their paper in [30], Moore et al. republished the model, correcting known issues with performance; particularly dealing with binaural and threshold perception. [32] In order to improve on the model's ability to predict binaural loudness summation, it now operates with the assumption that a signal presented binaurally will be perceived as twice as loud then if the signal were presented at each ear individually. The absolute threshold of binaural hearing was also adjusted in accordance with new experimental data. It was determined that when listening with both ears the threshold should be 1-2 dB lower than if one was listening monaurally. [32] The last revision to the model was to predict a greater than zero loudness level at the threshold levels. This change appears intuitive given that if one can detect a single tone, then it is expected that the signal would have some finite loudness level. The previous model predicted a zero loudness level corresponding to the value at threshold. Now sub-threshold values are possible as in the case of complex tones with individual sub-threshold components summing to audible levels.

Based on the improvements, the new Glasberg and Moore program better correlates with the latter equal loudness contours. The adjustments gave the model a steeper slope in the lower frequency regions, predicting contours that are in better agreement with those found in Suzuki and Takeshima's compiled study. [47] With all of the improvements listed here, the resulting model was a comprehensive loudness tool for analyzing stationary sound sources. Capable of performing measurements on a variety of listening conditions including: free-field, diffuse-field, binaural, monaural, or listening via headphones this model was applicable for a broader signal range than any of the previous models developed. An executable computer program was developed in accordance with the model, permitting an easier implementation given the variety of input types. However unlike the other models, no source code was included in the paper making the comparison of the electronic metric styles impossible. The new Glasberg and Moore model soon became the revised version of ANSI S3.4:2005.

In the transfer process from a published paper to a standardized method, the Glasberg and Moore 1997 model remained almost entirely the same. **[1]** The standard includes all the necessary definitions and the entire procedure of the 1997 model with only minor changes made to figures and data sets; while the information presented remained essentially the same. **[9]** The computer program for the model was also included in the standard as with the model, only now under the title of ANSILOUD.exe. Unfortunately the source code for the program is still not available but an analysis of the performance by DeFoe (2007) indicated minor discrepancies between the standard and the 1997 model results. **[8]** The deviations included formulae reproduction errors in the standard that did not correspond to the given sample results. Based on these findings the authors of the standard eventually revised the loudness model again, taking these errors into account.

Although the 2005 version of the ANSI standard performed adequately, an update was imminent as the model needed to be adjusted to better approximate new findings on human perception. In 2006, Glasberg and Moore updated their model in accordance with the newly accepted ISO 226:2003 equal loudness contours. **[18]** The authors recognised they needed to permit the revision of the threshold of hearing. At the time that the 1997 model was produced, the absolute threshold values had been based on the ISO 389-7:1996 standard. **[25]** The new values for the ISO 226 standard are based on the revised ISO 389-7:2005 for the "Reference threshold of hearing under free-field and diffuse-field listening conditions" **[26].** To account for the update in their model, the authors modified the assumed middle-ear transfer function to better fit the data. **[18]** The alterations provided the desired improvements as these results now provide a slightly better comparison to the ISO 226:2003 equal loudness contours. The new update led to a second revision of the standard, resulting in the currently available ANSI S3.4:2007. **[2]**

The 2007 version of the ANSI standard is the only known standardized loudness metric to match and account for the latest updates made to the ISO 226 equal loudness contours. As before, a variety of listening conditions are available for calculating loudness. The ANSI standard is suited for free-field, diffuse-field or listening via headphones; allowing for this model to have a wide range of applications. A new computer program was generated and included with the standard reflecting the 2006 improvements, (LOUD2006A.exe). Alternatively the standalone executable file is also available from the University of Cambridge – Auditory Perception Group website. **[19]** Even though there are currently three standardized loudness metrics available, the ANSI S3.4:2007 being the most recently updated is assumed to be the most likely to perform in accordance with the reference contours.

One other popular metric which warrants mentioning is often used to portray the magnitude of a sound and is commonly referred to as the A-Weighting scale. The A-weighted sound pressure level was originally derived from the 40-phon contour line of the Fletcher and Munson contours. It was meant to adjust recorded tones to the sensitivities of the human hearing spectrum by inverting and normalizing the 40-phon shape. A series of weighting values were then generated across the frequency spectrum which could be applied to any input signal. This resulted in an A-weighted contour which heavily attenuates the lower frequency SPLs and marginally reduces values above 6000 Hz. The transfer function process could easily be implemented into sound level meters generating results quickly in the field. It was this ease of use that made the A-weighting approach popular. However, critics of the model's use have been questioning its applicability to loudness for years. For instance as Schomer et al. Indicated that, although the A-weighting filters do vary with the human sensitivity to frequency, the filter set does not account for the sound pressure level of the signal; the filter values always remain constant regardless of amplitude. **[39]** As a result louder signals will be corrected in the same manner as lower noise sources. Observation of the equal loudness contours reveals that the human perception in these areas differs significantly. In other words, a lot of the important content could be inappropriately attenuated when presenting loudness as A-weighted decibels or dB(A). This is particularly true if the signal is outside of the range immediately surrounding the 40 phon curve. Due to the known errors associated with presenting loudness information using the A-weighting method, an in-depth look into the performance of this approach was not included in this comparison. Discussion of this method was included as it is a reoccurring focal point during the study of loudness and therefore should be mentioned in this discussion of available models.

2.4 Non-Stationary Loudness Metrics

Although the study of stationary signals is important for understanding the perception of loudness; the majority of signals encountered in practice tend to be temporal or non-stationary. Of the models studied in the previous section, two models have been adapted to include the effects present in temporal signals; the Glasberg and Moore model named simply the Time Varying Loudness (TVL) method and the continuation of the German stationary model with a draft entitled DIN 45631 – Amendment 1. **[16, 11]** Both of these models are still in draft form as no standardized method currently exists for temporal sounds. As this report is a study of the standardized stationary loudness metrics, it was decided to analyze these two extensions to compare and discuss their performance to the same stationary tones. This would provide an added investigation into loudness, and further the discussion on performance.

The additional characteristics present in non-stationary signals account for the application of temporal masking, (for an in-depth description refer to Chapter 3), and temporal averaging. To convert a stationary loudness model to a non-stationary model, one must apply the effects of these phenomena accordingly. In 2002, Glasberg and Moore developed an extension of the stationary loudness procedure with a goal of creating a more accurate model capable of handling the discrete spectral components which 1/3 octaves cannot. At the same time they wanted the model to be capable of handling non-steady sounds which are more common than stationary noise sources. **[16]** The model they developed as a result was the TVL model, capable of calculating two types of non-stationary loudness: both short-term and long-term loudness. Best described using examples of speech; the authors explained short-term loudness as the intensity of a syllable. Long-term loudness would be used to measure the intensity of a much longer noise sample, such as a sentence. **[16]** To accurately model the complex signals present in temporal samples, the model accepts 16-bit WAVE files with a sampling rate of 32 kHz. Using the WAVE file as an input rather than filtering through 1/3 octaves permits the information to be processed at a higher resolution, retaining as much information as possible. The calculations can then act on the time waveforms of the signals while calculating a running average of both the short term and long term loudness.

When considering the target performance of the model, the authors wanted to concentrate on predicting the loudness of two important trends in temporal signals. The first issue deals with the amplitude modulation of a carrier sinusoid. From the overview of literature the authors provided, it became clear that predicting the loudness level of amplitude modulated signals can be quite challenging as the trends can vary as the rate of modulation increases. The authors wanted to develop a model capable of predicting this complex relationship. Secondly, the authors wanted to be able to include the effect of the temporal masking which takes place after a fixed intensity signal burst. Loudness levels of short bursts can increase for durations up to 100-200 ms, after which the levels seem to remain roughly constant. This was just another factor which they hoped to describe. **[16]**

Aside from some minor modifications to the procedure, the majority of the loudness calculations remain consistent with the stationary loudness model from 1997. **[32, 16]** One important modification was the use of 6 parallel FFTs to calculate spectral information over six bandwidths which increase in frequency and decreasing in lengths of time. The ranges of the filters "are 20 to 80 Hz, 80 to 500 Hz, 500 to 1250 Hz, 1250 to 2540 Hz, 2540 to 4050 Hz, and 4050 to 15000 Hz for segment durations of 64, 32, 16, 8, 4, and 2 ms, respectively." **[16]** This was done in order to retain a high spectral resolution at lower frequencies as is present in the auditory system. The varying time segments were used to give adequate temporal resolution at higher frequencies; this turned out to be an effective method of detecting high frequency amplitude modulation. **[16]** The excitation pattern and instantaneous loudness levels are then calculated in the same fashion as the previous model. Short term loudness could then be obtained by temporally averaging the instantaneous levels, giving you a running average for the signal. Likewise, the long-term loudness is the result of a temporal average of the short-term loudness. After describing the procedure of the model, the authors went on to verify the performance of the model, even going as far as plotting the predicted equal loudness contours. Note that this was prior to the update in 2003 which was mentioned in the report, (see Fig. 2 in **[16]**).

The second time-varying model that will be looked at is the DIN 45631 – Amendment 1 or simply DIN 45631/A1. **[11]** Currently in the draft process for the DIN standard, this metric is the closest non-stationary loudness metric to being accepted as a standardized loudness model. Once again this model is based off of Zwicker's approach to loudness prediction.

In 1977, Zwicker created an in-depth calculation method for temporally variable sounds as an extension of the 1972 stationary model. **[52]** As in the derivation paper for Glasberg and Moore's TVL model, Zwicker begins his description by looking at all of the temporal loudness components and stating which characteristics the model will be designed to approximate. Some of the temporal issues considered include phase effect, physiological noise, amplitude modulation, and frequency modulation. Upon review, the phase effect in temporal analysis was determined to have minimal influence so it was ignored so as to not complicate the model further.

From an analysis of tone bursts (duration vs. loudness), it was determined that for tones less than 100ms in length, "a decrease of the burst duration by a factor of 10 decreases the perceived loudness by a factor of two." **[52]** A burst of at least 200 ms was determined to be a long-lasting burst, having the highest perceived level and the longest decay. This was similar to the findings in the development of the TVL model. The perceived loudness of tone bursts was determined to be represented by the peak loudness value found over the period of the burst.

Zwicker's model was also designed to ignore pre-masking as it was determined it was not nearly as influential as post-masking was on loudness, (See the masking comparison in Chapter 3). This choice was made to include the effect of the relatively slow speed of the decay for a signal compared to the quick rise of the perception. Due to the inclusive investigations, this model had been proven capable of describing tone bursts, amplitude and frequency modulated signals, narrow band noise, and speech. When building the model, the author made it clear that the method he used resulted in a design which was quite complicated. The procedure was carried out this way to remain compatible for the previous stationary loudness model procedure, the ISO 532B. This was necessary in order to produce the correct critical band levels needed. As the DIN 45631/A1 standard is still in draft form, the complete standard was not available for review. However, the procedure outlined in the Zwicker 1977 paper seems to be similar to the one outlined in the available loudness meter description of the DIN amendment, but a direct connection cannot be established. **[11]**

The majority of the metrics listed here have been available for several years and as a result, numerous authors have presented research findings comparing the performance and use of the models. The following is a detailed summary of their findings and potential areas for improvement that this project intends to correct. Although there are other loudness models varying in procedure and application, the purpose of this study is to examine and compare the performance of only those accepted by standardizing committees. Analyzing and comparing every model developed to date is beyond the scope of this investigation.

2.5 Loudness Metric Comparisons to Date

Of three stationary loudness models investigated, the ISO 532B appears to be the most well-known and prevalent model used. Being the first standardized loudness model, the ISO 532B has subsequently been compared against newly developed loudness descriptors for years, concerning both the performance and applicability in various settings. In 1987, Hellman and Zwicker compared the metric against the popular A-weighting approach. **[21]**

Hellman and Zwicker's study compared the performance of the ISO 532B and the A-weighting approach using complex noise-tone combinations, to locate the believed poor performance areas of the A-weighted sound level. Using subjective tests backed up by calculated loudness results they were able to show that using pink noise and pure tone combinations, a negative correlation can exist between the A-weighted sound pressure levels and the calculated loudness levels of ISO 532B. In this case a reduction of 6 dB(A) actually resulted in a doubling of the loudness level. The findings were contrary to what had been previously widely accepted; helping to bring forward the inadequacies of A-weighting for noise control purposes; particularly with complex noise sources.

The research of Hellman and Zwicker has since been verified by various other studies which further discredited the use of the A-weighting scale for loudness. In a 1994 in-depth study by Quinlan, the range of variations between the A-weighting levels and the ISO 532B calculated loudness was investigated. **[36]** Using a variety of spectral conditions with a fixed A-weighted level, Quinlan was able to derive the maximum and minimum possible loudness values that could potentially occur using the ISO 532B. The results indicated an extremely broad range of values were possible for a fixed A-weighted sound pressure level; including at one extreme, a loudness range extending from 2.2 to 45 sones. In this case a constant 70 dB(A) level was used where it was determined that a 20-fold increase in loudness values was possible, given a variety of spectrally different signals. The full range of the author's results can be seen in **Fig. 2** of **[36]**. As in Hellman and Zwicker's work, Quinlan also noted that increases in loudness were possible for decreasing dB(A) levels; at times observing a five-fold increase in loudness over 20 dB(A). To support his findings, Quinlan arranged a subjective test that was carried out to verify the accuracy of the approach; thereby proving the significance of the results.

The detailed approach taken by Quinlan had once again provided insight into why the ISO 532B method is considered a useful engineering tool where the A-weighting method can be considered as severely misleading. Although the findings were quite thorough, Quinlan's approach only verified the ISO 532B to specific areas of loudness; rather than over the entire frequency spectrum. It did however re-enforce the conclusions that A-weighting should not be used as a method of presenting loudness; providing further justification as to why this common descriptor will not be included in this comparison.

The most recent comparison available is a discussion paper written in 2007 by Hellman regarding a new loudness standard ANSI S3.4:2005, (now replaced with ANSI S3.4:2007). **[22]** The purpose of the investigation was twofold, to identify and discuss the improvements made for the current revision of the ANSI S3.4 standard and secondly to compare the performance of the new standard over the restrictions of the ISO 532 loudness model. According to the research, the previous 1980 version of the ANSI model (ANSI S3.4:1980 based on Stevens' approach), had three main limitations for applicability: the model was restricted to broadband signals with no applicability to tonal components, the model suffered an inability to depict the detailed shapes of the revised ISO 226:2003 equal loudness contours, and lastly it was only applicable for loudness levels down to 20 phons. **[22]** As shown above, each one of these points was rectified in the new ANSI S3.4:2005 model which Hellman addressed in her discussion. As a recap, the new loudness model is applicable for all types of stationary signals where it can now predict the new equal loudness contours with a good amount of agreement particularly below 500 Hz. The new improvements to the standard also allow for the prediction of loudness levels down to approximately the threshold levels of hearing. By discussing the advantages of the new 2005 version of the ANSI standard, Hellman made it apparent that the new version of the standard was a vast improvement over its predecessor. The discussion gave reference to several sources confirming the changes but did not include any numerical data of its own.

Hellman's analysis next targeted the performance of the now aged ISO 532B model, by using the same 'old versus new' approach. The author was once again able to identify three main shortcomings which the new ANSI model was able to overcome; the first of which related to the listening conditions for the calculation of loudness. While Zwicker's model only had the option of monaural loudness, the ANSI method is applicable for both monaural and binaural listening. In order to approximate the response of binaural listening, the ANSI method relies on the assumption that a binaural presentation of the same signal at both ears (diotic presentation), will result in an overall loudness that is twice as loud as if the signal were presented at each ear separately. This assumption was backed up Hellman's own past research, as well as that by Marks in a later study. **[20, 29]**

The second shortcoming of the Zwicker approach is that according to previous findings, the ISO 532B accuracy is limited to only mid-range frequency, noise-tone combinations. **[22]** In the 1997 update by Moore and Glasberg, this issue was corrected, allowing the resulting ANSI S3.4:2005 method to generate better predictions below 500 Hz. **[22].** To further improve on the model, the loudness conversion factors were also revised to generate more accurate results. As the ISO 532B method uses the obsolete method for converting sones to phons, the newer ANSI model performs markedly better below 1 sone; this aspect will be discussed in more detail in the theory of Chapter 3 in this report. Concluding her comparison, Hellman has clearly indicated the performance areas which the ANSI S3.4:2005 model excels over the ISO 532B standard. However, the existing DIN 45631 model was not mentioned during this comparison. This was surprising given the fact that the DIN method is another Zwicker modified approach which also improved on the ISO 532B; particularly over the low sone conversion.

Hellman's comparison of the ISO 532B and the ANSI S3.4:2007 stationary loudness models was the most recent comparison available at the time of this study. No mention of the DIN 45631 was found in any unbiased comparisons for loudness metrics. Furthermore, no comparisons have been found between the various temporal loudness models, either against the respective predecessors or against each other. Only the inclusive performance examples within the models give any indication as to their performance. Therefore, lack of an unbiased investigation indicates that this area of research still has room for added research and improvements before any conclusions can be reached.

2.6 What Is Missing Thus far

Upon reviewing the available performance comparisons of the standardized loudness models, it was clear there was a void in the research within this important area of psychoacoustics. Where a vast amount of literature is available discrediting the use of A-weighting use for loudness, few documented studies exist comparing the more popular standardized methods available.

With so many loudness models available, it has become difficult for engineers and acousticians to make informed decisions as to which model is best suited for a given situation. [36] As such, informative studies must be carried out continuously as new updates become available. Only then will the user be able justify applying one standard over another, rather than assuming that the newer method must be better. The selection may then be made based on the performance, ensuring accurate and relevant results.

Experimental data has proven that the equal loudness contours are slightly different than initially thought. Research now shows that the contours are steeper with more pronounced shaping. The ISO 226 data represents the actual characteristics of auditory perception which the various loudness models are intended to predict. When the reference values changed, it was expected that an update to the loudness predictors would be essential for them to continue to be accurate.

To verify that an update is indeed necessary and recognise the deviations from the reference contours, a study must be conducted comparing the various models against each other and the new standard of reference, the ISO 226:2003 equal loudness contours. Before describing the experimental procedure and results taken in this study, a section outlining and comparing the important concepts of loudness will first be presented. Such a discussion is necessary in order to understand the underlying concepts involved in each model indicated above.

III. THEORY

In the previous chapter, several fundamental concepts relating to the calculation of loudness were mentioned. These include filter bandwidths, masking effects and the loudness conversion function. As these acoustic concepts play a critical role in the way the various loudness metrics perform, a brief description of each characteristic is included here.

3.1 Filter Bandwidths

As previously discussed, the sensitivity of the hearing system is non-linear over the frequency spectrum. By experimentally testing subjects using sets of tones and varying widths of noise bands, acousticians have been able to quantify the limits of this sensitivity and derive models approximating the results. In 1940 Fletcher developed one of the first studies outlining the concept of "position coordinates" along the basilar membrane. **[14]** This study resulted in the first known auditory based filter bandwidths and mapping of the excitation sensitivity of the ear. As technology advanced, several theories surfaced as to what the shapes of these filter sets look like. The two most notable sets that have emerged are the critical bandwidths of Zwicker et al. and the equivalent rectangular bandwidths as developed by Glasberg and Moore. **[58, 17]**

Critical bandwidths refer to a set of frequency filters which have increasing bandwidths as frequency is increased. The critical values were defined through extensive auditory experiments by Zwicker et al. **[58]** According to Zwicker the ear subdivides itself into these various frequency bands to carry out an internal analysis of what we hear. **[50]** Expanding Fletcher's work, Zwicker and his colleagues collected information using a variety of jury tests targeting the bandwidth limits. Sets of tones and noise bands were used along with four sets of experiments which defined the limits "on the threshold for complex sounds, on masking, on the perception of phase, and (through) the loudness of complex sounds." **[50]** Using this data, they were able to generate their best estimate of the critical band shape as a function of frequency. However, even with a complete set of critical bandwidths, the position of bands along the frequency spectrum was not yet identified. In 1961, Zwicker published an editorial as a result of an ISO standards meeting on the subject. **[50]** It was determined that for convenience the bandwidths should resemble the "preferred frequencies" similar to the previously arranged 1/3 octave bands. As such, the lowest limit of the bands was set to 20 Hz and several of the values were generously rounded to match various other preferred centre frequencies. **[50]** This rounding was assumed to be acceptable as the measurements of critical bands was known to have errors associated with them.

The critical bandwidths at this point were available for use from a figure included in the editorial, but the values had to be collected directly from the plot; resulting in potentially different bandwidths depending on the user which defeated the purpose of establishing a universal bandwidth set. To define the bandwidth values in a user friendly manner, a second paper was later published in 1980 in which Zwicker included several mathematical formulae relating to the critical-band-rate function. **[55]** This function produces a critical bandwidth in unit Hz for a desired frequency given in kHz, see **Equation (5).**

$$\frac{CB_c}{Hz} = 25 + 75\left[1 + 1.4\left(\frac{f}{kHz}\right)^2\right]^{0.69} [55] \ldots \qquad \textbf{(5)}$$

The resulting critical bandwidths remain constant below approximately 300 Hz, and increase logarithmically with frequency from that point onward. The above formula approximates the tabulated data from **[50]** with an accuracy of ±10% as a result of the rounding errors. A complete set of critical bandwidths now exists for the 24 Bark bands ranging from centre frequencies of 25 Hz up to 12.5 kHz for use in loudness-calculation procedures.

The second filter method introduced by Glasberg and Moore uses a set of formulas to calculate the equivalent rectangular bandwidths (ERBs) of the auditory system. The ERB is an approximation of the measured auditory bandwidth from experimental investigations. Following the theories of Fletcher's 1940 work, Moore and Glasberg conducted research examining the use of the power-spectrum model to determine the auditory filter shapes. **[14, 31]** Upon their investigation, however, they determined that the power-spectrum model was known to result erroneous results when given specific masking patterns. In these instances, observers occasionally performed loudness comparisons across several auditory filters rather than targeting the individual filter information as intended. Therefore, the authors devised a method encouraging the use of only the target auditory filter while retaining the assumptions of the power-spectrum model. Based on their findings, they decided the best approach was to conduct their experiment using notched-noise masking data where noise bands are played in unison with a probe tone used to direct the listener's attention. By targeting the listener's attention to the notch in the noise band, the authors sought to minimize any 'off-frequency listening.' **[40, 31]** Through this approach, Moore and Glasberg were able to identify specific trends present in the auditory shapes based on their experimental results. It was determined that for a normal hearing individual, the auditory filter shape is quite asymmetrical, with the lower branch generally rising less sharply than the upper. From the summary of the auditory filter shape, the authors were able to derive the ERB values of the auditory filters across the audible frequency spectrum. The resulting relationship produced the ERB value in Hz for a given centre frequency given in kHz. The included equation was later updated in 1990 when Glasberg and Moore presented new findings on the subject. **[17]** In this update, the authors modified their previous procedure, increasing the accuracy of the filter shapes. By correcting small assumptions in their previous model, including an equal loudness contour correction and limiting the frequency shift to $0.2f_c$ (20% of the centre frequency), they were able to improve on their previous estimations. The new relationship, shown in **Equation (6)**, defines the ERB value in Hz for a given centre frequency (F) in kHz.

$$ERB = 24.7(4.37(F) + 1)\ [17] \ldots \qquad \textbf{(6)}$$

This relationship was based on an equation originally suggested by Greenwood, where following his original theories, the above equation locates specific distances along the basilar membrane and represents each segment as an individual ERB. **[17]**

A second equation was included in Glasberg and Moore's 1990 paper which allowed the user to scale the frequency coordinates as units of ERB; hence creating a scale for frequency comparable to the unit Bark developed by Zwicker above. **Equation (7)** is the resulting equation for calculating the ERB Number (unit-less) given a centre frequency value (F) in kHz. This is useful when one wishes to present the data in a way which better corresponds to the trends present in the auditory system. **[31]**

$$\#\ of\ ERB = 21.4 \log_{10}(4.37(F) + 1)\ [17] \ldots \qquad \textbf{(7)}$$

The two critical bandwidth models listed here are clearly different both in the manner in which they were derived and in the results obtained. In **Figure 3.1**, Seeber reproduced both critical bandwidth sets on a common plot where the specific differences can be easily compared. **[40]**

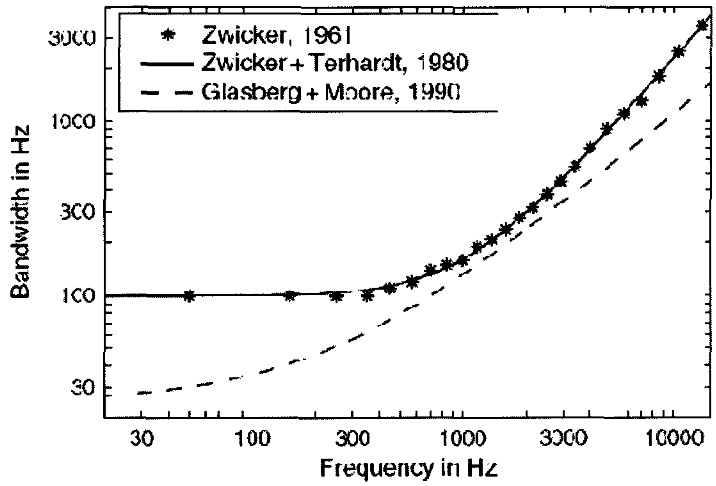


Figure 3.1 - Critical Bandwidth Comparison [40]

As indicated previously, these two sets of filters differ in both shape and slope across the frequency spectrum. At moderate frequency values, the two filter sets appear to be quite similar, but as values extend into the outer frequency levels, the similarities stop. The constant bandwidth trends depicted by Zwicker's critical filter set are notably dissimilar compared to Glasberg and Moore's downward slope at the low frequencies values. In this range, Sek and Moore indicate that Zwicker's approach was heavily influenced by critical modulation frequency (CMF) due to the use of the complex tone signals. **[41]** The derivation of the critical bandwidths was based on experimental work where a pair of tones was continually separated in frequency until an increase in loudness was noticed. **[58]** However, at low frequencies it was evident that the tones were influencing each other through modulation. **[41]** Based on their findings, Sek and Moore noticed that in this low frequencies region, the CMF flattens off due to sideband influences and the low frequency internal noise. While in this region, Sek and Moore's results indicated that auditory bandwidths actually continue to decrease, (as shown above). Glasberg and Moore's approach therefore seems to avoid the CMF effect with the use of the notched-noise test signal used in their approach. Using only one tone, there was no possibility of the tone modulation interference.

Another idea used by all three approaches is the concept of masking. In order to provide a complete discussion of the processes involved in loudness summation, a brief introduction for masking is included for reference purposes.

3.2 Masking

The phenomenon of masking is best described as one sound characteristic inhibiting the audibility of another. The two most prevalent forms of masking include simultaneous masking and temporal masking.

Simultaneous masking occurs when one source (the masker) is preventing another from being heard (the masked). A great deal of research has been done on the perception of tones in the presence of a masking background noise which has resulted in the development of informational plots such as that given in **Figure 3.2**, which are important to the understanding of the performance of the human ear. The figure is the summary of an experiment where a pure tone was played at varying levels and frequencies, while simultaneously a masker tone centered at 1 kHz (f_M) was held constant at three increments of 20 dB. The various levels of masking tones (20 dB, 40 dB, and 60 dB), were used in order to demonstrate the effect of amplitude while the frequency of the masking tone always remained constant.

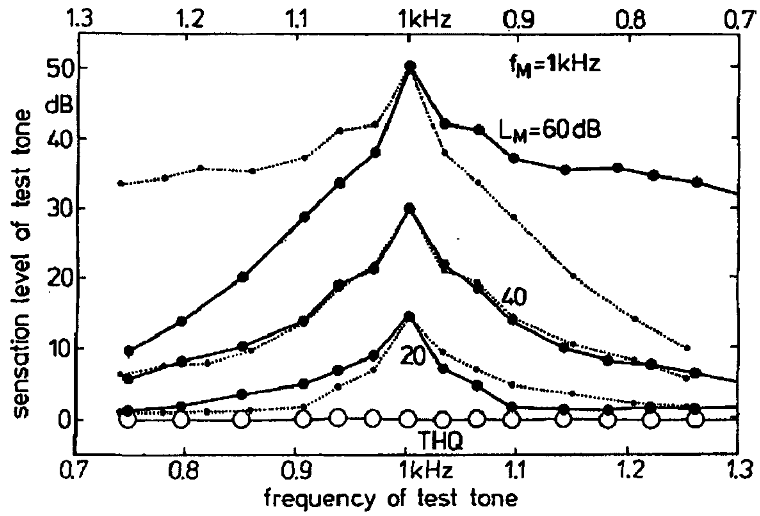


Figure 3.2 – Tone on tone simultaneous masking [13]

From this plot, one can see the upward and downward slopes of masking characteristics while discovering some interesting conclusions. Seeber indicates that as the test tone approaches the frequency of the masking tone, the amplitude must be increased to levels approaching that of the masker (*solid lines*) to become audible. The slope of the masker also appears to be heavily dependent on its amplitude. This is especially true for the higher sound pressure levels where the slope on the high-frequency side becomes much shallower. **[40]** Note that the plot was mirrored for discussion purposes, where the dotted, lighter lines represent the same data set with an inverted frequency scale (upper abscissa). This inversion was created to show the symmetry of the slopes, particularly for the low amplitude tones. Studies such as this have unlocked important discoveries as to the characteristics of human perception. Knowledge of the results indicates that the phenomenon of simultaneous masking plays an important role in industry, particularly for the removal of unwanted noise and sound quality analysis. In loudness models such as the ISO 532B and the DIN 45631, the use of simultaneous masking is especially apparent in the inclusive stencils. The downward sloping of decreases in spectral loudness is a direct result of this simultaneous masking effect. **[27, 10]**

When dealing with the measurements in most applications, temporal masking becomes the prominent characteristic due to the large amount of time-varying signals encountered. Temporal masking is an interesting phenomenon which has attracted much attention in recent years. The simplest example is the masking of a tone burst following a short noise-band. When a noise-band sample has played for a sufficiently long period of time and suddenly stops, the hearing response of the ear decays very slowly. Therefore, if a pure tone burst were to be played before the decay had finished, temporal masking may cause the tone to be partially or completely inaudible. To quantify this effect, experiments were conducted with various lengths of noise bands and short bursts of pure tones. As reported in Fastl and Zwicker's work, it was determined that based on the type masker present, pre-masking effects were evident up to 20 ms before the masker and post-masking effects were evident up to 200 ms after the masker had finished. **[13]** The two effects are best presented in **Figure 3.3,** where three types of masking are identified with respect to a noise band, plotted with respect to time. From this figure, one can easily see the transitions between masking patterns observed for a non-stationary sound sample.

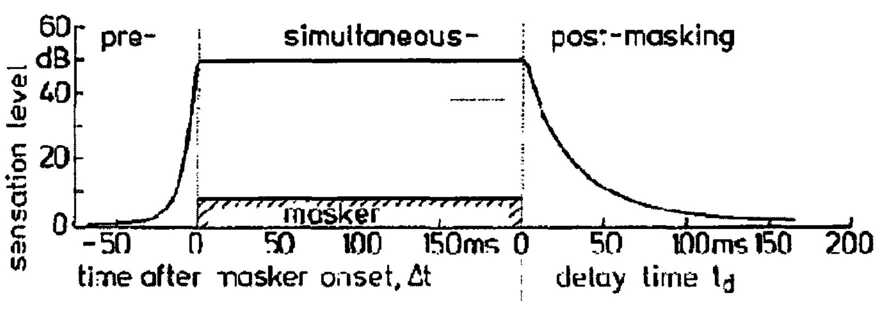


Figure 3.3 - Temporal Masking [13]

Initially, before the masker is turned on, an external signal would be fully audible. Based on experiments, 20 ms prior to the signal being initiated partial masking starts to occur rising steeply until 0 ms where the masking signal is turned on and the simultaneous masking starts to occur. The simultaneous masking effects remain present the entire duration of the noise band and until 5 ms after the masker signal is turned off. At this point, the postmasking stage is in effect, decaying slowly. Finally, after about 200ms there is no longer any masking present and external signals would be fully audible once again. Graphically, if any tone were played within the masking curves of **Figure 3.3**, the signal would be either partially or completely inaudible due to this temporal masking effect. **[13]** The application of this phenomenon has been included in the various loudness metrics indicated above, as masking plays a prominent role in the perception of tones and noise samples alike. As mentioned previously however, Zwicker's model has been designed to ignore pre-masking as it was determined to be not nearly as influential as post-masking on loudness; clearly visible here. **[52]**

There is one last concept used by the various loudness metrics that warrants discussing before moving on to the experimental procedure. As mentioned previously, there was some discrepancy between the models regarding the loudness conversion from sones to phons. As presented in the next section, this function plays an important part in the performance of the various loudness models.

3.3 Loudness Conversion

As originally derived, the loudness value of a signal was presented in the unit sone. One sone is defined to be the loudness value of a 1 kHz tone with a SPL of 40 dB relative to the reference value of 20 μPa. **[36]** Above this value, an increase of one sone is equivalent to a doubling of loudness; two sones indicated a signal four times as loud, and so on. However, in order to quantify loudness using a more direct scale, a conversion was available to convert from a loudness value of sones to a loudness level in unit phons. A loudness level in phons was then equivalent numerically to the sound pressure level in decibels of a tone at 1 kHz, (e.g. 40 dB at 1 kHz equals 40 phons, 60 dB at 1 kHz equals 60 phons, and so on). This is described in ISO/R131:1959, from which the ISO 532B and ANSI S3.4:1980 derived their conversion equations from. **[22]** As new information became available, the conversion factors below 1 sone changed (values below 40 phons), and the ISO standard was withdrawn. The ISO 532B was never updated since and thereafter the ISO 532B used an obsolete and known erroneous equation during its loudness calculation process, (an immediate indicator that this area would have problems during the following comparison). To analyze the effects of this change, the various conversion processes from the loudness models were plotted against each other for comparison.

The ISO 532B, as mentioned above, uses the since withdrawn conversion from ISO/R131:1959 as shown in **Equation (8)**. Based on a logarithmic relationship, it has been determined that when the loudness value of one or more sones is entered into the equation, the resulting loudness level is in fact the correct phons value. Below this value, however, the logarithmic relationship incorrectly approaches negative infinity, (see blue line in **Figure 3.4**).

$$Phons = 33.2 * \log_{10}(Sones) + 40 \qquad [27] \ldots \qquad \mathbf{(8)}$$

Prior the acceptance of the DIN 45631, another equation was derived for regions below one sone. The new equation permitted more appropriate levels of loudness to be derived near the threshold of hearing. Seen in **Equation (9)** is the relationship between loudness values less than one sone and loudness levels in phons.

[10]... **(9)**

The ANSI S3.4:2007 document uses a similar approach but presents a tabulated set of values accompanied by a graph for the corresponding conversion, rather than an equation. From a comparison plot in **Figure 8**, The ANSI values shown as **(A)**, closely correlate with the DIN **Equation (7)**; deviating up to 4 phons at the lower sone values.

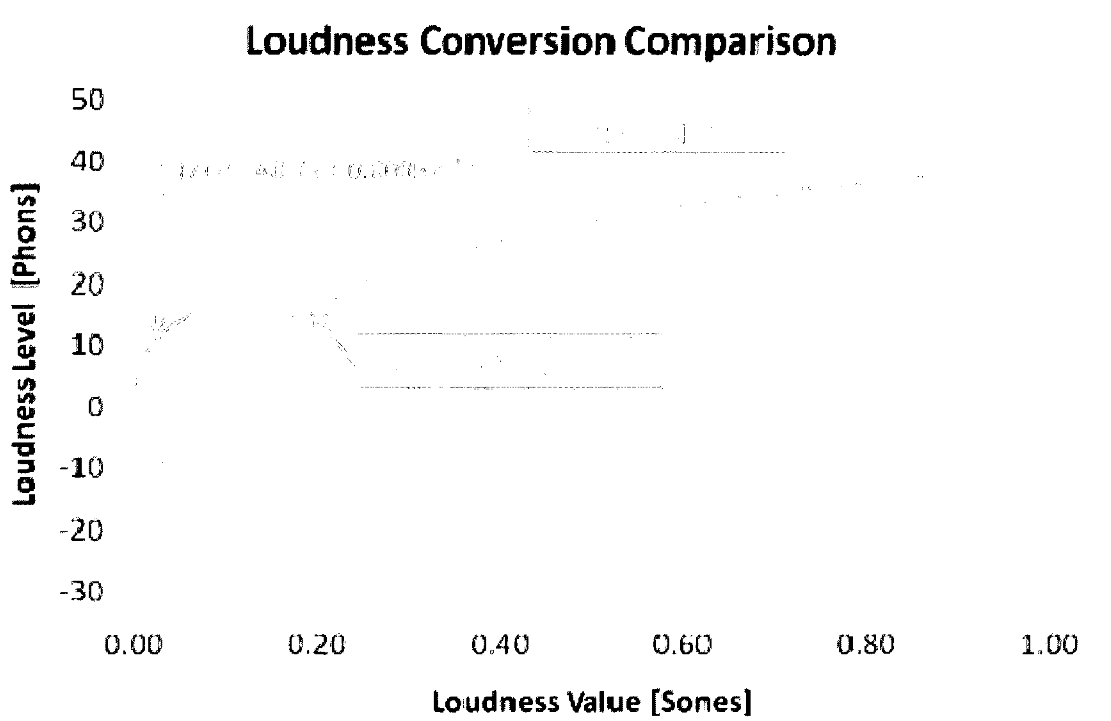


Figure 3.4 - Loudness Conversion Comparison

It can be seen from the graphical representation in **Figure 3.4** that the loudness conversion provided in **Equation (7)** creates a slightly shallower slope for loudness levels below 30 phons and ultimately drops down to 3 phons; corresponding to a loudness value of 0 sones. Likewise, the ANSI model depicts the 0 phons value as 0.0011 sones; a finite value as according to Glasberg and Moore's research where a zero loudness level shall be above the threshold allowing for lower loudness levels than the DIN. From the figure, it is clearly evident that the conversion factor presented in the ISO 532B model is unusable below approximately 0.2 sones. Below this point the curve quickly approaches loudness levels of negative infinity, values that simply do not make sense. From the definition of a threshold level, even the threshold loudness values must have a finite positive value. **[22]** To deal with this, Zwicker's 1984 BASIC program contains an initial value of 0.2 sones which corresponds to a minimum loudness level of 16.8 phons for this model. This difference in methods is important to note when comparing the different loudness models against one another, particularly in the lower amplitude regions. As a result of these findings, we now expect that the ISO 532B loudness model will perform quite poorly below 1 sone or for equal loudness contours below 40 phons.

To compare the remainder of the loudness models, a series of experiments have been conducted which test the performance of the metrics to stationary signals. The following is the experimental procedure used during this investigation.

IV. EXPERIMENTAL DETAILS

In order to perform a complete and informative comparison of the various loudness metrics available, an experimental procedure was devised which would test the performance of the models in response to stationary pure tones. The testing process was divided into two distinct experiments; these are the direct feed loudness measurements and the semi-anechoic loudness measurements. The following will provide a complete outline of the equipment and instrumentation used, as well as the environmental considerations, the design and preparation for the experiment, and the experimental procedure taken for each of the measurement procedures.

4.1 Direct Feed Measurement

Prior to any laboratory testing, preliminary experiments were set up to verify the discrepancies between the models. This "direct feed" approach describes the process used to completely model the various stationary loudness metrics in the absence of any background influences.

4.1.1 Equipment and Instrumentation

For the direct feed test, the equipment and instrumentation used was minimized to ensure simplicity. **Figure 4.1** illustrates the data acquisition hardware used during the direct feed approach. Signals were generated using a Brüel and Kjær (B&K) Portable PULSE 3560B (B-Frame) analysis system. **[64]** The front end system housed 1 output and 5 input BNC connections which provided both signal generation and acquisition. This permitted the generation and analysis of direct feed signals using only one piece of hardware. No microphone was necessary for this approach as the output BNC connector was directly fed into the BNC Input 1 of the front end using a short coaxial cable. The front end assembly was then connected to a Personal Computer (PC) located immediately next to the hardware, via an Ethernet crossover cable.

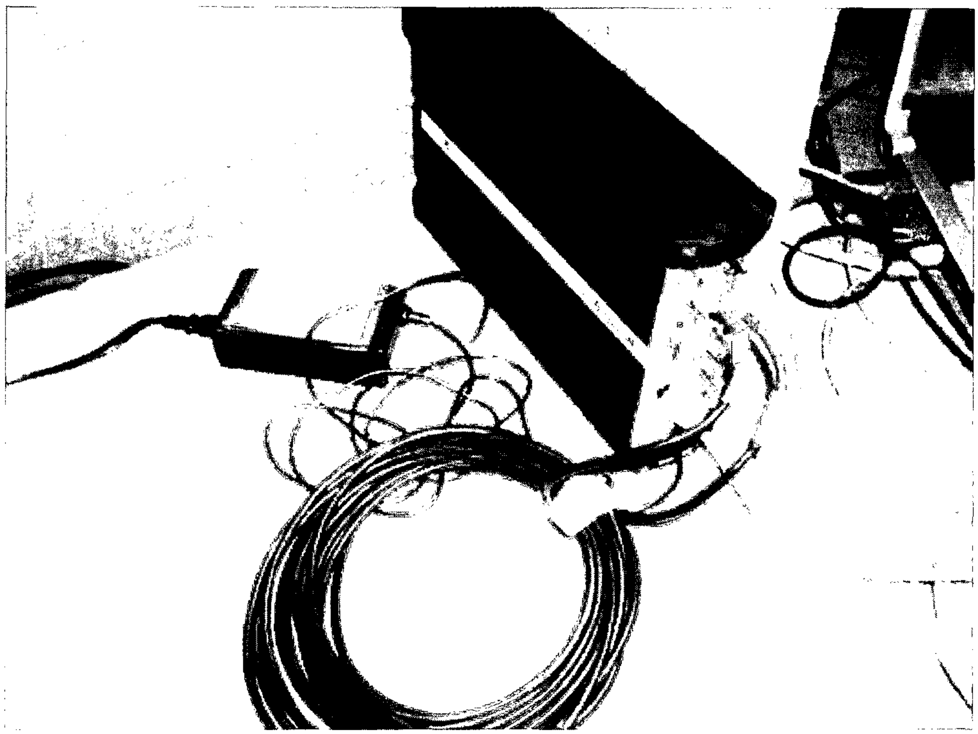


Figure 4.1 – Front end connection layout with added 2250 SLM analyzer

The real time acquisition and post processing system used here was PULSE LabShop Version 13.0.0.113 also created by B&K. [69] This software suite includes several individual software packages capable of performing a wide variety of acoustic and vibration functions including basic calculations such as the Fast Fourier Transform (FFT) and Constant Percentage Bandwidth (CPB) analysis. Alternatively it is capable of handling more complicated psychoacoustic models such as the sound quality loudness according to the DIN 45631; each on a real time or post processing basis.

As only the DIN 45631 model was included within the PULSE suite, other software packages had to be sourced for the calculation of the remaining loudness models. Mentioned previously, the ISO 532B model was available in the BASIC programming language of **[59]**. From this document Defoe (2007) generated a Microsoft EXCEL file containing the same model as a VISUAL BASIC calculator. This file was used in conjunction with the PULSE LabShop CPB values for targeting the loudness levels. To quicken this process, however, a small macro was created by the author to format and insert the data collected from PULSE into the calculator. This macro copied the necessary values from the constant percentage band (CPB) data and placed it into the input location for the calculator; only the data from 25 Hz to 12.5 kHz was used in the calculation.

In order to derive the ANSI S3.4:2007 loudness values, the previously mentioned Glasberg and Moore program LOUD2006A.exe was used once again with values recorded in PULSE. **[19]** The LOUD2006A.exe software was capable of several input options and methods. The listening conditions permitted either a free-field, diffuse-field or listening via headphones, the model applied various transfer function based on the selected option. Next the signal was presented as either a monaural or binaural signal, allowing the user to completely specify the physical environment of the tests. In order to approximate the equal loudness contours, it was decided that the best option would be to present the data as a free-field, binaural signal in order to follow how the experimental ISO 226 contours were derived, (this was also the default selection of the software). Implementing the CPB data into the model this way was a tedious process with many steps, so a macro was once again created to improve the efficiency of the calculation process. The macro copied the 50 Hz to 16 kHz range and format it to run as free field, binaural, 1/3 octave inputs in the LOUD2006A.exe program.

4.1.2 Experimental Design and Preparation

In setting up the PULSE LabShop software package for a direct feed approach, several considerations were taken into account to ensure all of the necessary data was acquired, while maintaining an accurate response. Shown in **Figure 4.2** are the input channel settings for the direct feed signal as seen from PULSE LabShop. As there was no microphone present, the sensitivity of the input signal was set manually to 1 V/Pa with a gain adjust value of 1, allowing the input signal to be adjusted directly with the voltage adjustments on the signal generator.

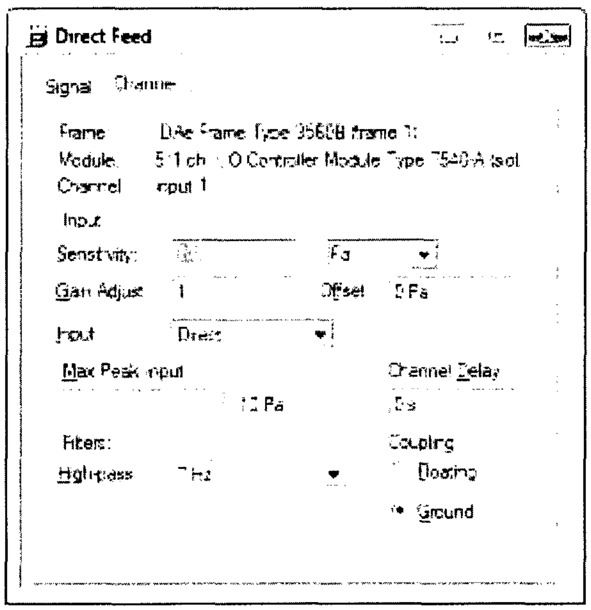


Figure 4.2 – Input signal settings for direct feed data

To monitor the data FFT and CPB data was collected simultaneously and recorded for post processing. The CPB data was collected as 1/3 octave data from

25 Hz up to 16 kHz; in order to cover the input range requirements of the various loudness models being studied. Recall that both the ISO 532B method and the DIN required 1/3 octave data from 25 Hz to 12.5 kHz, while the ANSI S3.4:2007 model was applicable from 50 Hz up to 16 kHz. By setting the range the same for every test, this would ensure applicability for all models while reducing the possibility for error.

Lastly the CPB and FFT settings were adjusted to ensure that the results indicated true stationary signals with the precise values desired. The CPB filter was set to a 1/3 octave bandwidth with a linear averaging time of 30 second. The length ensured that the pure tone had sufficiently settled before the loudness levels were recorded during the derivations. The FFT filter was set-up to include 400 lines over a span of 25.6 kHz; as the FFT filter was only used as an aid in determining the target SPL, this setting was determined to be sufficient without recording an unnecessary amount of excess data. The values from the FFT filter were then calculated from 10 averages determined exponentially.

To facilitate the time-varying loudness models, recordings of the various pure tones were made during the measurement process. These recorded files (.REC) saved all the information from the tone and could be altered using PULSE LabShop or an accompanying program PULSE Sound Quality (SQ). PULSE LabShop had a built in time-edit view where the recorded signals could be adjusted to minimize the effect of brief noise spikes or to simply shorten the signal to a desired length. REC files were made in PULSE in 5 second samples for all of the pure tones recorded. To reduce space and computational time, the signals were cropped down to 1 second for the time-varying loudness calculations. The signals were then transferred to PULSE SQ where a ramp-up filter was used to remove the 'pop' sound predicted by a loudness model if the signal were to start at full strength. Given the short time-spectrum of the signals, this would have created a loudness average much higher than the actual prediction intended. These steps were used for both time-varying models being studied.

While deriving the DIN 45631/A1 non-stationary loudness contours, it was decided that the best approach would be to use the same pure tone signals recorded for the stationary model, and simply input the REC files into the non-stationary model. It was hypothesized that if the procedure of the models closely resembled one another, the exact same results should occur. A draft version of the DIN 45631/A1 time-varying loudness model was a built in feature of PULSE SQ based on binaural information. Loudness levels were read directly from this program based on the filtered tones and recorded into an EXCEL file.

For the time varying loudness (TVL) model, a different approach had to be used as this model uses a scaling approach where a recorded signal is inserted into the model and calibrated as a full scale sinusoid at 100 dB. To collect target loudness information, the full-scale sinusoid was then scaled down to the desired level appropriately. For levels below 50 dB, the full-scale sinusoid was calibrated to 50 dB and scaled down from there. This was done so as not to lose valuable spectral information in the scaling process. Based on this procedure, the equal loudness contours for the TVL model were derived based on the findings from the stationary curves. The 100 dB full-scale tones were scaled down to levels matching the SPLs derived in the stationary model for each frequency targeted – once again, theoretically resulting in the same conclusions as the stationary model.

4.1.3 Environment Considerations

By performing the measurements through a direct feed approach, the intent was to derive an uninfluenced resultant signal, producing as pure of a tone as possible. With the absence of an ambient noise source, 'ideal' loudness levels could be predicted accurately down to near threshold levels; levels that would be impossible in most laboratory settings. **Figure 4.3** for example shows a 10 dB pure tone played at 1 kHz where it is clear from this sample that no disrupting background noise is present above 1 dB for the direct feed data. Therefore, no added consideration was needed for any environmental influences of the direct feed approach.

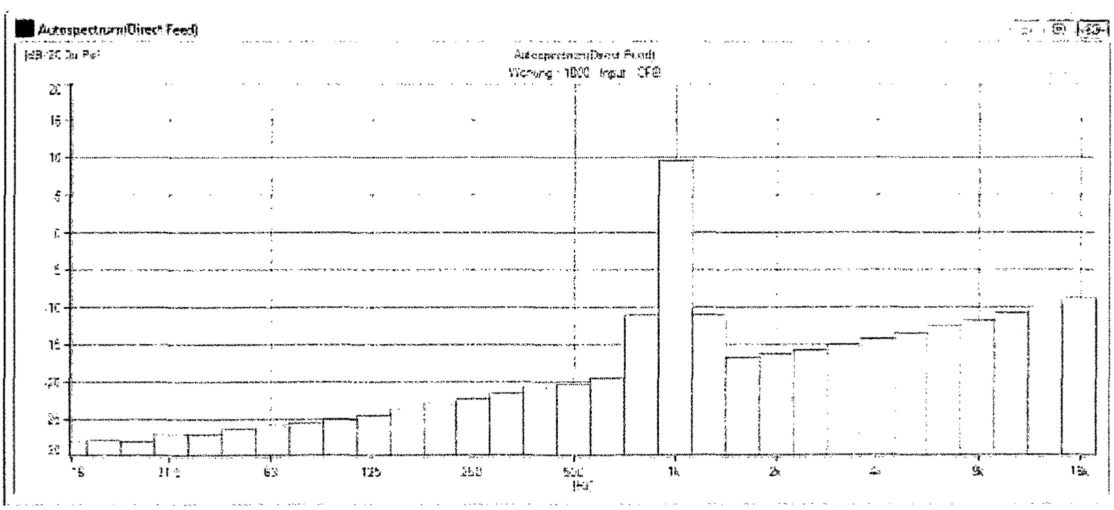


Figure 4.3 – A CPB example of a 10 dB direct feed pure tone at 1 kHz

4.1.4 Testing Procedure

As there was no simple way of collecting the equal loudness contours for the various metrics, a bracketing method was employed to locate and record the target SPL values. The following describes the procedure used:

4.1.4a - Stationary Loudness Procedure

1) Record a pure tone for the target frequency value.

2) Each individual loudness model had a particular bracketing approach, varying higher than lower until the desired level was reached:

   A. **DIN 45631**
   Judging from the loudness level indicated by PULSE LabShop, adjust the voltage signal level accordingly until the desired loudness level is met.

   B. ISO 532B
   Copy the SPL data from the CPB of PULSE LabShop and paste it into the EXCEL formatting macro. After formatting (which collects the data from 25 Hz to 12.5 kHz), proceed to calculating the ISO 532B loudness level and if necessary go back to PULSE to adjust the voltage level accordingly.

   C. ANSI S3.4:2007
   Copy the SPL data from the CPB of PULSE LabShop and paste it into the EXCEL formatting macro. Paste the resulting formatted data (from 50 Hz to 16 kHz), into a text file (.TXT) and run the LOUD2006A.exe program. If necessary based on the resultant loudness value, go back to PULSE and adjust the voltage level accordingly.

3) Once the target loudness level has been located, save the CPB and FFT data into PULSE and record the overall SPL and loudness level into an EXCEL file for plotting.

4) Save the pure tone recorded file (REC) for implementation into the time varying loudness models.

5) Continue the process until all possible loudness contours are derived to be compared against the reference contours of ISO 226:2003.

   NOTE: A secondary procedure was used for determination of the ANSI S3.4:2007 contours due to the multiple input methods. Aside from the 1/3 octave data, the contours were also derived using the pure tone specification method of the LOUD2006A.EXE program. Using this approach, pure tones were specified at each centre frequency and varied in SPL until the target loudness was observed. This process was carried out until the complete set of contours was derived from 25 Hz to 12.5 kHz.

4.1.4b - Time-Varying Loudness Procedure

As motioned above, separated steps had to be used for the time-varying loudness models. These steps are outlined as follows:

1) Open the pure tone REC file in PULSE time-edit and cut the sample size down to 1 second.

2) Apply a ramp-up filter to the tone to remove the "pop" from the calculations.

3) From here the two methods take on separate approaches:

   A. DIN 45631/A1

   i. Record resulting binaural loudness level according to DIN 45631/A1 DRAFT into an EXCEL sheet for plot.

   B. Glasberg and Moore's Time-Varying Loudness (TVL)

   i. Export the resulting signal as a wave audio file (.WAV)

   ii. Import the audio file into the included resampling editor from the TVL model, (the file resamples the WAV signal from 16 kHz to 32 kHz).

   iii. Import the adjusted WAV file into TVL.EXE and calibrate to a full scale signal of 100 dB, (or 50 dB if targeting levels below 50 dB).

   iv. Scale down the calibrated signal to the desired SPL values and record the short term average level into EXCEL for plotting.

4) This procedure is continued until all of the applicable loudness levels are derived and plot on a single graph. The resultant curves should resemble straight lines.

4.2 Semi-Anechoic Measurement

The primary focus of the semi-anechoic procedure was to back up and verify the results generated through the direct feed approach. In order to accomplish this, physical measurements were conducted through the use of loudspeakers, microphone transducers and a semi-anechoic environment.

4.2.1 Equipment and Instrumentation

A semi-anechoic room, like the one used in this study, is an acoustically treated testing environment which approximates a free field listening condition. In this simulated atmosphere, sound waves are free to travel without any obstructions other than the limits of the device being tested and the negligent reflections from the acquisition equipment. The room is lined with acoustical absorbing wedges where theoretically all of the acoustic energy will be absorbed rather than being reflected. The wedges also act as an insulator for ambient noise sources that may pass through from outside the room, (See **Figure 4.4**).

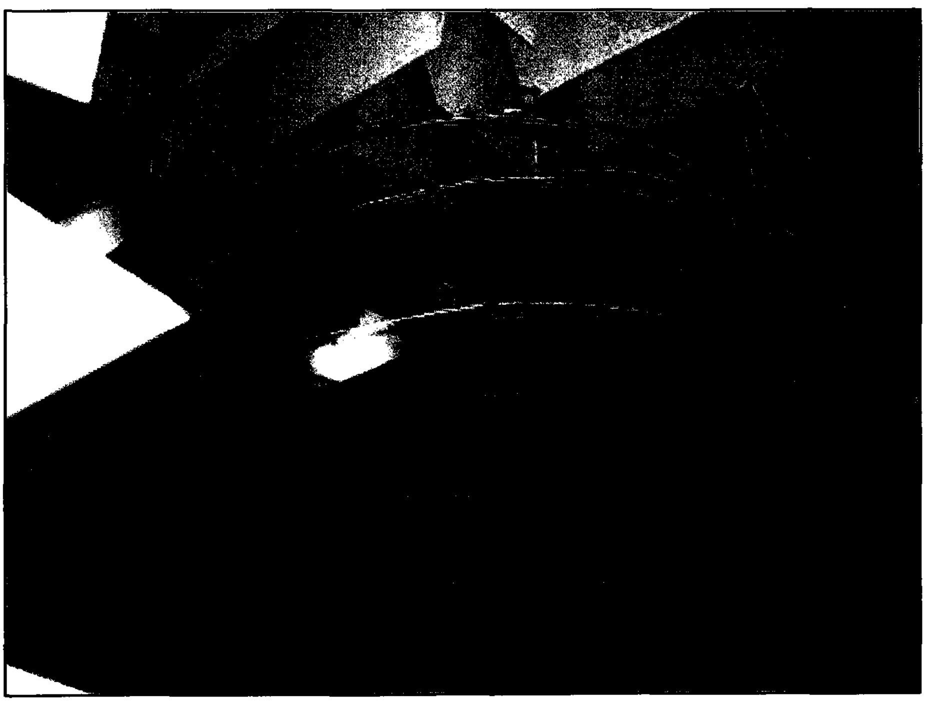


Figure 4.4 – Ceiling and walls of semi-anechoic room line with 1 metre wedges.

The acoustical wedges are design sufficiently large enough to create a desired cut-off frequency for testing. The wedge length should be one-quarter of the desired lower cut-off frequency, so the approximate 1 metre (m) wedges in the semi-anechoic room located at the University of Windsor were capable of a lower cut-off frequency of less than 100 Hz, (See **Appendix A** for calculation table).

Measurements were again conducted using B&K's B-Frame front end, only this time with the use of an OmniSource sound speaker as the output driver. Both a binaural head and an external microphone were used for collecting the signals.

To generate the desired levels of pure tones in an appropriate manner for the laboratory test, a Brüel and Kjær Type 4295 OmniSource™ Sound Source was used in accordance to a Brüel and Kjær Type 2716 Audio Power Amplifier. **[68, 63]** Using this combination, pure tones were accessible across the desired frequency range for nearly all of the required sound pressure levels. The highest SPLs could not be reached without risk of damage to the OmniSource™ sound source and an alternative model had to be used. While the OmniSource™ speaker was capable of emitting a sound power level of 105 dB, the next step up from this model is the Type 4292 OmniPower™ sound source, capable of a much higher sound power level at 122 dB. Due to the time constraints for borrowing the OmniPower™ sound source only the highest measurements were recorded with the louder speaker. At these levels, the Type 4295 model created unwanted, off-frequency noise when the speaker was pushed to higher amplitudes. This noise would have negatively influenced the results and was, therefore, avoided by using the Type 4292 model whenever possible. Each of the speakers listed above are capable of producing a multi-directional sound spectrum, radiating sound evenly in all directions. Based on the free-field environment, the pure tone signals could then be measured anywhere in the proximity to the source, at various locations simultaneously, (as long as the distance from the source remained constant).

The acoustic signals were acquired via two measurement devices, a Brüel and Kjær Type 4100-D Head and Torso Simulator (HATS) and an external Brüel and Kjær Type 4189 ½" prepolarized free-field microphone. **[65, 66]** In order to record binaural information, the HATS device was included in this experiment as it best approximates the human perception based on physical features. The simulator includes a formed manikin surface and moulded-silicone pinnae that approximate the physical geometry of the average adult head and torso. **[65]** To further adjust the diffraction of sound energy, the torso is covered with a damping fabric. Two microphones are placed at the entrances to the ear canals to acquire the sound signals where the result provides separate spatial information from both ears. The microphones are Brüel and Kjær Type 4189 ½" prepolarized free-field microphones with a nominal sensitivity of 50mV/Pa., (individually calibrated). As the HATS device includes the affects of the torso, head, and pinna on a sound signal, the resultant information gives the user an accurate three dimensional recording. **[65]** A HATS device is used to simulate the binaural recording of information. This includes the added influence of the upper torso on signals travelling to the ears, providing the best acoustic approximation of the human body available. To further simulate the effects of the ear, head and torso transfer functions are added to the recorded signals to account for these reflections when the signals are analyzed. Binaural information was required as some metrics being studied require binaural signals to calculate the loudness levels; particularly the time-varying loudness metrics. In this case, the DIN 45631/A1 time-varying loudness model was the only metric using the binaural response. For the remaining measurements, the binaural head served as an added measurement for quality assurance purposes.

A third signal was recorded with an external microphone mounted on a tripod. Again a Brüel and Kjær Type 4189 ½" prepolarized free-field microphone was used here as it is capable of a broad measurement spectrum and was designed for high precision. **[66]** With a dynamic range of 14.6 dB to 146 dB and a frequency range from 6.3 Hz to 20 kHz, the microphone was fully capable of recording the equal loudness contours from 20 to 100 phons. This signal was used as the monaural input for the remaining loudness metrics. All of the microphones remained in the same locations for the duration of the experiments to ensure consistency.

To calibrate the microphones, a Brüel and Kjær Pistonphone Type 4228 calibrator was used at the beginning and end of each testing run. **[67]** This was done as a constant check to ensure that there were no problems with the microphones being used. Over the three month testing period, the left and right ear had an average gain value of 1.03 each, while the external microphone had an average gain of 1.04. These values were consistently observed and only one data point was particularly off the norm, (See point 04/08/09 in **Figure 4.5**). However, this point was calibrated at the end of a test run as a check and most likely resulted from user error when calibrating. As the calibration for this data point was done after testing had been completed for that day, the gain value of 1.07 was not used to record any test results. Since the gain values returned to the "normal-range" the following morning, no errors were associated with this occurrence. See **Appendix B** for a complete table of the collected gain values.

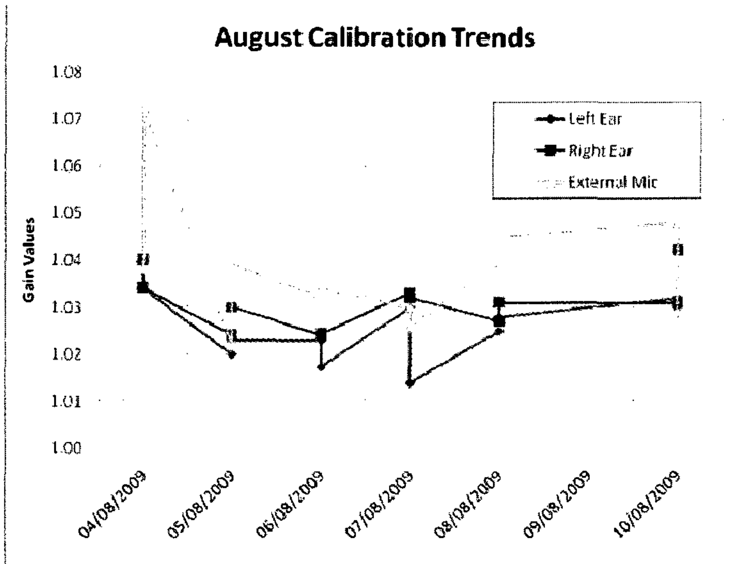


Figure 4.5 – August calibration trends

4.2.2 Experimental Design and Preparation

As with the direct feed method, PULSE LabShop was used for the signal generation and acquisition. The majority of the settings remained the same such as those for the FFT and CPB. However, for the semi-anechoic approach, there were now three separate inputs which had to be set up. Each signal was collected as both a FFT and a CPB, while the signal itself was recorded as a .REC file for insertion into the time-varying loudness metrics.

For the set-up, the measuring devices used in this experiment were placed at a distance of one metre, symmetrically on either side of the sound source. On one side sat the single external microphone transducer mounted on a tripod at 1.5 metres. On the other side sat Head and Torso Simulator with the ears levelled to 1.5 metres above the floor for consistency. This assembly was centered in the semi-anechoic room and remained essentially the same for all of the measurements taken, see **Figure 4.6**. The only change to the set up was the replacement of the OmniSource speaker with the larger OmniPower speaker for the higher amplitude measurements, see **Figure 4.7 and 4.8**.

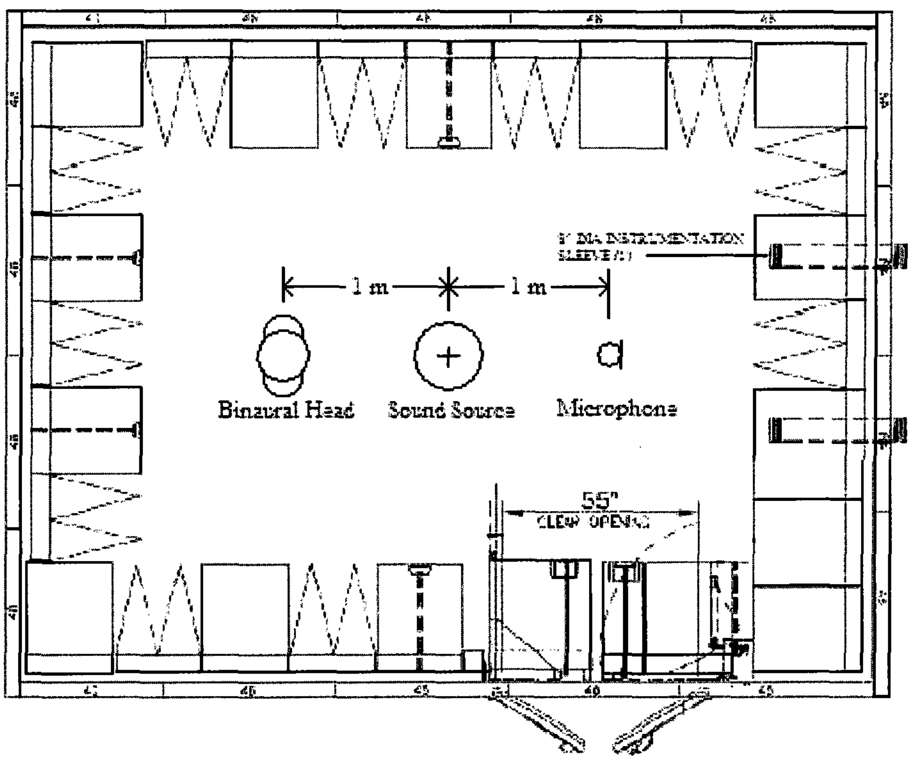


Figure 4.6 – Semi-anechoic room layout

Figure 4.7 – Binaural head, OmniSource sound source and external microphone

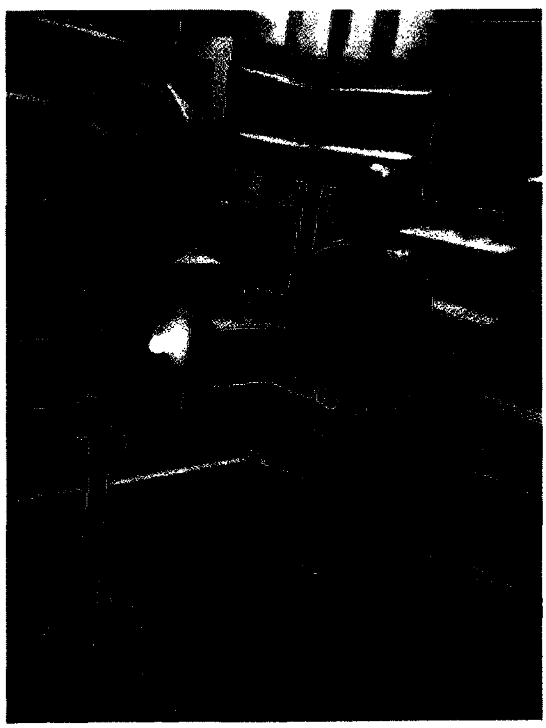


Figure 4.8 – Binaural head and OmniPower sound source

Outside of the semi-anechoic room, the acquisition equipment is separated from the testing environment. The equipment remains connected to the measurement devices through insulated 8 inch instrumentation sleeves, (seen in **Figure 4.7**). The signal path is as follows: the B-Frame produces the generated pure tones and feeds the signal into the signal amplifier. The amplified tone is then fed to the sound source via a speaker cable through the insulated sleeve. The recorded signal from microphones travels back through the wall and into the input BNC connections of the B-Frame and into the PC via an Ethernet cable. The exterior setup can be seen in **Figure 4.9**.

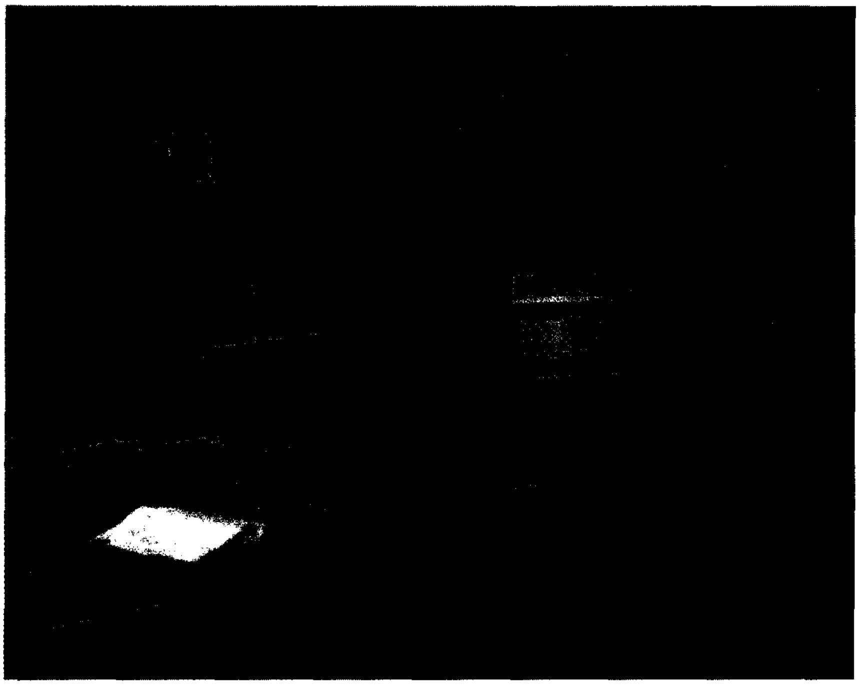


Figure 4.9 – Exterior acquisition equipment set-up

4.2.3 Environment Considerations

When recording acoustic signals for laboratory use, one must consider the effects of the environment on the acquired data. Therefore, as a daily check prior to collecting data the temperature and humidity of the semi-anechoic room was recorded. Measurements were made via the Kestrel® 4000 Pocket Wind Meter;

recording both the air temperature and the relative humidity. From the measurement results, it was confirmed that the environmental conditions were in fact favourable for experimental readings; the microphones used had an operating temperature of -30°C to +150°C and an operating humidity range of 0 to 100% (without condensation). The average values recorded during the testing times were a room temperature of 22.3°C and an absolute humidity of 47.2%.

Background noise is also an issue when targeting values in the low amplitude regions for acoustical tests. Therefore, to ensure the collection of the most accurate data, measurement times were staggered to work around other school activities that may negatively influence the results. Semi-anechoic tests were run only at times when the laboratory was free of students, and the traffic entering and exiting the building was minimal; reducing the amount of doors opening and closing, a common low frequency noise source. Therefore, the majority of the data gathered was collected between the hours of 10:00 P.M. and 6:00 A.M. when background noises levels were at their lowest.

Even with the staggered timeslot, a substantial amount of background noise was still present in the semi-anechoic room; values were generally recorded with excessive acoustic information located below the 100 Hz cut-off frequency as established by the wedge depth mentioned above, (see **Figure 4.10**). Therefore, to solve this problem, a weighting function was applied below 100 Hz reducing the high peak to more acceptable levels. This way the majority of the unwanted data was removed, permitting lower calculated loudness levels. This function substantially reduced the background noise levels, allowing for more accurate 'measured' values to be collected, as can be seen in **Figure 4.11**.

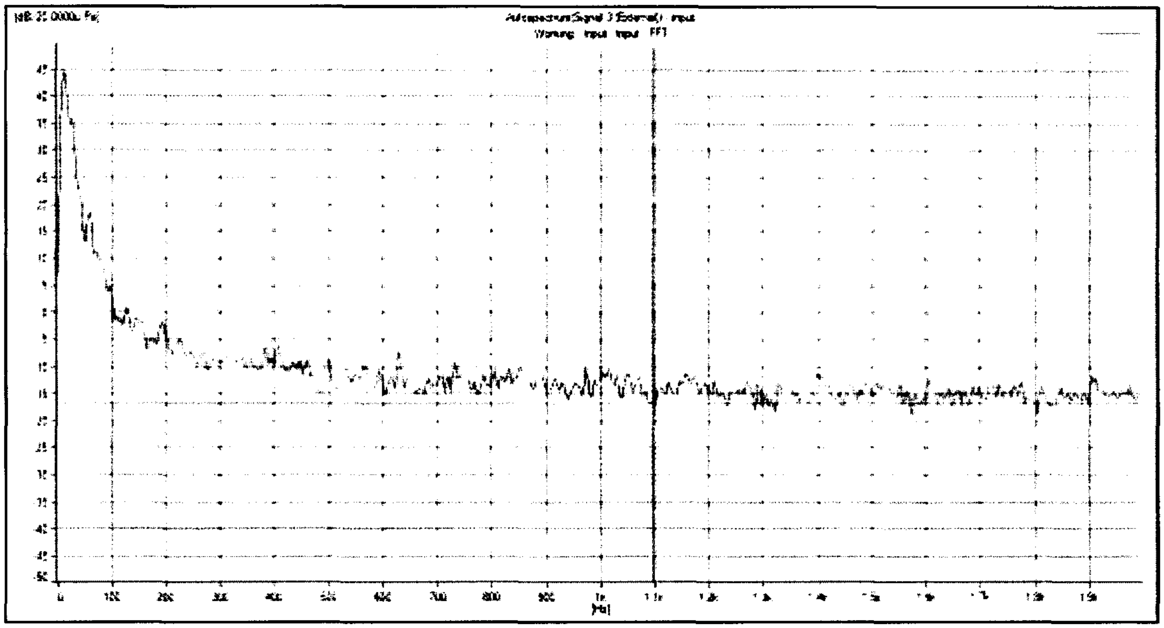


Figure 4.10 – FFT of background noise present in semi-anechoic room

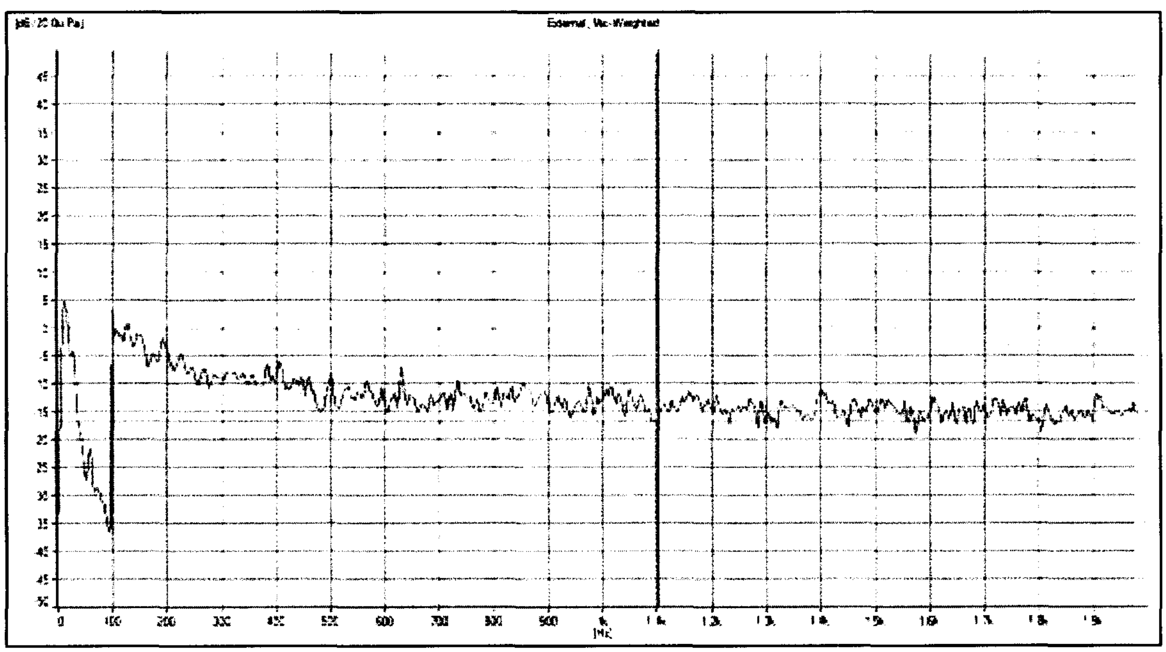


Figure 4.11 – Weighted FFT of background noise present in semi-anechoic room

After reducing the sound pressure levels below 100 Hz, the background noise was still measured at around 18-20 dB, creating a lower limit for what loudness levels could be recorded. This added another restriction to the areas of the equal loudness contours that could be explored using the semi-anechoic room process; the limitations now being a lower cut-off of frequency 100 Hz, a speaker dependant upper amplitude cut-off, and lower cut-off amplitude of 20 dB.

4.2.4 Testing Procedure

The procedure used for collecting loudness information in the semi-anechoic room was essentially the same as that of the direct feed method. Only a few small steps were added to permit boosting the signal through an amplifier and running daily calibrations and environment checks for quality control. The Procedure was as follows:

4.2.4a - Stationary Loudness Procedure

1) Calibrate all three microphones at the start of each testing run.
2) Collecting semi-anechoic room temperature and humidity levels.
3) Adjust the amplifier until pure tone levels are within the target range without distorting the signal.
4) Record a pure tone for the target frequency value.
5) For all three models, the external microphone information was used for targeting the loudness contours. This signal did not require a transfer function to account for the head and torso effects, and presented a monaural signal that the Zwicker models required. Each individual loudness model had a particular bracketing approach:

   A. **DIN 45631**
   Judging from the loudness level indicated by PULSE LabShop, adjust the voltage level accordingly until the desired loudness level is met.

B. **ISO 532B**
Copy the SPL data from the CPB of PULSE LabShop and paste it into the EXCEL formatting macro. After formatting (which collects the data from 25 Hz to 12.5 kHz), proceed to calculating the ISO 532B loudness level and if necessary go back to PULSE to adjust the voltage level accordingly.

C. **ANSI S3.4:2007**
Copy the SPL data from the CPB of PULSE LabShop and paste it into the EXCEL formatting macro. Paste the resulting formatted data (from 50 Hz to 16 kHz), into a text file (.TXT) and run the LOUD2006A.exe program. If necessary based on the resultant loudness value, go back to PULSE and adjust the voltage level accordingly.

6) Once the target loudness level has been located, save the CPB and FFT data into PULSE and record the SPL and loudness level into an EXCEL file for plotting.

7) Save the pure tone recorded file (REC) for implementation into the time varying loudness models.

8) Continue the process until all possible loudness contours are derived to be compared against the reference contours of ISO 226:2003.

9) Calibrate all three microphones again to ensure there are no problems with the recorded results.

4.2.4b - Time-Varying Loudness Procedure

The time-varying loudness procedure remains essentially the same as mentioned in the design. One step was added to the process concerning the background noise. When implementing the time-varying loudness models from the anechoic results, the weighting function indicated above only applied to the post process of the recorded signals. Therefore, when the REC file was made for each of the generated tones, this weighting was not included. To remove this noise, a high pass filter was included into all of the tones during the Sound Quality modifications and the procedure then carried on as before. Although all of the software settings remained essentially the same, the environment where the recordings were made warranted some unique considerations. These steps are outlined as follows:

1) Open the pure tone REC file in PULSE time-edit and cut the sample size down to 1 second.

2) Import the binaural sampled tone into PULSE Sound Quality using the HATS automatic transfer function to account for head and torso interaction.

3) Apply a ramp-up filter to the tone to remove the "pop" from the calculations.

4) Apply a High Pass filter to remove the content below 100 Hz.

5) From here the two methods take on separate approaches:

   A. DIN 45631/A1

      i. Record resulting binaural loudness level according to DIN 45631/A1 DRAFT into an EXCEL sheet for plot.

   B. ANSI Time-Varying Loudness (TVL)

      i. Export the resulting signal as a wave audio file (.WAV)

      ii. Import the audio file into the included resampling editor from the TVL model, (resamples the WAV file to 32 kHz).

      iii. Import the adjusted WAV file into TVL.EXE and calibrate to a full scale signal of 100 dB, (or 50 dB if targeting levels below 50 dB).

      iv. Scale down the calibrated signal to the desired SPL values and record the short term average level into EXCEL for plotting.

6) This procedure is continued until all of the applicable loudness levels are derived and plot on a single graph. The resultant curves should resemble straight lines.

V. ANALYSIS OF DATA AND OBSERVED RESULTS

The results realized using the procedure described in the previous chapter are discussed in regard to the performance observations. The intention here is to provide an in-depth comparison of the various loudness metrics studied and to criticise each on their performance when compared against the standardized set of reference curves of the ISO 226:2003 document. As in the order of collected data, the direct feed derivations will be compared first. This will include plots of both the stationary and non-stationary loudness models for comparison, plot on a common graph with the reference contours. Lastly, the verification plots as recorded in the semi-anechoic room will be analyzed and compared to the results of the direct feed approach. This will be done in order to check the repeatability of the experiment while confirming the initial results. The comparisons of these plots will be followed by a discussion prior to any conclusions.

5.1 Direct Feed Results

A direct feed approach was chosen as an initial investigation since it was determined to be the best approach for gathering a wide range of data while removing the risk of extraneous noise sources contaminating the raw data. The performance of three stationary loudness models and two non-stationary loudness models is compared against the reference contours chosen at the beginning of this study. When arranging the order for the comparison, the stationary models will be compared according to the date they were accepted as this provides the reader with a perspective on the improvements resulting from research conducted over the years. The direct feed time-varying loudness models will then be compared to their stationary predecessors on performance. Recall that it is expected that they should produce identical results since the time-varying calculations were derived as a result of the stationary methods.

5.1.1 ISO 532B (1975)

As one of the first standardized stationary loudness models accepted, the ISO 532B has a long history of use and subsequently has roots in all of the loudness models which followed thereafter. Using 1/3 octave inputs from 25 Hz to 12.5 kHz, the equal loudness contours were derived and plotted against the ISO 226:2003. **Figure 5.1** illustrates the contours as predicted by the ISO 532B for direct feed pure tones.

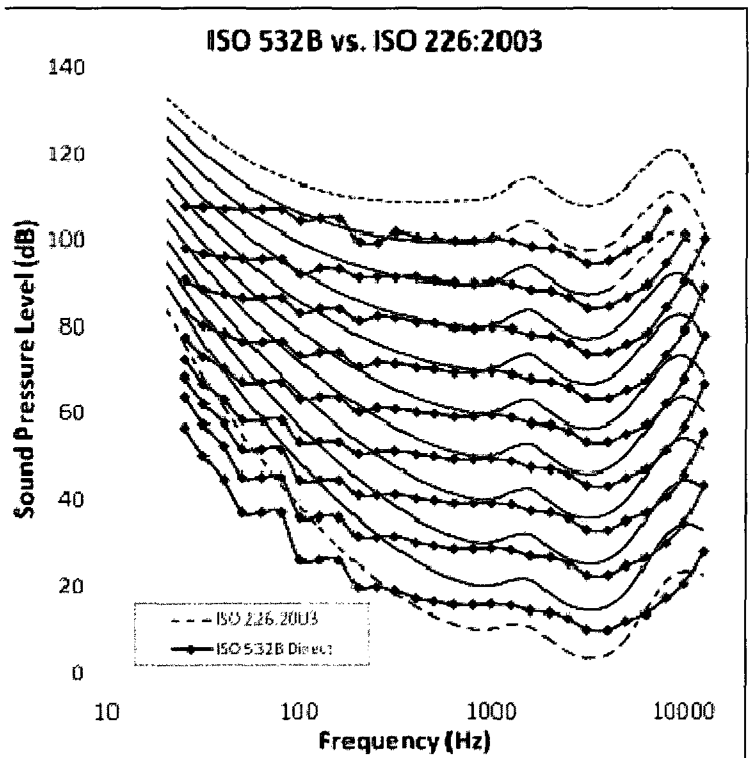


Figure 5.1 – ISO 532B compared against the ISO 226:2003 reference

From the figure it is clearly evident that the 1975 model for stationary loudness was not intended to predict contours of this shape. The loudness metric predicts trends of a very shallow slope and does not acknowledge the added hump above 1 kHz. Below 300 Hz, the ISO 532B model takes on an interesting reaction to the tones, wavering roughly with plateaus and steep jumps in loudness. Hellman in 2007 explains the jump in data as a result of using a set of tabulated values rather than equations. **[22]** This wavering makes the application of the ISO 532B model unsuitable as an accurate prediction model; especially below 300 Hz for any level of loudness.

As mentioned previously, the loudness conversion used by the ISO loudness model is not correct for levels below 40 phons. This is again quite evident in the figure as the two lowest loudness curves, the 20 and the 30 phon, exhibit greater deviation as the levels decrease. Thus, below a loudness level of 40 phons, the ISO 532B document should not be used.

It is important to point out that if one were to compare the ISO 532B loudness model to the equal loudness contours of 1987, it is evident that for the time it was developed, the model predicted the contours very well above 300 Hz and 40 phons (see **Figure 5.2**). As mentioned previously, the 1987 version of the contours were a lot shallower with distinctly less curvature in the mid-frequency range; shapes that the 1975 version of Zwicker's method was able to follow nicely. Therefore, prior to the update, the ISO 532B document was an accurate model based on the current information available at that time, but there was an obvious need for improvements below 300 Hz.

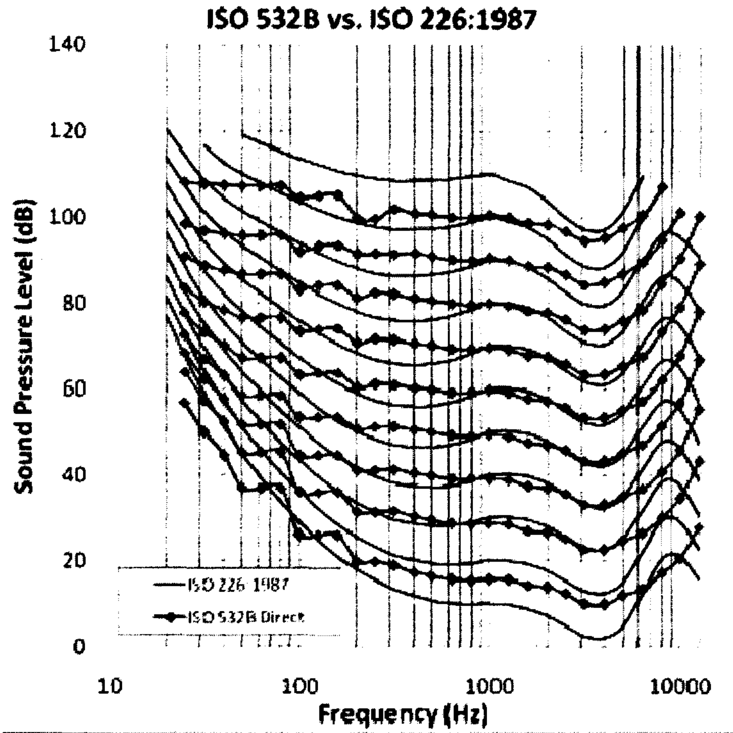


Figure 5.2 – ISO 532B compared against ISO 226:1987 contours

To quantify the relationship between the reference loudness contours of ISO 226:2003 and the ISO 532B model, correlation coefficients were calculated for each contour line. **Table 5.1** is a compilation of all the correlation coefficients for both the 2003 equal loudness contours and the 1987 version for comparison.

Table 5.1 – Overall correlation coefficients for comparison between ISO 532B and ISO 226 equal loudness contours

Phon Contour	10	20	30	40	50	60	70	80	90	100
ISO vs. ISO226:1987	--	0.973	0.949	0.907	0.838	0.763	0.681	0.638	[redacted]	
ISO vs. ISO226:2003	--	[redacted]							0.801	0.819

The correlation coefficients determine how well predicted values approximated the actual or past derived values. A value of 1 indicates a perfect match while a value of 0 would indicate no relationship at all. The correlation coefficients determined here demonstrate how well the ISO 532B model predicts the contour shapes of the ISO 226 equal loudness contours. From the values of **Table 5.1**, it is interesting to note the specific trends present upon a closer inspection. For instance, a common occurrence between each set of the reference contours is the fact that, as levels increase in loudness, the slope of the low frequency portions of the curves tend to become slightly shallower. While the ISO 532B model tends to get shallow as well, it does so more rapidly; almost to the point where it seems to progress horizontally. This would partially explain why the correlation coefficients progressively diminish with increasing loudness levels. Secondly, even though the predicted loudness contours from the ISO 532B appear to be better correlated with the 1987 version of the ISO 226 contours, the overall trends of decreasing and increasing slopes better match with the revised set of contours. For instance, while the 1987 version of the contours level off at around 400 Hz, and at times have a positive slope, the ISO 532B model is still sloping slightly negative which better corresponds to the more recently derived trends of the ISO 226:2003. These trends help explain the better correlation between the newer reference curves, even though the previous model appears to be a better match. Based on these observations, comparisons cannot be made using a visual comparison or correlation coefficient alone. Both of these observations will be used together to determine the best performance overall.

5.1.2 DIN 45631 (1991)

The DIN 45631 stationary loudness model is essentially an updated version of the ISO 532B method indicated above. As such, it was expected to perform at least as good, if not better than its predecessor. **Figure 5.3** is a plot of the experimental results generating the predicted equal loudness contours according to the DIN 45631 stationary loudness model.

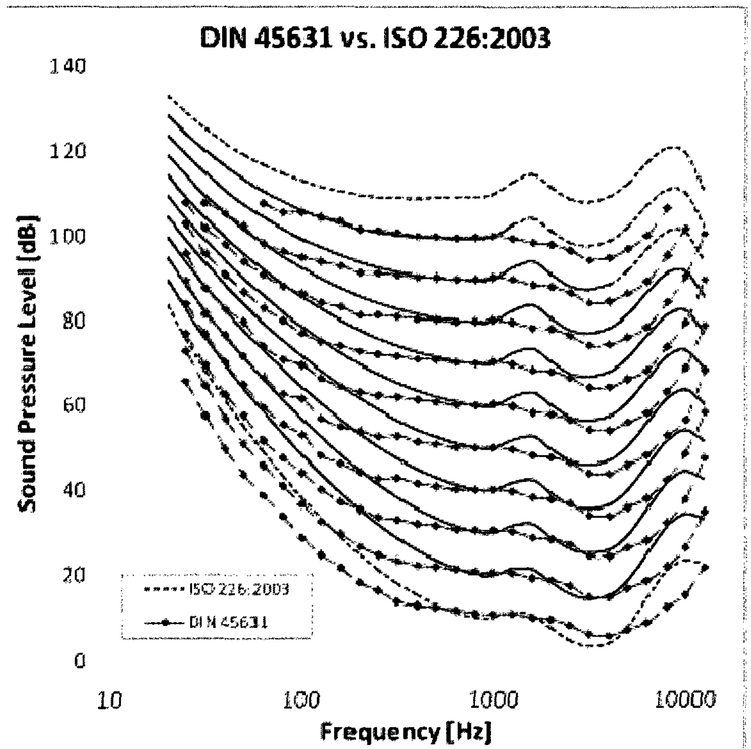


Figure 5.3 – DIN 45631 compared against the ISO 226:2003 reference

At first glance it is evident that this model is a large improvement over the ISO 532B version. The low frequency wavering has been, for the most part, completely corrected and the smooth curves retain their slope at the higher amplitudes as well; an area where the previous model levelled off. The DIN 45631's improved loudness conversion enables the lowest contours to correlate well with the expected values at the 1 kHz point. From the modification, the DIN model is also capable of determining loudness levels well below its predecessor version, permitting the derivation of the 10 phon line shown here as the lowest contour. Recall that the ISO 532B was only capable of prediction loudness levels down to a lowest level of 16.8 phons due to the programs starting value of 0.2 sones. In the update, however, there was still no improvement for the loudness 'bump' after 1 kHz, and although the DIN version of the Zwicker method is greatly improved over the ISO 532B version, the contours are still much too shallow for the ISO 226:2003 data set.

As with the previous Zwicker model, the DIN 45631 metric was created to approximate the 1987 version of the equal loudness contours. Therefore it is no surprise that it performs better to the previous standard as seen in **Figure 5.4**.

Here it clearly visible how much the improvements to the Zwicker approach resulted in a better curve match, particularly in the lower loudness regions. The DIN 45631 model proved to be an excellent approximation of the previous set of contours. Where the ISO 226:1987 contours tend to plateau in the 400 Hz range, the DIN model still did not approximate the dip well, but as seen in **Figure 5.3** the model accurately follows the trends of recent experimental data in this range.

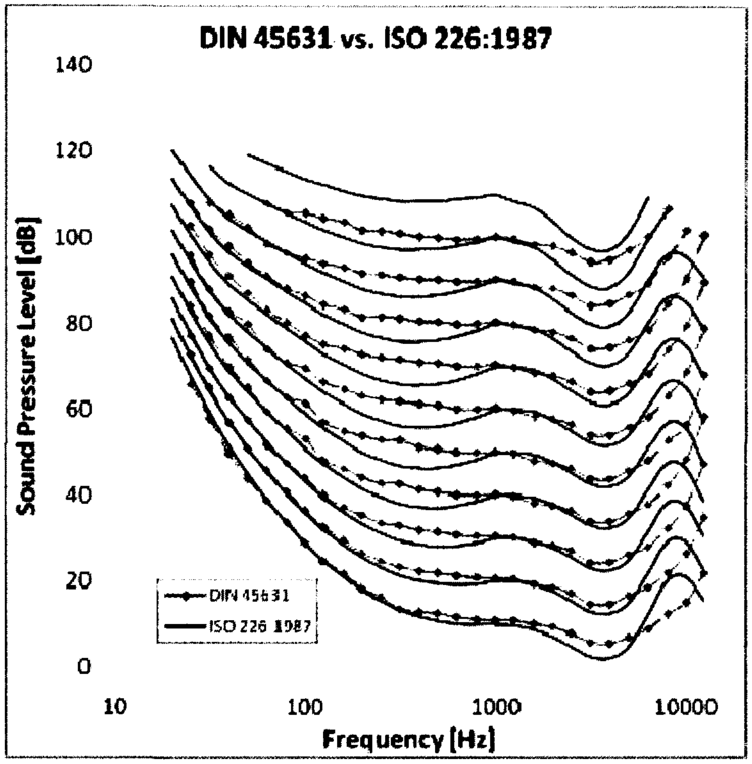


Figure 5.4 – DIN 45631 compared against the ISO 226:1987 reference

When comparing the overall correlation coefficients of the DIN 45631, it is apparent that as with the ISO 532B model, the contours better correlate to the updated reference values, although the difference between the two coefficients is not as large.

Table 5.2 – Overall correlation coefficients of DIN 45631

Phon Contour	10	20	30	40	50	60	70	80	90	100
DIN vs. ISO226:1987	[redacted]	[redacted]	0.966	0.951	0.933	0.913	0.900	0.884	[redacted]	
DIN vs. ISO226:2003	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	0.946	0.805

The DIN model has very high correlation values, particularly for the lower loudness contours for both reference curve data sets. Once again it is interesting to note that visually the DIN 45631 model is almost an exact match to the ISO

226:1987 contours below 20 phons yet the same correlation coefficient value is calculated for the 2003 contours. The identical calculated value is of course a coincidence based on the slope trends present in the model, as opposed to the actual performance of the predictor. Again, this shows why both forms of comparison are needed to make an informed decision.

5.1.3 ANSI S3.4:2007

As indicated previously, the ANSI S3.4:2007 stationary loudness model was the only metric studied that was updated to account for recent changes to the ISO 226:2003 document.

When examining the performance of the ANSI standard, two separate methods were applied. For the purpose of the comparison, the pure tone specification method using the LOUD2006A.EXE software will henceforth be labelled as the "Program" approach and the 1/3 octave input method will be labelled as the "Direct" approach. Using these labels, each input method will have individual comparisons against the ISO 226 reference contours before being compared against each other. The separate graphs were generated in order to reduce confusion in the plots.

Figure 5.5 represents the results of the Program approach for the ANSI stationary loudness model. Looking at the results, the ANSI S3.4:2007 equal loudness contours span the entire frequency range of the ISO 226 reference contours, (from 20 Hz up to 12.5 kHz).

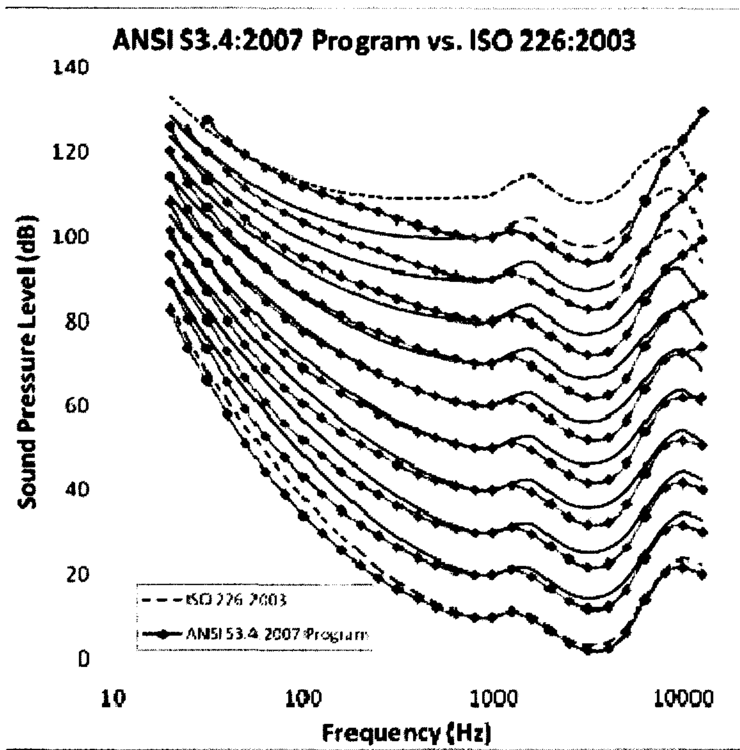


Figure 5.5 – LOUD2006A.exe predicted equal loudness contour comparison.

From **Figure 5.5** it can be seen how well the Program method correlates with the equal loudness contours of ISO 226:2003. This is particularly true in the low frequency, high amplitude regions of the plot. For the 70 phon equal loudness contour, the low frequency data is essentially on top of the target values. Although the derived contours from the program seem to acknowledge the bump at 1 kHz, they still do not follow it completely at the higher amplitudes. One important trend to note is that the ANSI S3.4:2007 model appears to be the only model that is able to predict the high frequency drop-off after 10 kHz. Although the drop is only present below loudness levels of 30 phons, all other models examined thus far predicted concave, upward slopes in this region; an area where even the 1987 version of the equal loudness contours sloped downward. It is then possible to conclude that the ANSI S3.4:2007 model appears to be the best performing model above 5 kHz.

The alternative input approach, the 1/3 octave inputs from 50 Hz to 16 kHz, substantially restricts the application of this model in the low frequency regions, (the Program method was applicable down to 20 Hz – a reduction of four 1/3 octave bands). As this approach used the 1/3 octave information from PULSE as its input, it was determined that the Direct approach was the most appropriate input method for this comparison. The Direct method relies on the experimental data rather than the internal equations within the LOUD2006A.exe software. Therefore, it results in a more "accurate" representation of the ANSI S3.4:2007 predicted contours as it would be used in the real world applications. As a result, **Figure 5.6** is the experimental derivation of the Direct approach which will be used for this comparison.

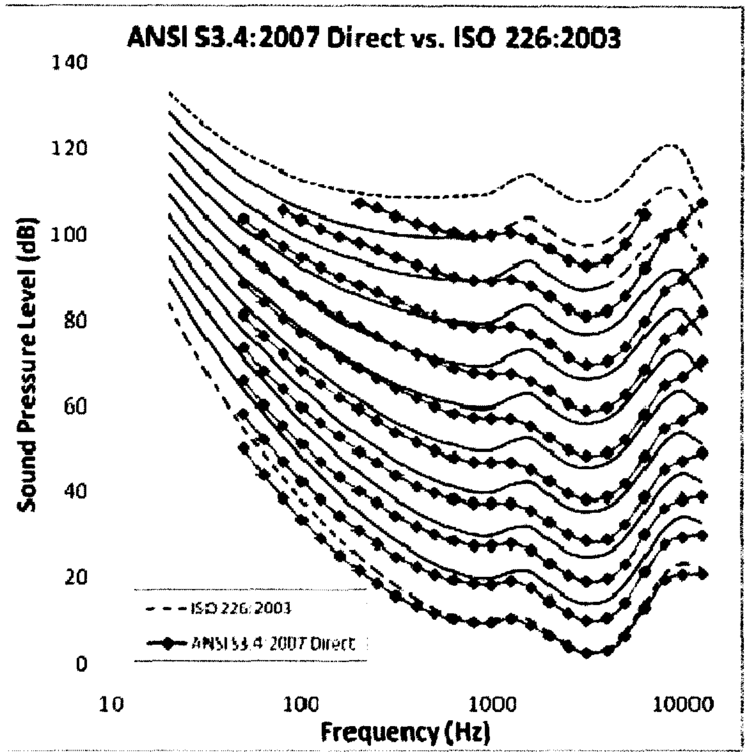


Figure 5.6 – ANSI S3.4:2007 as derived from 1/3 Octave inputs.

Aside from the restriction in applicability, the Direct approach appears to drop in performance with increasing frequency values. This deviation from the Program method reaches values of 4 dB in the 1 kHz range and can differ by levels that are up to 6 dB lower than the Program results at the highest frequencies. The effect is particularly troublesome on the performance in the 1 kHz-bump region. In this area, the Program results were already below the reference values. The contours visible from the plot in **Figure 5.6** reflect a substantial drop above 1 kHz, almost to a point where the contour shapes seem to line up with the reference curve below the target. Aside from the lower than expected values, the overall trends remain convincingly close to the reference contours which will result in reassuring correlation coefficient values.

To compare the input methods directly, **Figure 5.7** was included showing the discrepancies between the two.

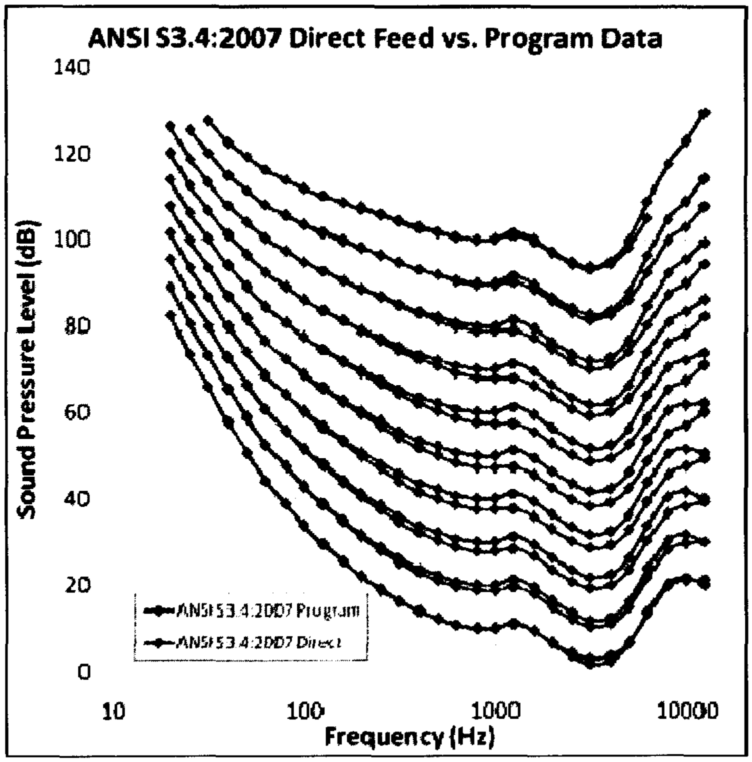


Figure 5.7 – ANSI Program method compared against the Direct method.

From the figure, the difference between the methods is easily observed. While the response initially starts at common values, as the frequency of the tones is increased, the mid-range loudness levels between 30 phons and 80 phons diverge in performance; with differences exceeding 6 dB in the worst cases. As seen in **Figure 5.6**, this deviation results in poor curve matching at the higher frequency levels for the Direct approach. Quantifying this effect, **Table 5.3** includes all of the overall correlation coefficient values for the ANSI S3.4:2007 model. The contours are only compared to the reference contours of ISO 226:2003 as the model was never intended to compare to the 1987 version of the reference standard.

Table 5.3 – Overall correlation coefficients of ANSI S3.4:2007

Phon Contour	10	20	30	40	50	60	70	80	90	100
Program vs. ISO226:2003	0.999	0.997	0.997	0.997	0.996	0.992	0.985	0.967	0.914	0.809
Direct vs. ISO226:2003	0.999	0.998	0.996	0.991	0.983	0.969	0.947	0.911	0.795	0.478

It is easy to see how even though the contours of the Direct approach can deviate substantially from the target values, the general shape of the predicted loudness results in correlation coefficient values above 0.9 for essentially all of the contours. Both methods indicated better correlation values than any of the loudness prediction methods examined so far.

With the equal loudness contours predicted from the various stationary loudness metrics available, the analysis techniques provided a platform on which to extend this investigation using the available time-varying loudness metrics based on the stationary signals used above. The first of which was Amendment 1 for the DIN 45631 stationary loudness model, (DIN 45631/A1).

5.1.4 DIN 45631 / Amendment 1 (2007)

Due to the number of programs and data conversions involved, a different approach was used to compare the time-varying models. Here, the pure tones used to derive the stationary DIN 45631 equal loudness contours were inserted into the time-varying loudness model. Given this, it is expected that similar conclusions will be realized as before. **Figure 5.8** is the resulting plot from this procedure. If the time-varying loudness model had in fact given the same results, the loudness levels would have resulted in a straight horizontal line of constant loudness levels for each centre frequency. The results indicated by **Figure 5.8** signify that the two models are very close in their predictions.

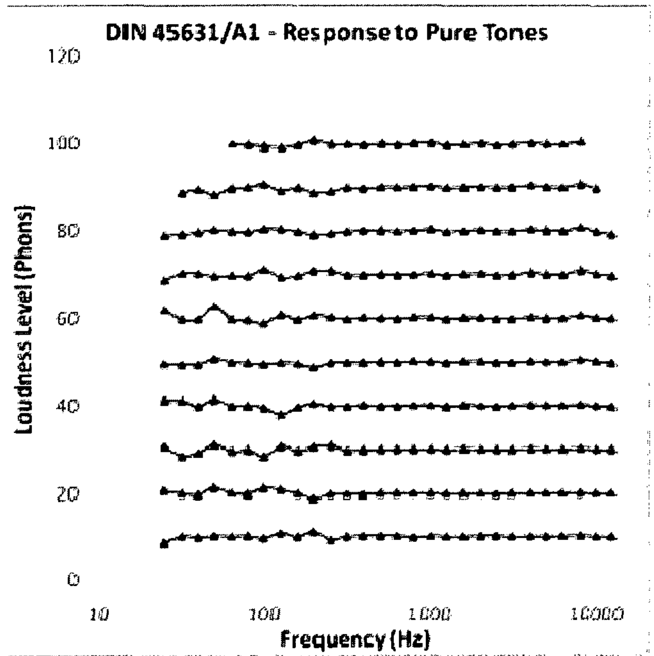


Figure 5.8 – DIN 45631/A1 response to stationary signals.

From the figure, the two metrics produce essentially identical results for pure tones above 200 Hz. Below this point, only minor fluctuations are observed deviating by no more than 3 phons at the maximum response. This demonstrates that the amendment for time-varying loudness has essentially the same performance level as the stationary DIN 45631 model. The performance is slightly reduced in the lower frequency regions, but the range of values is only plus or minus three phons. Extrapolating from these results, it is expected that the DIN 45631/A1 would have a set of equal loudness contours nearly identical to those determined above; a good match for the 1987 reference contours but too shallow for the new ISO 226:2003 data set.

5.1.5 Glasberg and Moore's Time-Varying Loudness Model (2002)

The second time-varying loudness model studied was the Glasberg and Moore Time-Varying Loudness (TVL) model available from the University of Cambridge's Auditor Perception Group website. **[19]** Provided as a set of executable files, the application of this standard was more involved than that of the DIN time-varying loudness model. Again, pure tones were recorded in PULSE LabShop and trimmed down to size using the time-edit software. The signals were then filtered to reduce the 'pop' sensation using Brüel and Kjær's Sound Quality software and saved as a 16-bit WAVE file. The WAVE signals generated were re-sampled to 32 kHz and calibrated to a full scale sinusoid of either 100 dB or 50 dB depending on the target SPL. Finally, the signals were scaled down to the appropriate values derived above to calculate the loudness. The target sound pressure levels used for the scaling were the same levels predicted by the stationary loudness software.

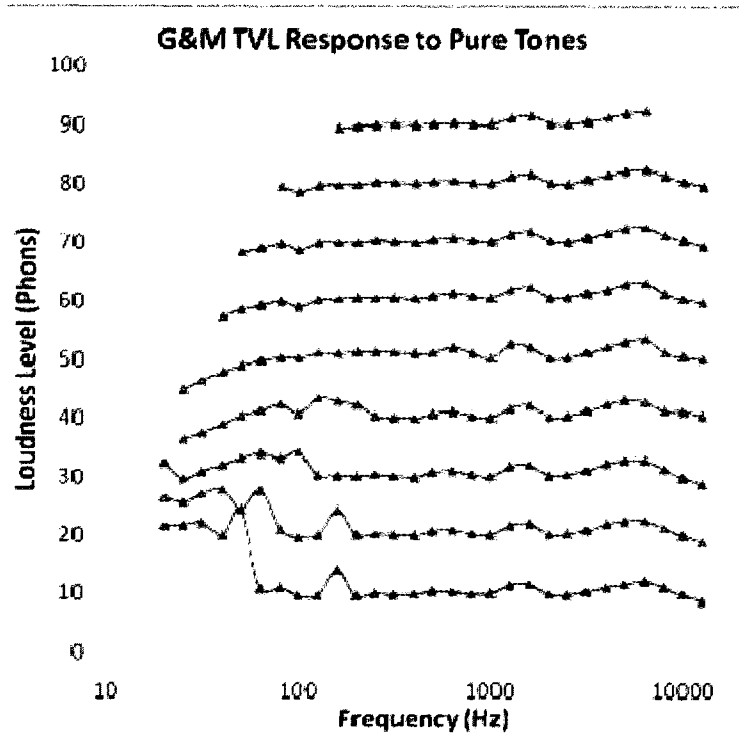


Figure 5.9 – Glasberg and Moore's TVL.exe software response to stationary signals.

Figure 5.9 illustrates the output of the TVL.exe program to the steady pure tones according to ANSI S3.4:2007. Although the contours resemble the straight lines expected, there are specific areas where the two models do not predict the same values; most notably the area below 63 Hz for almost every loudness contour. It appears that for areas which the ANSI S3.4:2007 model predicts a loudness of 10-20 phons the time-vary loudness model actually predicts a tone much louder with one instance as much as a 14 phon deviation from the expected. Above the 30 phon level, the TVL.exe model under predicts the loudness below 50 Hz resulting in lower levels than expected. It is not clear what might cause these low frequency discrepancies as the model seems to perform quite well above 200 Hz.

In the frequency range from 1 kHz to 2 kHz a reoccurring bump is predicted by the time-varying model, and again a smaller bump from 4 kHz to 8 kHz. These boosts in loudness values would actually improve the stationary model as it was at these points in the contours where the ANSI S3.4 metric was too low for a good correlation. It is possible that the TVL model improved on the stationary metric in this particular area. Although the extent of the improvements was not reported by the software authors, changes would account for discrepancies.

Overall, the TVL model produced by Glasberg and Moore produced a similar result to that of the preceding stationary loudness model. The time-varying model does appear to have specific local differences when compared to the previous model; in areas not only where the previous model needed improvements, but also where the stationary metric excelled.

5.2 Semi-Anechoic Results

To verify the data derived in the direct feed results, experiments were conducted using the semi-anechoic chamber and pure tones produced using a sound source. This experimental set-up was intended to produce 'real-world' samples of the stationary signals which included the effects of moderate background noise and the simulated collection of acoustic energy via a microphone. Tests were conducted at late hours, using every precaution necessary to ensure the acquisition of the best results. Three samples of each tone were recorded to ensure the repeatability of this experiment.

As mentioned previously, the semi-anechoic room has a lower cut-off frequency of approximately 100 Hz and an ambient noise level of approximately 18 dB. These limitations will therefore be present for all of the contours collected. The following are the averaged results of this secondary study intended to provide a thorough comparison of the various standardized metrics available. The purpose of the plots is to prove the repeatability and accuracy of the initial direct feed results. Therefore, only the direct feed and semi-anechoic data sets will be presented on the common plots; not the reference curve comparisons from the previous sections. For a complete collection of the reference plot comparisons, please refer to the **Appendix D and E** of this thesis.

5.2.1 ISO 532B (1975)

Using the approach outlined in the experimental details, **Figure 5.10** illustrates the semi-anechoic comparison as predicted by ISO 532B. The plot is provided using consistent scaling from the reference contour data set to facilitate easier comparisons between the two.

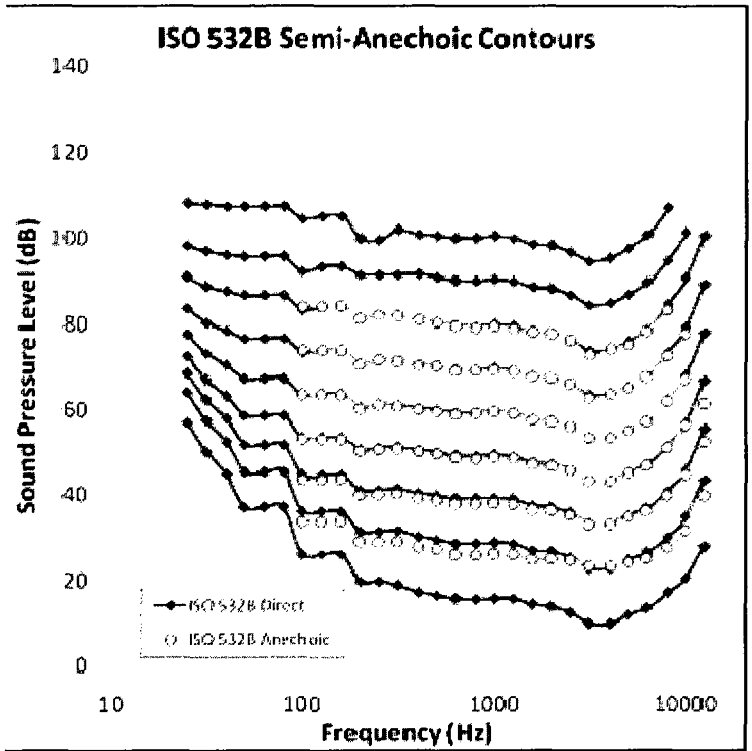


Figure 5.10 – Direct feed versus semi-anechoic data for ISO 532B

From the figure, the semi-anechoic data appears to mimic the direct feed approach quite well. For the 50 phon to 80 phon contours, the match between the recorded signals and the direct results is practically identical. Below this point some deviations are present where the anechoic approach seems to prematurely predict the target loudness level. This was however expected due to the influence of the ambient sound pressure levels in the semi-anechoic room.

5.2.2 DIN 45631 (1991)

As with the ISO 532B, the DIN 45631 contours predicted during the direct feed approach are reinforced from the semi-anechoic results. Once again, above the 40 phon contour line the results of **Figure 5.11** indicate a near exact match.

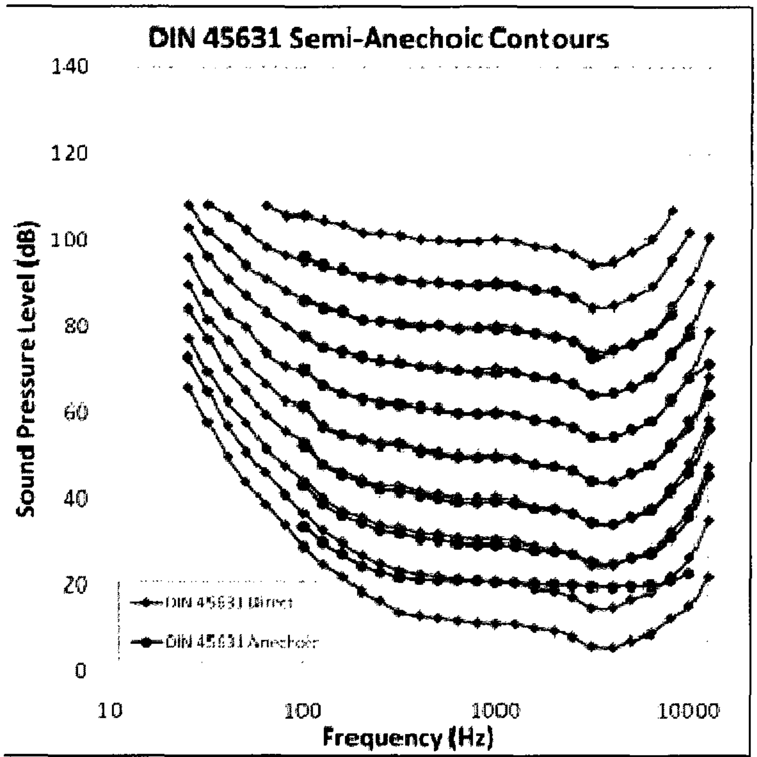


Figure 5.11 – Semi-anechoic contours of the DIN 45631 stationary loudness model.

The same trends appear with very good correlation above 50 phons while separating steadily as levels decrease. Note that for the DIN model, the anechoic results of the 20 phon line were partially obtained but severely affected by the ambient noise above 1 kHz.

5.2.3 ANSI S3.4:2007

The ANSI S3.4:2007 model appears to be the most heavily affected model by the presence of ambient noise. From **Figure 5.12**, it can be seen that contours as high as the 50 phon line are affected by the added background information. Meanwhile, a similar trend remains where decreasing loudness levels result in increasing deviations, particularly at the lower frequency levels. The 30 phon contour appears to be affected the most, dropping down to levels previously predicted by the 20 phon line.

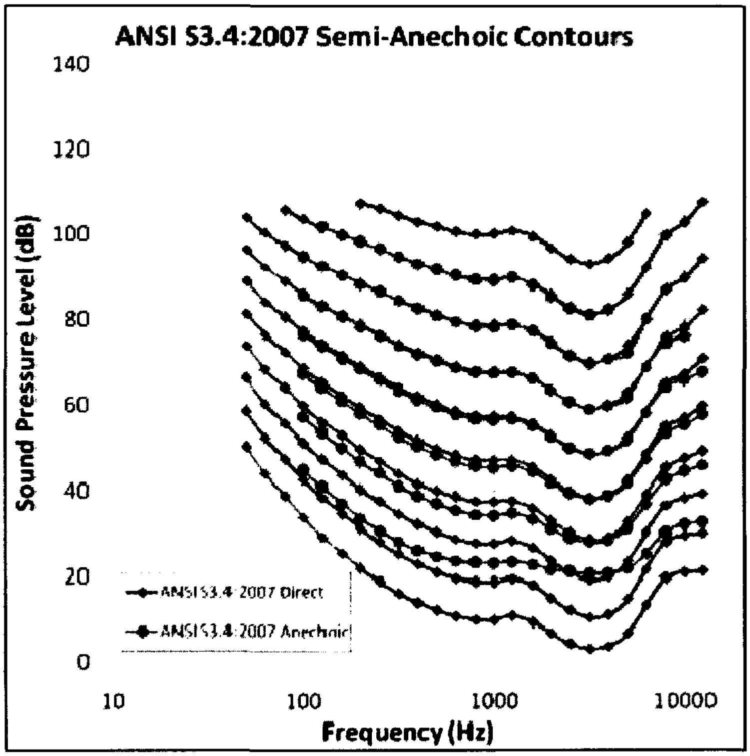


Figure 5.12 – ANSI S3.4:2007 direct versus anechoic comparison.

From these results, the ANSI S3.4:2007 stationary loudness model is very sensitive to the influence of background noise. Strangely, in all three models the ambient influence appears to not affect the 2 kHz to 5 kHz range where the dip is present in the loudness models. It is not clear what may have caused this as the ambient SPLs were approximately consistent at -13 dB for all frequencies above 500 Hz.

5.2.4 DIN 45631/Amendment 1 (2007)

The response of the time-varying loudness models to the ambient noise was different than that of the stationary models. The DIN 45631 amendment as shown in **Figure 5.13** is a good example of this.

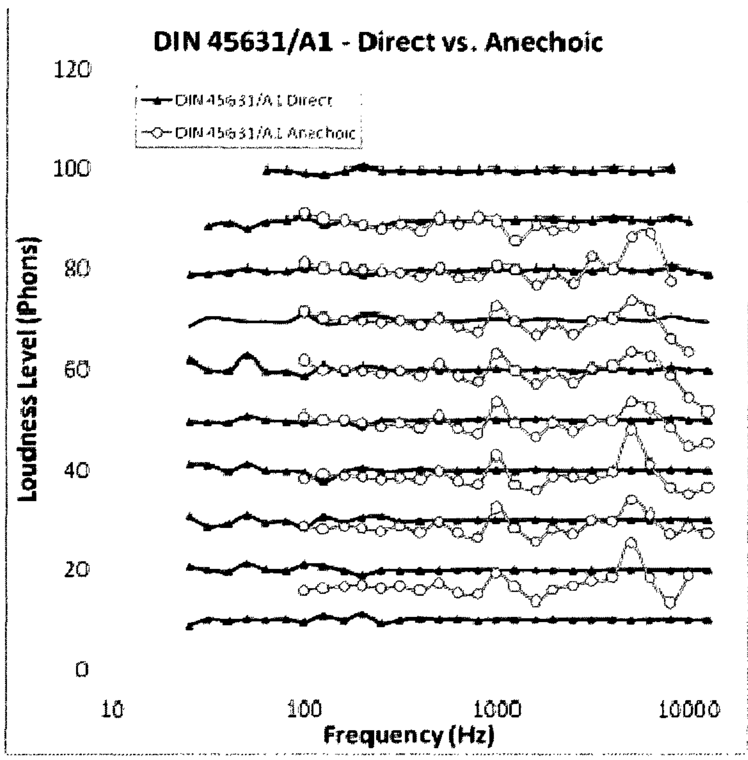


Figure 5.13 – DIN 45631/A1 influence from ambient noise.

As the DIN time-varying loudness model performed well in the direct feed trials, it was surprising to notice the fluctuations present during the same procedure using the semi-anechoic information. In areas where the amendment provided the same results as the stationary model, the effects of ambient noise caused the values to vary with some notable trends and frequency patterns. For pure tones at 1 kHz and 5 kHz, the non-stationary model produced peaks which are not explained in the either of the stationary model comparisons (neither the direct nor anechoic plots). At the higher levels, **Figure 5.11** indicated identical results between the two and no influence from the ambient levels as indicated here.

Although the overall result is consistent with the expected response, the trends present indicate a more thorough investigation is necessary. As this research project focused on the stationary loudness metrics, a more thorough analysis was beyond the scope of this study and should be considered as an examination goal for future work.

5.2.5 Glasberg and Moore's Time-Varying Loudness Model (2002)

The last comparison performed provided similar results. The Glasberg and Moore Time-Varying Loudness model appears to mimic the previously derived relationship, but once again, new trends and patterns are present. Note that the applicable range of the TVL model was significantly reduced. For the most accurate results, the model required full scale sinusoids recorded at 100 dB based on the procedure included with the program. This was only possible between 100 Hz and 1.6 kHz using the available sound source without risking damage to the equipment.

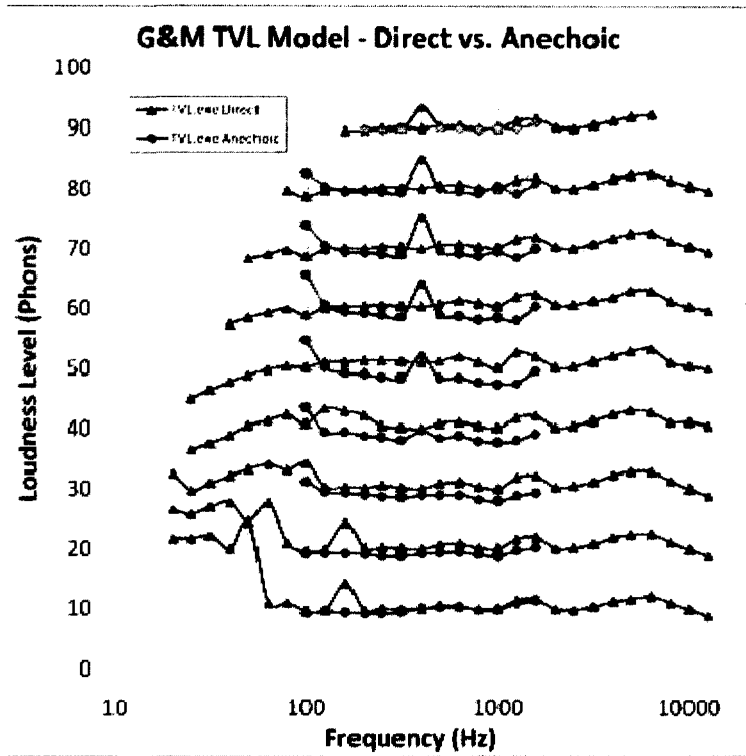


Figure 5.14 – TVL.exe model response to semi-anechoic data.

From the data collected, **Figure 5.14** illustrates the TVL response to pure tones as recorded in a semi-anechoic environment. As with the DIN time-varying model, unexplained fluctuations are present where none existed before. Unlike the DIN amendment, this model indicated a loudness jump in values at 400 Hz. Overall, however, the data correlates well with the previous observations, again though with unexplainable discrepancies.

VI. DISCUSSION

From the analysis described in the previous chapter, a summary of the observations is described here, as well as a critical comparison between the loudness models. These discussions will take into account the ease of use of the models, as well as any apparent limitations before any conclusions are made.

6.1 Performance Summary

Based on the analysis of the various models compared, it is clear that performance differences exist for all of the standardized metrics examined. As expected, the model with the most discrepancies was the outdated ISO 532B. While this model was the oldest metric compared, it is also still a current and much used loudness model standard. When compared to the ISO 226:2003 experimental data, the ISO 532B model did not correlate well with the target values, thus indicating a poor performance.

The improved Zwicker method as given by the DIN 45631 does indeed improve on the performance in the low frequency response. Being another older loudness model, the DIN method for calculating loudness fails to approximate the trends present in the new target loudness data, and therefore, should be updated.

Lastly, the performance of the ANSI S3.4:2007 metric can be described as both good and bad. On the good side, the Program approach of the model gives an excellent correlation to the reference contours of ISO 226:2003. However, when the recorded data is used in place of the 'specified' tones in the program, the observed performance decreases in the higher frequency range. This decrease extends to a point where the plotted values predicted are 10 dB below the expected response. While the slopes of the contours appear to better correlate with the reference contours, this drop causes a misrepresentation of the perceived loudness.

6.2 Ease of use

The ease of use of an engineering tool is almost as important as its performance. A model that is too complex may be subject to user error while an over simplified method may be limited in its application. Therefore, when comparing the various loudness models, usability and the manner in which data could be entered should be considered.

The ISO 532B was a very simple model to use once the program code was available. Originally written in BASIC code, the program was first converted into a more usable format such as a Visual Basic code in Microsoft Excel, as was done by Defoe (2007). [9] Once in this format, the use of the model was easy and straight forward, provided that the 1/3 octave data was available. Simply imputing the 28 third-octave data points from 25 Hz to 12.5 kHz into the model and specifying the recorded field type provided a corresponding loudness level based on the original Zwicker method.

The DIN 45631 stationary loudness model was a little different. As the standard was written in German with no translation available, it was not possible to follow exactly what process the standard was following. From the inclusive program code, it is apparent that the application of the model appears to be almost identical to the method above with only the tabular values altered, thus improving the overall performance of the model. As before, the input variables continue to be the 28 third-octave band elements and the specification of the type of sound field.

The last model analysed was the ANSI S3.4:2007 model for stationary loudness. This program included the most options regarding types of inputs and sound field settings. By allowing the input data to be entered as either 1/3 octave information or instead by specifying the spectral elements individually, the ANSI approach is the most widely applicable model. While providing the user with more options for measurement analysis, this model can be tailored to a variety of applications with a simple selection in the software. Provided the user has access to the exact 'broken-down' spectral information of the stationary signal being analyzed, the model portrays loudness levels in good agreement with the reference contours of ISO 226:2003. However, as not many users would have access to this data, the alternative 1/3 octave inputs provided an accurate approximation, particularly well in the low frequency regions while overestimating high frequency content.

6.3 Limitations

While all of the loudness models appeared to be capable of handling both free-field and diffuse-field listening conditions, the applicable frequency ranges vary with each individual metric.

From the experimental results, it was determined that the ISO 532B was the most heavily restricted model. From the program starting point, the lowest loudness value obtainable is 0.2 sones, or approximately 16.8 phons according to the loudness conversion factor used. This lower limitation effetely reduces the number of contours that could be derived using this model. Secondly the low frequency performance of the ISO 532B severely impacted the applicability of this model below 300 Hz. Below this point, misleading values could result in erroneous predictions in loudness.

The DIN 45631 model improved on both of the limitations of the ISO 532B metric. Using a corrected loudness conversion factor and an improved procedure below 300 Hz, the DIN loudness model is applicable over the entire hearing spectrum.

The Program approach of the ANSI S3.4:2007 model had the largest applicable frequency range of the metrics studied. By predicting loudness levels from 20 Hz up to 16 kHz, this particular approach has a slightly wider scope than the previous Zwicker methods. When the Direct approach is used, however, the applicable frequency range is strictly limited to a range from 50 Hz to 16 kHz. This produces a much more restricted area when compared to the procedures above. Based on the 1/3 octave band inputs, the DIN 45631 model appears to be the least restricted model when predicting loudness.

6.4 Uncertainty Analysis

In order to ensure the accuracy of the results above, a full uncertainty analysis was carried out on the calculation procedure and the equipment used. Adhering to a very detailed approach taken by Defoe (2007), this analysis will account for the random uncertainties associated with the testing equipment and the systematic uncertainties from the resulting experimental data. As Defoe's project dealt with targeting loudness levels as well, the uncertainty analysis conducted by the author was followed almost identically, resulting in similar results. For the complete procedure, please refer to Defoe's description in **[9]**.

The approach is based on the theory of error propagation as presented in Wheeler and Ganji's book in **[49]**. The total uncertainty associated with a measurement set is a combination of the propagated errors which result from each element used to derive the data. When considering most engineering related uncertainty values, these elements may be grouped into two separate categories; the systematic uncertainty (u_R) related to the measurement process, and the random uncertainty (p_R) involved in each trial set. This thesis uses Quinlan's assumption that the various loudness models may be treated as simple mathematical functions with one output value (loudness level) resulting from the combination of several variables (band pressures). **[34].** Using these ideals and the process outlined by Defoe, a complete uncertainty analysis was carried out resulting in the error bar plots of **Appendix F**. As Defoe pointed out, the uncertainty associated with direct feed data is essentially insignificant due to the procedure taken and the lack of physical measurements. Therefore, an uncertainty analysis was only carried out for the semi-anechoic room data, which is discussed as follows.

6.4.1 Discussion of Overall Uncertainty Results

The uncertainty of the various equal loudness contours was heavily dependent on the sound pressure level and frequency. This was largely due to the sensitivity of the various loudness metrics to small variations in sound pressure level; an effect which worsened with decreasing levels of sound pressure. To illustrate this effect, **Figure 6.1** represents the sensitivity of the DIN 45631 loudness calculations when pure tones are increased by only 1 dB. With unit Sone/Pa, the trends show that for low pressure, high frequency tones, the loudness calculation is very sensitive to the minor fluctuations. For convenience, the trends investigated were generalized into 10 dB increments. This was done to take into account the varying level sensitivity, while not overly complicating the process. As each equal loudness contour derived above remained essentially horizontal above 100 Hz, it was assumed this generalization was appropriate.

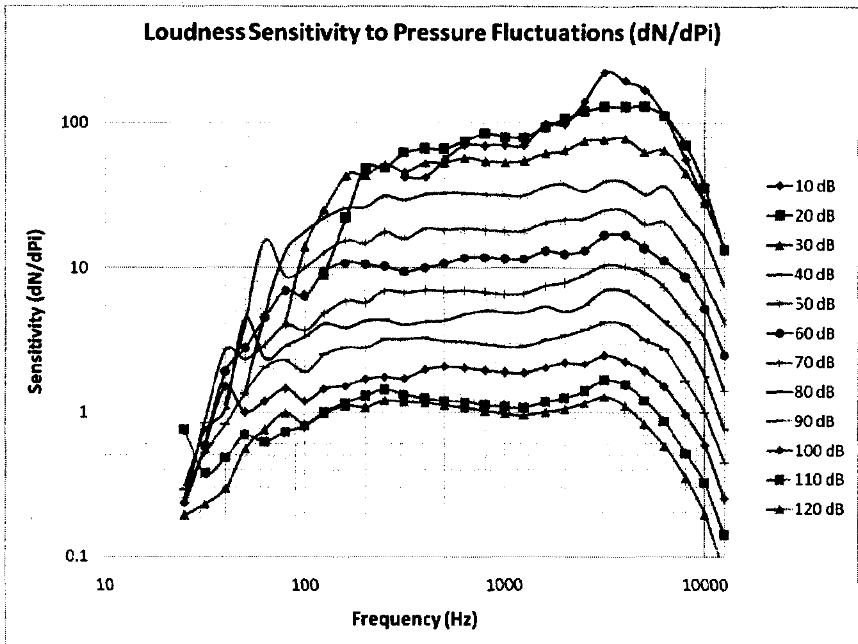


Figure 6.1 – Loudness Sensitivity of the DIN 45631

The other models indicated similar sensitivity plots (see **Appendix F**), where it appears that the lowest sound pressure levels exhibit the largest loudness sensitivities. This effect is again apparent in the uncertainty analysis, as the lower equal loudness contours result in greater levels of uncertainty; an effect that was not present in Defoe's study as he only examined high SPL trends in loudness.

To graphically show the effects of the trends, **Figure 6.2** and **Figure 6.3** illustrate the error bars associated with the ANSI S3.4:2007 – 30 phon and 90 phon equal loudness contours; two separate ends of the spectrum. To be able to clearly see the error bars, different scales had to be used on the plots as the resulting uncertainty levels varied markedly. From the figures, one can see the difference that the level of sound pressure plays in the amount of uncertainty detected. For lower level tones, the uncertainty is primarily a result of the systematic uncertainty, or the error associated with the calculation process. In the higher loudness levels, the random uncertainty associated with the variations between trials is the main contributor; while remaining less than one percent of the recorded value. This trend was present in all three of the standardized stationary loudness models studied. For a complete set of contour plots with the associated error bar analysis, again please refer to **Appendix F**.

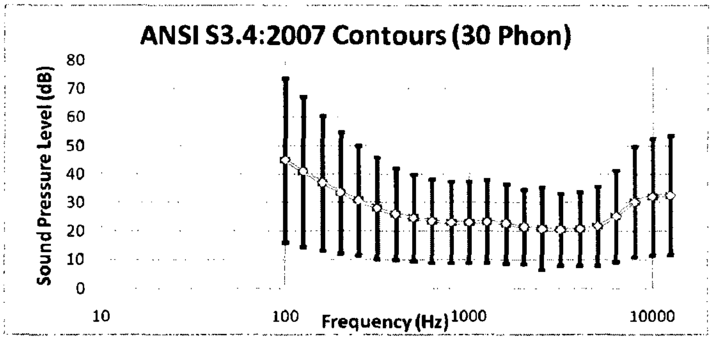


Figure 6.2 – ANSI S3.4:2007 – 30 Phon Error Bars

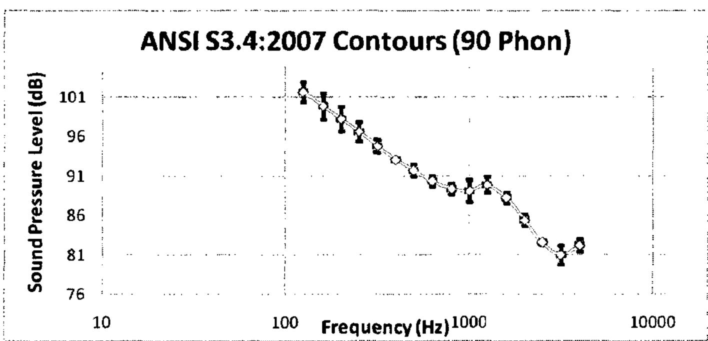


Figure 6.3 – ANSI S3.4:2007 – 90 Phon Error Bars

VII. CONCLUSIONS AND RECOMMENDATIONS

For this study, a detailed comparison of various loudness metrics was performed and documented. Targeting the performance of the standardized stationary loudness models, the procedure successfully identified the limitations and performance aspects of each model using a high resolution approach. Pure tones were used in this study as the stationary signal input for the loudness models. The signals were generated using a computer program and recorded both directly via an input-output or were acquired using a microphone-speaker set-up in a semi-anechoic room. The resulting plots from the above procedure produced contours of equal loudness according to the various loudness models investigated. Using a set of contours taken from perception experiments as the targeted trends, the derived results were critically compared to the experimental values and those of the various loudness models. Using the same pure tones, a secondary experiment was conducted to analyze the performance of the available non-stationary loudness models with regard to pure tone stationary signals. A comparison was then made between the two available models and their respective stationary predecessors as to the relationships that exist between the two. After reviewing the predicted shapes of the various metrics, several conclusions were made regarding the performance as well as some recommendations for further work in the related field. The following is a compilation of the conclusions based on those observations.

7.1 Conclusions

While investigating the available stationary loudness models, it became apparent that several trends were present in the performance of the loudness models. In order to compare the metrics against relevant experimental data, the predicted loudness contours were compared against the internationally accepted ISO 226:2003 equal loudness contours which acted the reference for deciding a best performing stationary model. The performance, limitations and user conclusions about each model is presented as follows prior to the selection of the most applicable model for use in industry:

1) The ISO 532B model being the oldest calculation tool for stationary loudness was expected to have lapses in performance based not only on its age, but based on the intended experimental results of the replaced ISO 226:1987. When compared against the newer equal loudness contours of ISO 226:2003, the ISO 532B model was simply too shallow in the trends present and neglected the important sensitivities above 1 kHz. The model was severely limited in the fact that performance below an amplitude of 40 phons (approximately 40 dB across the frequency spectrum) or below a frequency of 300 Hz results in known erroneous results due to calculation errors and tabulated coefficients. Implementation of the model was relatively simple so long as the program code included within the standard was available to the user in a usable format. For the results included in this study, the model was available as a visual basic macro imbedded into a Microsoft EXCEL spreadsheet. Based on the observed results, this stationary loudness model is not recommended for use as a calculation tool for determining loudness due to the fact that more usable and accurate models currently exist causing this model to be considered obsolete. The semi-anechoic data supports this claim by predicting the same response with pure tones as the direct approach.
2) The DIN 45631 performed considerably better than the ISO 532B in two respects. First, due to more accurate readings from the coefficient plots, the predicted contours below 300 Hz are considerably more realistic allowing the loudness to

be better predicted down to the frequency of 25 Hz. Secondly, the loudness conversion equation from sones to phons was modified for values below one sone. This permits the DIN method to be applied to loudness levels near threshold. By greatly increasing the applicable range of this loudness model, while retaining the easy input method of ISO 532B, the DIN 45631 stationary loudness model predicted accurate pure tone loudness based on the reference of the previous ISO 226:1987 document. However, once again the model's age is apparent when comparing to the newer target contours of ISO 226:2003 where the Zwicker approach proves to be too shallow to match the new standard perception contours. The 1 kHz sensitivity was neglected while both the low and high frequency loudness estimates continued to be much lower than the target values. Again, the response of the DIN 45631 model to the semi-anechoic data supported the direct comparisons with only minor influences from background noise.

3) The last standardized metric examined was the Glasberg and Moore ANSI S3.4:2007 stationary loudness model. The investigation of this approach was implemented using several of the input options available within the executable software. Initially the loudness contours were predicted using the internal pure tone specification approach which yielded accurate results. The predicted contours followed the trends and amplitudes of the ISO 226:2003 equal loudness contours over the entire frequency spectrum. Above 1 kHz, the model's performance dropped slightly as the model predicted loudness levels lower than expected. However, when 1/3 octave inputs were used, the performance of the model dropped again, resulting in loudness levels predicted well below the

expected. This was particularly evident as frequency levels increased where the two methods appeared to increasingly separate from one another. When verified using the semi-anechoic data, the results were further affected, dropping in loudness levels with pure tone amplitude; more so than any previous loudness metric. This indicates that the ANSI S3.4:2007 stationary loudness model is quite sensitive to ambient sounds at low amplitudes; it is unclear at this time whether or not this sensitivity is accurate or not as subjective experimental results were not available from the semi-anechoic room used.

7.2 Identified Best Overall Stationary Loudness Model

Selecting an overall best model was not a straight forward decision as each model has its own performance shortcomings. However, due to the age of both the ISO 532B and the DIN 45631, the two Zwicker methods predict overly shallow trends compared to the current standard of the ISO 226:2003 equal loudness contours. As such, the best use model from the above analysis has been identified here as the ANSI S3.4:2007 stationary loudness model.

Although the ANSI model has some high frequency performance issues when using the 1/3 octave band inputs, the overall trends of the model remain close to the trends present in the reference contours. The sensitivity of the 1 kHz bump as well as the high frequency drop are both identified in this model; areas which the previous loudness metrics appear to neglect. The applicability of the ANSI program for a variety of listening conditions and input methods adds to the depth of the Glasberg and Moore model, allowing for easier use given a variety of recording conditions. Text files may be written for value implementation to speed up the calculation process or pure tones may be specified on a hypothetical basis without the need for measured samples.

To ensure consistency throughout the various areas of acoustical modelling, it is essential to select one best-use model to avoid the confusion from various methods portraying the same units. Based on the conclusions summarized above, the ANSI S3.4:2007 model is recommended for use where loudness levels are measured for sounds which are stationary in nature.

7.3 Time-Varying Loudness Results

From the time-varying loudness investigations only brief observations can be identified from this investigation without further analysis into the performance of the specific trends. As before, each model discussed individually based on the performance using pure tones as inputs, but no best model is selected due to the incompleteness of the investigation conducted.

1) The DIN 45631/Amendment 1 (Draft) was the first time-varying model examined. The analysis was performed by inserting the pure tones generated in the derivation of the DIN 45631 stationary loudness model directly into the time-varying amendment. Based on the direct feed response, the two loudness models are quite similar in their approach resulting in nearly identical values. The amendment approach was therefore concluded to have a similar equal loudness contour set to that of the stationary model; a set which is too shallow compared to the ISO 226:2003. This cannot be verified though unless a more thorough investigation was to be carried out which completely mapped the equal loudness contours of the time-varying approach. For semi-anechoic data results, the

outcomes fluctuated substantially from the direct feed samples. Trends in peaks and valleys became apparent in higher frequencies while remaining somewhat consistent with the previous results. The amendment method appears to be quite sensitive to the additional ambient noise without a known cause for the peaking trends.

2) The second time-varying loudness model investigated was the Glasberg and Moore TVL metric. Overall the direct feed results were as consistent as the DIN amendment model. However, below 200 Hz the similarities between the stationary and the non-stationary model deviate significantly, particularly at the lower loudness levels. While continuing to follow the expected contours, the results depict specific trends with relation to frequency which may be a result of improvements over the stationary loudness metric. Once ambient noise data was inserted into the model via the semi-anechoic experiments, specific trends were once again present in the results but his time to a lesser degree than the DIN time-varying loudness model. It is again uncertain what caused these trends without a more thorough investigation.

The two time-varying loudness models were examined based on the pure tone information recorded for their predecessor stationary models. As such, the comparison conducted was only an indication of the similarities between the stationary and non-stationary metrics. From the information available, it is only possible to observe that both of the models are indeed similar to their respective stationary model and apparently susceptible to large influences from ambient noise sources. Without further investigations, the above observations cannot be verified and no concrete conclusions may be reached regarding the metric performance.

7.4 Recommendations

The analysis of the available stationary loudness metrics has generated the predicted equal loudness contours for clear comparisons with verified results. There are areas where the investigation could be carried further for future work in the area. The recommendations are listed as follows:

1) The verification of the direct feed results was hindered by the ambient sounds of the semi-anechoic room available. If a fully anechoic room were available with a lower cut-off frequency, the direct feed data may be confirmed for the entire frequency spectrum applicable.
2) Only the standardized stationary loudness models were used in this study. Other non-standardized methods exist which have become available since the ISO 226:2003 update which have not been compared via a unbiased third party. In order to truly establish a best use model, a next step would be to compare the ANSI S3.4:2007 metric to the various non-standardized metrics.
3) The above investigation compared the performance of the various models to pure tones across the frequency spectrum. Several other forms of stationary signals exist including complex signals which would include the effects of simultaneous masking. To further test the performance of each model, an experiment may be derived to compare complex tone results between models to completely test the applicability of each metric.

4) The response of the time-varying loudness models to ambient sounds is unsettling. While the models performed adequately in direct feed testing, the fluctuations present during 'real tests' indicate that a more complete investigation is necessary into this area of psychoacoustics. For example, a thorough unbiased analysis of the time-varying loudness models similar to the one done here (for the stationary models) could map out the response to pure tones and be compared to the reference contours of ISO 226:2003.

7.5 Contributions

The completed experimental results above have provided a thorough and extensive investigation into the selection of a superior performing stationary loudness model. Due to the growing importance of loudness measurements in industry, this research project has provided a meaningful comparison to the acoustic community, allowing for a more educated decision in the selection of a loudness model. The results indicated above have been published in literature under the following references **[5, 6, and 33]**.

VIII. REFERENCES

[1] American National Standards Institute. *ANSI S3.4:2005 - American National Standard Procedure for the Computation of Loudness of Steady Sounds.* Standard, Melville, NY: Acoustical Society of America, 2005.

[2] American National Standards Institute. ANSI S3.4:2007 - American National Standard Procedure for the Computation of Loudness of Steady Sounds. Standard, Melville, NY: Acoustical Society of America, 2007.

[3] Bauch, H. "Die Bedeutung der Frequenzgrupp fur die Lautheit von Krangen." *Acustica*, 1956: 6, 40-45.

[4] Chalupper, Josef. "Aural Exciter And Loudness Maximizer: What's Psychoacoustic About "Psychoacoustic Processors" ?" *AES 109th Convention.* Los Angeles: Journal of the Audio Engineering Society., 2000 Sept. 22-25. 1-20.

[5] Charbonneau, Jeremy, Colin Novak, and Helen Ule. "Comparison of Loudness Calculation Proceedure Results to Equal Loudness Contours." *Internoise 2009 The 38th International Congress and Exposition on Noise Control Engineering.* Ottawa, Ontario: INCE-USA, 2009.

[6] Charbonneau, Jeremy, Colin Novak, and Helen Ule. "Loudness Prediction Model Comparison Using the Equal Loudness Contours." *Canadian Acoustics.* Niagara-On-The-Lake, Ontario: CAA, 2009. 37(3), 64-65.

[7] Churcher, B. G., and A. J. King. "The performance of noise meters in terms of the primary standard." *J. Inst. Electr. Eng.*, 1937: 81, 57-90.

[8] Defoe, J. *ANSI S3.4:2005 standard.* Standard Editorial, A.S. Committee, R. Hellman, and B. Glasberg, Editors, 2007, p. 1.

[9] Defoe, J. *Evaluation of Loudness Calculation Techniques with Applications for Product Evaluation.* Windsor, Ontario: University Of Windsor, 2007.

[10] Deutsches Institut für Normung, (German Institute for Standardization). *DIN 45631 - Procedure for calculating loudness level and loudness.* Standard, Berlin, Germany: DIN, 1991.

[11] Deutsches Institut für Normung, (German Institute for Standardization). *DIN 45631/A1 - Calculation of loudness level and loudness from the sound spectrum - Zwicker method - Amendment 1: Calculatoin of the loudness of time-variant sound.* DRAFT Standard, DIN, 2007.

[12] Everest, F. Alton. *Master Handbook of Acoustics - Fourth Edition.* New York: McGraw-Hill, 2001.

[13] Fastl, H., and E. Zwicker. *Psychoacoustics, Facts and Models.* Berlin Heidelberg New York: Springer, 3rd edition, 2007.

[14] Fletcher, H. "Auditory patterns." *Revs. Modem Phys.*, 1940: 12, 47-65.

[15] Fletcher, Harvey, and W.A. Munsson. "Loudness, Its Definition, Measurement and Calculation." *The Journal of the Acoustical Society of America, Volume 5, Issue 2,* 1933: 82-108.

[16] Glasberg, B. R., and B. C. J. Moore. "A model of loudness applicable to time-varying sounds." *Journal of the Audio Engineering Society*, 2002: 50(5), 331-42.

[17] Glasberg, B. R., and B. C. J. Moore. "Derivation of auditory filter shapes from notched-noise." *Hearing Research*, 1990: 47(1-2), 103-138.

[18] Glasberg, B. R., and B. C. J. Moore. "Prediction of absolute thresholds and equal-loudness contours using a modified loudness model (L)." *J. Acoust. Soc. Am.*, 2006: 120(2), 585-588.

[19] Glasberg, B.R., and B.C.J. Moore. "LOUD2006A.exe - Loudness model calculated according to ANSI S3.4-2007." *Auditory Perception Group University of Cambridge.* 2006. http://hearing.psychol.cam.ac.uk/Demos/demos.html (accessed April 09, 2009).

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

[21] Hellman, Rhona, and Eberhard Zwicker. "Why can a decrease in dB(A) produce an increase in loudness?" *J. Acoust. Soc. Am.*, 1987: 82(5), 1700-1705.

[22] Hellman, Rhonda P. "A New ANSI Loudness Standard." *Acoustics Today*, 2007: 40-42.

[23] International Organization for Standardization. *ISO226 Acoustics – Normal Equal-Loudness Contours.* Standard, Geneva: International Organization for Standardization, 1987.

[24] International Organization for Standardization. *ISO226 Acoustics – Normal Equal-Loudness Contours.* Standard, Geneva: International Organization for Standardization, 2003.

[25] International Organization for Standardization. *ISO 389-7 Acoustics - Reference zero for the calibration of audiometric equipment. Part 7: Reference threshold of hearing under free-field and diffuse-field listening conditions.* Standard, Geneva: International Organization for Standardization, 1996.

[26] International Organization for Standardization. *ISO 389-7 Acoustics - Reference zero for the calibration of audiometric equipment. Part 7: Reference threshold of hearing under free-field and diffuse-field listening conditions.* Standard, Geneva: International Organization for Standardization, 2005.

[27] International Organization for Standardization. *ISO532 Acoustics - Method for calculating loudness level.* Standard, Geneva: International Organization for Standardization, 1975.

[28] Kingsbury, B.A. "A Direct Comparison of the Loudness of Pure Tones." *Phys. Rev., 29*, 1927: 588.

[29] Marks, L. E. "Binaural summation of the loudness of pure tones." *J. Acoust. Soc. Am.*, 1978: 64, 107-113.

[30] Moore, B. C. J., and B. R. Glasberg. "A revision of Zwicker's loudness model." *Acustica - Acta Acustica*, 1996: 82(2), 335-345.

[31] Moore, B. C. J., and B. R. Glasberg. "Formulae describing frequency selectivity as a function of frequency and level, and their use in calculating excitation patterns." *Hearing Research*, 1987: 28, 209-225.

[32] Moore, B. C. J., B. R. Glasberg, and T. Baer. "A model for the prediction of thresholds, loudness, and partial loudness." *Journal of the Audio Engineering Society*, 1997: 45(4), 224-240.

[33] Novak, Colin, Jeremy Charbonneau, and Helen Ule. "Comparison of Non-Stationary Loudness Results to Equal Loudness Contours." *Canadian Acoustics.* Niagara-On-The-Lake, Ontario: CAA, 2009. 37(3), 66-67.

[34] Paulus, E., and E. Zwicker. "Programme Zur Automatischen Bestimmung Der Lautheit Aus Terzpegeln Oder Frequenzgruppenpegeln. (Computer Programs for Calculating Loudness from Third-Octave Band Levels or from Critical Band Levels)." *Acustica*, 1972: 27(5), 253-266.

[35] Peirce, B. "Criterion for the rejection of doubtful observations." *Astronomical Journal II*, 1952, 45: 161-163.

[36] Quinlan, Daniel A. "A Comparitive Study of Zwicker's Method for Calculating the Loudness of Continuous Sounds." *Noise Control Eng. J.*, 1994: 42(3), 117-126.

[37] Robinson, D.W., and R.S. Dadson. "A re-determination of the equal-loudness relations for pure tones." *Br. J. Appl. Phys*, 1956: 7, 166-181.

[38] Ross, S. M. "Peirce's criterion for the elimination of suspect experimental data." *Journal of Engineering Technology*, 2003, (Fall): 1-12.

[39] Schomer, Paul D: Suzuki, Yoiti, and Fumitaka Saito. "Evaluation of loudness-level weightings ro assessing the annoyance of environmental noise." *J. Acoust. Soc. Am.*, 2001: 110(5), 2390-2397.

[40] Seeber, Bernhard U. "Masking and Critical Bands." In *Handbook of Signal Processing in Acoustics Volume 1*, 229-240. New York: Springer, 2008.

[41] Sek, Aleksander, and B. C. J. Moore. "The critical modulation frequency and its relationship to auditory filtering at low frequencies." *J. Acoust. Soc. Am.* , 1994 : 95(5), 2606-2615.

[42] Stevens, S. S. "A scale for the measurement of psychological magnitude: loudness." *Psychological Review*, 1936: 43, 405-416.

[43] Stevens, S. S. "Calculation of loudness of complex-noise." *Acoust. Soc. Am.*, 1956: 28(5), 807-829.

[44] Stevens, S. S. "Procedure for Calculating Loudness: Mark VI." *J. Acoust. Soc. Am.* , 1961: 33(11), 1577-1585.

[45] Stevens, S. S. "Perceived Level of Noise by Mark VII and Decibels (E)." *J. Acoust. Soc. Am.*, 1972: 51(2), 575-601.

[46] Stevens, S. S. *The psychophysics of sensory function, Sensory Communication.* N.Y.: W. Rosenblith. Wiley and Sons, 1961.

[47] Suzuki, Yôiti, and Hisashi Takeshima. "Equal-loudness-level contours for pure tones." *J. Acoust. Soc. Am. 116 (2)*, 2004: 918-933.

[48] Vér, István L., and Leo L. Beranek. *Noise and Vibration Control Engineering: Principles and Applications.* Hoboken, New Jersey: John Wiley & Sons, Inc., 2006.

[49] Wheeler, A. J., and A. R. Ganji. *Introduction to Engineering Experimentation. 2 ed.* Upper Saddle River, New Jersey: Pearson Prentice Hall, 2004.

[50] Zwicker, E. "Subdivision of the Audible Frequency Range into Critical Bands (Frequenzgruppen)." *J. Acoust. Soc. Am., 33(2)*, 1961: 248.

[51] Zwicker, E. "Ein Verfahren zur Berechnung der Lautstärke." *Acustica* , 1960: 10, 304-308.

[52] Zwicker, E. "Procedure for calculating loudness of temporally variable sounds." *J. Acoust. Soc. Am.*, 1977: 62(3), 675-82.

[53] Zwicker, E. "Erratum: "Procedure for calculating loudness of temporally variable sounds" [J. Acoust. Soc. Am. 62, 675-682, (1977)]." *J. Acoust. Soc. Am.*, 1978: 63(1), 238.

[54] Zwicker, E. "Über psychologische und Methodische Grundlagen der Lautheit." *Acustica*, 1958: 8, 237-258.

[55] Zwicker, E., and E. Terhardt. "Analytical expressions for critical-band rate and critical bandwidth as a function of frequency ." *J. Acoust. Soc. Am.* , 1980: 68(5), 1523-1525.

[56] Zwicker, E., and H. Fastl. *Psychoacoustics: Facts and Models.* Berlin: Springer, 1990.

[57] Zwicker, E., and R. Feldtkeller. "Uber die Lautstarke von Gleichformigen Gerauschen (On the loudness of stationary noises)." *Acustica*, 1955: 5, 303-316.

[58] Zwicker, E., G. Flottorp, and S. S. Stevens. "Critical band width in loudness summation." *Acoust. Soc. Am.*, 1957: 29(5), 548-557.

[59] Zwicker, E., H. Fastl, and C. Dallmayr. "BASIC-program for calculating the loudness of sounds from their 1/3-oct band spectra according to ISO 532B." *Acustica*, 1984: 55(1), 63-7.

[60] Zwicker, E., H. Fastl, E. Widmann, K. Kurakata, S. Kuwano, and S. Namba. "Program for calculating loudness according to DIN 45631 (ISO 532B)." *J. Acoust. Soc. Jap. (E)*, 1991: 12(1), 39-42.

[61] 2270 PRODUCT DATA Hand-held Analyzer Type 2270 with Sound Level Meter Software BZ-7222 Optional Frequency Analysis Software BZ-7223, Logging Software BZ-7224, Enhanced Logging Software BZ-7225 and Sound Recording Option BZ-7226 [.PDF file] [cited 2009 November 11]; Available from: http://www.bksv.com/doc/bp2199.pdf

[62] 2669 PRODUCT DATA Falcon™ Range ½" Microphone Preamplifier — Type 2669 [.PDF file] [cited 2009 December 18]; Available from: http://www.bksv.com/doc/bp1422.pdf

[63] 2716 PRODUCT DATA Audio Power Amplifier 100W Stereo Type 2716-C [.PDF file] [cited 2009 November 12]; Available from: http://www.bksv.com/doc/bp1923.pdf

[64] 3560 SYSTEM DATA IDAe Hardware Configurations for PULSE – Types 3560-B, 3560-C, 3560-D and 3560-E. [.PDF file] [cited 2009 November 11]; Available from: http://www.bksv.com/doc/bu0228.pdf

[65] 4100 PRODUCT DATA Sound Quality Head and Torso Simulator — Types 4100 and 4100D [.PDF file] [cited 2009 November 12]; Available from: http://www.bksv.com/doc/bp1436.pdf

[66] 4189 PRODUCT DATA ½" Prepolarized Free-field Microphone Type 4189 [.PDF file] [cited 2009 November 12]; Available from: http://www.bksv.com/doc/bp2210.pdf

[67] 4228 PRODUCT DATA Brüel & Kjær Pistonphone — Type 4228 [.PDF file] [cited 2009 November 15]; Available from: http://www.bksv.com/doc/bp0881.pdf

[68] 4295 PRODUCT DATA Sound Sources for Building Acoustics: OmniPower™ Sound Source — Type 4292, OmniSource™ Sound Source — Type 4295, Tapping Machine — Type 3207, and Power Amplifiers — Type 2734-A and 2734-B [.PDF file] [cited 2009 November 12]; Available from: http://www.bksv.com/doc/bp1689.pdf

[69] 7700 SYSTEM DATA Software for PULSE™ 14 incl. Types 7700, 7705, 7707, 7709, 7764, 7770, 7771, 7773, 7789 and 7797 [.PDF file] [cited 2009 November 11]; Available from: http://www.bksv.com/doc/bu0229.pdf

IX. APPENDIX

Appendix A – Wedge Length versus Lower Cut-off Frequency

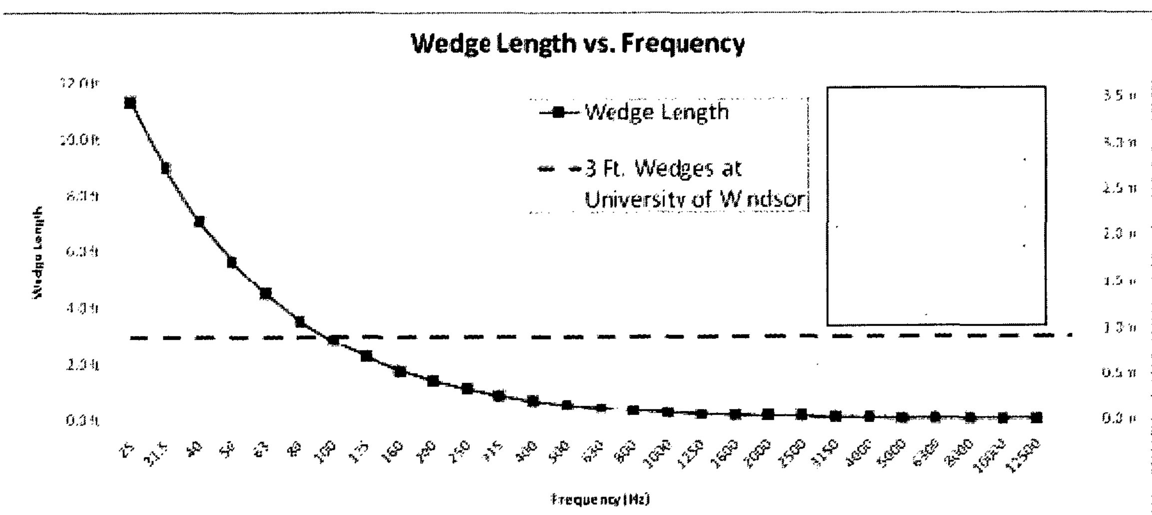
Frequency (Hz)	Wavelenth	1/4 Wavelength (m)	1/4 Wavelength (ft)
20	17.201 m	4.3 m	14.1 ft
25	13.761 m	3.4 m	11.3 ft
31.5	10.921 m	2.7 m	9.0 ft
40	8.600 m	2.2 m	7.1 ft
50	6.880 m	1.7 m	5.6 ft
63	5.461 m	1.4 m	4.5 ft
80	4.300 m	1.1 m	3.5 ft
100	3.440 m	0.9 m	2.8 ft
125	2.752 m	0.7 m	2.3 ft
160	2.150 m	0.5 m	1.8 ft
200	1.720 m	0.4 m	1.4 ft
250	1.376 m	0.3 m	1.1 ft
315	1.092 m	0.3 m	0.9 ft
400	0.860 m	0.2 m	0.7 ft
500	0.688 m	0.2 m	0.6 ft
630	0.546 m	0.1 m	0.4 ft
800	0.430 m	0.1 m	0.4 ft
1000	0.344 m	0.1 m	0.3 ft
1250	0.275 m	0.1 m	0.2 ft
1600	0.215 m	0.1 m	0.2 ft
2000	0.172 m	0.0 m	0.1 ft
2500	0.138 m	0.0 m	0.1 ft
3150	0.109 m	0.0 m	0.1 ft
4000	0.086 m	0.0 m	0.1 ft
5000	0.069 m	0.0 m	0.1 ft
6300	0.055 m	0.0 m	0.0 ft
8000	0.043 m	0.0 m	0.0 ft
10000	0.034 m	0.0 m	0.0 ft
12500	0.028 m	0.0 m	0.0 ft
	**Assumed Temperature of 21.4°C **		

$$\lambda = \frac{c}{f}$$

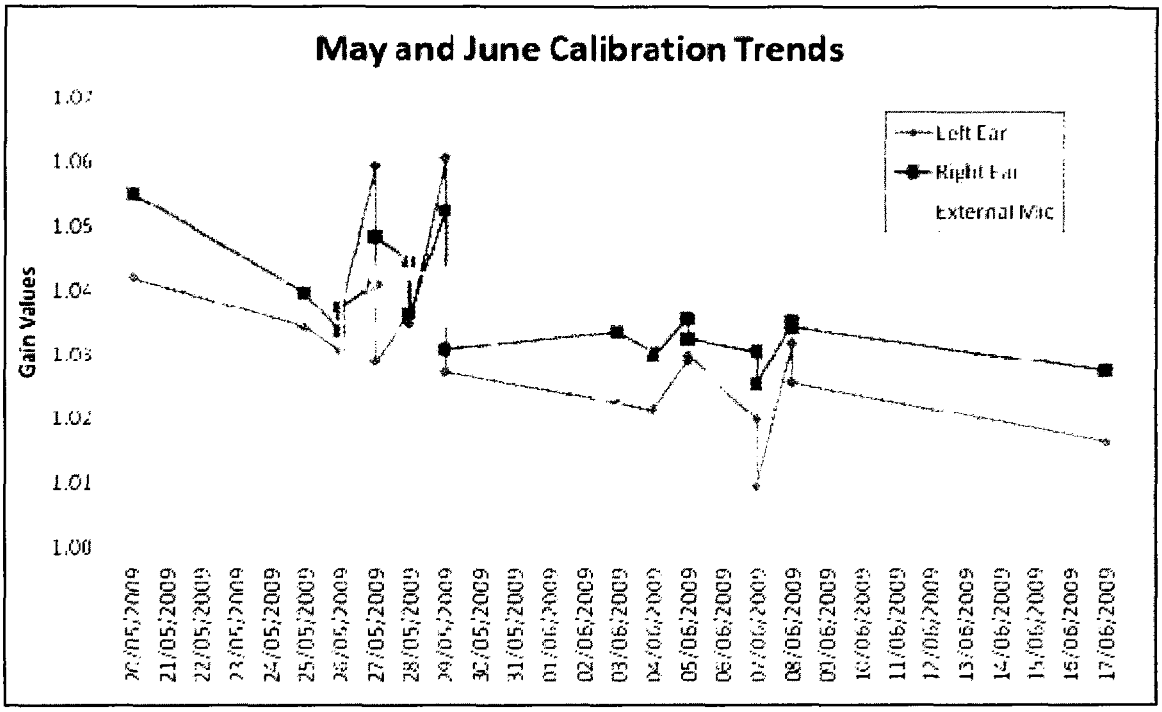
c – Speed of sound (m/s)
λ – Wavelength (m)
f – Frequency (Hz) or (sec^{-1})

$$c = 20.05\sqrt{T(°K)}$$

c – Speed of sound (m/s)
T – Air Temperature (°Kelvin)



Appendix B – Calibration Trends



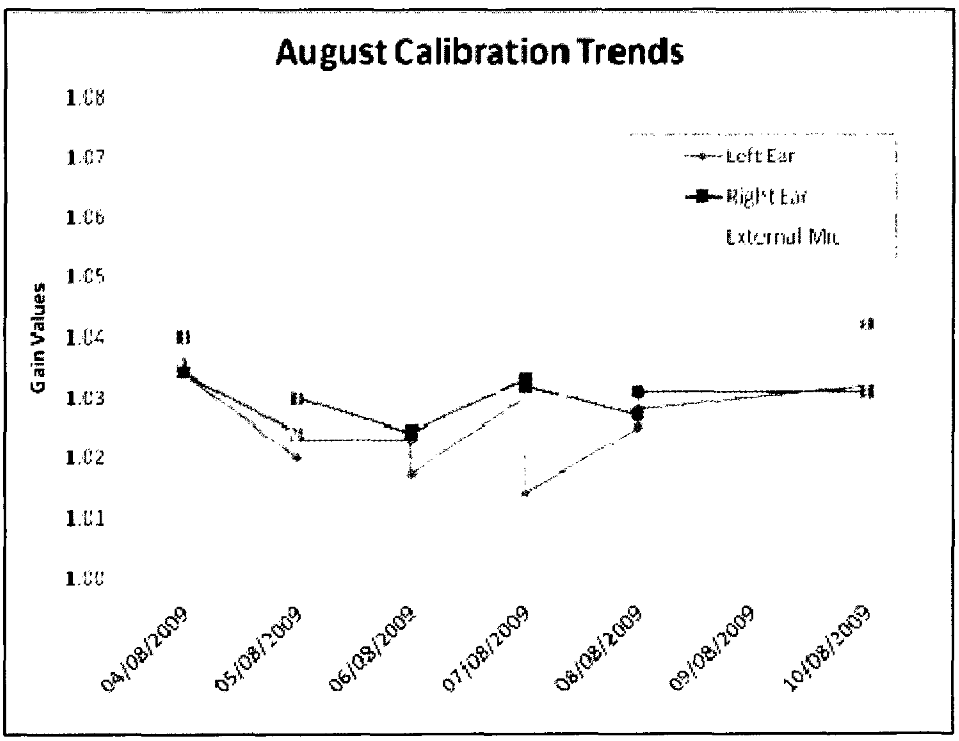


Table A-2 – Microphone Calibration Gain Values

	Date		Microphone Left Ear 2637736	Microphone Right Ear 2637735	Microphone External 2591370
May 2009	20/05/2009	Start	1.04	1.06	1.03
	20/05/2009	End			
	25/05/2009	Start	1.03	1.04	1.04
	25/05/2009	End			
	26/05/2009	Start	1.03	1.03	1.04
	26/05/2009	End	1.04	1.04	1.03
	27/05/2009	Start	1.06	1.04	1.04
	27/05/2009	End	1.03	1.05	1.04
	28/05/2009	Start	1.04	1.04	1.05
	28/05/2009	End	1.03	1.04	1.04
	29/05/2009	Start	1.06	1.05	1.04
	29/05/2009	End	1.03	1.03	1.04
June 2009	03/06/2009	Start	1.02	1.03	1.02
	03/06/2009	End			
	04/06/2009	Start	1.02	1.03	1.03
	04/06/2009	End	1.02	1.03	1.05
	05/06/2009	Start	1.03	1.04	1.04
	05/06/2009	End	1.03	1.03	1.05
	07/06/2009	Start	1.02	1.03	1.04
	07/06/2009	End	1.01	1.03	1.04
	08/06/2009	Start	1.03	1.04	1.04
	08/06/2009	End	1.03	1.03	1.05
	17/06/2009	Start	1.02	1.03	1.05
	17/06/2009	End			
August 2009	04/08/2009	Start	1.04	1.04	1.04
	04/08/2009	End	1.03	1.03	1.07
	05/08/2009	Start	1.02	1.02	1.02
	05/08/2009	End	1.02	1.03	1.04
	06/08/2009	Start	1.02	1.02	1.03
	06/08/2009	End	1.02	1.02	1.03
	07/08/2009	Start	1.03	1.03	1.03
	07/08/2009	End	1.01	1.03	1.03
	08/08/2009	Start	1.03	1.03	1.04
	08/08/2009	End	1.03	1.03	1.05
	10/08/2009	Start	1.03	1.03	1.05
	10/08/2009	End	1.03	1.04	1.03
	AVERAGE		1.03	1.03	1.04

Appendix C – Test Condition Trends

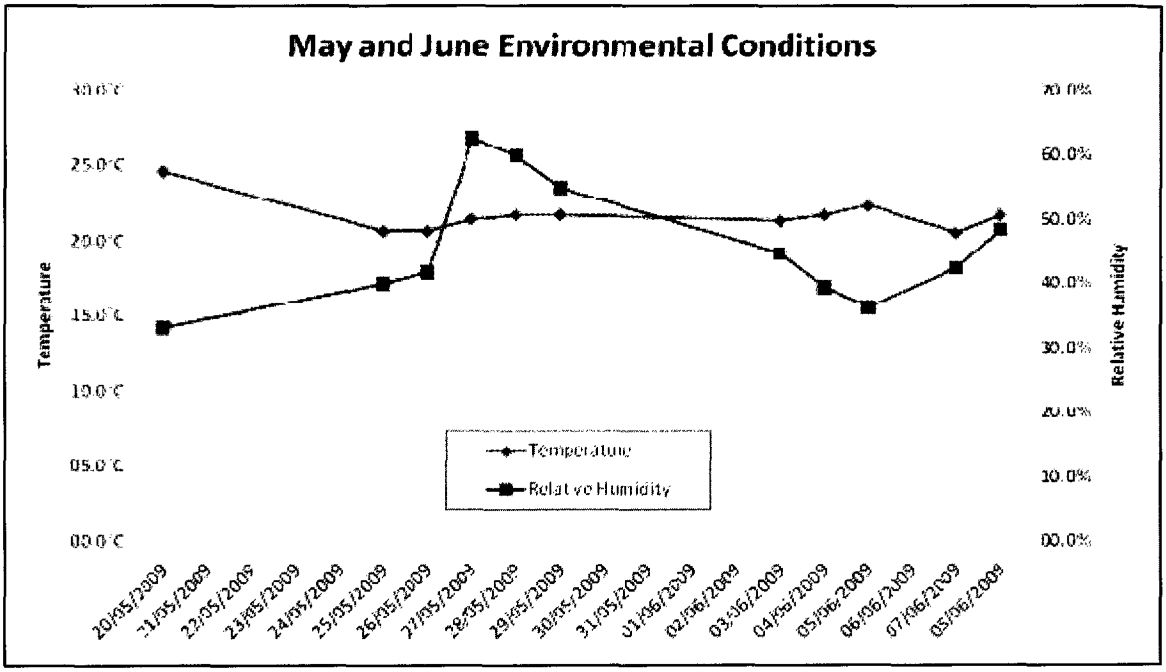


Figure F-1 – May and June environmental condition trends

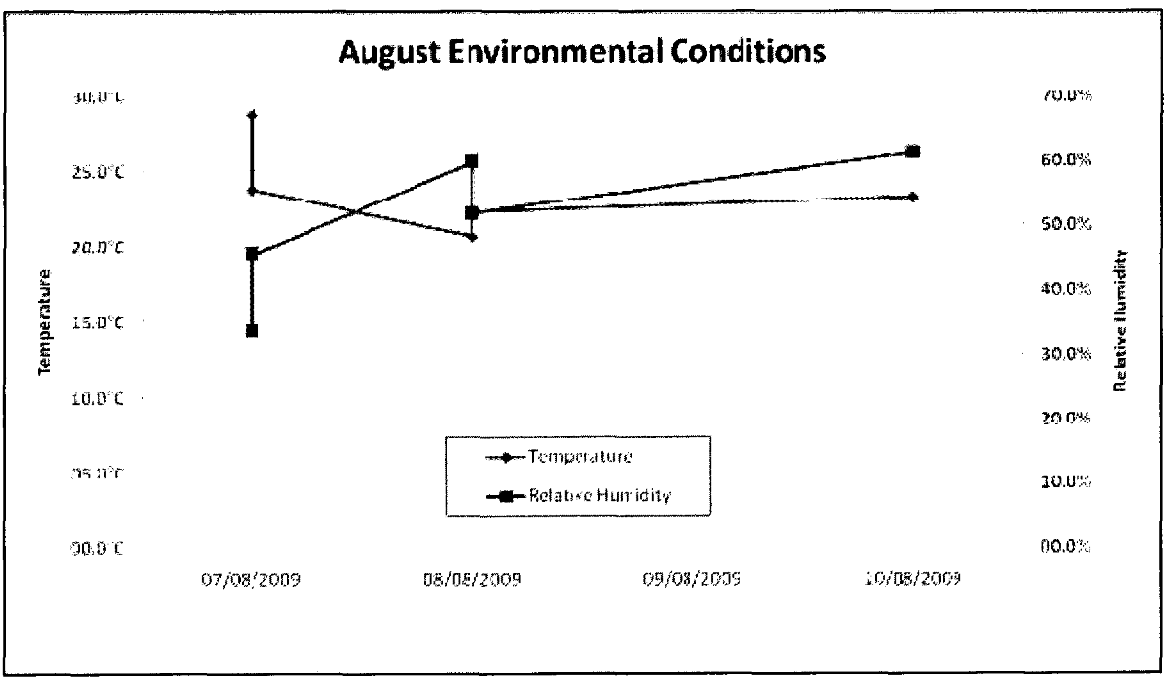
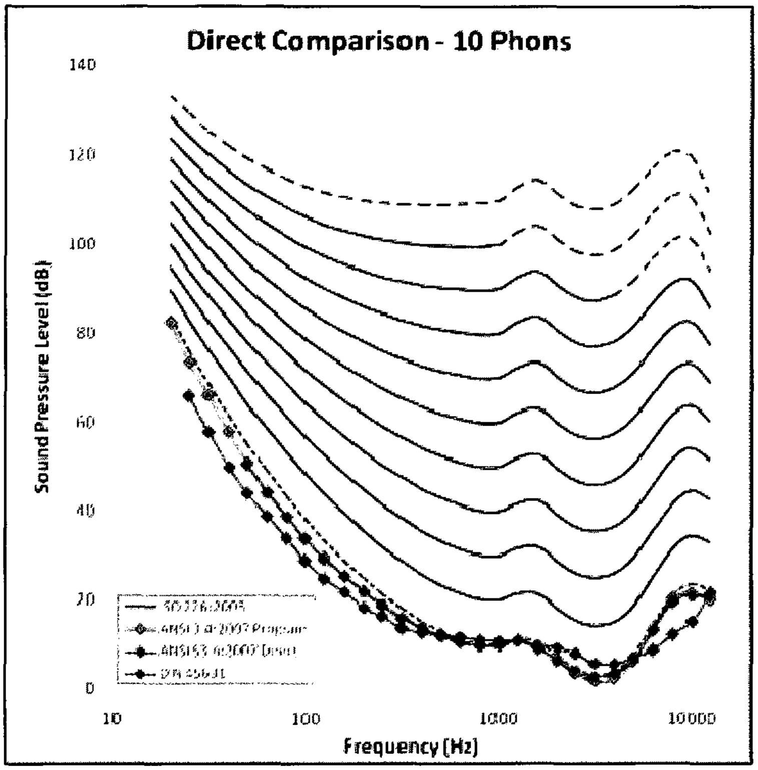
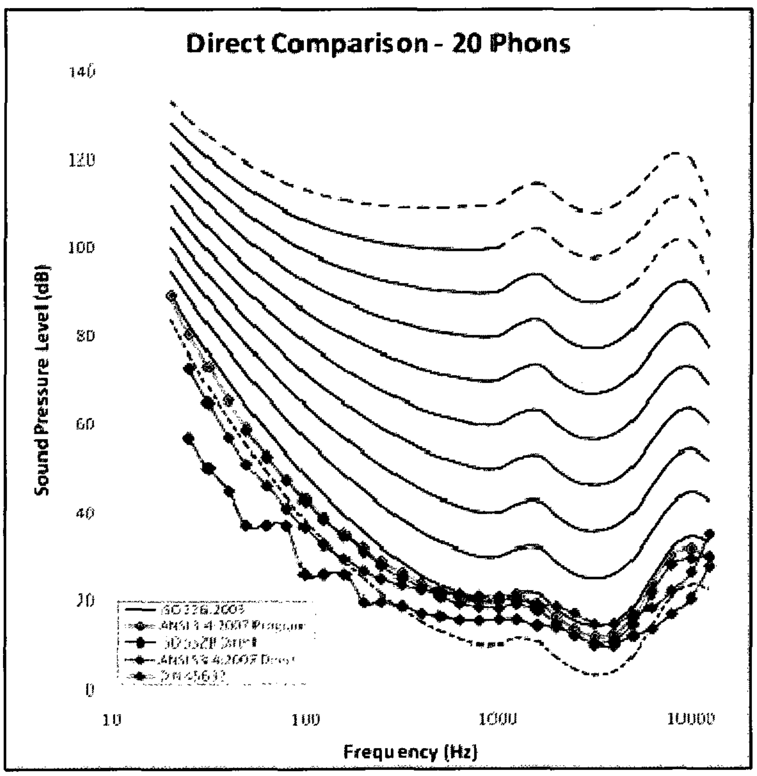


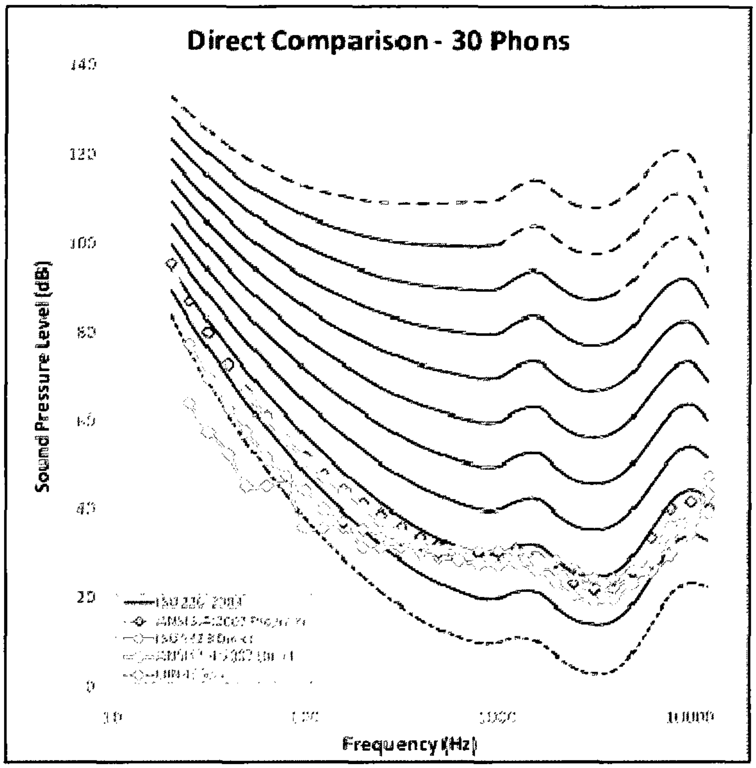
Figure F-2 – August environmental condition trends

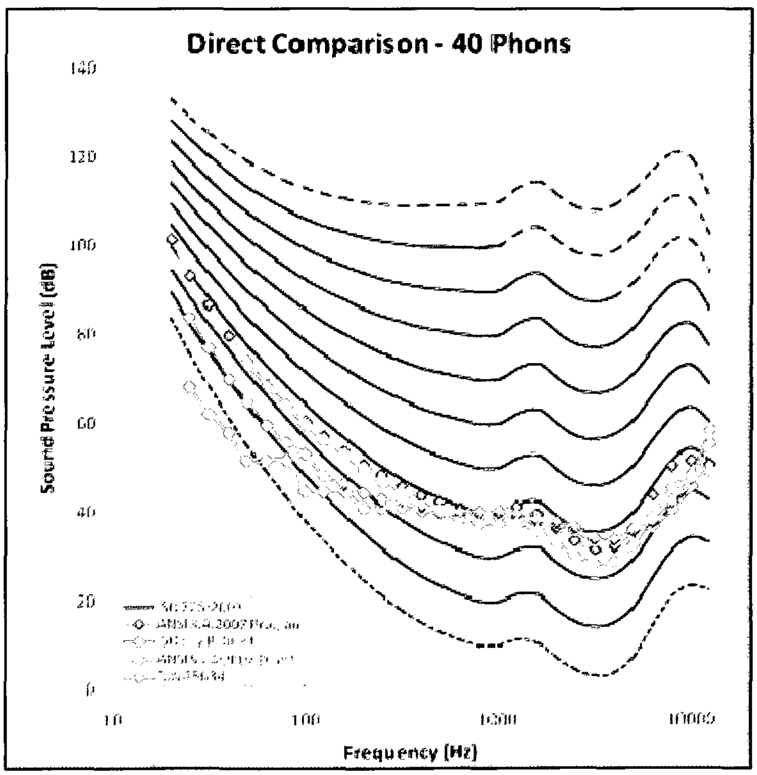
Date	Temperature	Relative Humidity
20/05/2009	24.6°C	33.0%
25/05/2009	20.6°C	39.9%
26/05/2009	20.6°C	41.6%
27/05/2009	21.4°C	62.6%
28/05/2009	21.7°C	59.9%
29/05/2009	21.7°C	54.9%
03/06/2009	21.3°C	44.6%
04/06/2009	21.7°C	39.3%
05/06/2009	22.3°C	36.2%
07/06/2009	20.5°C	42.3%
08/06/2009	21.7°C	48.4%
07/08/2009	28.8°C	33.7%
07/08/2009	23.8°C	45.4%
08/08/2009	20.7°C	59.8%
08/08/2009	22.3°C	51.9%
10/08/2009	23.3°C	61.3%
AVG	22.3°C	47.2%

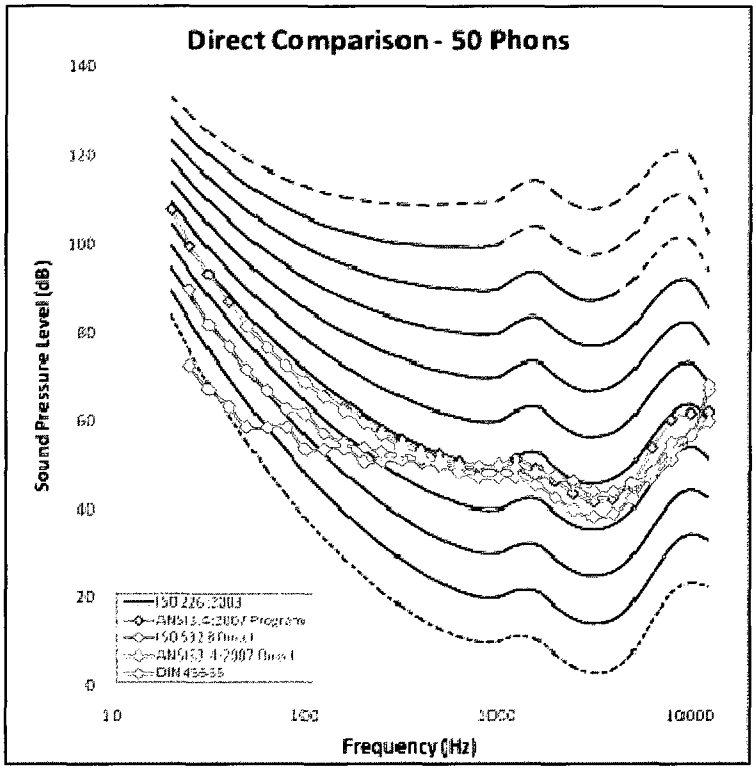
Appendix D – Direct Feed Common Plot Comparisons

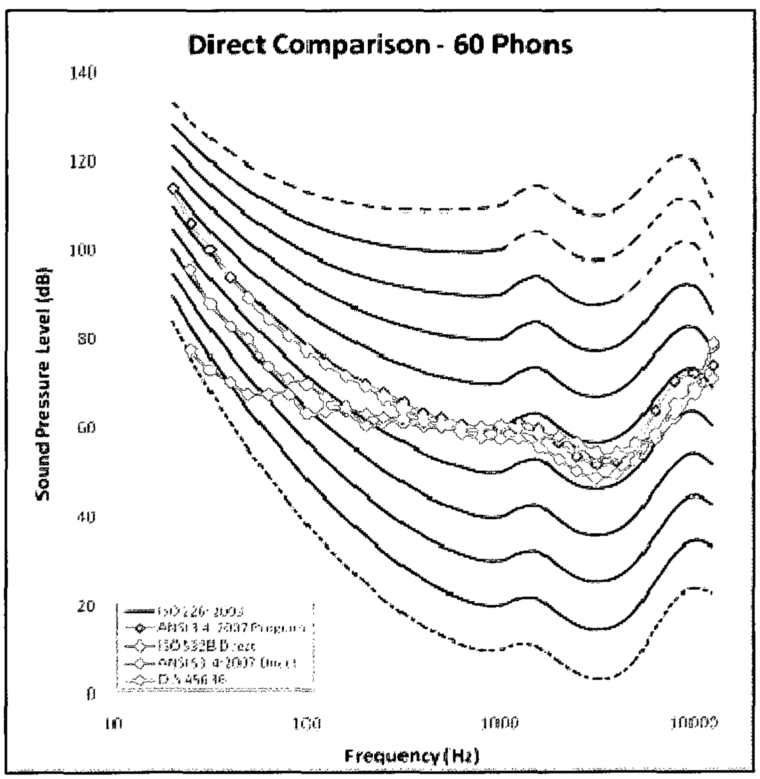


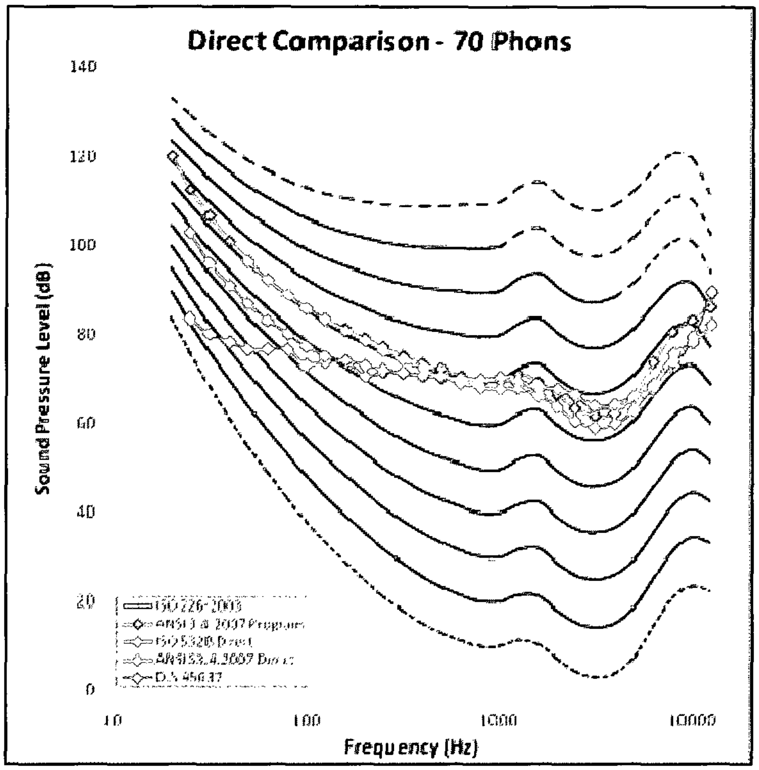


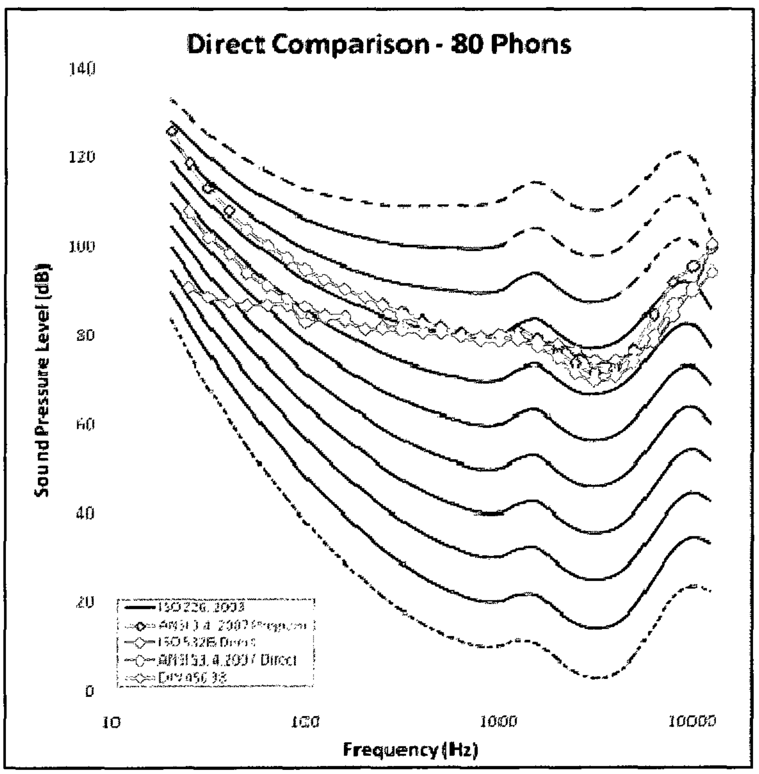

Direct Comparison - 30 Phons
Sound Pressure Level (dB)
Frequency (Hz)

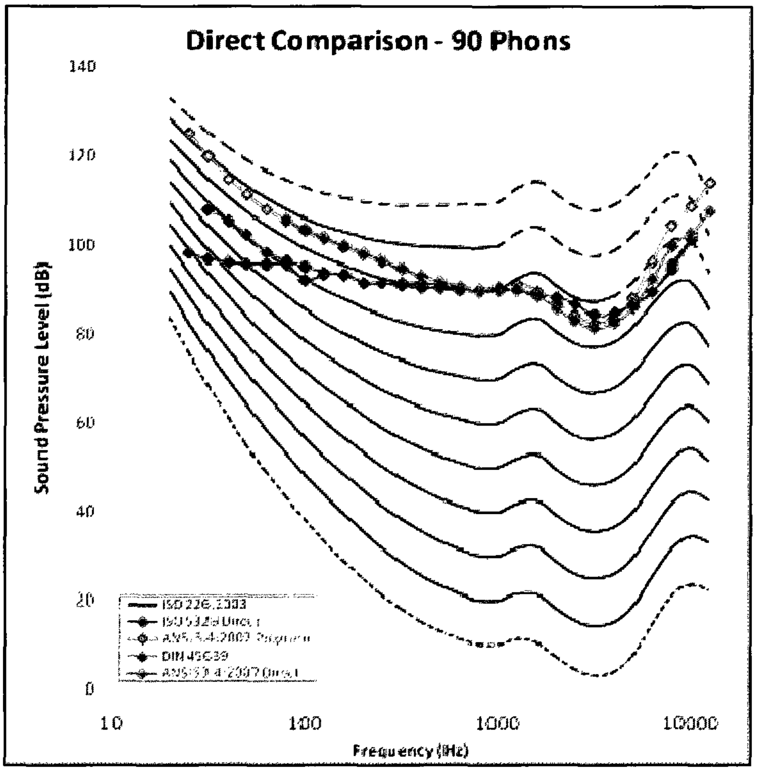

Direct Comparison - 40 Phons
Sound Pressure Level (dB)
Frequency (Hz)

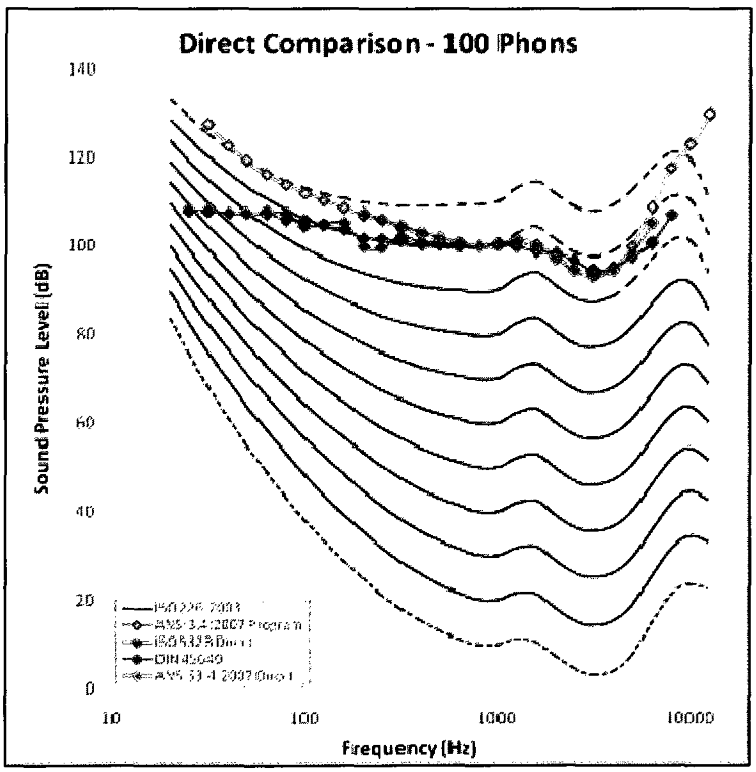

Direct Comparison - 50 Phons
Sound Pressure Level (dB)
140
120
100
80
60
40
20
0
ISO 226:2003
10
100
1000
10000
Frequency (Hz)


Direct Comparison - 60 Phons
Sound Pressure Level (dB)
140
120
100
80
60
40
20
0
10
100
1000
10000
Frequency (Hz)

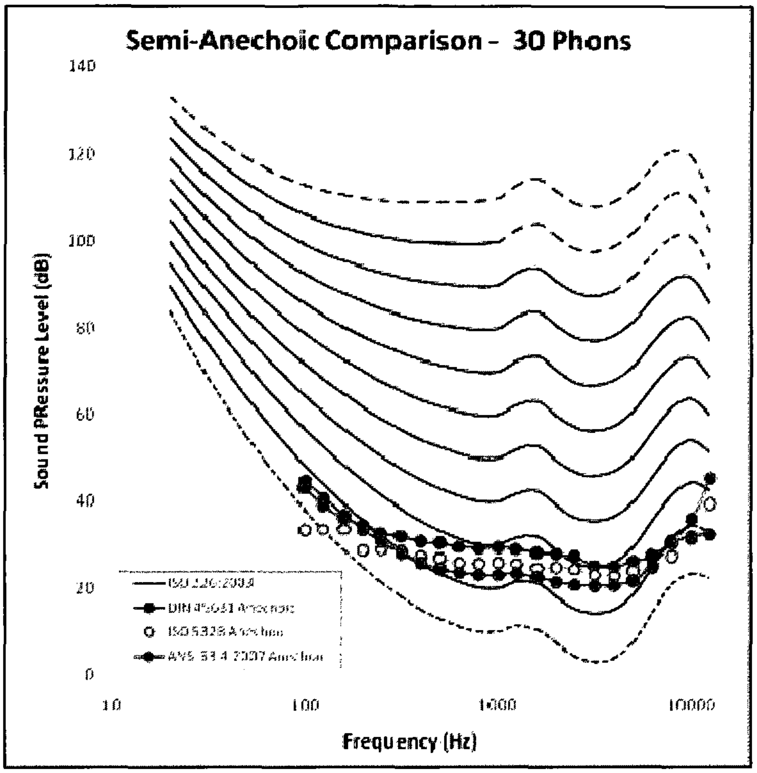

Direct Comparison - 70 Phons
Sound Pressure Level (dB)
140
120
100
80
60
40
20
0
10
100
1000
10000
Frequency (Hz)

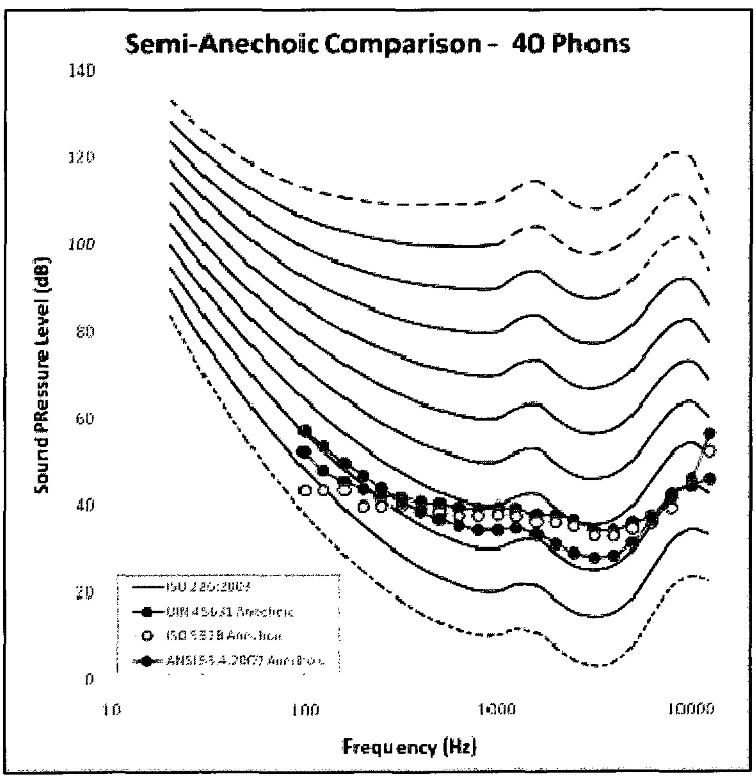

Direct Comparison - 80 Phons
Sound Pressure Level (dB)
140
120
100
80
60
40
20
0
10
100
1000
10000
Frequency (Hz)

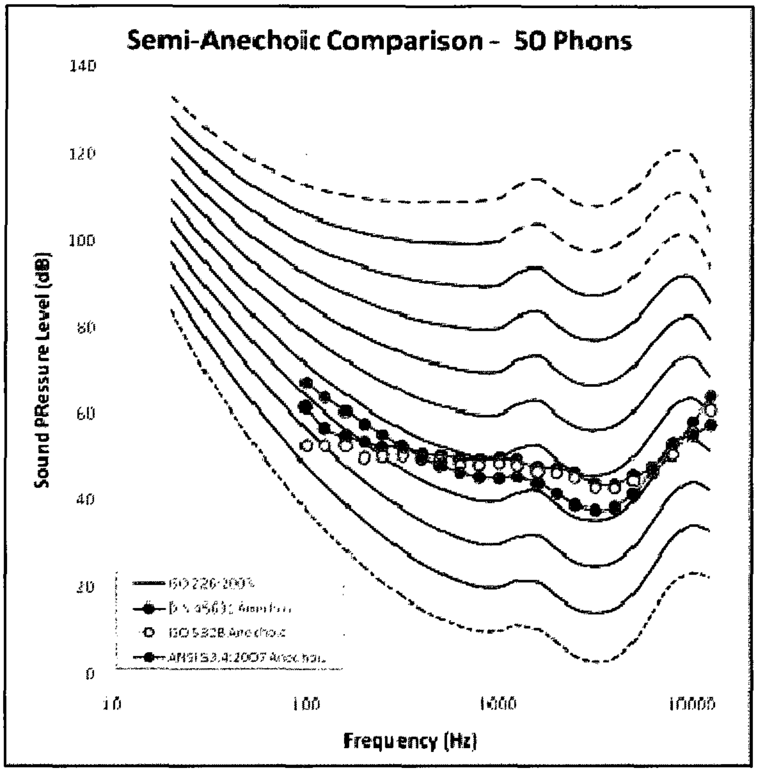

Direct Comparison - 90 Phons
Sound Pressure Level (dB)
140
120
100
80
60
40
20
0
10
100
1000
10000
Frequency (Hz)

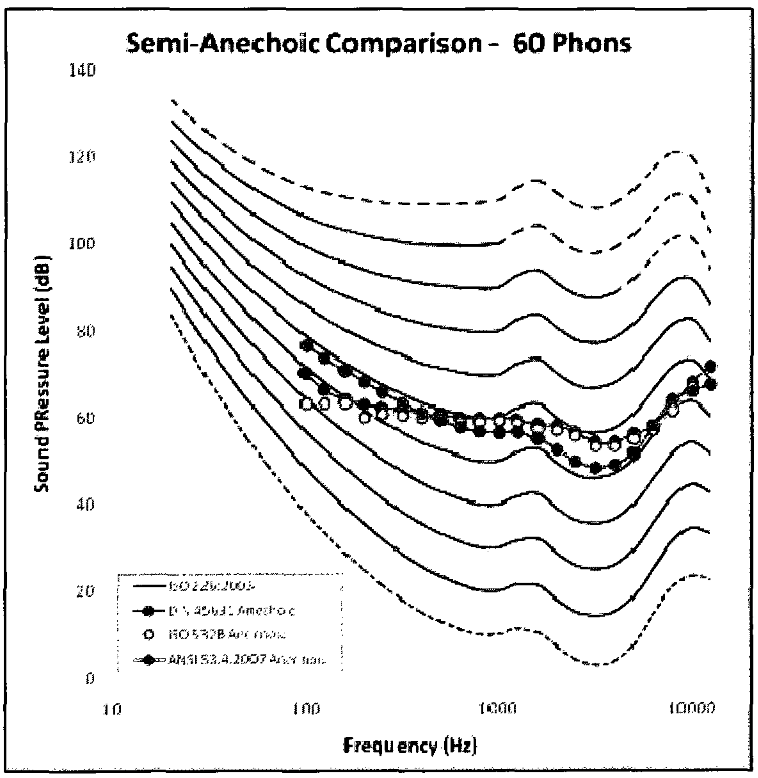

Direct Comparison - 100 Phons
Sound Pressure Level (dB)
140
120
100
80
60
40
20
0
10
100
1000
10000
Frequency (Hz)

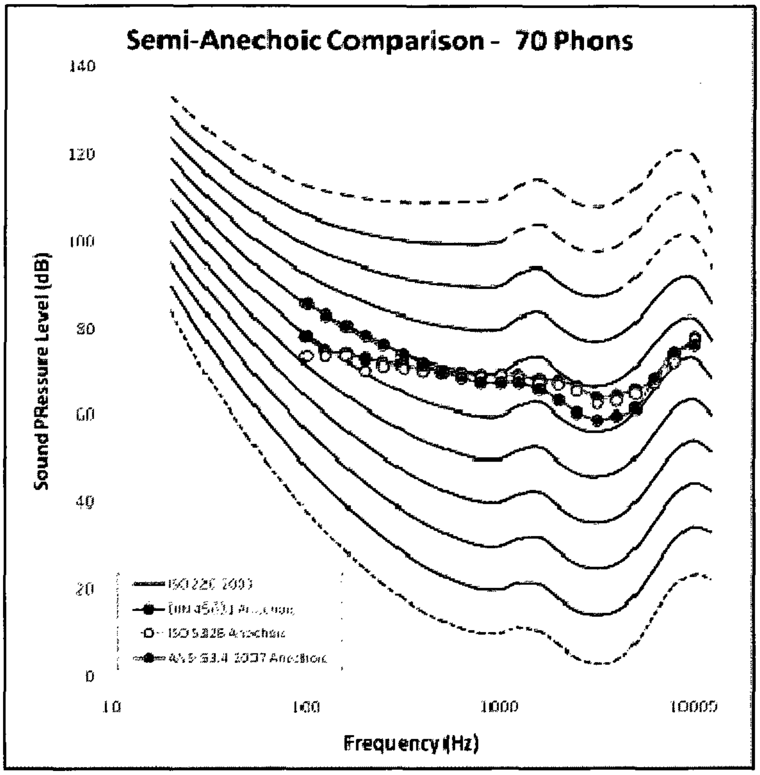
Appendix E – Semi-Anechoic Contour Comparison Plots



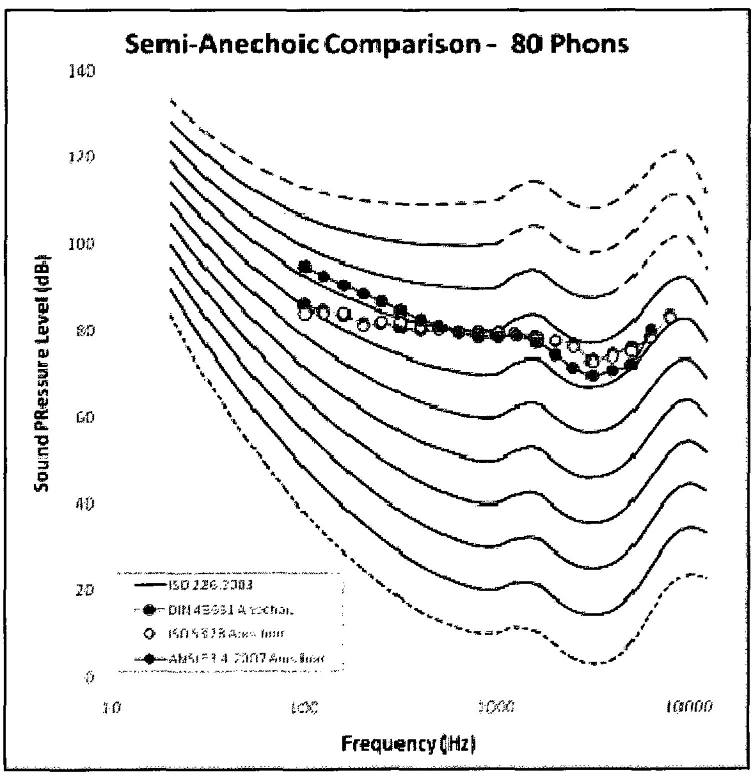



Semi-Anechoic Comparison - 50 Phons
Sound PRessure Level (dB)
140
120
100
80
60
40
20
0
10
100
1000
10000
Frequency (Hz)

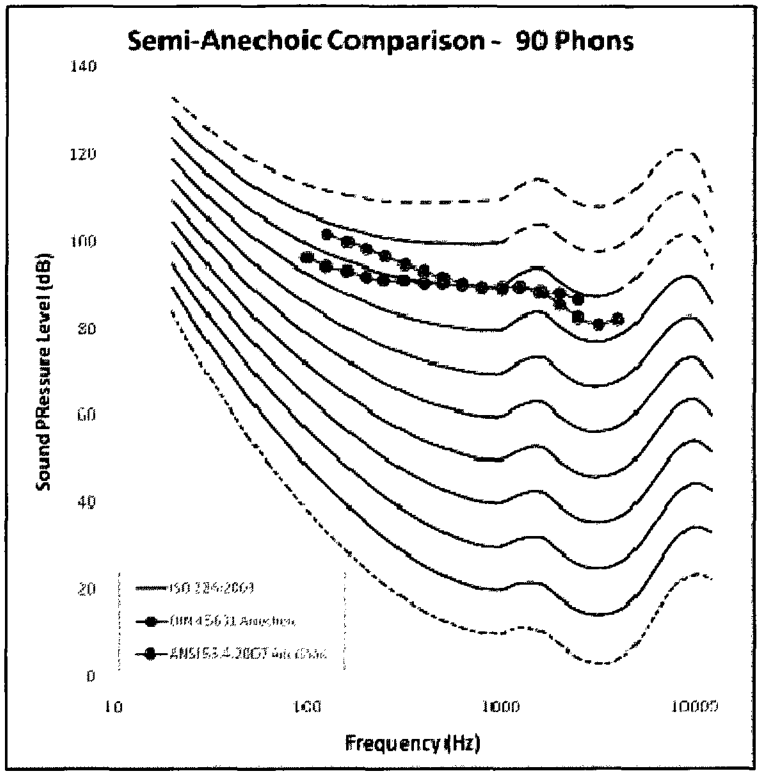

Semi-Anechoic Comparison - 60 Phons
Sound PRessure Level (dB)
140
120
100
80
60
40
20
0
10
100
1000
10000
Frequency (Hz)



Semi-Anechoic Comparison - 70 Phons
Sound PRessure Level (dB)
Frequency (Hz)



Semi-Anechoic Comparison - 80 Phons
Sound PRessure Level (dB)
Frequency (Hz)


Semi-Anechoic Comparison - 90 Phons
Sound PRessure Level (dB)
140
120
100
80
60
40
20
0
10
100
1000
10000
Frequency (Hz)

Appendix F – Overall Uncertainty Analysis Results

A complete uncertainty analysis was carried out on all of the stationary loudness metric data derived in this study. Prior to calculating the results, the data was first checked for any outlier values using Pierce's Criterion as presented in Ross' work in **[38].** From the check, no outlier values were located, indicating that all of the values were consistent for each trial. The following is a summary of the Uncertainty analysis conducted.

Appendix F.1 – Uncertainty Procedure (Reproduced from [9])

The following is a condensed procedure as produced in Defoe's dissertation (2007) of [9]. For a complete description of the procedure taking place please refer to the Defoe's work or that of Wheeler and Ganji for fundamental uncertainty concepts. [49]

Overall Uncertainty (w_R) is a combination of both systematic (u_R) and random (p_R) uncertainty which are analyzed separately as follows.

$$w_R = (u_R{}^2 + p_R{}^2)^{1/2}$$

Systematic Uncertainty (u_R)

The systematic uncertainty was determined as the magnitude of uncertainty associated with the digital signals used for processing. This value was represented as an integer between 0 and 32767 which represents the range of possible positive values encoded into a 16-bit data file. The value for systematic uncertainty was determined using the following sets of equations.

Uncertainty in the Digital Signal (u_{DS})

$$u_{DS} = \left[\left(\frac{\partial DS}{\partial v} \cdot u_v\right)^2 + \left(\frac{\partial DS}{\partial v} \cdot u_{res,ADC}\right)^2\right]^{1/2}$$

$$\rightarrow \quad \frac{\partial DS}{\partial v} = \frac{32767}{v_{FS}}$$

The sensitivity of the digital signal with respect to voltage is calculated by dividing the maximum digital signal value (32767) by the full scale voltage (v_{FS}).

$$\rightarrow \quad v_{FS} = \frac{\partial v}{\partial P} \cdot P_{FS}$$

Full scale voltage is the product of the microphone sensitivity ($\frac{\partial v}{\partial P}$) resulting from calibration, multiplied by the full scale sound pressure capable of being produced (P_{FS}).

$$\rightarrow \quad \frac{\partial v}{\partial P} = 0.0526564 \text{ V/Pa} \quad \text{(Taken From PULSE)}$$

→ $P_{FS} = P_{ref} \cdot 10^{\frac{1}{20}(L_{P,FS})}$

→ $L_{P,FS}$ = 102 dB (Collected From PULSE)

→ P_{ref} = 0.00002 Pa (Acoustics Constant)

Uncertainty in the Analog Voltage (u_v)

The uncertainty value in the analog voltage is a result of the microphone sensitivity and the uncertainty associated with both the microphone (u_P) and the preamp ($u_{FR,Preamp}$) used for collecting data. The values for the acoustic equipment were taken from their respective product data sheets.

$$u_v = \left[\left(\frac{\partial v}{\partial P} \cdot u_P\right)^2 + \left(\frac{\partial v}{\partial P} \cdot u_{FR,Preamp}\right)^2\right]^{1/2}$$

→ $u_{FR,preamp} = P \cdot \left(10^{\frac{1}{20}(U_{LFR,preamp})} - 1\right)$

→ $U_{LFR,preamp}$ = ±0.5 dB (Taken from Product Data **[62]**)

→ $u_p = P \cdot \left(10^{\frac{1}{20}(U_{Lp,mic})} - 1\right)$

→ $U_{Lp,mic}$ = ±2.0 dB (Taken from Product Data **[66]**)

→ $P = P_{ref} \cdot 10^{\frac{1}{20}(L_P)}$

Uncertainty Resulting from Quantisation Error ($u_{res,ADC}$)

The quantisation error results from the analog to digital conversion process and may be expressed as follows.

$$u_{res,ADC} = \frac{1}{2}\left(\frac{v_{FS}{}^2}{32767 \cdot v}\right)$$

→ $v = \left(\frac{\partial v}{\partial P}\right) \cdot P$

Uncertainty in the Loudness Level (u_{LN})

The uncertainty in the loudness level and loudness function is a large portion of the uncertainty measured in the system. The calculation for the loudness level uncertainty relies on the sensitivity of the loudness level conversion to small fluctuations in loudness values ($\frac{dLN}{dN}$) and the uncertainty of the various loudness metric calculations (U_N).

A $\Delta N = 0.1$ sone was used in the calculation of the sensitivities as was done in Defoe [9].

$$u_{LN} = \left|\frac{dLN}{dN} \cdot U_N\right|$$

$$\rightarrow \quad \frac{dLN}{dN} = \frac{LN(N+\Delta N) - LN(N)}{\Delta N}$$

Depending on the loudness metric being analyzed combinations of the following two equations were used to derive the resulting sensitivity. Note that as no equation was available for the ANSI S3.4:2007 model it was assumed that the equations for the DIN method were sufficient, (the two methods produced similar results).

$$\rightarrow \quad LN(N \geq 1) = 33.2 * \log_{10}(N) + 40$$

$$\rightarrow \quad LN(N < 1) = 40 * (N + 0.0005)^{0.35}$$

$$\rightarrow \quad u_N = \left[\sum_{i=1}^{m} \left(U_{Pi} \cdot \frac{\partial N}{\partial Pi}\right)^2\right]^{1/2}$$

The uncertainty in the loudness calculation is related to the sum of uncertainties of the band pressures (U_{Pi}) and the respective loudness function sensitivity to band pressure fluctuation ($\frac{\partial N}{\partial Pi}$). The number of bands used is represented by the variable (m).

$$\rightarrow \quad u_{Pi} = \left|u_{DSi} \cdot \frac{dPi}{dDS}\right|$$

For the band pressure uncertainty, the value depends heavily on the fraction of digital signal uncertainty associated with each pressure band (u_{DSi}) and the sensitivity of the full scale pressure to the digital signal.

$$\rightarrow \quad \frac{dPi}{dDS} = \frac{P_{FS}}{32767}$$

$$\rightarrow \quad u_{DSi} = \frac{P_i}{P}(U_{DS})$$

The sensitivity of the loudness functions to small fluctuations in band pressure was analyzed for each loudness model using the following function. Again following Defoe's procedure for consistency, a (ΔP_i) value was used to correspond to an increase of 1 dB. [9]

$$\rightarrow \quad \frac{\partial N}{\partial Pi} \cong \frac{N(P_1,P_2,\ldots,P_i+\Delta P_i,\ldots,P_m)-N(P_1,P_2,\ldots,P_i,\ldots,P_m)}{\Delta P_i}$$

$$\rightarrow \quad \Delta P_i = P_{ref} \cdot 10^{\frac{1}{20}(L_{Pi}+1)}$$

$$\rightarrow \quad P_i = P_{ref} \cdot 10^{\frac{1}{20}(L_{Pi})}$$

Random Uncertainty (p_R)

The random uncertainty associated with the experimental data was calculated using the method outlined in Wheeler and Ganji [49]. The random uncertainty is the product of a Student's t-value (t_S) taken from a table in [49] and the standard deviation of the result (S_R).

$$p_{\bar{R}} = t_S \cdot S_R$$

$$\rightarrow \quad S_R = \left[\sum_{i=1}^{m}\left(S_{Pi} \cdot \frac{\partial N}{\partial P_i}\right)^2\right]^{1/2}$$

$$\rightarrow \quad S_{Pi} = \left(\sum_{i=1}^{n}\frac{(P_i-\bar{P})^2}{(n-1)}\right)^{1/2}$$

$$\rightarrow \quad \bar{P} = \frac{\sum_{i=1}^{n} P_i}{n}$$

→ t_S = Taken From [49]

In order to locate a Student's t-value from the chart, a value for the degree of freedom was specified based on the following relationship for low sample measurements. The Welch-Satterthwaite formula results in 2 degrees of freedom for each of the measurements of three trials.

$$\rightarrow \quad v_{fN} = \frac{\left(S_N{}^2\right)^2}{\sum_{i=1}^{m}\left(\frac{1}{v_i}\left[\left(\frac{\partial N}{\partial P_i}S_{Pi}\right)^2\right]^2\right)} \quad \text{Taken From } \textbf{[49]}$$

Appendix F.2 – ISO 532B Error-Bar Plots

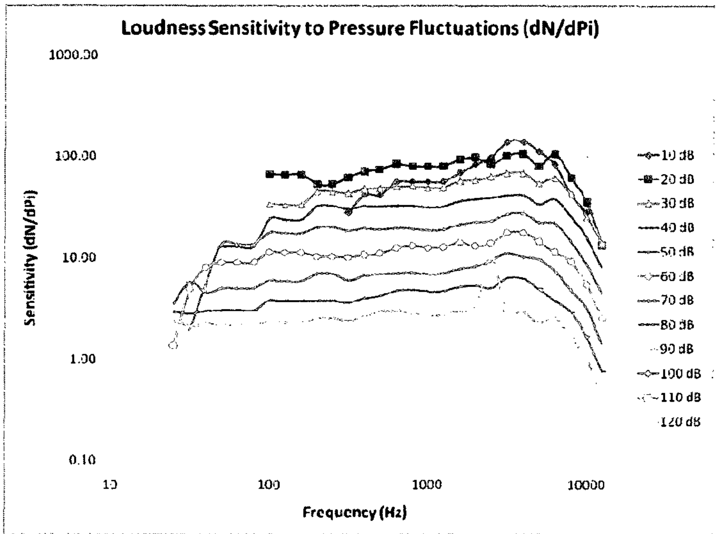


Exhibit F1 – Loudness Sensitivity of the ISO 532B Stationary Loudness Model

	ISO 532B Error Bar Values					
	30 Phon	40 Phon	50 Phon	60 Phon	70 Phon	80 Phon
20						
25						
31.5						
40						
50						
63						
80						
100	34 ± 19.0 dB	44 ± 8.3 dB	53 ± 3.0 dB	63 ± 1.1 dB	74 ± 0.4 dB	84 ± 0.3 dB
125	34 ± 21.8 dB	43 ± 8.1 dB	53 ± 3.0 dB	63 ± 1.1 dB	74 ± 0.4 dB	84 ± 0.1 dB
160	34 ± 18.5 dB	44 ± 7.9 dB	53 ± 2.9 dB	63 ± 1.0 dB	74 ± 0.4 dB	84 ± 0.1 dB
200	29 ± 15.7 dB	40 ± 7.2 dB	50 ± 2.7 dB	60 ± 1.0 dB	70 ± 0.3 dB	81 ± 0.1 dB
250	29 ± 15.4 dB	40 ± 7.0 dB	50 ± 2.6 dB	61 ± 1.0 dB	71 ± 0.4 dB	82 ± 0.1 dB
315	29 ± 15.4 dB	40 ± 6.9 dB	51 ± 2.6 dB	61 ± 0.9 dB	71 ± 0.3 dB	82 ± 0.3 dB
400	27 ± 14.4 dB	39 ± 6.7 dB	50 ± 2.5 dB	60 ± 0.9 dB	70 ± 0.3 dB	81 ± 0.1 dB
500	27 ± 14.2 dB	39 ± 6.5 dB	49 ± 2.5 dB	60 ± 0.9 dB	70 ± 0.3 dB	80 ± 0.1 dB
630	26 ± 13.6 dB	38 ± 6.4 dB	49 ± 2.4 dB	59 ± 0.8 dB	69 ± 0.3 dB	80 ± 0.1 dB
800	26 ± 13.6 dB	38 ± 6.4 dB	49 ± 2.4 dB	59 ± 0.8 dB	69 ± 0.2 dB	79 ± 0.1 dB
1000	26 ± 13.9 dB	38 ± 6.6 dB	49 ± 2.5 dB	59 ± 0.8 dB	69 ± 0.2 dB	79 ± 0.2 dB
1250	26 ± 13.8 dB	38 ± 6.5 dB	48 ± 2.5 dB	59 ± 0.8 dB	69 ± 0.2 dB	79 ± 0.2 dB
1600	25 ± 13.2 dB	36 ± 6.3 dB	47 ± 2.4 dB	57 ± 0.8 dB	68 ± 0.2 dB	78 ± 0.2 dB
2000	25 ± 13.3 dB	36 ± 6.2 dB	47 ± 2.5 dB	57 ± 0.9 dB	67 ± 0.3 dB	78 ± 0.1 dB
2500	25 ± 13.1 dB	35 ± 6.2 dB	46 ± 2.4 dB	56 ± 0.8 dB	66 ± 0.2 dB	76 ± 0.0 dB
3150	23 ± 12.5 dB	33 ± 5.9 dB	43 ± 2.3 dB	53 ± 0.8 dB	63 ± 0.3 dB	73 ± 0.6 dB
4000	23 ± 12.5 dB	33 ± 5.9 dB	43 ± 2.4 dB	53 ± 0.9 dB	63 ± 0.3 dB	74 ± 0.1 dB
5000	24 ± 13.4 dB	35 ± 6.3 dB	45 ± 2.5 dB	55 ± 0.9 dB	65 ± 0.2 dB	75 ± 0.1 dB
6300	25 ± 13.9 dB	36 ± 6.7 dB	47 ± 2.7 dB	57 ± 0.9 dB	67 ± 0.3 dB	78 ± 0.3 dB
8000	28 ± 15.3 dB	40 ± 7.5 dB	51 ± 3.0 dB	62 ± 1.0 dB	73 ± 0.2 dB	83 ± 0.0 dB
10000	31 ± 17.8 dB	45 ± 8.5 dB	56 ± 3.1 dB	67 ± 0.8 dB	78 ± 0.1 dB	
12500	40 ± 22.1 dB	52 ± 7.6 dB	61 ± 1.4 dB			

Exhibit F2 – Error Values of the ISO 532B Stationary Loudness Model

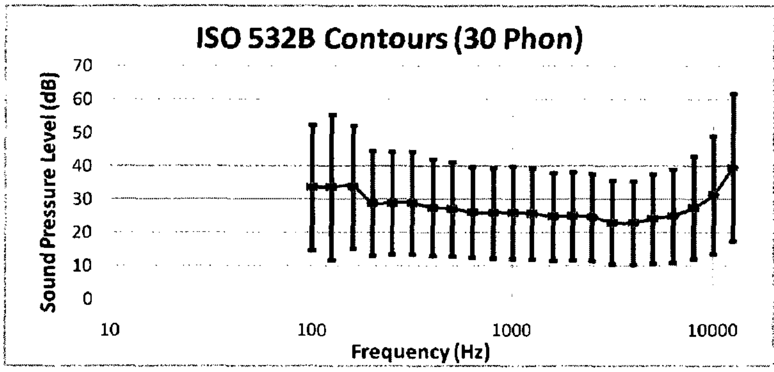


Exhibit F3 – 30 Phon Error-Bar Plot of ISO 532B

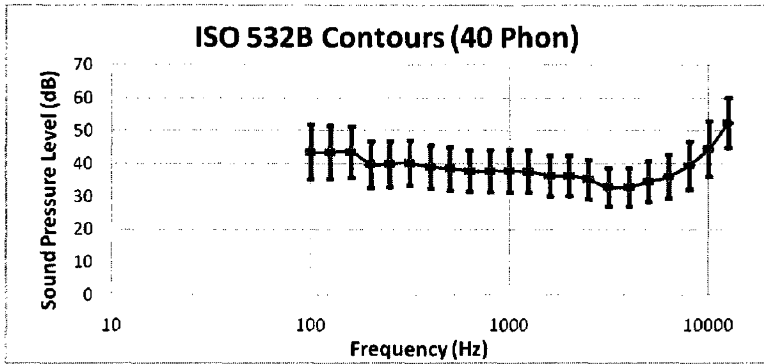


Exhibit F4– 40 Phon Error-Bar Plot of ISO 532B

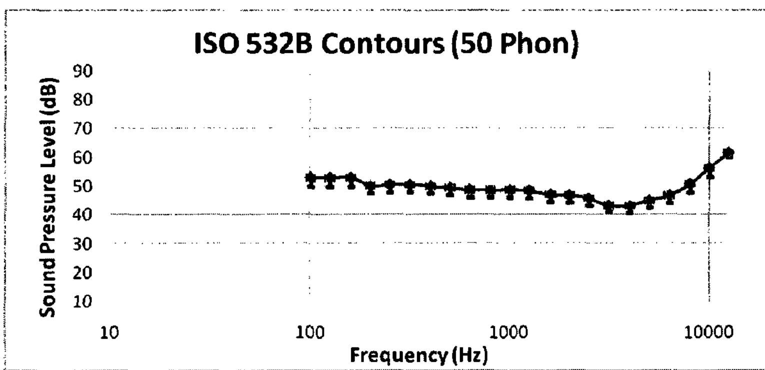


Exhibit F5– 50 Phon Error-Bar Plot of ISO 532B

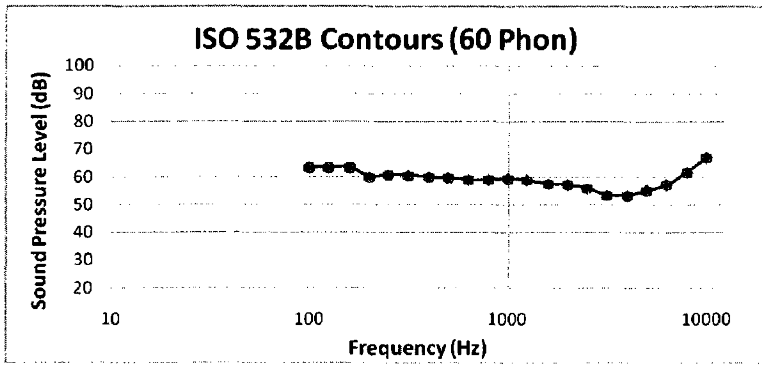


Exhibit F6– 60 Phon Error-Bar Plot of ISO 532B

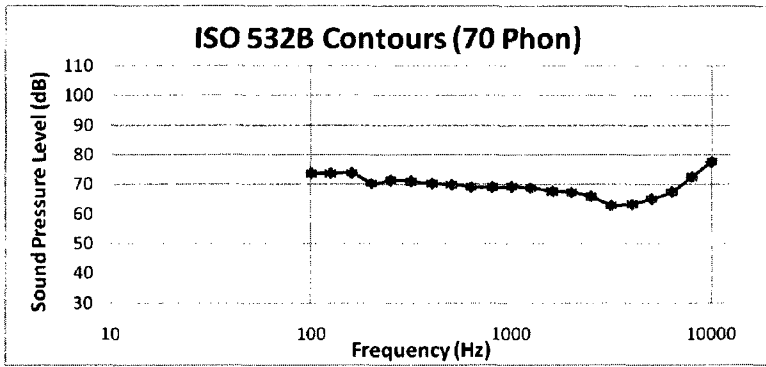


Exhibit F7– 70 Phon Error-Bar Plot of ISO 532B

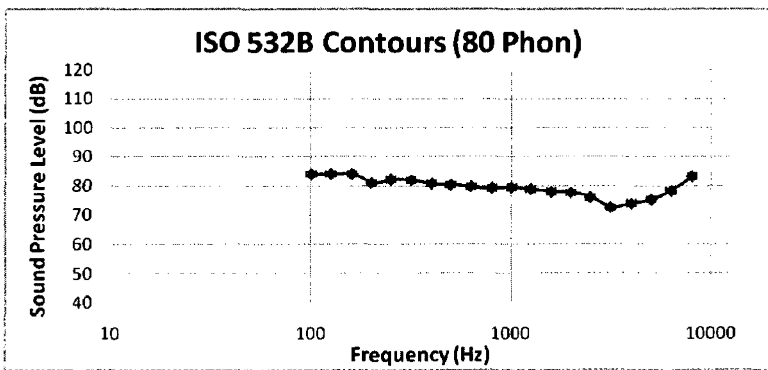


Exhibit F8– 80 Phon Error-Bar Plot of ISO 532B

Appendix F.3 – DIN 45631 Error-Bar Plots

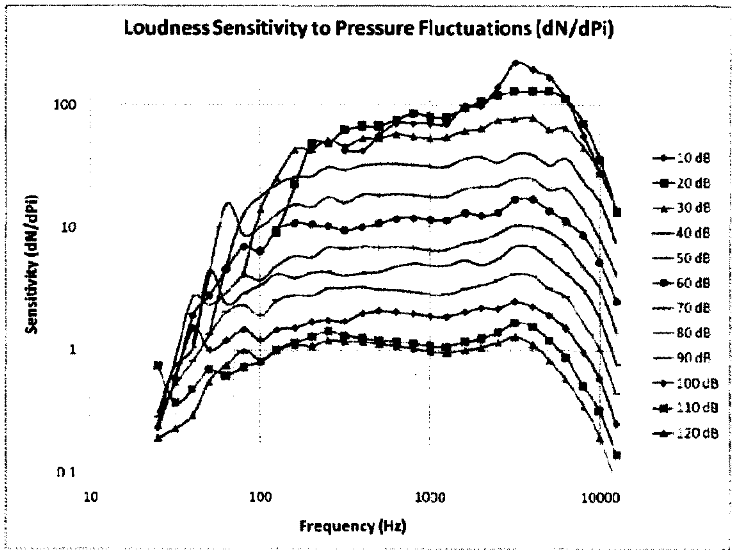


Exhibit F9 – Loudness Sensitivity of the DIN 45631 Stationary Loudness Model

	DIN 45631 Error Bar Values						
	30 Phon	40 Phon	50 Phon	60 Phon	70 Phon	80 Phon	90 Phon
20							
25							
31.5							
40							
50							
63							
80							
100	43 ± 23.0 dB	52 ± 9.6 dB	62 ± 3.1 dB	70 ± 1.1 dB	78 ± 0.3 dB	86 ± 0.3 dB	
125	39 ± 20.8 dB	48 ± 13.5 dB	57 ± 3.0 dB	66 ± 1.1 dB	75 ± 0.4 dB	84 ± 0.5 dB	94 ± 0.1 dB
160	36 ± 18.6 dB	45 ± 8.1 dB	55 ± 2.8 dB	64 ± 1.0 dB	74 ± 0.4 dB	83 ± 0.1 dB	93 ± 0.2 dB
200	34 ± 17.6 dB	44 ± 7.6 dB	54 ± 2.7 dB	63 ± 0.9 dB	73 ± 0.4 dB	81 ± 0.1 dB	92 ± 0.1 dB
250	32 ± 16.5 dB	42 ± 7.2 dB	52 ± 2.5 dB	62 ± 0.9 dB	72 ± 0.3 dB	81 ± 0.1 dB	91 ± 1.2 dB
315	32 ± 15.8 dB	42 ± 6.9 dB	52 ± 2.4 dB	62 ± 0.8 dB	72 ± 0.6 dB	81 ± 0.3 dB	91 ± 0.2 dB
400	31 ± 15.1 dB	41 ± 6.6 dB	51 ± 2.3 dB	61 ± 0.8 dB	71 ± 0.3 dB	80 ± 0.1 dB	90 ± 0.2 dB
500	30 ± 14.7 dB	40 ± 6.4 dB	51 ± 2.2 dB	61 ± 0.8 dB	70 ± 0.6 dB	80 ± 0.3 dB	90 ± 0.3 dB
630	30 ± 14.4 dB	39 ± 6.4 dB	50 ± 2.2 dB	60 ± 0.8 dB	70 ± 0.2 dB	80 ± 0.2 dB	90 ± 0.2 dB
800	29 ± 14.3 dB	39 ± 6.4 dB	50 ± 2.2 dB	60 ± 0.6 dB	69 ± 0.3 dB	80 ± 0.1 dB	89 ± 0.4 dB
1000	29 ± 14.5 dB	39 ± 6.6 dB	50 ± 2.2 dB	60 ± 0.6 dB	69 ± 0.1 dB	79 ± 0.1 dB	90 ± 0.7 dB
1250	29 ± 14.4 dB	39 ± 6.6 dB	49 ± 2.3 dB	60 ± 0.7 dB	69 ± 0.2 dB	79 ± 0.3 dB	89 ± 0.2 dB
1600	28 ± 14.0 dB	38 ± 6.4 dB	48 ± 2.3 dB	58 ± 0.8 dB	68 ± 0.2 dB	78 ± 0.1 dB	88 ± 0.2 dB
2000	28 ± 13.9 dB	38 ± 6.5 dB	48 ± 2.3 dB	58 ± 0.8 dB	68 ± 0.2 dB	78 ± 0.0 dB	88 ± 0.0 dB
2500	27 ± 13.5 dB	37 ± 6.4 dB	47 ± 2.2 dB	57 ± 0.8 dB	67 ± 0.2 dB	77 ± 0.0 dB	87 ± 0.1 dB
3150	25 ± 12.8 dB	34 ± 6.0 dB	44 ± 2.2 dB	54 ± 0.8 dB	64 ± 0.3 dB	74 ± 0.4 dB	
4000	25 ± 12.9 dB	34 ± 6.1 dB	44 ± 2.3 dB	54 ± 0.8 dB	64 ± 0.3 dB	74 ± 0.1 dB	
5000	26 ± 13.7 dB	36 ± 6.4 dB	46 ± 2.3 dB	56 ± 0.8 dB	66 ± 0.2 dB	76 ± 0.9 dB	
6300	27 ± 14.4 dB	38 ± 6.9 dB	48 ± 2.5 dB	58 ± 0.9 dB	68 ± 0.4 dB	78 ± 0.3 dB	
8000	31 ± 16.3 dB	42 ± 7.7 dB	53 ± 2.8 dB	63 ± 1.0 dB	74 ± 0.3 dB	83 ± 0.6 dB	
10000	36 ± 18.6 dB	46 ± 5.8 dB	58 ± 2.9 dB	68 ± 0.7 dB	78 ± 1.1 dB		
12500	45 ± 23.3 dB	56 ± 7.6 dB	64 ± 1.3 dB	72 ± 0.4 dB			

Exhibit F10 – Error Values of the DIN 45631 Stationary Loudness Model

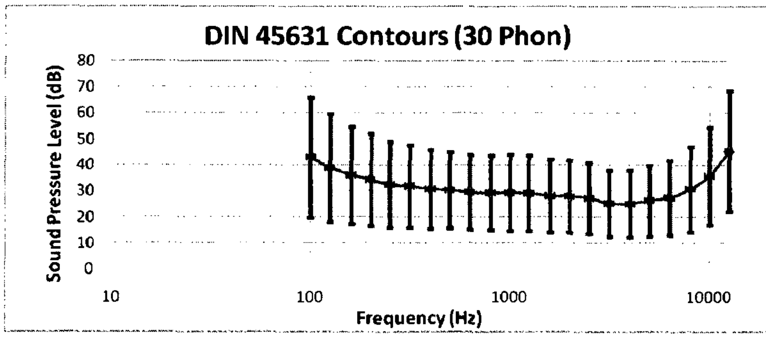


Exhibit F11 – 30 Phon Error-Bar Plot of DIN 45631

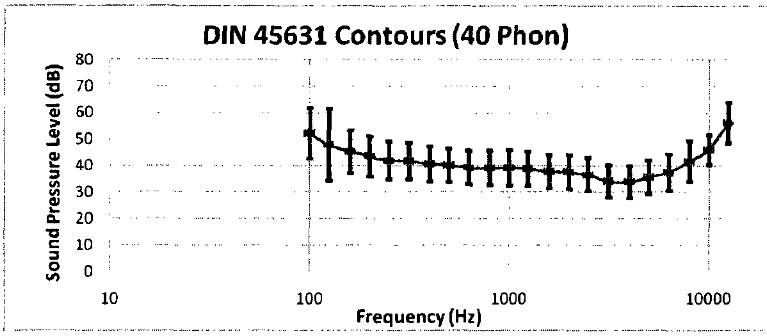


Exhibit F12 – 40 Phon Error-Bar Plot of DIN 45631

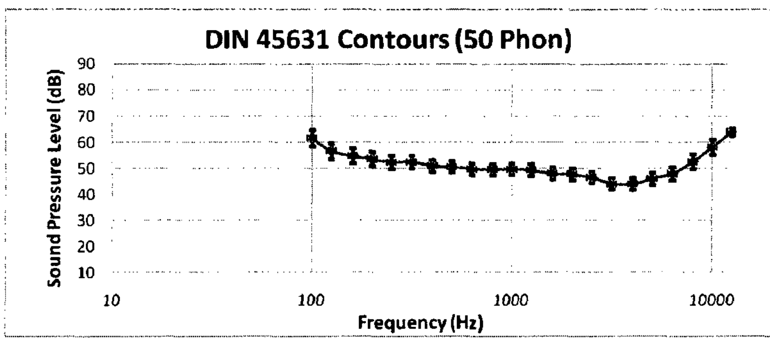


Exhibit F13 – 50 Phon Error-Bar Plot of DIN 45631

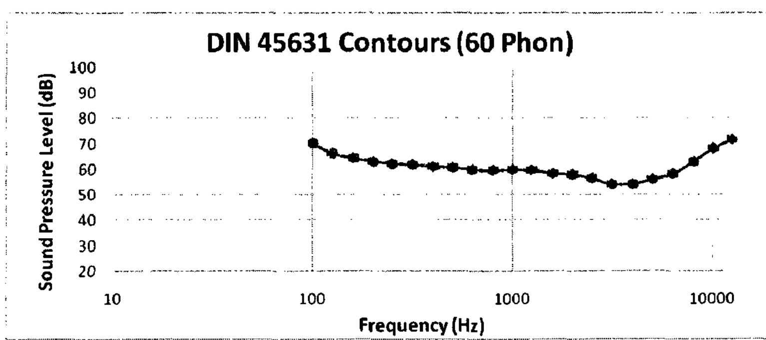


Exhibit F14 – 60 Phon Error-Bar Plot of DIN 45631

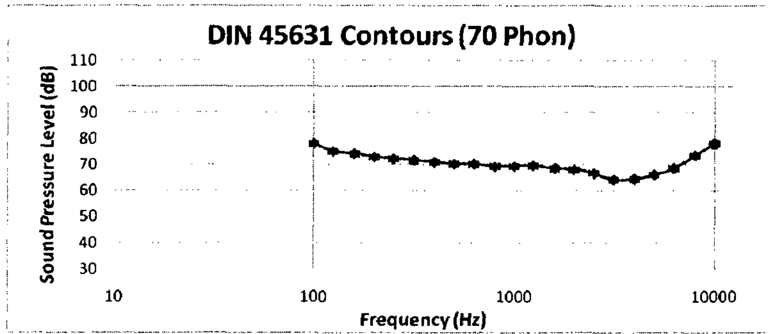


Exhibit F15 – 70 Phon Error-Bar Plot of DIN 45631

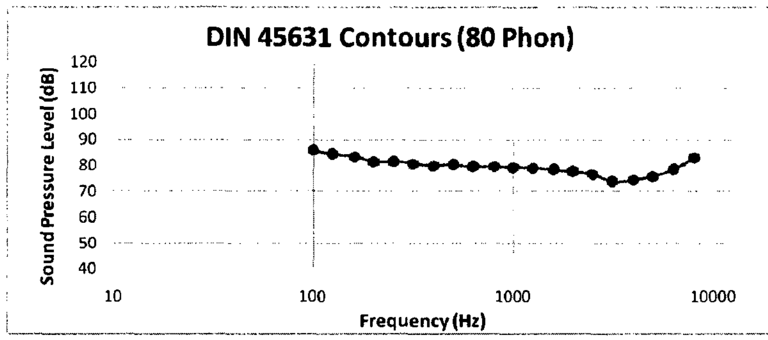


Exhibit F16 – 80 Phon Error-Bar Plot of DIN 45631

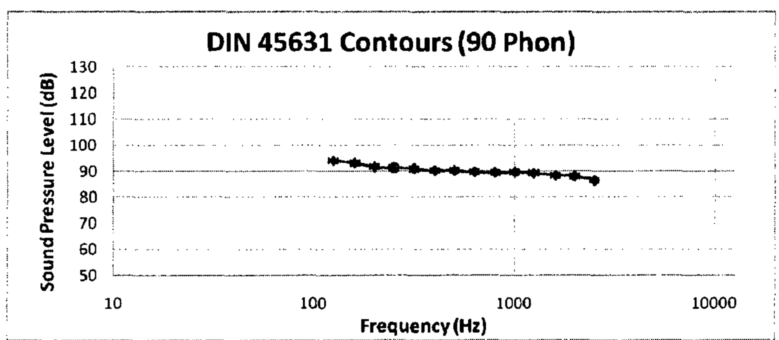


Exhibit F17 – 90 Phon Error-Bar Plot of DIN 45631

Appendix F.4 – ANSI S3.4:2007 Error-Bar Plots

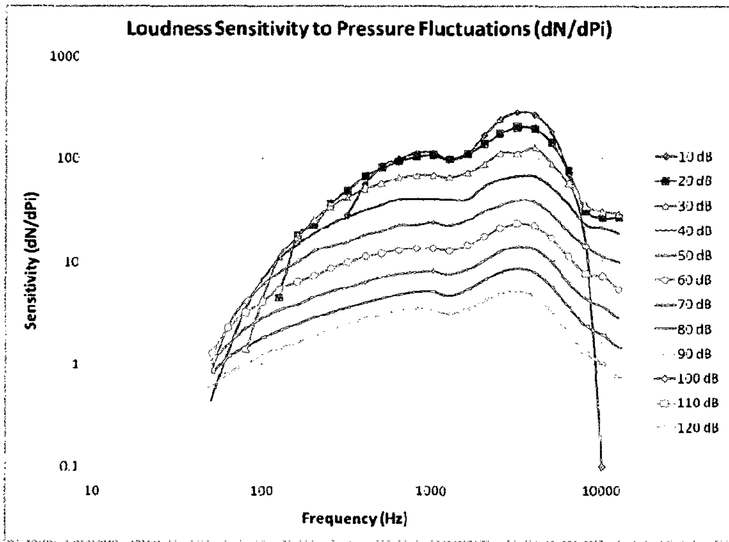


Exhibit F18 – Loudness Sensitivity of the ANSI S3.4:2007 Stationary Loudness Model

	ANSI S3.4:2007 Error Bar Values						
	30 Phon	40 Phon	50 Phon	60 Phon	70 Phon	80 Phon	90 Phon
20							
25							
31.5							
40							
50							
63							
80							
100	45 ± 28.9 dB	57 ± 13.1 dB	67 ± 4.4 dB	77 ± 1.4 dB	86 ± 0.6 dB	95 ± 0.2 dB	
125	41 ± 26.3 dB	54 ± 12.3 dB	64 ± 4.3 dB	74 ± 1.4 dB	83 ± 0.6 dB	93 ± 0.6 dB	102 ± 1.3 dB
160	37 ± 23.7 dB	50 ± 11.2 dB	61 ± 4.0 dB	71 ± 1.3 dB	81 ± 0.6 dB	90 ± 0.5 dB	100 ± 1.7 dB
200	34 ± 21.3 dB	47 ± 10.4 dB	58 ± 3.7 dB	68 ± 1.2 dB	78 ± 0.5 dB	88 ± 0.6 dB	98 ± 1.6 dB
250	31 ± 19.4 dB	44 ± 9.8 dB	55 ± 3.5 dB	66 ± 1.2 dB	76 ± 0.4 dB	87 ± 0.4 dB	97 ± 1.2 dB
315	28 ± 17.8 dB	41 ± 9.0 dB	53 ± 3.3 dB	63 ± 1.1 dB	74 ± 0.4 dB	85 ± 0.4 dB	95 ± 0.8 dB
400	26 ± 16.0 dB	39 ± 8.3 dB	50 ± 3.0 dB	61 ± 1.0 dB	72 ± 0.4 dB	83 ± 0.5 dB	93 ± 0.2 dB
500	25 ± 15.2 dB	37 ± 7.8 dB	48 ± 2.9 dB	59 ± 1.0 dB	70 ± 0.4 dB	81 ± 0.1 dB	92 ± 0.8 dB
630	24 ± 14.5 dB	35 ± 7.3 dB	47 ± 2.7 dB	58 ± 0.9 dB	69 ± 0.3 dB	80 ± 0.3 dB	90 ± 0.7 dB
800	23 ± 14.2 dB	34 ± 7.1 dB	46 ± 2.7 dB	57 ± 0.9 dB	68 ± 0.3 dB	78 ± 0.4 dB	89 ± 0.7 dB
1000	23 ± 14.2 dB	34 ± 7.2 dB	45 ± 2.7 dB	57 ± 0.9 dB	68 ± 0.4 dB	78 ± 0.3 dB	89 ± 1.4 dB
1250	23 ± 14.4 dB	35 ± 7.3 dB	46 ± 2.6 dB	57 ± 0.9 dB	68 ± 0.2 dB	79 ± 0.3 dB	90 ± 1.0 dB
1600	23 ± 13.9 dB	33 ± 6.9 dB	44 ± 2.5 dB	55 ± 0.8 dB	66 ± 0.2 dB	77 ± 0.3 dB	88 ± 0.7 dB
2000	22 ± 13.1 dB	31 ± 6.4 dB	42 ± 2.4 dB	53 ± 0.8 dB	64 ± 0.2 dB	74 ± 0.1 dB	85 ± 0.7 dB
2500	21 ± 14.4 dB	29 ± 5.9 dB	39 ± 2.2 dB	50 ± 0.8 dB	61 ± 0.2 dB	71 ± 0.7 dB	83 ± 0.3 dB
3150	21 ± 12.6 dB	28 ± 5.8 dB	38 ± 2.2 dB	48 ± 0.7 dB	59 ± 0.2 dB	70 ± 0.1 dB	81 ± 1.2 dB
4000	21 ± 12.9 dB	28 ± 6.0 dB	39 ± 2.3 dB	49 ± 0.9 dB	60 ± 0.3 dB	71 ± 0.2 dB	82 ± 0.8 dB
5000	22 ± 13.8 dB	31 ± 6.8 dB	42 ± 2.6 dB	52 ± 0.6 dB	62 ± 0.1 dB	72 ± 0.3 dB	
6300	25 ± 16.1 dB	37 ± 8.2 dB	48 ± 3.0 dB	58 ± 1.0 dB	69 ± 0.4 dB	80 ± 0.5 dB	
8000	30 ± 19.4 dB	43 ± 9.8 dB	54 ± 3.6 dB	64 ± 1.0 dB	74 ± 0.3 dB		
10000	32 ± 20.6 dB	44 ± 10.2 dB	55 ± 3.7 dB	66 ± 1.0 dB	76 ± 0.2 dB		
12500	33 ± 20.9 dB	46 ± 10.5 dB	58 ± 3.4 dB	68 ± 0.7 dB			

Exhibit F19 – Error Values of the ANSI S3.4:2007 Stationary Loudness Model

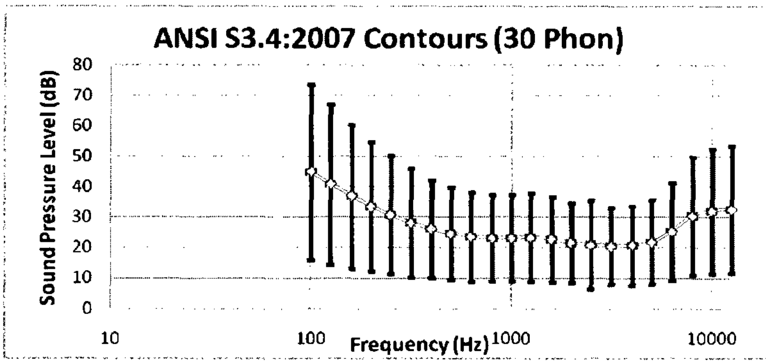


Exhibit F20 – 30 Phon Error-Bar Plot of ANSI S3.4:2007

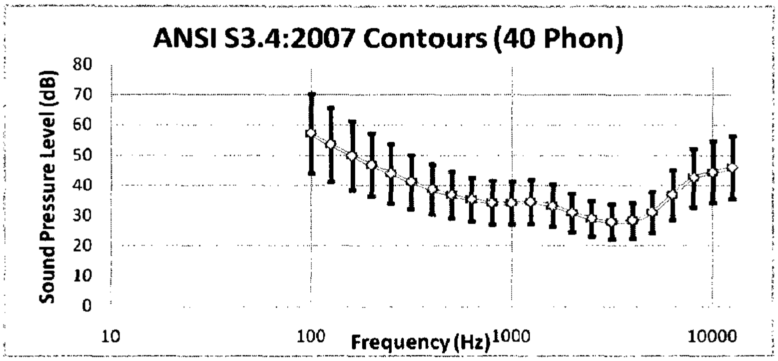


Exhibit F21 – 40 Phon Error-Bar Plot of ANSI S3.4:2007

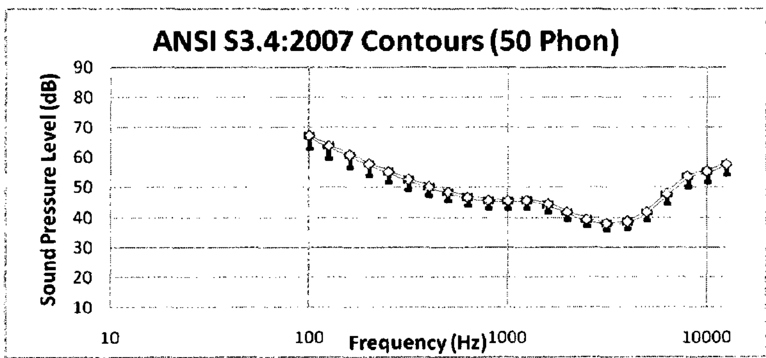


Exhibit F22 – 50 Phon Error-Bar Plot of ANSI S3.4:2007

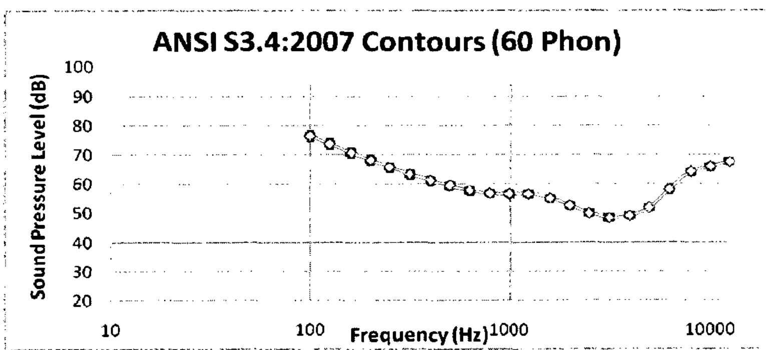


Exhibit F23 – 60 Phon Error-Bar Plot of ANSI S3.4:2007

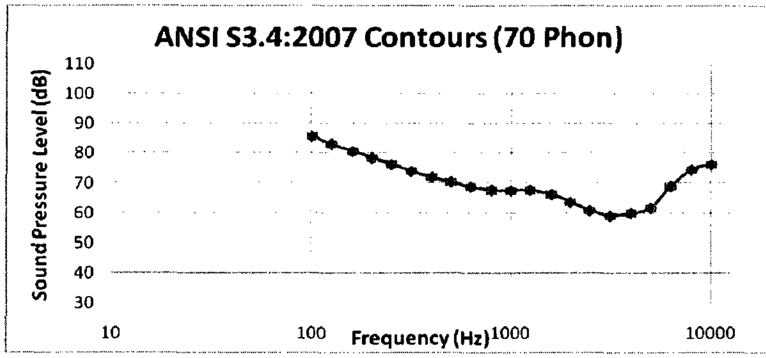


Exhibit F24 – 70 Phon Error-Bar Plot of ANSI S3.4:2007

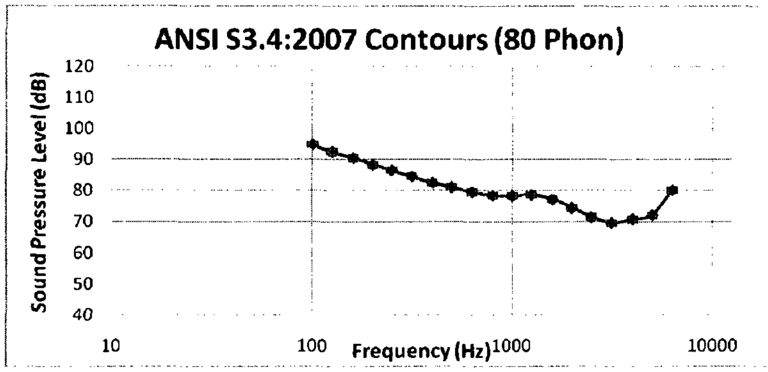


Exhibit F25 – 80 Phon Error-Bar Plot of ANSI S3.4:2007

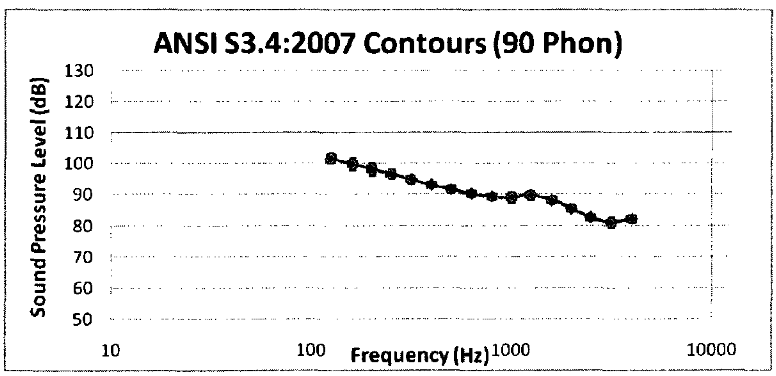


Exhibit F26 – 90 Phon Error-Bar Plot of ANSI S3.4:2007

X. VITA AUCTORIS

Jeremy Charbonneau was born in Chatham, Ontario on February 12th, 1985. He graduated from Blenheim District High School in 2003. From there he went on to the University of Windsor Ontario where he obtained a B.Sc. in Mechanical Engineering in 2008. He is currently a candidate for the Master's degree in Mechanical Engineering at the University of Windsor and hopes to graduate in the Spring of 2010.