2021

Development of the Carbon Nanotube Thermoacoustic Loudspeaker

Troy Bouman
Michigan Technological University, tmbouman@mtu.edu

DEVELOPMENT OF THE CARBON NANOTUBE THERMOACOUSTIC LOUDSPEAKER

By

Troy M. Bouman

A DISSERTATION

Submitted in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

In Mechanical Engineering – Engineering Mechanics

MICHIGAN TECHNOLOGICAL UNIVERSITY

2021

This dissertation has been approved in partial fulfillment of the requirements for the Degree of DOCTOR OF PHILOSOPHY in Mechanical Engineering – Engineering Mechanics.

Department of Mechanical Engineering – Engineering Mechanics

Dissertation Advisor:	*Andrew Barnard*
Committee Member:	*Charles Van Karsen*
Committee Member:	*Jason Blough*
Committee Member:	*Christopher Plummer*
Department Chair:	*William Predebon*

Table of Contents

List of figures

List of tables

Author contribution statement

Chapter 2 of this dissertation is reproduced in its entirety from:

Bouman, T., Barnard, A., and Asgarisabet, M.. "Experimental quantification of the true efficiency of carbon nanotube thin-film thermophones." The Journal of the Acoustical Society of America 139.3 (2016): 1353-1363. https://doi.org/10.1121/1.4944688

with permission of the Acoustical Society of America. Chapter 3 of this dissertation is reproduced in its entirety from:

Bouman, T., Barnard, A., and Alexander, J.. “Continued Drive Signal Development for the Carbon Nanotube Thermoacoustic Loudspeaker Using Techniques Derived from the Hearing Aid Industry.” SAE Technical Paper No. 2017-01-1895 (2017). https://doi.org/10.4271/2017-01-1895

with permission of SAE International. Chapter 4 of this dissertation is reproduced in its entirety from a submission to the Journal of the Audio Engineering Society. That submission happened on February 11, 2021 and is currently under review. Copyright documentation for Chapters 2 & 3 are in Appendix A.

Acknowledgements

No great achievement is ever accomplished alone. It requires hard work from the individual, those actively supporting them, those who came before them, and a little luck. This accomplishment is no different.

Thank you to my parents for showing me what work ethic looks like. To my mom for always working so hard to provide for my family and for showing me how to use that work ethic to achieve my goals. To my dad for thrusting me, usually wrench first, into the physical world around us and teaching me that solving the world's engineering problems would not be easy. I feel very lucky to be your kid.

Thank you to my wife and daughter. This achievement came as a sacrifice to your time with me and time is something we can't get back. Thank you for that selflessness.

Thank you to my advisor. It has always been obvious to me that you view my success as paramount to your own. I wish academia was lucky enough to have more advisors like you. Your significant donation of time, talent, and treasure to my dissertation is second to none.

Additionally, I feel very fortunate to be able to have a long list of family, mentors, friends, colleagues, and others who supported me in this effort. I am forever grateful.

To those who came before me, while my work my not feel as momentous as some of yours (e.g. Fourier, Shapton, etc…) thank you for clearing the path so I could walk my own. I feel very lucky to have been able to walk this path and I hope my luck continues!

Nomenclature

μ	Micro 10^{-6}
η	True efficiency
$\prod$	Acoustic power (watts)
π	Pi 3.14159
ω	Frequency (radians/second)
ρ_0	Density (kg/m^3)
β_0	Convection heat trans. coef. (W/(m^2 K))
A	Signal amplitude (volts)
ABAB	Paired comparison question duplicates where sample A then B where played and then sample A then B were repeated
ABBA	Paired comparison question repeatability test where sample A and then B were played and at a later time the opposite sampled B then A were played
AC	Alternating Current
A_c	Amplitude of the carrier frequency in amplitude modulation
adjR^2	The adjusted coefficient of determination to account for varying degrees of freedom
AM	Amplitude Modulation CNT drive signal processing method
A_m	Amplitude of the modulation signal in amplitude modulation
AMAC	This is the same method as AM
ANSI	American National Standards Institute
AP	Adaptive Predistortion drive signal processing method
Avg	Average or mean
B	DC offset (volts)
c	Speed of sound (m/s)
CNN	Convolution Neural Network machine learning model
CNT	Carbon Nanotube
CntUP	Unprocessed CNT drive signal. This is the same as UP
c_p	Specific heat (J/(kg K))
CVD	Chemical Vapor Deposition
dB	Decibel
DC	Direct Current offset CNT drive signal processing method
DCAC	This is the same method as DC
DI	Directivity Index
DV	Dependent Variable in a regression analysis
f	Frequency (hertz)
FA	Factor Analysis
F_c	Carrier frequency in amplitude modulation
FCAC	Spectral Envelope Decimation. This is the same method as SED
FCAC	Spectral Envelope Decimation
FFT	Fast Fourier Transform
F_m	Modulation frequency in amplitude modulation
HMM	Hidden Markov machine learning Model
IV	Independent Variable in a regression anaysis
Logistic Regression	Regression where the independent variable is categorical

L_p	Sound pressure (decibels)
L_w	Sound power (decibels)
MWNT	Multi-walled Nanotube
NaN	Not a Number
OLS	Ordinary Least Squares regression
OTO	One Third Octave
p	Pico 10^{-12}
PAM	Pulse Amplitude Modulation CNT drive signal processing method
PC	Percent Correct
PCA	Principle Component Analysis
P_{input}	Electric power (watts)
PSD	Power Spectral Density
r	Radius (meters)
R^2	Coefficient of determination
S	Surface area (m^2)
SED	Spectral Envelope CNT drive signal processing method. This is the same as FCAC.
SII	Speech Intelligibility Index
SPL	Sound Pressure Level
STFT	Short Term Fourier Transform
STI	Speech Transmission Index
STOI	Short Term Objective Intelligibility
t	Time (seconds)
T_0	Ambient temperature (K)
T_a	Average temperature of thin film (K)
TCAC	Dynamic Linear Frequency Compression CNT drive signal processing method
THD	Total Harmonic Distortion
THDN	THD + Noise
THDNSI	THDSI + Noise
THDSI	Total Harmonic Distortion for Speech Intelligibility
TradUP	This is the traditional moving coil loudspeaker with an unprocessed drive signal
W_{rms}	Root Mean Square power in Watts

Abstract

Traditional speakers make sound by attaching a coil to a cone and moving that coil back and forth in a magnetic field (aka moving coil loudspeakers). The physics behind how to generate sound via this velocity boundary condition has largely been unchanged for over a hundred years. Interestingly, around the time moving coil loudspeakers were first investigated the idea of using heat to generate sound was also known. These thermoacoustic speakers heat and cool a thin material at acoustic frequencies to generate the pressure wave (i.e. they use a thermal boundary condition). Unfortunately, when the thermoacoustic principle was initially discovered there was no material with the right properties to heat and cool fast enough. Carbon nanotube (CNT) loudspeakers first generated sound early in the 21st century. At that time there were many questions unanswered about their place in the sound generation toolbox of an engineer.

The main goal of this dissertation was to continue the development of the CNT loudspeaker with focus on practical usage for an acoustic engineer. Prior to 2014, when this effort began, most of the published development work was from material scientists with objective acoustic performance data presented that was not useful beyond the scope of that particular publication. For example, low sound pressure levels in the nearfield at low power inputs was a common metric. Therefore, this effort had three main objectives with emphasis placed on acquiring data at levels and in nomenclature that would be useful to acoustic engineers so they could bring the technology to market, if adequate.

i) Investigation into the true power efficiency of CNT loudspeakers
ii) Investigation into alternative methods to linearize the pressure response of CNT loudspeakers
iii) Investigation into the sound quality of CNT loudspeakers

Overall, it was found that CNT loudspeakers are approximately four orders of magnitude less power efficient than traditional moving coil loudspeakers. The non-linear pressure output of the CNT loudspeakers can be linearized with a variety of drive signal processing methods, but the selection of which method to use depends on a variety of factors (e.g. amplification architecture available). In general, all methods studied are on the same order of magnitude power efficiency, but the direct current offset and amplitude modulation drive signal processing methods are superior in terms of sound quality.

1 Introduction

1.1 Motivation of research

The use of heat to generate sound (i.e. the thermoacoustic effect) was first discovered by Braun 1898 [1] and was later elaborated on by Arnold 1917 [2] to described the perfect material required to use this effect. Unfortunately, at that time the closest available material was 700nm thick platinum and its frequency response was below the human audible range. This lack of sufficient material caused thermoacoustic loudspeaker development to fall behind that of the modern moving coil loudspeaker. In 2006, Yu et. al used the relatively new material, carbon nanotubes, to generate acoustic waves up to 3kHz [3]. This demonstration set in motion the need for further development of these thermoacoustic transducers to understand their place in the market. From 2006 to the early 2010s most of the development was being done by material scientists. While their work was important for the development of the underlying structure, their results were not acoustic engineer compatible. The results often showed objective performance data in the near field at low power levels where the sound pressure level (SPL) would likely be in the background moving more than a few meters away. This drove the need for development of these devices from the acoustics perspective. If these transducers were going to be commercially viable there were certain topics that needed more investigation.

Looking back to fall of 2014, it was obvious that the next objective metric needed was true power efficiency. Barnard et. al were able to generate 111 dBA at 1m at 2kHz, but that required 6kW$_{pk}$ of input power [4]. This was an important step in acknowledging the output pressure capability, but it also emphasized the potential efficiency concerns. Additionally, the data, as with all other publications at that time, was SPL output for a given electrical power input. That is not a watts-to-watts energy comparison. A comparison of electrical input power to acoustic power (i.e. a true power efficiency) was needed. The true power efficiency data was needed for any future modeling effort. Additionally, at that time there was limited effort put into dealing with the nonlinear pressure output of the CNT loudspeaker (i.e. the frequency doubling issue). Therefore, efficiency and linearization were prioritized at the beginning of this dissertation.

As the efficiency and linearization effort was concluded, there were many other aspects of the technology that needed to be addressed (e.g. durability). After considering the dissertation’s scope and what the rest of the research group was studying, it made sense to focus the final portion of this dissertation on the subjective quality of the CNT loudspeaker. There has never been mention of their sound quality in any published paper, but to anyone who has ever heard them in person it is obvious that they are inferior to traditional moving coil loud speakers. To better understand the reason for the difference required a subjective evaluation.

The purpose of this introduction is to set the stage for why efficiency, linearization, and sound quality were selected as the topics to focus on. At the beginning of each of the following chapters there will be a detailed introduction for that topic. This introduction is simply an introduction to the dissertation.

1.2 Objectives

The main objective of this dissertation was to characterize CNT loudspeakers with a focus on generating information (e.g. data and knowledge) that would be applicable to the acoustics community. The end result being information that the community could use to determine if CNT loudspeakers could be brought to market for their specific application. The specific information that this effort planned to discover was power efficiency data, knowledge about sound pressure linearization methods that do not require expensive amplification, and sound quality data.

1.3 Explanation of chapters

Chapter 2 is a reproduction of a Journal of the Acoustic Society of America publication titled *Experimental quantification of the true efficiency of carbon nanotube thin-film thermophones* [5]. This publication is the initial work of quantifying the true power efficiency of the CNT loudspeaker. Prior to this publication only sound pressure at known distances was published and most of those distances were in the near field. This led to data that could not be generalized. In order to model performance of a CNT loudspeaker, the true electrical energy input to output sound power is needed. The directivity was not measured, but can be assumed to be a monopole for low frequencies. The power efficiency data along with total harmonic distortion (THD) was presented for drive signal processing methods amplitude modulation (AM), direct current offset (DC), and unprocessed (CntUP). Data for a traditional moving coil loudspeaker was also presented (TradUP).

Chapter 3 is a reproduction of a SAE Technical Paper titled *Continued Drive Signal Development for the Carbon Nanotube Thermoacoustic Loudspeaker Using Techniques Derived from the Hearing Aid Industry* [6]. This publication expanded the work of Chapter 2 to include two additional linearization methods obtained from the hearing aid industry. A frequency domain method called spectral envelope decimation (SED/FCAC) and a time domain method called dynamic linear frequency compression (TCAC) were demonstrated and their power efficiency and THD were quantified. The importance of the expansion to these methods was that they do not require a class AB amplifier like DC nor do they require a high frequency amplifier like AM. These methods can be used with an inexpensive class D amplifier.

Chapter 4 is currently under review at the Journal of the Audio Engineering Society. This paper is titled *Subjective evaluation of carbon nanotube loudspeaker drive signal processing methods using single word techniques*. This effort used single word spoken text to evaluate the drive signal processing methods in order to determine which drive signal processing method is more intelligible on a relative scale and if CNT loudspeakers are intelligible on an absolute scale. The study performed a drive-up jury study comparing the drive signal processing methods AM, CNT loudspeaker unprocessed (CntUP), DC, Pulse amplitude modulation (PAM), SED, and TradUP. This work will be the first published subjective performance data for CNT loudspeakers.

Chapter 5 takes the results from the Chapter 4 jury study and compares it to traditional psychoacoustic metrics. The goal was a “golden metric” that could be used in place of a full jury study to predict CNT subjective performance. It takes a significant amount of effort to conduct a jury study. Therefore, having an idea of which psychoacoustic metric(s) approximate subjective performance can help save development time.

Chapter 6 summarizes the results of the complete effort and attempts to pull out primary themes and connections between the chapters. Additionally, recommended next steps are outlined.

2 Experimental quantification of the true efficiency of carbon nanotube thin-film thermophones

2.1 Abstract

Carbon Nanotube thermophones can create acoustic waves from 1 Hz to 100 kHz. The thermoacoustic effect that allows for this non-vibrating sound source is naturally inefficient. Prior efforts have not explored their true efficiency (i.e. the ratio of the total acoustic power to the electrical input power). All previous works have used the ratio of sound pressure to input electrical power. A method for true power efficiency measurement is shown using a fully anechoic technique. True efficiency data are presented for three different drive signal processing techniques: standard alternating current (AC), direct current added to AC (DCAC), and amplitude modulation of an AC signal (AMAC). These signal processing techniques are needed to limit the frequency doubling non-linear effects inherent to carbon nanotube thermophones. Each type of processing affects the true efficiency differently. Using a 72 W_{rms} input signal, the measured efficiency ranges were 4.3 E-6 – 319 E-6, 1.7 E-6 – 308 E-6, and 1.2 E-6 – 228 E-6 percent for AC, DCAC, and AMAC, respectively. These data were measured in the frequency range of 100 Hz to 10 kHz. In addition, the effects of these processing techniques relative to sound quality are presented in terms of total harmonic distortion.

2.2 Introduction

Carbon nanotube (CNT) thermophones create sound with heat, as opposed to a traditional moving coil loudspeaker, which uses a magnet to push and pull a metal coil of wire attached to a cone. This velocity boundary condition of a traditional speaker's cone creates the pressure wave that propagates to the listener's ear. In contrast, CNT thermophones use a thin-film that can oscillate its surface temperature at acoustic frequencies, creating a varying temperature boundary condition. With every heating cycle the air near the thin-film expands. When the current is removed from the thin-film, it cools, contracting the surrounding air. The repeated expansion and contraction of the adjacent air due to the thermal boundary condition creates the pressure wave that propagates to the listener's ear. This type of thermoacoustic device is called a thermophone.

The thermoacoustic effect was first published in 1898 by Braun, demonstrating how heat can create sound [1]. In the early 1900s, Arnold and Crandall explored this phenomenon using 700nm platinum, which could only heat and cool at frequencies less than 16 Hz, below the human audible range [2]. A material that could heat and cool quickly enough did not exist until 1991, when CNT thin-film was discovered [7]. In 2006, Yu et al. were the first to use the thermoacoustic effect with CNT thin-films and create sound in the audible range [3].

Carbon nanotubes have a very low heat capacity per unit area and have been shown to oscillate their surface temperature at frequencies up to 100 kHz [8]. Without the heavy magnet of a traditional moving coil loudspeaker, CNT thermophones are useful for applications where a lightweight speaker is desired. In addition, rare-earth metals, commonly used to reduce weight of traditional moving coil loudspeakers, are unnecessary. This makes CNT thermophones a good choice for sustainable loudspeakers. Application areas may include automotive, aerospace, and defense systems, where weight is at a premium. CNT thermophones are also flexible and stretchable, which allows them to be placed over complex geometric surfaces.

Several authors have analytically explored CNT thin-film thermophones [4], [9]–[12]. Xiao et al., were the first to develop a theoretical model of the CNT thermophone's true efficiency, given as

$$\eta = \frac{\Pi}{P_{input}} = \frac{\pi f^2 P_{input}}{2\rho_0 c C_P^2 (T_0 + T_a)^2} \qquad \text{EQ2-1}$$

where η is the efficiency, Π is the sound power (watts), P_{input} is the total input power (watts), f is the frequency (Hz), ρ_0 is the density of the surrounding gas (kg/m^3), c is the speed of sound in the surrounding gas (m/s), C_P is the specific heat of the surrounding gas (J/kg K), T_0 is the ambient temperature (K) of the surrounding gas, and T_a is the mean temperature (K) of the thin film [13]. This model assumes the acoustic wavelength is much larger than the size of the source (i.e. it radiates as a monopole).

Prior to this effort, however, there has been minimal work measuring the efficiency of CNT thermophones [13]–[15]. Previous efficiency measurements compared the measured sound pressure level (SPL) at 1 meter to the total electrical input power into the CNT. However, in some experiments, the sound pressure level was not measured at 1 m, but instead measured in the nearfield and estimated at 1 m using spherical spreading. In addition, previous studies have focused on the low input power regime of CNT thermophones, on the order of 1 to 10 W$_{rms}$. True efficiency is defined as the ratio of acoustic output power (watts) to the input electrical power (watts). Experimentally measuring this true efficiency over a range of realistic input power levels is the goal of this study.

CNT thermophones are non-linear transducers. The non-linearity occurs because the output SPL is proportional to the square of the input electrical current. This causes a doubling of frequency between the input and output signals [11], resulting in significant distortion for broadband content (e.g. speech, music, etc.). Signal processing techniques such as DC offset, amplitude modulation, and single-sided pulse width modulation have been shown to significantly reduce this distortion, but these methods require additional input power [4], [16]. These processing techniques are used to modify the drive signal going into the CNT thermophone. This work will show a test method for measuring the true efficiency of thermophones and explore that efficiency using alternating current

(AC), direct current offset with alternating current (DCAC), and amplitude modulation of an alternating current (AMAC).

Because pressure is proportional to power (voltage or current squared), the AC input method produces a doubled output frequency. It is trivial to show this using the power reduction trigonometric identity. For the case of DCAC, this non-linearity results in an output pressure of

$$\mathrm{P(t)} \approx B^2 + 2BAsin(\omega t) + A^2[\frac{1-\cos{(2\omega t)}}{2}] \qquad \text{EQ2-2}$$

where *P* is the pressure (pascals) as a function of time, *t* (seconds), *A* is the peak amplitude of the signal (volts), *B* is the amount of DC offset (volts), and *ω* is the frequency of the signal (rad/s). The doubled frequency is observed in the third term, the fundamental frequency appears in the second term, and the first term contributes to waste DC heating. For AMAC, the input voltage signal is

$$V(t) = \left(1 + A_M cos(2\pi F_M t)\right) * A_C \sin{(2\pi F_C t)} \qquad \text{EQ2-3}$$

which is squared due to the non-linearity of the system. In EQ2-3, *V* is the voltage (volts) as a function of time, *t* (seconds), A_M is the amplitude of the modulated signal (volts), F_M is the frequency of the modulated signal (rad/s), A_C is the amplitude of the carrier signal (volts), and F_C is the frequency of the carrier (rad/s). The resulting components when this input signal is squared are F_M, $2F_M$, $2F_C$, $2F_C$-F_M, $2F_C$+F_M, $2F_C$+$2F_M$, and $2F_C$-$2F_M$. It is interesting to note the presence of the $2F_M$ peak and second side lobes at $2F_C$+$2F_M$ and $2F_C$-$2F_M$, as these are not created in a linear loudspeaker's response to AMAC input.

The relative amplitudes of the modulated and carrier signal can also affect the response. This is typically described with Modulation Index, or the ratio of the modulated to carrier amplitude. Modulation depth is commonly used to describe modulation index as it is the percent representation of modulation index. For example, if a 1 Vpk 1000 Hz signal was modulated by a 2 Vpk 40 kHz carrier signal, the resulting signal would have a 0.5 modulation index or a 50% modulation depth.

Sound quality is also important for loudspeakers and can be a competing parameter with efficiency in thermophone design [4]. This work evaluates total harmonic distortion (THD) of the CNT thermophone as a function of many input parameters, such as frequency, the ratio of signal amplitude to amount of DC offset, the ratio of modulation frequency to carrier frequency, and modulation index. THD is the ratio of the sum total acoustical pressure of the 2-6th harmonics to the pressure of the fundamental, or

$$THD = \frac{Sum(pressure\ of\ 2-6th\ harmonics)}{pressure\ of\ fundamental} \qquad \text{EQ2-4}$$

High THD results in an audio signal that is distorted and unintelligible. Therefore, the lowest possible THD as efficiency allows is desired for a high quality sound.

2.3 Carbon nanotube description

The CNT thermophone used for this work was composed of multi walled nanotubes (MWNT) roughly 100 nm in length, grown on a silicon substrate. The CNT forests were grown by NanoWorld Laboratories at the University of Cincinnati using a chemical vapor deposition (CVD) technique [17], [18]. These CNTs were grown in a forest and dry drawn over two copper rods by researchers at Michigan Tech. The CNT was not wrapped around the copper rods to prevent destructive interference at high frequencies. Structurally, the thermophone had six ribbons of CNT, each overlaid with five layers of thin-film, as shown in Figure 2-1. The total size was 9 cm high by 4.5 cm wide.

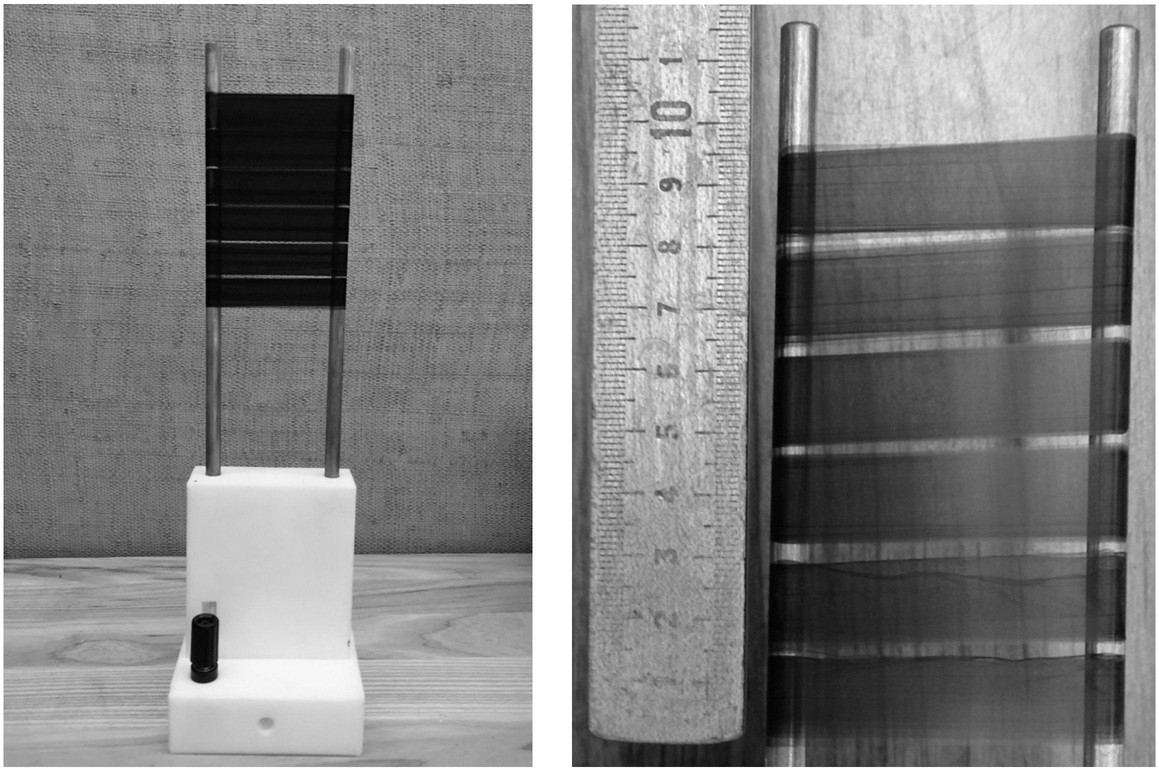


Figure 2-1: Picture of the CNT fixture used in this study (left) and a close up of the multiwalled CNT (right). Six ribbons, each five layers thick, were laid over two 101 copper rods. The CNT was not wrapped around the copper rods to prevent destructive interference at high frequencies.

2.4 Methods

To measure true efficiency, it was necessary to determine the acoustic power output and electrical power input to the CNT thermophone. ANSI S12.54 was used to measure the sound power level (L_W), which was then converted to watts of acoustic power using a reference power of 1 picowatt [19]. The standard measurement was implemented in a fully anechoic chamber. The chamber has dimensions of 2.16 m long x 1.5 m wide x 2.16 m high. This limited the radius of a typical hemisphere to below 1 m, so the CNT thermophone was placed on a rotating table, controlled by a stepper motor, and four microphones were located in a 90° elevation arc at a radius of 1 meter from the CNT thermophone base as shown in Figure 2-2. Rotating the source in this configuration allowed for a 1 m radius measurement hemisphere. Data were acquired six times for each test with a 60 degree azimuth spacing to measure the entire hemisphere around the source. To illustrate the process at a single frequency: a sine wave was played through the CNT thermophone, data were then acquired simultaneously for five seconds (25 averages) at four elevation angles, the CNT thermophone was rotated 60 degrees in azimuth, data were again acquired, and this was repeated for six total azimuth locations. Once all of the locations had been recorded, a single sound power value was calculated. Because the input signal was a stationary sinusoid, the electrical power was found by measuring the time-averaged RMS input voltage and current on the leads to the CNT thermophone. For the AC and DCAC signal processing techniques, PCB 130A23 microphones were used to measure sound pressure. Signal conditioning was provided internally from a National Instruments PXIe-4497 data acquisition (DAQ) module. For AMAC and THD measurements, PCB 378C01 high frequency microphones were used with external signal conditioners providing gain values of 100. All tests were conducted in air with a temperature range of 21-29 °C and ambient pressure of 1014-1031 hPa. Ambient temperature and pressure were monitored throughout all testing to make the appropriate corrections when computing the sound power correction factor, per the standard.

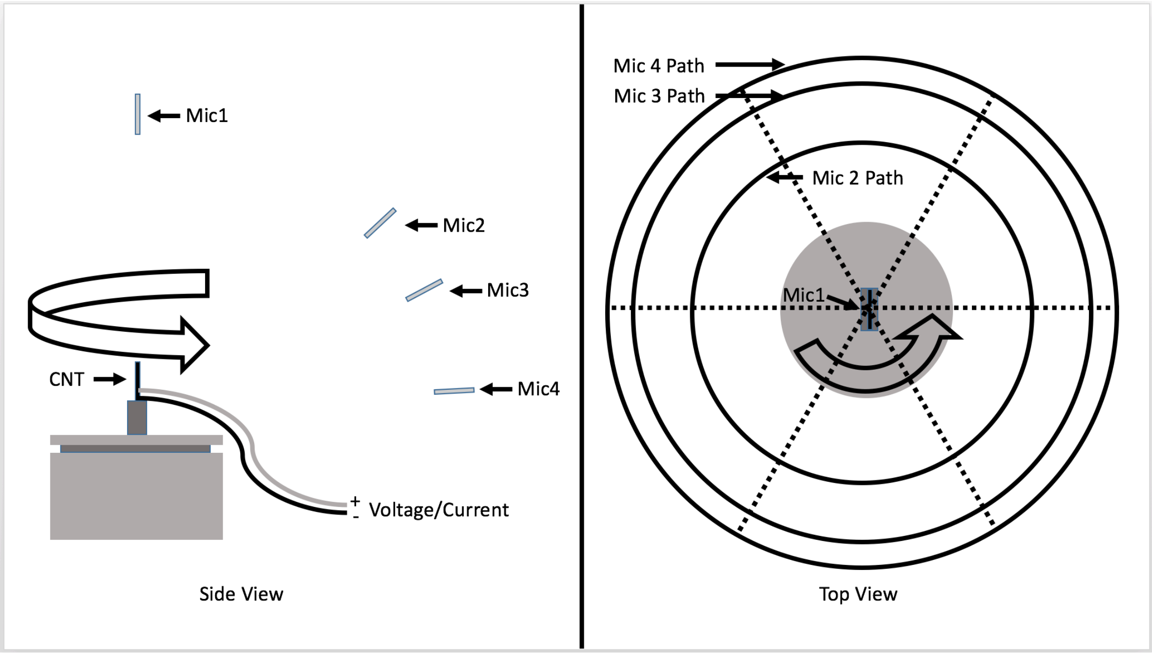


Figure 2-2: Test setup illustrating the implementation of ANSI S12.54 to measure average pressure around the CNT thermophone. Four elevation microphones took data at six azimuth locations (i.e. every 60 degrees-dashed lines) for each test.

Per ANSI S12.54, section 8.1.1b, if the source emits an A-weighted directivity index (DI) exceeding 5 dB in any direction, more microphones should be localized in that area. For example, the A-weighted DI in the elevation angle (i.e. between mic 4 and mic 1) is shown in Table 2-1. To account for this potential source of error, more microphones were localized in the area of high SPL for a single test. Figure 2-3 shows the standard 20 microphone locations for the ANSI S12.54 and the modified test locations. Due to testing time and equipment limitations, the modified test was only completed once and an L_w correction factor for each frequency was computed (Table 2-1). The correction factor was applied to all other data which were acquired with the standard locations shown in Figure 2-3. Because the source geometry and, therefore, its directivity were unchanged throughout the testing, this correction process produced repeatable results, while minimizing testing time.

Table 2-1: Sound pressure level between microphone locations 4 and 1 for a total input power of 72 Wrms and the correction factor applied to all sound power results to correct for the error from the standard microphone locations in ANSI S12.54 while testing a directional source.

Low Frequency Region											
Frequency (Hz)	**100**	**125**	**160**	**200**	**250**	**315**	**400**	**500**	**630**	**800**	
SPL Difference (dBA re 20μPa)	-5.2	-3.6	-1.4	-2.3	-1.2	3.3	-1.1	-5.4	-1.5	2.3	
Lw Correction (dB re 1e-12W	-0.2	-0.2	-0.2	-0.2	-0.2	-0.3	-0.2	-0.1	-0.1	-0.2	
High Frequency Region											
Frequency (Hz)	**1k**	**1.3k**	**1.6k**	**2k**	**2.5k**	**3.2k**	**4k**	**5k**	**6.3k**	**8k**	**10k**
SPL Difference (dBA re 20μPa)	1.3	3.8	5.2	6.6	8.5	17.1	32.4	28.1	28	26.4	29.9
Lw Correction (dB re 1e-12W	-0.1	-0.4	-0.4	0.3	0.3	0.4	0.7	-0.6	-0.6	-2.3	-1.8

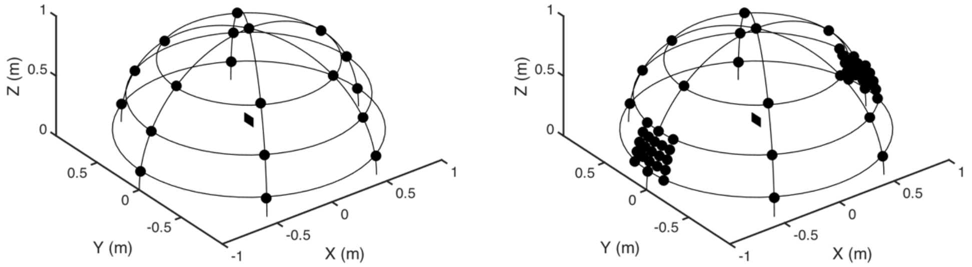


Figure 2-3: An isometric view of the standard 20 microphone locations outlined in ANSI S12.54 Annex B (left) and an isometric view showing the microphone locations used to compute the correction factor (right). The CNT thermophone is represented as a small square in the center of the hemisphere.

To measure the input power, the same PXIe-4497 module was connected to a 111.5x attenuator to acquire voltage and a Fluke 80i-110s clamp-on current probe was used to measure current. Because CNT thermophones are not pure resistors above 10 kHz (Figure 2-4), measuring the crosspower spectrum of these two signals allowed for easy computation of the true power at all frequencies. Figure 2-4 shows an example of the electrical impedance of the CNT thermophone used in this study.

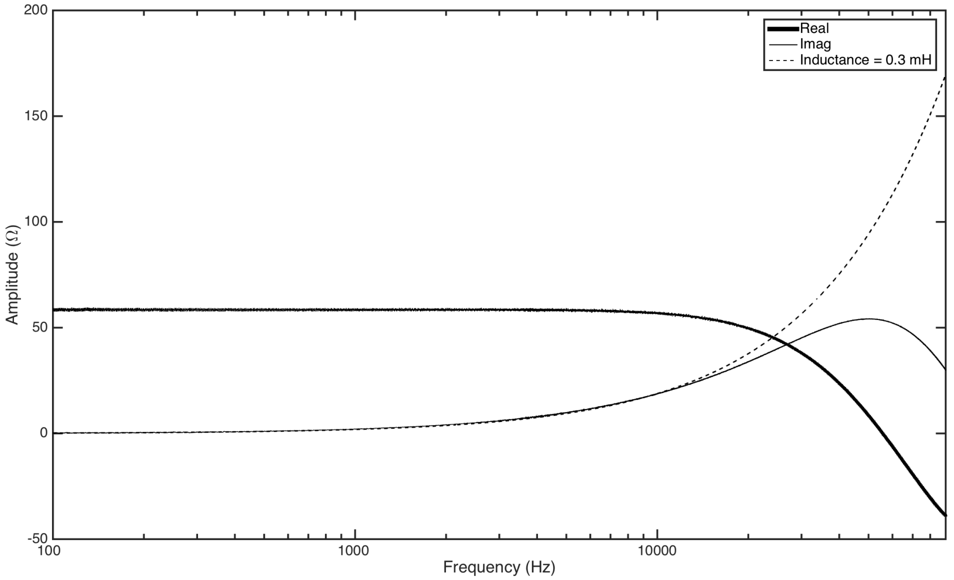


Figure 2-4: Impedance for the CNT thermophone used in this work showing the deviation from pure resistance above 10 kHz. Inductance plays an important role in the 10-20 kHz range, while a more complicated impedance model must exist at frequencies greater than 20 kHz. White noise 10 Hz to 100 kHz was played through the thermophone with total input power of 10 Wrms. 100

averages were taken and the resulting inductance was estimated at 0.3 mH for frequencies less than 20 kHz.

A LabView code was written to run an automated ANSI S12.54 sound power test using a .wav file input. The sound power level output (L_w) and electrical input (watts) were stored. MATLAB was used to process the data. For the AC signal processing technique, data were obtained using pure sine wave inputs at one–third-octave (OTO) band center frequencies ranging from 100 Hz to 20 kHz. Frequency and total input power were varied, because these are the two most important independent variables in Xiao's efficiency equation (EQ2-1) [13]. Since the sound pressure generated from CNT thermophones is proportional to the square of the input voltage signal, the efficiency for this signal processing technique was computed as the acoustic power (watts) in the second harmonic divided by the electrical input power in the fundamental,

$$AC\ Efficiency = \frac{Acoustic\ Power\ at\ The\ Second\ Harmonic}{Electrical\ Input\ Power\ at\ The\ Fundamenntal} * 100 \qquad \text{EQ2-5}$$

For DCAC, data were acquired at the same frequencies, but with varying amplitude ratios of DC current (B) to alternating current (A). These parameters were varied because of their influence in EQ2-2. For the constant amplitude case, the AC amplitude (A) was unchanged and the DC amplitude (B) was varied to obtain different ratios of B/A. For the constant input power case, both B and A were manipulated to obtain different ratios of B/A, all with the same amount of total electrical input power to the CNT thermophone. The efficiency for DCAC was computed using

$$DCAC\ Eff. = \frac{Acoustic\ Power\ Out\ at\ the\ Fundamental}{Sum\ of\ Electrical\ Power\ Into\ The\ Fund.\ and\ the\ DC\ offset} * 100 \qquad \text{EQ2-6}$$

For AMAC, data were acquired at the same frequencies but for varying ratios of the carrier frequency (F_c) to modulated frequency (F_m). The efficiency for AMAC was computed as

$$AMAC\ Efficiency = \frac{Acoustic\ Power\ Out\ at\ the\ Fundamental}{Sum\ Of\ All\ Power\ Into\ The\ CNT\ Thermophone} * 100 \qquad \text{EQ2-7}$$

noting that the denominator is the sum of all frequencies. Additionally, modulation depth was studied by looking at the effects of the ratio of the carrier signal amplitude (A_c) to the modulated signal amplitude (A_m).

THD was not computed for the AC signal processing technique as no acoustic waves are produced at the fundamental. Thus THD is theoretically infinite for this processing technique (i.e. the denominator is approximately zero, to within the noise floor of the data acquisition system, for EQ2-4. THD was calculated for the DCAC and AMAC using the 2-6th harmonics because there is no significant contribution to the total power from the higher harmonics.

2.5 Results and discussion

The results from the low and high input power AC case are shown in Figure 2-5. The true efficiency of a CNT thermophone varies from 4.3 E-6 to 319 E-6 percent between 100 Hz and 10 kHz for 72 W_{rms} total input power. This is theoretically the peak efficiency case for this device at this input power, because all of the acoustical power in the second harmonic (i.e. the doubled frequency) is directly from the electrical power in the fundamental frequency with no signal processing. DCAC requires DC electrical power to shift the signal and AMAC requires high frequency electrical power to produce the carrier frequency. Therefore, both of these processing techniques were expected to decrease the efficiency of the thermophone.

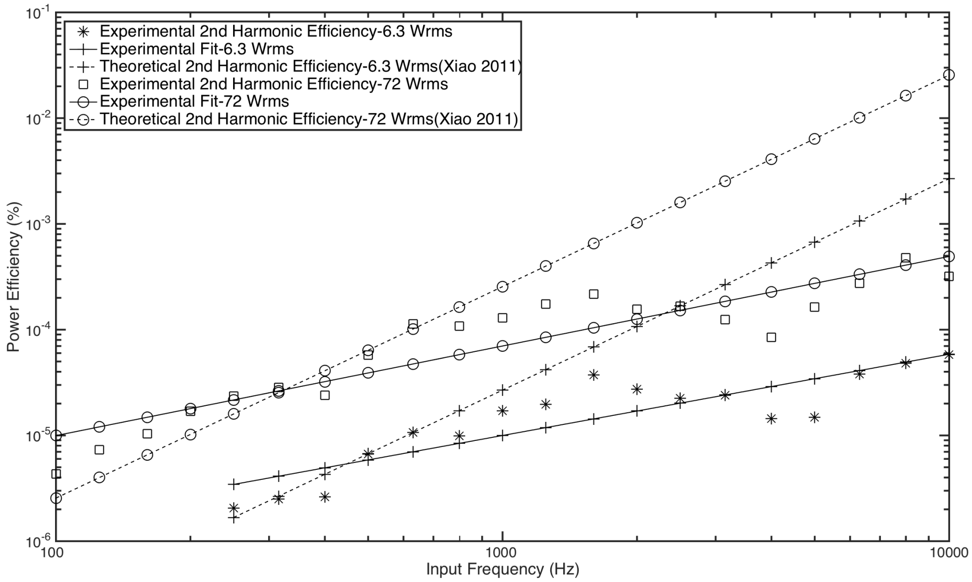


Figure 2-5: AC true efficiency data for total input power of 6.3 Wrms and 72 Wrms. This is the ratio of acoustic power generated in the second harmonic divided by the electrical power in the fundamental (EQ2-5). The resulting fits of the experimental data are shown in EQ2-8 & EQ2-9. The experimental data is consistent with the theoretical model from Xiao for lower frequencies [13]. Note: the lower power 6.3 Wrms data was only taken from 250 to 20,000 Hz.

The fit for the AC case with 6.3 W_{rms} input power (Figure 2-5) is

$$Power\ Efficiency\ (\%) = 50\text{E} - 9 * f^{0.77} \qquad \text{EQ2-8}$$

where f is the frequency in Hz and the R^2 value is 0.76. The fit for the AC case with 72 W_{rms} input power (Figure 2-5) is

$$Power\ Efficiency\ (\%) = 201\text{E} - 9 * f^{0.85} \qquad \text{EQ2-9}$$

where f is the frequency in Hz and the R^2 value is 0.84. The values used to compute the Xiao efficiency, from EQ2-1, are shown in Table 2-2.

Table 2-2: Values used to compute the Xiao efficiency. Convective heat transfer coefficient, β_0, was obtained from Xiao et al. for a stack of 5 thin films as it was not obtained experimentally [13].

ρ_0 (kg/m^3)	c (m/s)	C_P (J/(kg K))	T_0 (K)	T_a (K)	β_0 (W/(m^2 K))	S (m^2)
1.1764	343	1.00643E3	297.15	$\frac{P_{input}}{2\beta_0 S}$	66	0.017

The experimental data agreed well with EQ2-1 while the source radiated in a monopole-like pattern at frequencies below 1,600 Hz. At frequencies higher than 1,600 Hz, the height of the source, 9 cm, is large with respect to a wavelength and the source begins to become directional. When comparing the two power level efficiencies in Figure 2-5 it was observed that increasing power increases efficiency, as expected from EQ2-1.

A standard moving coil loudspeaker was tested as a baseline and the results are shown in Table 2-3. The moving coil loudspeaker was a custom-made PVC pipe speaker with an Axon 6s1 6-1/2" Shielded Midbass, an Audax DTW100TI25 4 Ohm 1" Dome tweeter, and a crossover frequency of approximately 4 kHz [20]. Efficiency for this test was calculated using

$$Standard\ Efficiency = \frac{Acoustic\ Power\ Out\ at\ The\ Fundamental}{Electrical\ Power\ In\ The\ Fundamental} * 100 \qquad \text{EQ2-10}$$

Table 2-3: Efficiency & THD results for a standard moving coil loudspeaker. Efficiency was calculated using EQ2-10. Total input power was 0.6 Wrms. THD was calculated with Eqn. EQ2-4.

Low Frequency Region											
Frequency (Hz)	**100**	**125**	**160**	**200**	**250**	**315**	**400**	**500**	**630**	**800**	
Efficiency (%)	0.41	0.38	0.38	0.27	0.23	0.31	0.32	0.39	0.45	0.67	
THD (%)	1.65	1.37	1.34	1.10	0.98	0.60	0.51	0.89	0.69	0.77	
High Frequency Region											
Frequency (Hz)	**1k**	**1.25k**	**1.6k**	**2k**	**2.5k**	**3.15k**	**4k**	**5k**	**6.3k**	**8k**	**10k**
Efficiency (%)	0.21	0.38	0.22	0.20	0.18	0.10	0.11	0.15	0.14	0.08	0.07
THD (%)	1.02	0.96	0.85	0.46	0.54	0.40	0.59	0.26	0.91	1.81	0.87

The standard speaker had an efficiency ranging from 7 E-2 to 67 E-2 percent. In approximate terms, the CNT thermophone was four orders of magnitude less efficient than the traditional moving coil loudspeaker.

For the second signal processing technique, DCAC, Figure 2-6 and Figure 2-7 show the results for constant amplitude and constant input power, respectively.

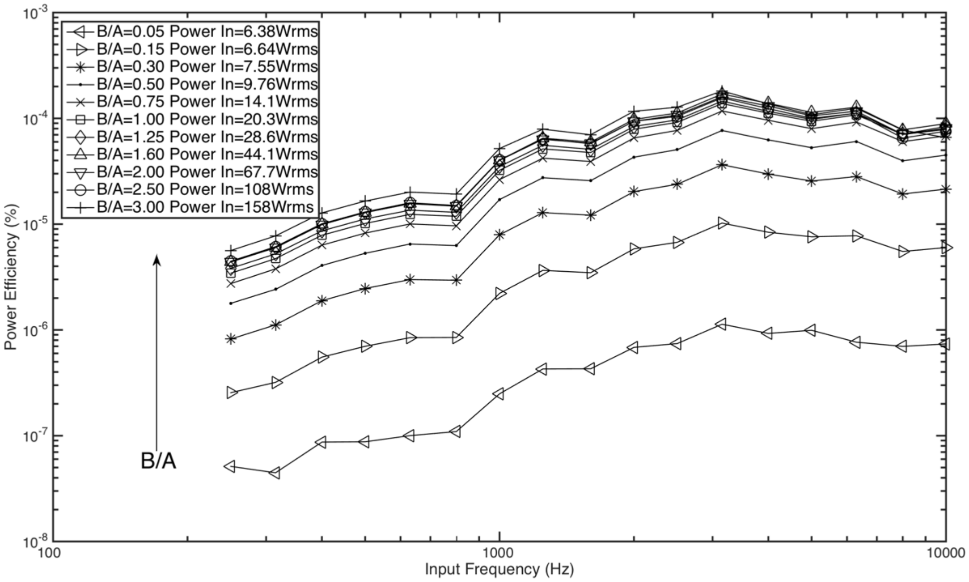


Figure 2-6: DCAC true efficiency data for a constant amplitude. The signal amplitude (A) was held constant while the amount of DC offset (B) was varied. Efficiency was computed with EQ2-6. Efficiency is shown to increase significantly with increased power, as expected.

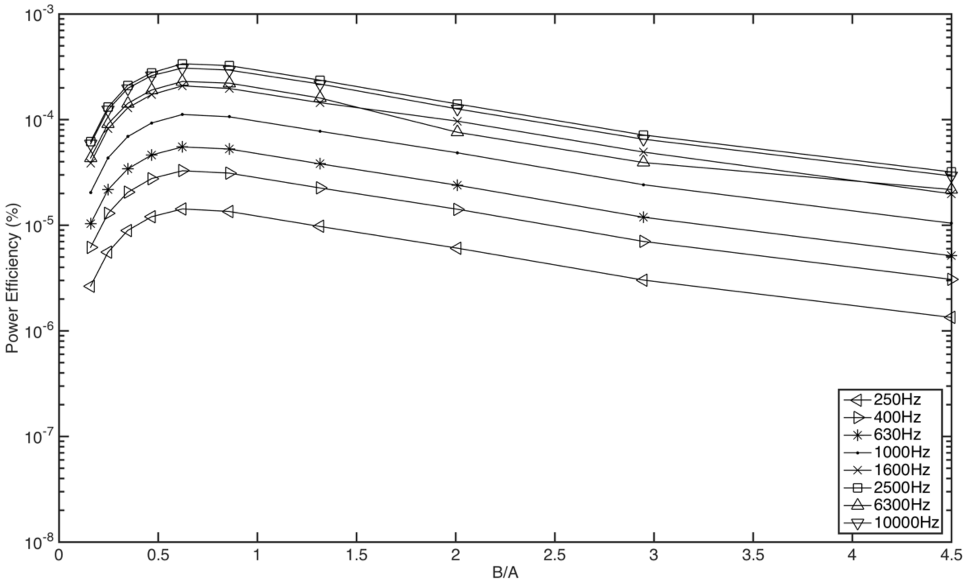


Figure 2-7: DCAC efficiency data for a constant power. The signal amplitude (A) and amount of DC offset (B) were both varied to get different values of B/A while keeping the total power constant at 72 Wrms. Efficiency was computed using EQ2-6. Here an optimal ratio of B/A, in terms of maximum efficiency, is shown at a value of about 0.62.

Figure 2-6 illustrates a diminishing return on increasing the amount of DC offset (*B*). Once the ratio of *B*/*A* reaches 0.75, the increase in efficiency for the added power is marginal. Based on Figure 2-7, for a constant input power, a *B*/*A* ratio of 0.62 is the most optimal ratio for efficiency. The efficiency for this ratio varies from 1.69 E-6 to 308 E-6 percent between 100 Hz and 10 kHz with 72 W_{rms} total input power.

Upon exploring Figure 2-6 & Figure 2-7, a more distinct comparison between the effects of varying *B* vs *A* was desired. To achieve this, a single 1 kHz sine wave was input into the thermophone for two scenarios: holding *A* constant while changing *B* and holding *B* constant while changing *A*. Figure 2-8 demonstrates that increasing *B* for a constant *A* does not increase the efficiency of the CNT thermophone. Instead, increasing *A* for a constant *B* is a more efficient way of increasing the true power efficiency. Ultimately, DCAC in application would be hindered because it requires a class A/B amplifier to satisfy the need for DC offset.

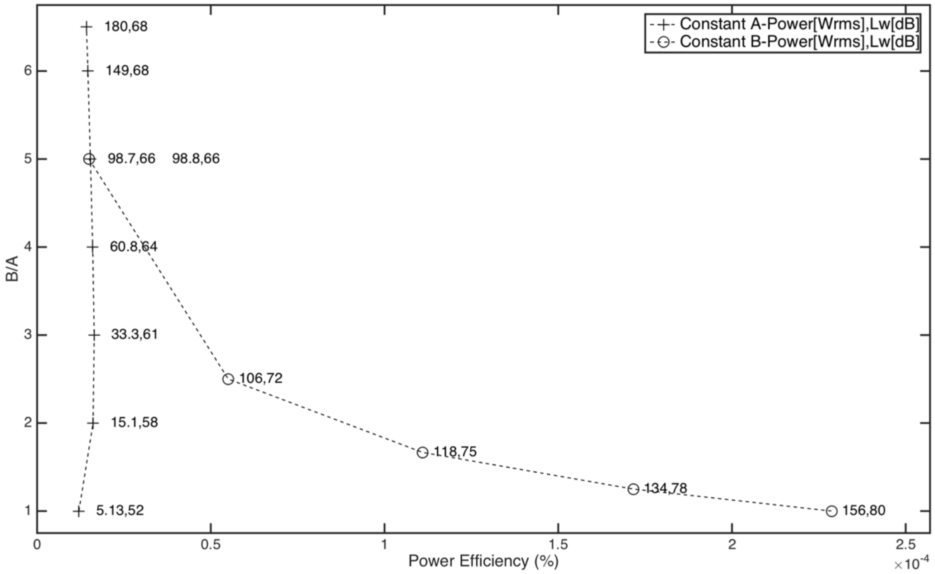


Figure 2-8: Data comparing the efficiency effects of holding the signal amplitude (A) constant while changing the amount of DC offset (B) vs holding B constant and changing A. The first value for each data point is the amount of power into the CNT thermophone and the second value is the sound power output in the 1 kHz band. Efficiency was computed using EQ2-6. All data points were obtained using a 1 kHz sine wave.

For the AMAC technique, Figure 2-9 demonstrates the frequency domain acoustic output of the CNT thermophone with the frequency axis normalized by the modulation frequency. The modulated signal and its second harmonic are shown at values of F/F_m equal to 1 and 2, respectively. The carrier frequency in this example is 15 times higher than the modulation frequency. The carrier frequency is doubled and is seen at a normalized frequency of 30 with four dominant side lobes. The fundamental at $F/F_m = 15$ and fourth harmonic at $F/F_m = 60$ are not predicted by theory, but are assumed to be artifacts of imperfect signal recreation.

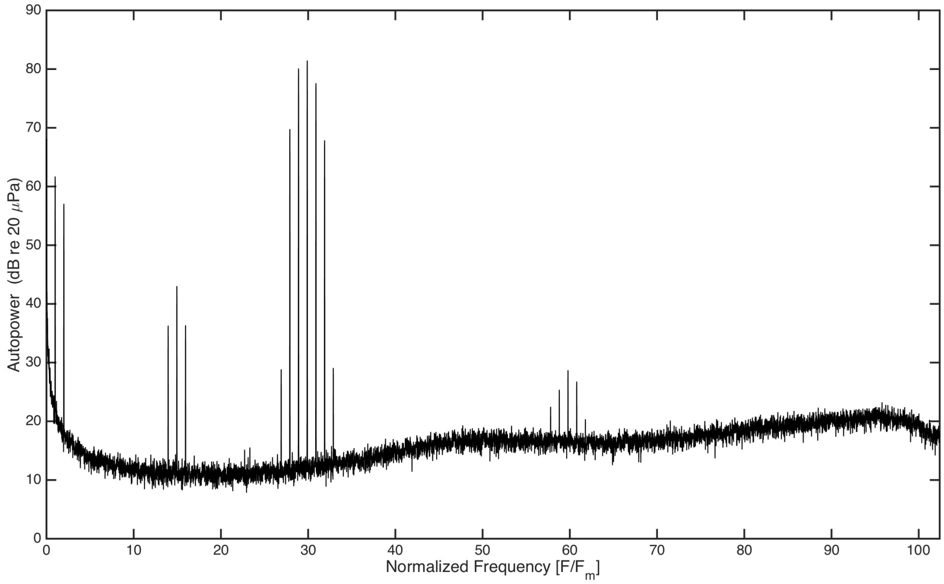


Figure 2-9: An example of the acoustic response of a CNT thermophone normalized to the modulation frequency. In this example, the carrier frequency is 15 times larger than the modulation frequency.

Figure 2-10 shows the AMAC efficiency of a CNT thermophone varies from 1.24 E-6 to 228 E-6 percent with 72 W_{rms} input power. It was found that varying the carrier frequency had no effect on the efficiency. Practically, amplitude modulation is difficult to use, because it requires an amplifier with high enough frequency output to power the carrier frequency. The human hearing range extends to 20 kHz, meaning the AMAC carrier frequency should be greater than 20 kHz to be out of the range of hearing. Many common class D amplifiers limit their output frequency to 20 kHz, which means AMAC's utility is limited in the current market.

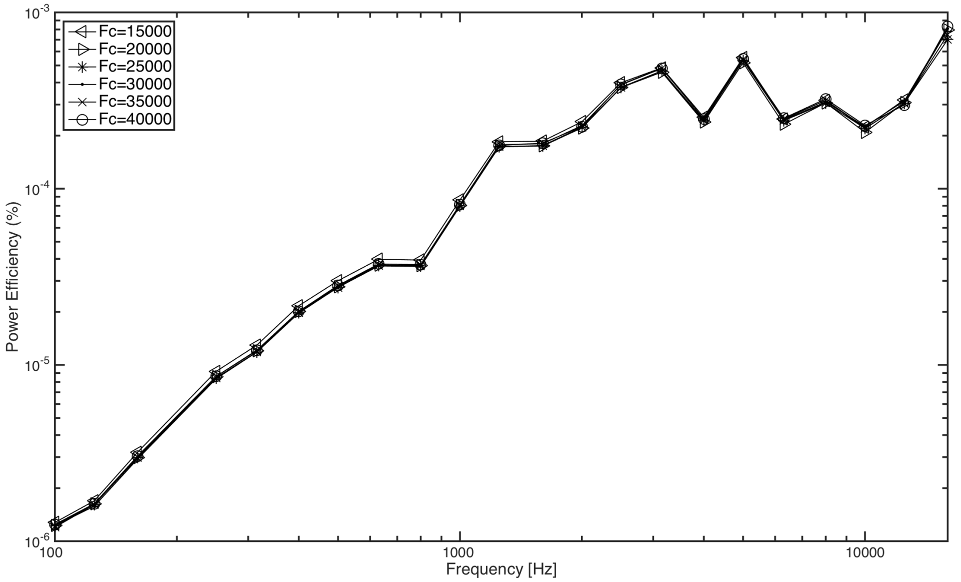


Figure 2-10: AMAC efficiency data. A modulated signal (Fm) was varied with carrier frequency (Fc). The modulation index for all tests was 1 and had a total input power of 72 Wrms. Efficiency was calculated with EQ2-7 and was not affected by varying the carrier frequency (Fc).

Figure 2-11 illustrates the effects of modulation depth. The optimal efficiency is found at an amplitude modulation ratio of 1.5; however, THD effects also need to be taken into account.

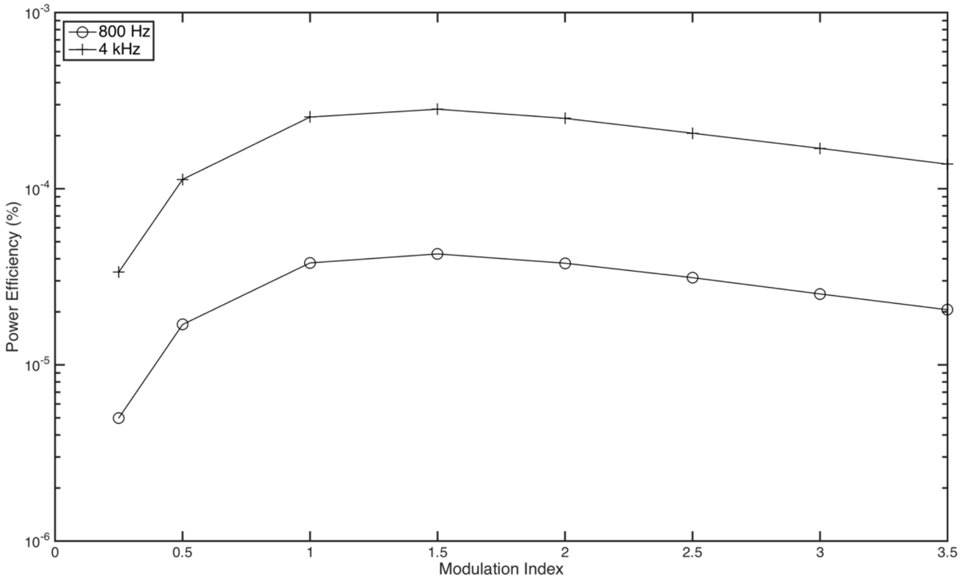


Figure 2-11: Experimental data illustrating the effects of varying modulation index. 72 Wrms total input power was used and efficiency was calculated with EQ2-7.

Figures 12 & 13 compare the THD for the DCAC method. They demonstrate that increasing *B*/*A* decreased THD, but there was a diminishing return; the more *B*/*A* increased the less reduction in THD was observed. Since THD does not have a threshold level where content becomes intelligible, the value of *B*/*A* required for an acceptable level of THD will be subjective. Based on optimal efficiency and EQ2-4, a *B*/*A* level of 0.62 produces THD in the 43-93% range. A *B*/*A* ratio of 0.62 created subjectively intelligible content for the author, but the THD was roughly 65 times higher than a standard moving coil loudspeaker (Table 2-3). It should be noted that intelligibility and high fidelity are not the same thing.

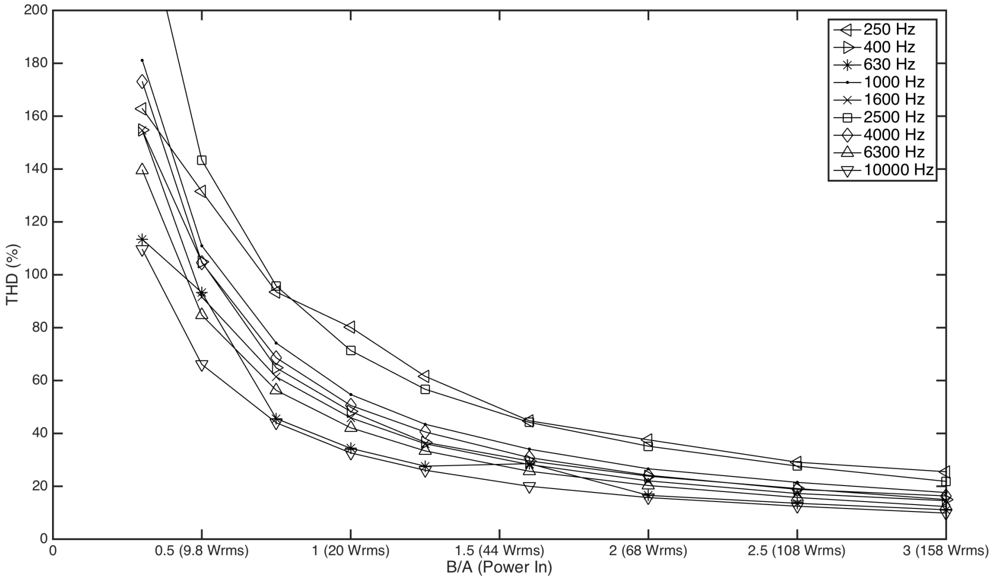


Figure 2-12: Data comparing THD for different frequencies and ratios of B/A for different input power levels. A was held constant and B was increased. THD was computed with EQ2-4.

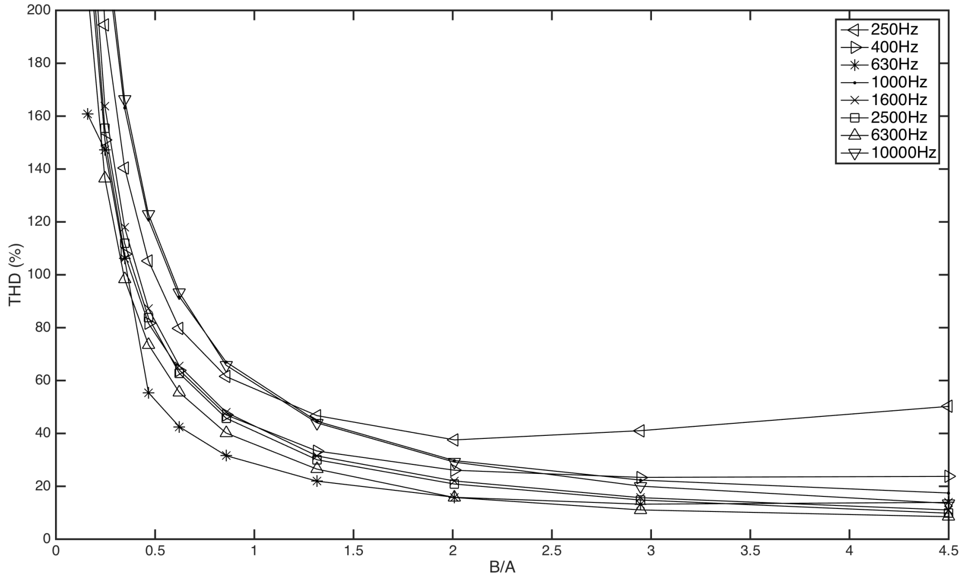


Figure 2-13: Data comparing THD for varying frequencies and ratios of B/A. In this case B and A were manipulated to get a constant power of 72 Wrms input to the CNT thermophone for each case. THD was computed with EQ2-4.

Figure 2-14 demonstrates that THD for AMAC varies from 22-95%. For certain higher frequencies where the carrier was a harmonic of the modulated frequency, THD was significantly higher, but this should not cause any practical issues as long as the carrier is above 20 kHz. From a modulation index perspective, THD increased rapidly as modulation index was increased (Figure 2-15). Therefore, while the optimal modulation index for efficiency is 1.5, the THD increased significantly from 1 to 1.5. A modulation index of 1.0 is the best compromise between efficiency and THD.

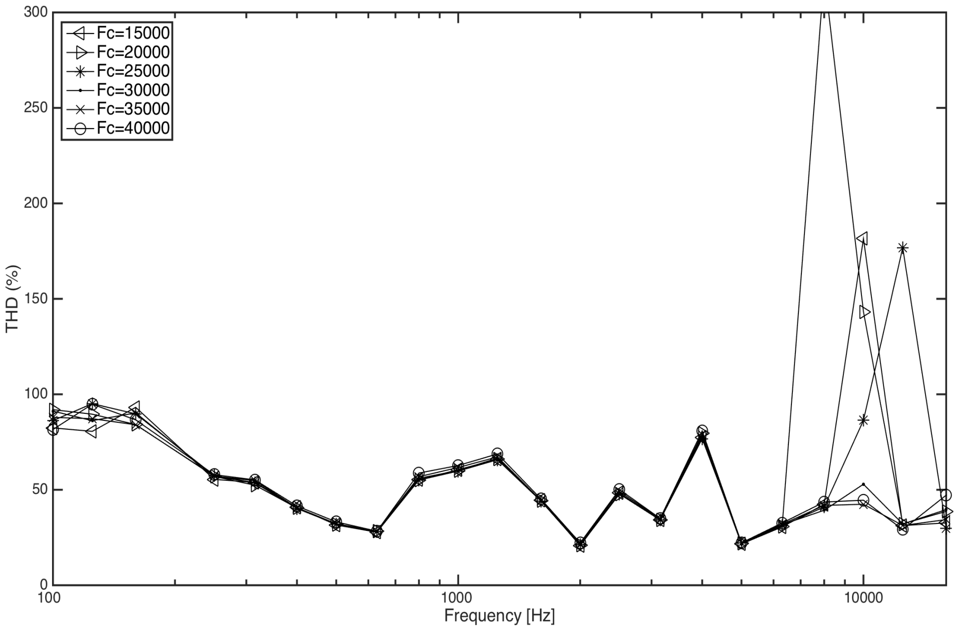


Figure 2-14: THD data for AMAC. The lack of correlation in the high frequency region is a result of the carrier frequency being at a harmonic of the fundamental. Therefore, the THD was artificially increased by the carrier. THD was computed with EQ2-4.

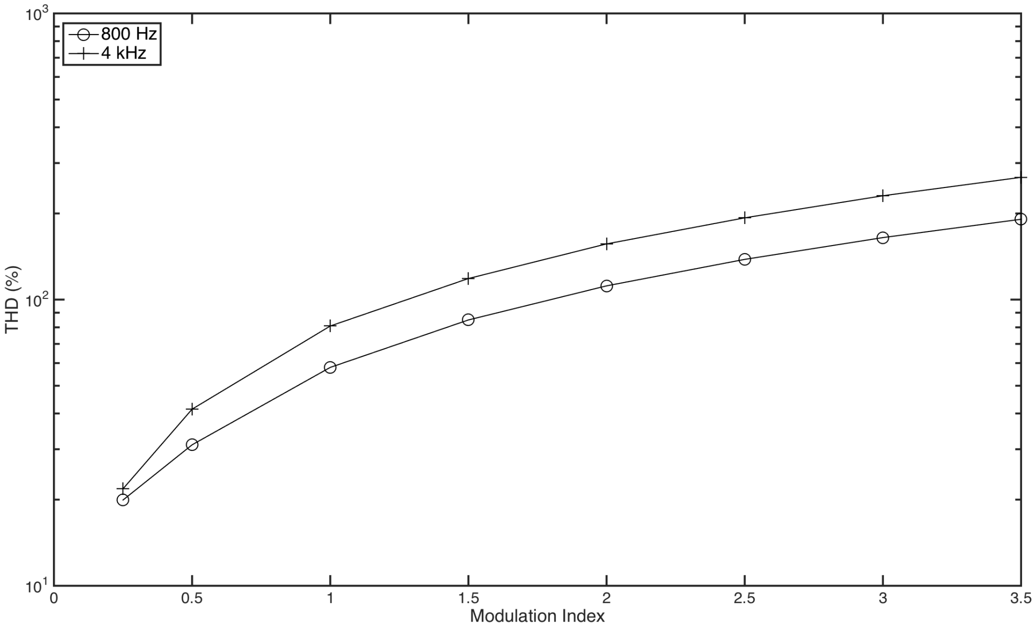


Figure 2-15: Data showing the effects on THD for varying modulation index. THD was computed with EQ2-4.

A summary comparison of AC, DCAC, and AMAC is shown in Table 2-4. As expected, the AC case is the most efficient, but DCAC & AMAC efficiencies are on the same order of magnitude. In terms of THD, AMAC created slightly lower THD, but was the least efficient.

Table 2-4: Summary of experimental data for AC, DCAC, and AMAC signal processing techniques. The total input power for all tests was 72 Wrms with frequency ranges of 100 Hz to 10 kHz. Note that the efficiency for the AC case is the second harmonic efficiency.

	Efficiency (μ%)	THD (%)
AC	4.3 - 319	$\approx \infty$
DCAC (B/A=0.62)	1.69 - 308	43 - 93
AMAC	1.24 - 228	22 - 95

2.6 Conclusions

The fundamental true efficiency of an AC signal is approximately zero due to the non-linearity of CNT thermophones. The second harmonic efficiency of a CNT thermophone is 4.3 E-6 to 319 E-6 percent for 72 W_{rms} input. Experimentally, the efficiency is directly proportional to the input power, which supports the theoretical model created by Xiao et al. Additionally, the Xiao et al. model matched experimental efficiency data for frequencies below 1,600 Hz, where the sound source radiates as a monopole. For DCAC, the optimal efficiency ratio of DC offset to signal amplitude was found to be 0.62. The fundamental true efficiency with that ratio is 1.69 E-6 to 308 E-6 percent for 72 W_{rms} input. This ratio had a THD varying from 43-93%. In terms of AMAC, the fundamental true efficiency is 1.24 E-6 to 228 E-6 percent. It was found that varying the carrier frequency had no effect on efficiency. Additionally, the optimal modulation index in terms of efficiency is 1.5, but when considering THD an index of 1.0 gives the best efficiency for the least amount of THD of 22-95%. Therefore, AMAC has better THD than DCAC with slightly lower efficiencies. Ultimately, DCAC and AMAC are less efficient than a method that would not require additional input power, but the overall efficiency loss is small, so these methods may prove to be sufficient. Their main limitation is the requirement of special amplifiers. DCAC required a class A/B amplifier that can apply a DC offset, and AMAC requires an amplifier that can output frequencies as high as the sum of the carrier and modulated frequencies. The development of a signal processing method that does not require any special equipment, and does not reduce power efficiency would allow for easier loudspeaker market acceptance of CNT thermophones.

2.7 Acknowledgements

The authors would like to acknowledge the contributions of many who made this work possible, including Dr. Vesselin Shanov and his employees at NanoWorld Laboratories at the University of Cincinnati for growing the CNT forests. The CNT thermophones were produced at Michigan Tech with the help of Stephania Vaglica. The fixtures for the thermophone were made at Michigan Tech’s Mechanical Engineering-Engineering Mechanics Department by Steve Lehmann, Bill Langdon, Marty Toth, and Michael Goldsworthy. Finally, we would like to acknowledge PCB Piezotronics® for their donation of the measurement microphones used in this effort.

3 Continued drive signal development for the carbon nanotube thermoacoustic loudspeaker using techniques derived from the hearing aid industry

3.1 Abstract

Compared to moving coil loudspeakers, carbon nanotube (CNT) loudspeakers are extremely lightweight and are capable of creating sound over a broad frequency range (1 Hz to 100 kHz). The thermoacoustic effect that allows for this non-vibrating sound source is naturally inefficient and nonlinear. Signal processing techniques are one option that may help counteract these concerns. Previous studies have evaluated a hybrid efficiency metric, the ratio of the sound pressure level at a single point to the input electrical power. True efficiency is the ratio of output acoustic power to the input electrical power. True efficiency data are presented for two new drive signal processing techniques borrowed from the hearing aid industry. Spectral envelope decimation of an AC Signal operates in the frequency domain (FCAC) and dynamic linear frequency compression of an AC signal operates in the time domain (TCAC). Each type of processing affects the true efficiency differently. Using a 72 Wrms input signal, the measured efficiencies in the frequency range from 100 Hz to 10 kHz were 1.01 – 1083 E-6 and 1.26 – 388 E-6 percent for FCAC and TCAC, respectively. In addition, the effects of these processing techniques relative to sound quality were evaluated in terms of total harmonic distortion (THD). It was shown that although the different signal processing techniques affected the true efficiency, none of them increased the efficiency of the CNT loudspeaker to the level of current moving coil loudspeakers. Additionally, THD as the only sound quality metric is incomplete because these processing methods can be optimized for pure tones but highly distort complex signals like speech and music. Therefore, a sound quality metric for complex signals is needed. Overall, CNT loudspeakers show promise for specific applications where weight savings and complex geometries are required.

3.2 Introduction

Carbon nanotube loudspeakers create sound with heat, not vibration. Their extremely low heat capacity per unit area allows them to heat and cool up to 100,000 times per second [8]. Therefore, they have frequency response from 1-100 kHz. This phenomenon is synonymous with how lightning creates thunder. The energy in the lightning bolt heats the adjacent air, causing expansion, and therefore a pressure wave propagation. The main advantages of using CNT to create sound is that it is extremely light weight, flexible, and slightly transparent. These benefits have peaked interest for their use in automotive, aerospace, and defense applications.

Braun, Arnold, and Crandall in the late 1800s to early 1900s documented that heating and cooling a material rapidly creates sound [1], [2]. This phenomenon is known as the thermoacoustic effect. Once carbon nanotubes where discovered in the early 1990s [7], sound generation followed in 2006 [3], but researchers have observed that the output frequency of the loudspeaker is twice the input [11]. For example, if a 1 kHz sine wave is input into the CNT loudspeaker a 2 kHz pressure wave is output.

CNT loudspeakers are non-linear transducers. The non-linearity occurs because the output SPL is proportional to the input power (i.e. voltage squared) not voltage like traditional loudspeakers. This causes a doubling of frequency between the input and output signals [11], resulting in significant distortion for broadband content (e.g. speech, music, etc.). Signal processing techniques such as DC offset, amplitude modulation, and single-sided pulse width modulation have been shown to significantly reduce this distortion, but these methods require additional input power [4], [14]. These processing techniques are used to modify the drive signal going into the CNT loudspeaker.

The result of pressure being proportional to power is that a standard AC input signal produces a doubled output frequency. It is trivial to show this using the power reduction trigonometric identity. For the example case of using a DC offset, the input signal is

$$\mathrm{V(t)} = B + Asin(\omega t) \qquad \text{EQ3-1}$$

where V is the input voltage (volts) as a function of time, t (seconds), A is the peak amplitude of the signal (volts), B is the amount of DC offset (volts), and ω is the frequency of the signal (rad/s). This non-linearity results in an output pressure proportional to power (i.e. V^2). Squaring EQ3-1 gives

$$\mathrm{P(t)} \approx B^2 + 2BAsin(\omega t) + A^2[\frac{1-\cos{(2\omega t)}}{2}] \qquad \text{EQ3-2}$$

where P is the pressure (pascals) as a function of time, t (seconds), A is the peak amplitude of the signal (volts), B is the amount of DC offset (volts), and ω is the frequency of the signal (rad/s). The doubled frequency is observed in the third term, the fundamental frequency appears in the second term, and the first term contributes to waste DC heating. From this one could conclude that having a high amount of DC offset, B, will solve the problem as the second term would become dominate with respect to the third. Unfortunately, using DC offset or amplitude modulation requires additional power as well as more expensive class A/B amplifiers to be able to create a DC offset or frequency response above 20 kHz. This additional power requirement reduces the efficiency.

Several authors have analytically explored CNT thin-film loudspeakers [4], [9]–[12], [21], [22]. Xiao et al., were the first to develop a theoretical model of the CNT loudspeaker's true efficiency, given as

$$\eta = \frac{\Pi}{P_{input}} = \frac{\pi f^2 P_{input}}{2\rho_0 c C_P^2 (T_0+T_a)^2} \qquad \text{EQ3-3}$$

where η is the efficiency, Π is the sound power (watts), P_{input} is the total input power (watts), f is the frequency (Hz), ρ_0 is the density of the surrounding gas (kg/m^3), c is the speed of sound in the surrounding gas (m/s), C_P is the specific heat of the surrounding gas (J/kg K), T_0 is the ambient temperature (K) of the surrounding gas, and T_a is the mean temperature (K) of the thin film [13]. This model assumes the acoustic wavelength is much larger than the size of the source, i.e. it radiates as a monopole.

Based on strong correlation between Xiao's model and work by Bouman et al. [5], CNT loudspeakers have a true efficiency on the order of 10^{-6} percent using no drive signal processing, DC offset, and amplitude modulation. For comparison, a modern moving-coil driver is on the order of 10^{-2} percent efficient [5]. This is a significant difference, but the main conclusion from Bouman et al.'s work is that the drive signal alone cannot greatly increase the efficiency as the efficiency without any signal processing is still on the order of 10^{-6}. While the drive signal cannot increase the efficiency, it does play a large role with respect to the sound quality of the loudspeaker and the required amplifier [4], [10]. For example, a drive signal method that does not require a DC offset or frequency response above 20 kHz would allow the CNT loudspeaker to be used with a class D amplifier making it much easier and less expensive to enter into a wider market of applications.

Drawing from the hearing aid industry, different possible solutions for solving the frequency doubling issue without using additional power were explored. Specifically, dynamic linear frequency compression [23], a time domain method, and spectral envelope decimation [24], a frequency domain method, allow the frequency content of a signal to be lowered by an octave. Dynamic linear frequency compression by AVR Sonovation was the first commercial hearing aid with frequency lowering in 1991. It works by sampling a signal at the input by a factor of 2 times the sampling rate of the output and then discarding the additional samples over short windows (e.g., ~10 ms). Spectral envelope decimation was first used by Alexander in 2013. It takes a Fourier transform with 75% overlap, decimates the amplitude values by a factor of 2 with respect to frequency while not modifying the phase of each spectral component, and then inverse Fourier transforms to reconstruct the time domain signal.

This work will follow the method set by Bouman et al. to measure the true efficiency of CNT loudspeakers [5] to explore the efficiency using spectral envelope decimation (FCAC) and dynamic linear frequency compression (TCAC). Additionally, sound quality with total harmonic distortion (THD) of the CNT loudspeaker will be studied in this paper. THD is defined as the ratio of the sum total acoustical pressure of the 2nd-6th harmonics to the pressure of the fundamental, or

$$THD = \frac{Sum(pressure\ of\ 2 - 6th\ harmonics)}{pressure\ of\ fundamental} \qquad \text{EQ3-4}$$

High THD results in an audio signal that is distorted. Therefore, the lowest possible THD as efficiency allows is desired for a high quality sound.

While this work specifically explores CNT thin films for use as thermoacoustic loudspeakers, its application can be applied to any loudspeaker using the thermoacoustic effects as the pressure will always be proportional to power. Therefore, the recent work by Aliev et al. and Dashchewski et al. on a variety of all thermoacoustic loudspeaker materials can still use these methods [25], [26].

The automotive industry could see great benefit from this technology. These loudspeakers are ultra-lightweight, can conform to any geometry, have no moving parts, and do not depend on rare earth magnets. These features may allow CNT speakers to replace traditional moving coil speakers while providing significant weight savings. More importantly, they enable the placement of speakers in locations not previously possible, such as on windows or in the headliner. These transducers could also be used in cabin active noise control because they can be placed in optimal locations due to their small size and weight. They also present opportunities for active noise control in exhaust systems, due to their resilience in high temperature environments. Additionally, the heat generated from these loudspeakers could be recycled for other purposes, such as windshield deicing. With increased investment in research and development, thermoacoustic loudspeakers show significant promise for the automotive industry.

3.3 Methodology

The CNT loudspeakers used for this work were composed of multi-walled nanotubes (MWNT) roughly 100 nm in length, grown on a silicon substrate. The CNT forests were grown by NanoWorld Laboratories at the University of Cincinnati using a chemical vapor deposition (CVD) technique. These CNTs were grown in a forest and dry drawn over two copper rods by researchers at Michigan Technological University. The CNT was not wrapped around the copper rods to prevent the formation of two sources, one on each side of the copper rod, creating cancelling pressure waves at high frequency. In order to ensure a good electrical connection, the CNT was densified onto the copper rods using denatured alcohol. Figure 3-1 shows an example CNT loudspeaker. Structurally, each loudspeaker had six ribbons of CNT, each overlaid with five layers of thin-film. The total size was 9 cm high by 4.5 cm wide.

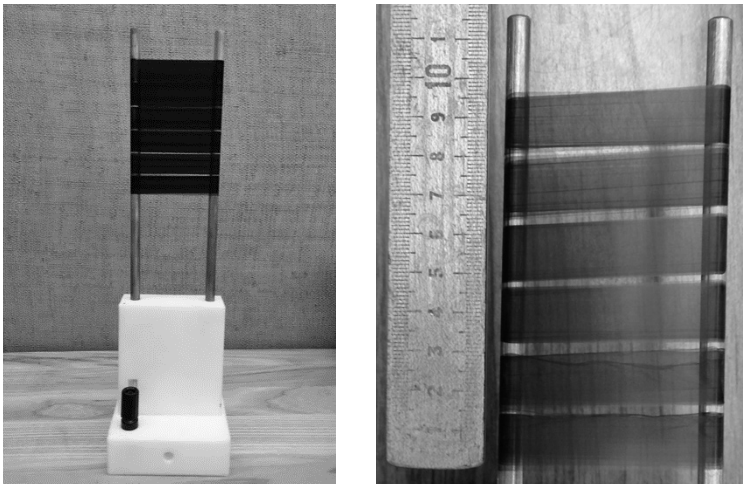


Figure 3-1: Picture of the CNT fixture (left) and a close up of the multi-walled CNT (right). Six ribbons, each five layers thick, were laid over two 101 copper rods. The CNT was not wrapped around the copper rods to prevent the formation of two sources, one on each side of the copper rod, creating cancelling pressure waves at high frequency. [5]

To measure the true efficiency, it was necessary to determine the acoustic power output and electrical power input to the CNT loudspeaker. Following Bouman et al.'s method [5], ANSI S12.54 was used to measure the sound power level, which was then converted to watts of acoustic power using a reference power of 1 picowatt [19]. Per ANSI S12.54 sound power is calculated as

$$L_w = \bar{L}_P - 10\log_{10}\frac{1}{2\pi r^2} - 10\log_{10}(\frac{\rho_0 c}{400}) \qquad \text{EQ 3-5}$$

Where L_w is the sound power (dB re 20 pW), $\bar{L}_P$is the average sound pressure from all measurement locations (dB re 20 µPa), r is the radius of the hemisphere (m), ρ_0 is the density of area (kg/m^3), and c is the speed of sound in air (m/s).

The standard measurement was implemented in a fully anechoic chamber. The chamber has dimensions of 2.16 m long x 1.5 m wide x 2.16 m high. This limited the radius of a typical hemisphere to below 1 m, so the CNT loudspeaker was placed on a rotating table, controlled by a stepper motor, and four microphones were located in a 90° elevation arc at a radius of 1 meter from the CNT loudspeaker base as shown in Figure 3-2. Rotating the source in this configuration allowed for a 1 m radius measurement hemisphere. Data were acquired six times for each test with a 60 degree azimuth spacing to measure the entire hemisphere around the source. To illustrate the process at a single frequency: a sine wave was played through the CNT loudspeaker, data were then acquired simultaneously for five seconds (25 averages) at four elevation angles, the CNT loudspeaker was then rotated 60 degrees in azimuth, data were again acquired, and this was repeated for six total azimuth locations. Once all of the locations had been recorded, a single sound power value was calculated. Because the input signal was a stationary sinusoid, the electrical power was computed by measuring the time-averaged root-mean-square input voltage and current on the leads to the CNT loudspeaker.

PCB 130A23 microphones were used to measure sound pressure. Signal conditioning was provided internally from a National Instruments PXIe-4497 data acquisition (DAQ) module. All tests were conducted in air with a temperature range of 21-29 °C and ambient pressure of 1014-1031 hPa. Ambient temperature and pressure were monitored throughout all testing to make the appropriate corrections when computing the sound power correction factor, per the standard.

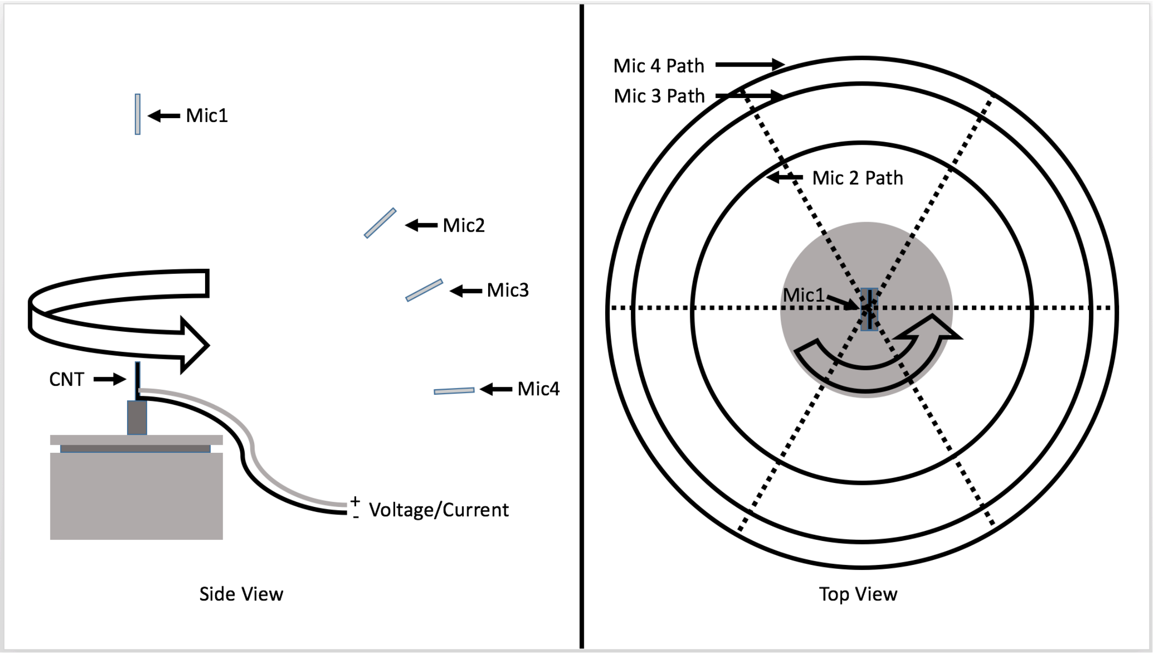


Figure 3-2: Test setup illustrating the implementation of ANSI S12.54 to measure average pressure around the CNT loudspeaker. Four elevation microphones took data at six azimuth locations (i.e. every 60 degrees-dashed lines) for each test. [5]

Per ANSI S12.54, section 8.1.1b, if the source emits an A-weighted directivity index (DI) exceeding 5 dB in any direction, more microphones should be localized in that area. To account for this potential source of error, more microphones were localized in the area of high SPL for a single test. Figure 3-3 shows the standard 20 microphone locations for the ANSI S12.54 and the modified test locations. Due to testing time and equipment limitations, the modified test was only completed once and a sound power correction factor for each frequency was computed (Table 2-1). The correction factor was applied to all other data that were acquired with the standard locations shown in Figure 3-3. Because the source geometry and, therefore, its directivity were unchanged throughout the testing, this correction process produced repeatable results, while minimizing testing time.

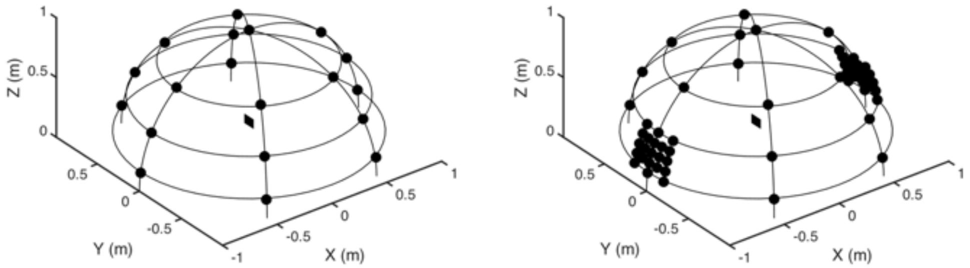


Figure 3-3: An isometric view of the standard 20 microphone locations outlined in ANSI S12.54 Annex B (left) and an isometric view showing the microphone locations used to compute the correction factor (right). The CNT loudspeaker is represented as a small square in the center of the hemisphere. [5]

To measure the input power, the same PXIe-4497 module was connected to a 111.5x attenuator to acquire voltage and a Fluke 80i-110s clamp-on current probe was used to measure current. Because CNT loudspeakers are not pure resistors above 10 kHz [5] measuring the crosspower spectrum of these two signals allowed for easy computation of the true power (taking phase difference into account) at all frequencies. A LabVIEW code was written to run an automated ANSI S12.54 sound power test using a wav file input. The sound power level output and electrical input (watts) were stored. MATLAB was used to process the data.

For FCAC and TCAC, data were obtained using pure sine wave inputs at one–third-octave (OTO) band center frequencies ranging from 100 Hz to 20 kHz. Efficiency was calculated using

$$FCAC\ \&\ TCAC\ Eff. = \frac{Acoustic\ Power\ Out\ at\ the\ Fundamental}{Electrical\ Input\ Power\ at\ Half\ the\ Fund.} * 100 \tag{EQ3-6}$$

The acoustic power is created at the fundamental, but the input electrical power is an octave below the fundamental. Therefore, the efficiency is the ratio of the fundamental acoustic response to the electrical input at half of the fundamental.

THD was calculated for FCAC and TCAC using the 2nd-6th harmonics because there is no significant contribution to the total power from the higher harmonics.

3.4 Results

Figure 3-4 shows the efficiency of FCAC and TCAC compared to the second harmonic AC efficiency. This shows that the FCAC and TCAC processing methods produced an efficiency of 1.01 E-6 to 1083 E-6 percent and 1.26 E-6 to 388 E-6 percent with 72 W_{rms} input power, respectively. The FCAC appears to be artificially high for frequencies above 1 kHz. The maximum efficiency should be the second harmonic AC efficiency because all of electrical energy goes into the second harmonic. For FCAC, there is some energy dispersed during the decimation process and therefore it is expected that its efficiency would be slightly less than the AC second harmonic efficiency. Regardless, the FCAC and TCAC methods are not orders of magnitude more efficient than the other signal processing techniques. Their main benefit is that with these pre-processing techniques a standard off-the-shelf amplifier can be used to power CNT loudspeakers.

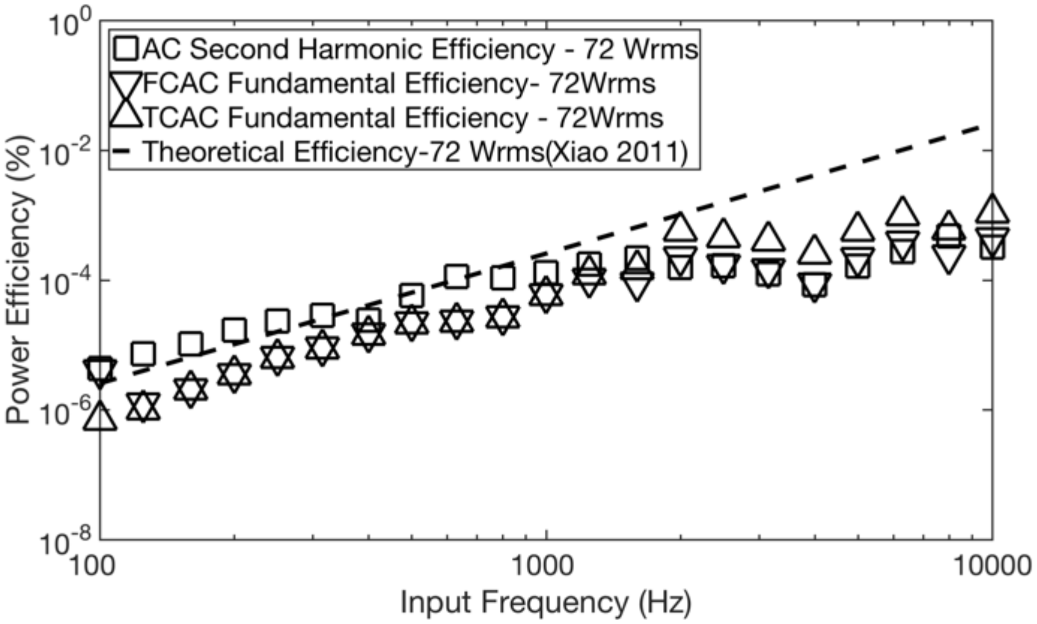


Figure 3-4: Experimental data comparing second harmonic AC efficiency to fundamental FCAC and TCAC efficiency. 72 wrms total input power was used and efficiency for AC was taken from Bouman et al. [5] while the efficiency for FCAC and TCAC was calculated using EQ3-6. [27]

Figure 3-5 demonstrates that the THD for FCAC and TCAC vary from 0.68-59% and 1.7-11%, respectively. This is better than the DCAC and AMAC processing techniques [5], but it should be noted that these are for single frequencies. When the FCAC and TCAC algorithms are optimized for single frequencies they can create perfect half frequency content. When processing complex signals these methods are limited. For example, subjectively using speech and music the FCAC and TCAC produced subjectively low quality reproduction. Based on that observation, THD is not the best sound quality metric and a more robust metric is needed that can be used with complex signals.

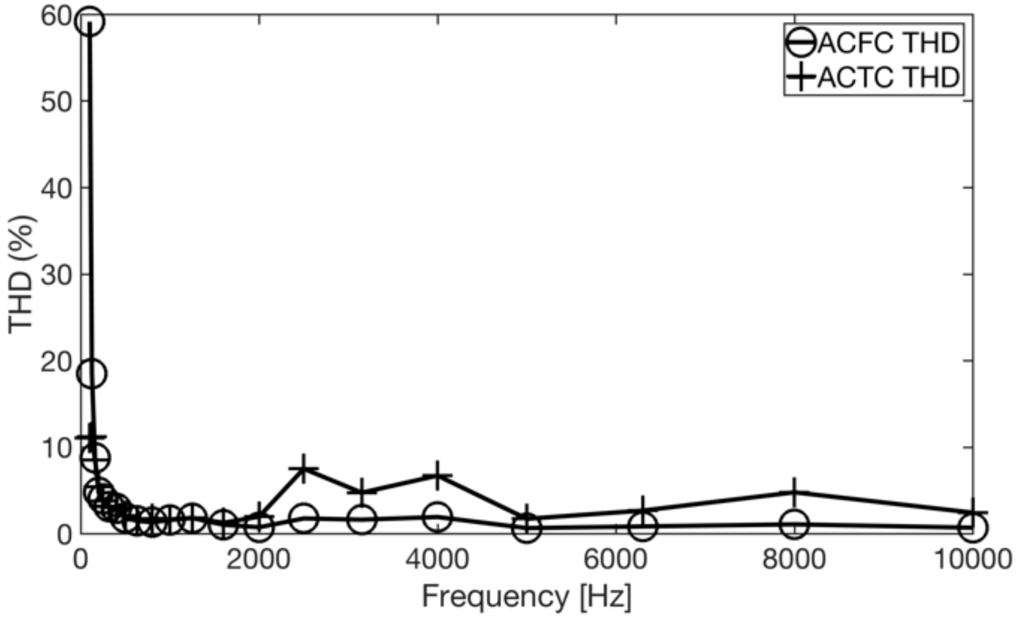


Figure 3-5: Data showing THD for FCAC and TCAC. THD was computed using EQ2-4. [27]

A summary comparison of FCAC, and TCAC is shown in Table 3-1.

Table 3-1: Summary of experimental data for FCAC, and TCAC signal processing techniques. The total input power for all tests was 72 Wrms with frequency ranges of 100 Hz to 10 kHz.

	Efficiency (μ%)	THD (%)
FCAC	1.01 - 1083	0.68 - 59
TCAC	1.26 - 388	1.7 - 11

3.5 Conclusions

Two new methods for thermoacoustic loudspeaker drive signal processing were leveraged from the hearing aid industry. Spectral envelope decimation of an AC Signal (FCAC) and dynamic linear frequency compression of an AC signal (TCAC) are methods that can be used with class D amplifiers. There efficiencies were 1.01 E-6 to 1083 E-6 percent and 1.26 E-6 to 388 E-6 percent with 72 W_{rms} input power, respectively. These efficiency levels are on the same order of magnitude as previously published methods that require class A/B amplifiers. FCAC and TCAC had THD of 0.68-59% and 1.7-11%, respectively. While these THD levels are significantly lower than previously published methods, THD was found to be a poor sound quality metric for complex signals. This study used single tone signals, but when complex signals (e.g. speech, music) were used the result was subjectively poor. A new sound quality metric is needed to be able to objectively compare thermoacoustic drive signal processing techniques.

3.6 Acknowledgments

I would like to thank the contributions of my advisor, Dr. Andrew R. Barnard, as well as my Masters/PhD committee members Professor Christopher N. Plummer, Professor Charles D. Van Karsen, and Dr. Jason R. Blough. Thank you Dr. Joshua Alexander from Purdue University for your assistance in understanding the time and frequency domain hearing aid algorithms. Thank you Dr. Vesselin Shanov and your employees at NanoWorld Laboratories at the University of Cincinnati for growing the CNT forests. Finally, I would like to acknowledge PCB Piezotronics® for their donation of the measurement microphones used in this effort.

4 Subjective evaluation of carbon nanotube loudspeaker drive signal processing methods using single word techniques

4.1 Abstract

Carbon nanotube loudspeakers make sound by generating heat as opposed to vibration. In this evaluation, paired comparison and modified rhyme test jury study techniques were used to evaluate the intelligibility of spoken single words processed with different drive signal processing methods. A jury study was conducted using a novel in-vehicle drive up format. Ultimately, direct current offset was found to be the most intelligible processing method for the carbon nanotube loudspeaker.

4.2 Introduction

While the theory for creating sound with heat, as opposed to vibration, was laid out as early as the late 19th century [1], [2], there have been limited physical devices that utilize the technology. Well known examples of this include plasma speakers and the generation of thunder by a lightning strike, but both are rather unwieldy to tame. Fortunately, in 2006 Yu et al. demonstrated a new material, carbon nanotubes (CNT), which could use the thermoacoustic principle to generate sound [3]. With benefits including that CNT loudspeakers are extremely lightweight, have formable geometry (i.e. directivity), no moving parts, no reliance on rare earth metals, and frequency response from DC to 100kHz [8], [13], [28], [29], CNT loudspeakers show much promise in a variety of applications [30]–[36].

Since the initial work done by Yu et al., much has been revealed about these acoustic transducers [9], [12], [16], [37]–[49] including development of their base physical properties [17], [18], [50]–[57] and the exact makeup of the carbon structure to be used (e.g. graphene, multi-walls nanotubes, etc..) [25], [58]–[68]. A notable discovery was that the CNT loudspeaker's radiated sound pressure is directly proportional to input electrical power rather than input electrical voltage such as moving coil loud speakers [11]. This results in a doubling of frequency content (e.g. if you input a 1kHz electrical voltage signal into a thermophone, you will hear a 2 kHz sound pressure wave). Additionally, they are inefficient when compared to traditional moving coil loudspeakers (~1e-6% power efficiency) [5], [21], [69], [70], the frequency response for open thermophones is not flat, but logarithmically increases with frequency [11], and they damage easily if not supported [71]. The CNT loudspeaker remains a positive addition to the limited sound generation toolbox of audio engineers, but unfortunately comes with significant restriction on its practical application, primarily to applications where weight savings and/or custom directivity are important and energy is abundant.

To date, all of the evaluation of CNT loudspeakers has been objective (e.g. sound pressure at a known distance, sound power, and total harmonic distortion). This work evaluates subjective sound quality of these transducers whereby single word spoken text is used to study the drive signal processing methods. The study utilized a novel drive-up jury study format that proved to be an effective and safe alternative to traditional in-lab studies.

4.3 Background

4.3.1 Drive signal processing methods

One of the prominent concerns with CNT loudspeakers is that they double all input frequency content. This is a trivial issue in the single sine wave case, because if, for example, a 400Hz sound pressure wave is desired, then a 200Hz electrical sine wave can be input. Yet what about complex transient signals such as speech or music? To solve this problem, drive signal processing of the desired audio is required before amplification. To date, there have been many drive signal processing methods used that all have their own positives and negatives.

In alphabetical order, the common processing methods are amplitude modulation (AM) [14], direct current (DC) offset [4], [72], pulse amplitude modulation (PAM) [73], and spectral envelope decimation (SED) [24].

AM processing is the same as AM radio, where the signal is modulated by a high frequency carrier wave such that the envelope of the final waveform is that of the original signal. Important variables for AM processing are the frequency of the carrier wave and the modulation depth (i.e. the ratio of signal to carrier amplitudes). AM processing can be used with a class D amplifier, but must have a maximum frequency response of the carrier plus maximum signal frequency content. For example, the carrier is typically above 20kHz so humans can’t hear it. Therefore, if a 1kHz sine wave is modulated with a 20kHz carrier, the amplifier needs to have frequency response to 21kHz. This can be limiting since most commercially available class D amplifiers have low pass filters incorporated at or near 20kHz. The modeling equation for AM is shown in EQ4-1,

$$y(\mathrm{t}) = (1+A_M x(\mathrm{t}))*A_C \sin(2\pi F_C \mathrm{t}) \qquad \text{EQ4-1}$$

where y is the signal played into the amplifier (volts) as a function of time, *t* (seconds), A_M is the amplitude of the modulated signal (volts), x(t) is the modulated signal, A_C is the amplitude of the carrier signal (volts), and F_C is the frequency of the carrier (Hz).

DC processing takes the signal and applies a static offset. The important parameter for this processing is the ratio of the offset to the signal amplitude. DC offset requires a class A/B amplifier which is typically more expensive than a class D amplifier. The modeling equation for DC is shown in EQ4-2,

$$y(t) = B + x(t) \qquad \text{EQ4-2}$$

where y is the signal played into the amplifier (volts) as a function of time, t (seconds), B is the amount of offset (volts), and x(t) is the signal to be modulated.

PAM processing, which is different than pulse width modulation (PWM) [4], takes a constant duty cycle square wave and varies the amplitude of the pulses such that the envelope of the final signal replicates the original signal. The important parameters for this processing method are duty cycle and pulse square wave frequency. This processing method requires a very high frequency amplifier (e.g. Radio Frequency) which can be very expensive especially with high power output requirements. The modeling equation for PAM is shown in EQ4-3,

$$y(t) = A_p p(t) * x(t) \qquad \text{EQ4-3}$$

where y is the signal played into the amplifier (volts) as a function of time, *t* (seconds), A_p is the amplitude of the pulse train, p(t) is the unity square wave pulse train at a certain duty cycle (i.e. percent high versus low), and x(t) is the signal to be modulated.

SED processing was originally developed in the hearing aid industry to pitch shift content lower so those with high frequency hearing damage could hear the content at a lower register. It uses the results from 75% overlap Fourier transforms to decimate the amplitude values by a factor of two while keeping the phase of the spectral lines the same. The result is then inverse transformed back into the time domain. This method can be use with common class D amplifiers.

4.3.2 Jury study methods

There were two main goals of this subjective effort:

1) Are CNT loudspeakers intelligible?

(Absolute – Yes/No)

2) Which processing method is the best to use for intelligibility?

(Relative - Ranked)

The above questions are subjective, so in order to answer them a subjective test is required. A jury study was conducted and in order to limit the scope and complexity, single word spoken text was chosen for evaluation. To investigate intelligibility (goal #1 above), a modified rhyme test (MRT) was conducted [74].

In a MRT test, subjects (i.e. jury participants or jurors) listen to speakers pronounce words (e.g. “The word is send”) and select the word spoken from a list of similar words. The common table (Table 4-1) has 6 columns and 50 rows with each row being a set of words with similar phonetic properties. All rows/sets have the same consonant-vowel-consonant style while varying either the initial or final consonant [75]. While the subject listens to the word spoken, the six similar words, including the one spoken, are shown on the screen. The subject then has to select which word was spoken.

Table 4-1: The first three rows of the MRT 300 word list [75]

Went	Send	Bent	Dent	Tent	Rent
Hold	Cold	Told	Fold	Sold	Gold
Pat	Pad	Pan	Path	Pack	Pass
…	…	…	…	…	…

A paired comparison test was chosen to rank the processing techniques (goal #2 above). The method has been used many times and the methods of processing the data are well documented [76]. This method requires the subject to listen to two paired samples and then select an answer based on a prompt. In this case the subject was to answer which one is more intelligible, but paired comparison can be used for many other prompts (e.g. which audio clip is more harsh, which audio clip would you prefer your car sound like, which food sample tastes better, etc.). With this method, it is important to avoid biasing the subject. For example, choosing which sample is played first is important. It is recommended to repeat at least a subset of the samples in the opposite order. For example, if sample “A” was played first and followed by sample “B,” at some point later in the study the subject should hear B followed by A and answer consistently. Additionally, it is recommended that there are only two options for the subject to choose from when answering (e.g. A>B or B>A). Including an option stating that they are the same has been known to cause poorer juror performance in any challenging comparison because they are not forced to make a decision [76, p. 139].

4.4 Methodology

4.4.1 Drive signal processing & Transducers

The drive signals compared in this study were Amplitude Modulation (AM) into a CNT loudspeaker, Direct Current (DC) offset into CNT, Pulse Amplitude Modulation (PAM) into CNT, Spectral Envelope Decimation (SED) into CNT, and Unprocessed into both a traditional moving coil loudspeaker (TradUP) and a CNT loudspeaker (CntUP). For AM, a carrier frequency of 45kHz was used so that the side bands would be above the audible range even if modulating 20kHz content. To set the modulation index of the transient spoken word files, the carrier amplitude was set at a fixed level of the fast a-weighted level maximum (i.e. LAFmax) for the signal to be modulated. This means that the modulation index was 1 at the moment when the signal had its maximum LAF level, but varied throughout the other parts of the signal.

For DC offset, the amount of offset was also set to the LAFmax. For PAM, duty cycle was set to 10% with a 25kHz square wave carrier frequency. For SED, a blocksize of 1,024 was used on the 48kHz sampling frequency audio with a decimation factor of two.

The unprocessed method was used on a traditional moving coil loudspeaker (TradUP) and a CNT loudspeaker (CntUP), meaning there was no pre-amplification processing done. TradUP is linear with output pressure frequency equal to input voltage frequency while CntUP is nonlinear with output pressure frequency twice that of input voltage frequency. The traditional loudspeaker used in the study was the author's estimate of a well-known lower end monitor. The transducer selected was an Avantone Pro Active MixCube with built in amplifier (Figure 4-1). The MixCube is an unprocessed full range driver.

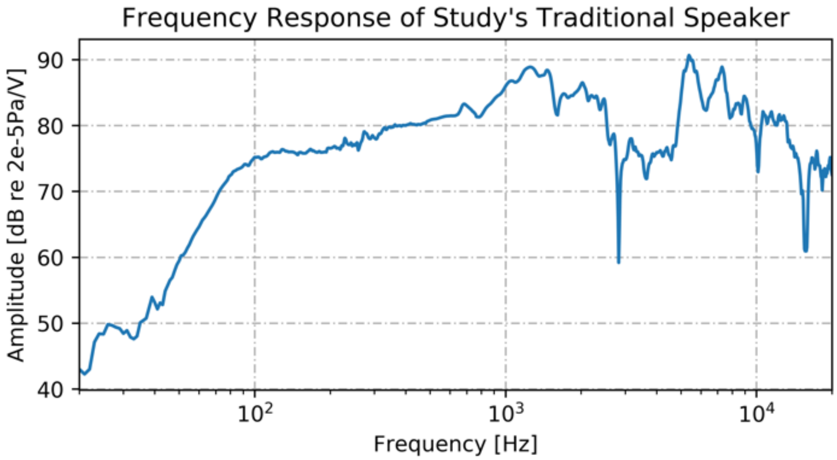


Figure 4-1: Acquired data showing the frequency response (Pa/Volt) for the Avantone Pro Active MixCube. The data was acquired at 1000 points logarithmically spaced between 20 and 20kHz. The input was a single sine wave for 20 seconds while the output was averaged for noise reduction. The distance from loudspeaker to microphone was 1m.

For the CNT loudspeaker, multiwalled carbon nanotube layers with ~ 15nm tube diameters and 500μm tube lengths were drawn from a forest approximately 12mm tall. Each layer was laid across two 6mm diameter copper rods five layers thick. There were six stacks of five layers in total over the copper rods for an overall dimension of ~9cm tall by 4.5cm wide by 75nm thick (Figure 4-2 & Figure 4-3).

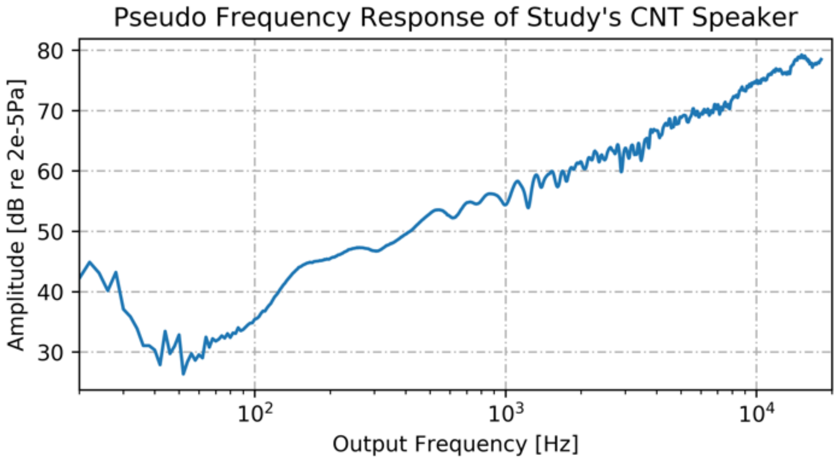


Figure 4-2: Acquired data showing the pseudo frequency response (Pa) for the CNT loudspeaker used in this study. As a result of the frequency doubling, the input voltage is at half the frequency of the output pressure wave so the common averaged frequency response measurement cannot be computed. Therefore, the authors decided to present what is more commonly referred to as the sound pressure linear autopower with constant voltage input as the pseudo frequency response. The data was acquired at 1000 points logarithmically spaced with input voltage from 10Hz to 9kHz and output pressure between 20Hz to 18kHz. The input was a single sine wave for 20 seconds while the output was averaged for noise reduction. The distance from loudspeaker to microphone was 1m.

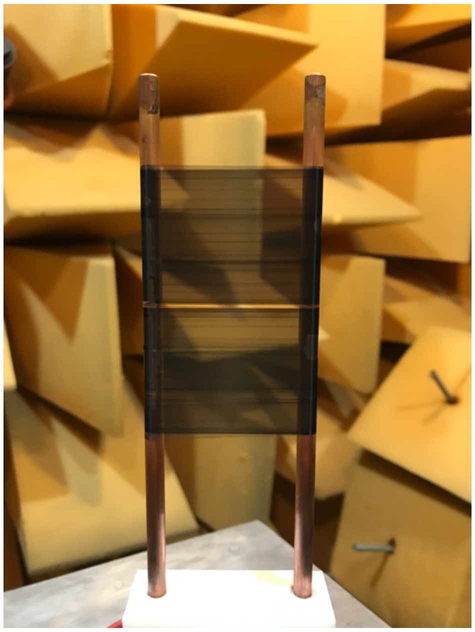


Figure 4-3: Picture of the CNT speaker used in this study. Note the CNT was not wrapped on both sides of the copper rods (i.e. it was single sided).

Comparing the speakers another way, their step responses were taken (Figure 4-4). The obvious difference is the incredibly short duration of the CNT loudspeaker pulse. This was expected. CNT loudspeakers have no moving parts and a frequency response to 100kHz. They can respond faster than a moving coil loudspeaker. Additionally, the MixCube has an enclosure so it has reflections and resonances as well at better low frequency response that contribute to a wider step response.

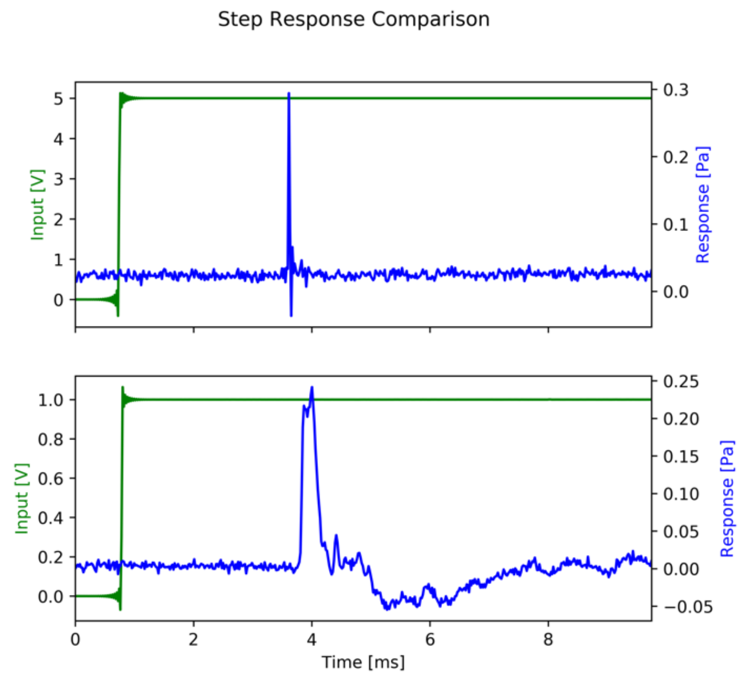


Figure 4-4: Step response comparing the CNT loudspeaker (a) to the traditional moving coil loudspeaker (b). Note the input voltages (y-axis) are different, but they generated a similar peak output pressure (~83dB) for their current amplification settings. Both measurements were taken at 1 meter distance. The CNT speaker was amplified with a Techron 7224. The Mixcube used its internal amplifier. The data was acquired at room temperature (~21C).

4.4.2 Sample recording

The spoken MRT word recordings for this study came from the National Institute of Standards and Technology (NIST), which has high quality recordings of nine speakers (4 females/5 males of varying age) speaking the complete 300 word list as "the word is X" [77].

To illustrate the complete signal preparation process, the 2700 audio files (9 speakers * 300 words) were acquired from NIST. They were de-noised with spectral noise gating as the default files had noticeable broadband noise on them and the author did not want to bring that noise into the processing. The less noisy audio files were processed by all of the previously described methods. Then those files were played into the CNT speaker via an AE Techron 7224 1kW A/B amplifier fed by a National Instruments 9269 analog output card running a custom file playing program. The playback was recorded by a Head Acoustics HMM II.1 Aachen head 1m away (0.5m for PAM) in a full anechoic chamber. The data acquisition unit was a Siemens Test Lab SCADAS III with PQA II analog input cards. For the unprocessed traditional moving coil loudspeaker (TradUP), the de-noised files were played directly into the built in amplifier on the device. Then both channels of the recorded binaural files were normalized so that their LAFmax was 0.03 Volts for export into a +/-1 Volt WAV file. 0.03V was chosen because it was the value that allowed for export without clipping of all files. The normalized WAV files were what was played back to the subject's headphones during the study.

Throughout the recording process, a top priority was to make sure there was good signal to noise for each recording. Some of the processing methods (e.g. PAM) were not able to be amplified well by the AE Techron 7224, because the 25kHz square wave and its harmonics get attenuated quickly even with an amplifier with ~500kHz roll off. In cases such as this, the speaker had to be moved closer than 1m in order to get good SNR by the instrument grade class A 12.7mm microphones in the Aachen head. The high dynamic range microphones and 24 bit Analog to digital converter on the data acquisition units helped with this.

To ensure there was no risk of hearing damage to the subject, all 2700 files were played through the study tablet and headphone pair with Windows system volume at 100%. During this playback, the headphone output A-weighted sound pressure level was recorded with a calibrated Larson Davis AEC 206 Headphone Test System. This level was below 85dBA for all files. Since the study was only ~40 minutes long and the levels were less than 85 dBA, being a part of the study was safe for the subjects. The OSHA noise dose limit is 85dBA for 16 hours.

4.4.3 Jury study

As described previously, the jury study had two main sections: a modified Rhyme Test (MRT) and a paired comparison (PC). Unfortunately, due to the COVID-19 pandemic, the study was not able to be conducted in a lab. The authors understand that performing a jury study in a controlled setting with minimal background noise and distraction is desired [78], but use of the planned lab space was not possible when this study was to be executed in March of 2020. Therefore, the authors adapted the study to be performed on a tablet inside of a vehicle and conducted it in August of 2020. The subjects signed up for a time, drove up to the proctor who was in a parking lot, and the proctor would ask the required COVID-19 screening questions and then hand the participant a sanitized tablet with headphones in single-use covers. The tablet used was an i3 Microsoft Surface and the headphones were Sennheiser HD598SE open-backed over ear headphones. This tablet amp/headphone pair were quantified (Figure 4-5). The subject completed the study in their car after which they received $20 in compensation. The subjects were not required to do anything in regard to their vehicle. They were only highly encouraged to leave their windows shut and HVAC fan as low as possible, but maintaining comfort was the most important. However, as the study took place in a parking lot, there were many more visual distractions than there would have been in an ideal setting with a closed sound quality test room.

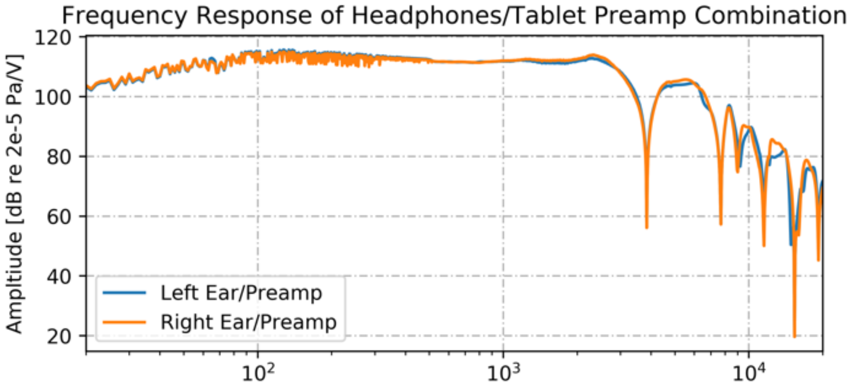


Figure 4-5: Acquired data showing the frequency response (Pa/Volt) for the headphones and computer preamp combination used in this study. The data was acquired at 1000 points logarithmically spaced between 20 and 20kHz. The input was a single sine wave for 20 seconds while the output was averaged for noise reduction. The headphones were attached to and measured with a Larson Davis AEC 206 Headphone Test System connected to a National Instrument 9234 24 bit ADC.

The software the subject interacted with was a custom written National Instruments LabVIEW program that guided the subject through the study, played the files, and exported the results. The software itself had six main parts:

1) Demographics questions

2) Five seconds of background data collected using the tablet microphone

3) Practice rounds with two PC and two MRT example questions. The files chosen for the PC section were very different for the first round and vary similar for the second round to help prepare and train the subjects.

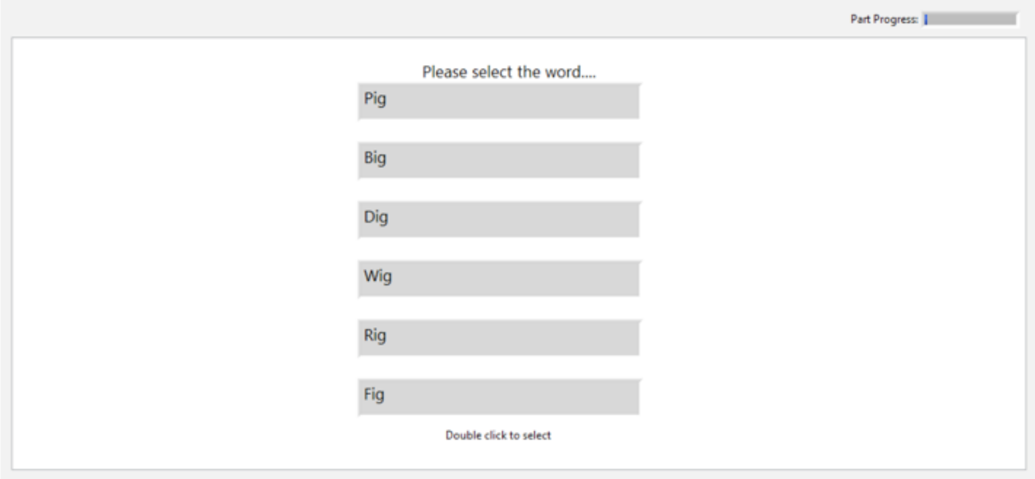
4) Paired Comparison - 155 questions

5) Modified Rhyme - 270 questions

6) Feedback section

There was no formal training or demographic requirements of the subjects. The only requirements were age between 18 and 65 and not be a non-resident alien for tax purposes for the compensation. Jury members were recruited from a broad demographic of people to represent a scenario where there is widespread adoption of this technology in the audio industry. Additionally, the subjects were not required to take a hearing test before participating. While hearing tests are a popular choice for jury studies, subject quality for this test was established with different methods described below in the PC section. This allowed the authors to avoid collecting sensitive medical data.

The initial goal was a study that would take approximately 30-45 minutes to complete. For the MRT portion of the study there were 270 questions. 270 divided by the six methods meant that there could be 45 randomly selected questions from the 2700 pool of MRT words for each method, making sure they were evenly selected from the nine speakers. An example of what the screen looked like during the MRT section is shown in Figure 4-6. Note, the subjects could not repeat hearing the sample in the MRT portion of the test.

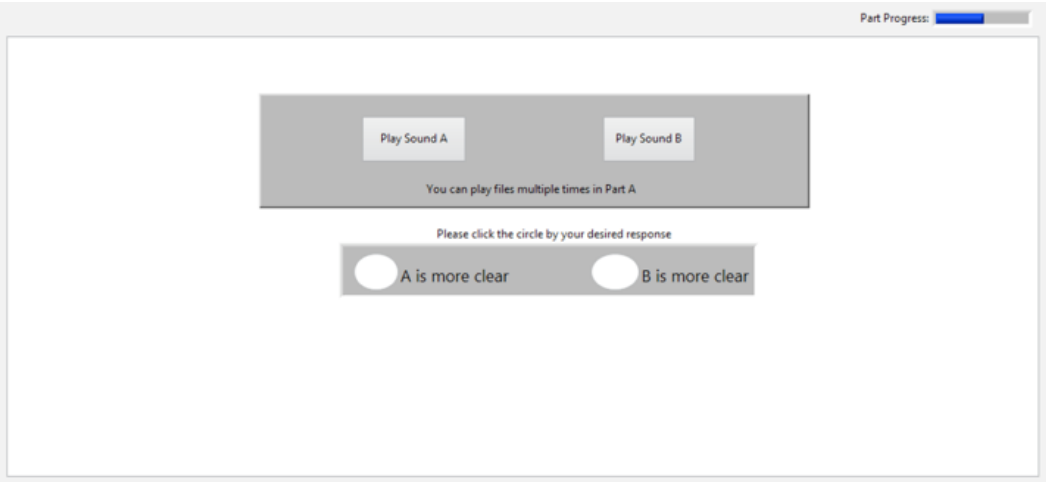


Please select the word....

Double click to select

Figure 4-6: Screenshot of the MRT portion of the study

For the PC section, there were six methods to compare (AM, CntUP, DC, PAM, SED, and TradUP). These methods can be thought of like sports teams that need to play each other in a tournament. In order for all of the teams to play each other there would need to be 15 games or in this case 15 pairs where method A competes against method B. As previously discussed, in paired comparisons the order of samples played matters. Continuing the sports analogy, this is equivalent to home field advantage. The authors wanted to test for A being played first (A→B) and also B being played first (B→A) to check for consistency of subject response as a juror quality metric. This was especially important since no hearing tests were done with the subjects. Additionally, a few duplicate pairs were used to test for consistency. This led to 155 questions for the PC portion. Six methods leads to 15 AB pairs per word so that allowed for six words or 75 total AB pairs. Including 100% duplication of AB and BA that brings the 75 pairs to 150. On top of that the authors included 5 ABAB duplications. So there were 155 total questions. The six words were randomly selected from the 2700 options, making sure each one was a different speaker. An example of what the screen looked like during the PC section is shown in Figure 4-7. Note, the subjects could repeat hearing the sample in the PC portion of the test.



You can play files multiple times in Part A

Please click the circle by your desired response

○ A is more clear ○ B is more clear

Figure 4-7: Screenshot of the PC portion of the study

4.5 Results

4.5.1 Demographics and juror quality

Overall, 47 subjects participated in the study. Using the PC portion of the data, subject quality was evaluated to see which subject datasets should be removed. Six were removed due to inconsistent ABC response. Meaning, for example, if the subject said A was better than B which was better than C. Then they also needed to say A was better than C. Six of the subjects did not do that enough to make their dataset usable. It is interesting to note that the six removed were all inconsistent with how they ranked CntUP versus SED versus PAM, hinting that CntUP/SED/PAM likely have similar intelligibility.

Two subject datasets were removed due to insufficient repeatability in AB vs BA responses. Finally, one additional dataset was subjectively removed by the author due to a low ABAB duplicate response of 20% correct and low ABBA accuracy of 74% consistent. In total that left 38 usable datasets.

The resulting demographics are shown in Figure 4-8. The subjects were mostly college age males with English as their first language, no known hearing issues, and no prior jury study experience. When asked how they would rank their typical audio listening experience, where 1 was never listen critically and 5 was always listen critically, a majority of the time the subjects ranked themselves as a 3 or 4.

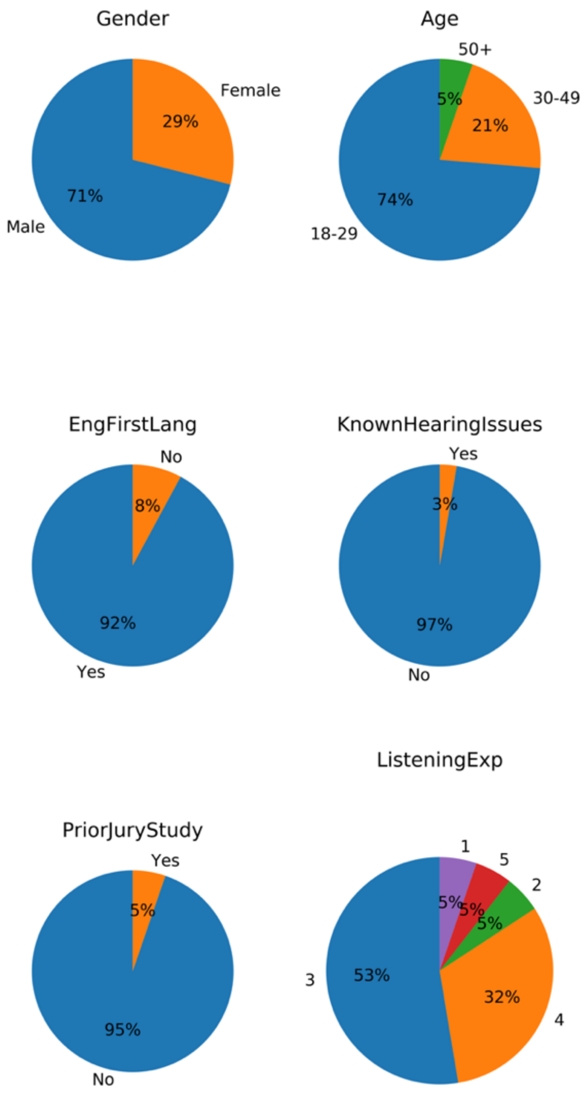


Figure 4-8: Pie charts showing the demographic distribution of the n=38 subjects used in this study. For listening experience, the higher the number the more critical the subject ranked their typical speaker listening.

4.5.2 Modified rhyme test

The first step in processing the MRT datasets was to determine the number of correct answers each method had for each subject. The percent correct for that method for that subject was then computed. The averaged percent correct for each method over all subjects was then determined. The results are shown in Figure 4-9. For example, on average for all subjects the word was correctly selected 80% of the time for CntUP.

In order to determine which method performed better (i.e. had a higher mean correct selection percentage), the data was analyzed with a Shapiro test and found not to be normally distributed. Therefore, a non-parametric Dunn's test was used with Bonferroni correction for the p-values. With a Dunn's test the null-hypothesis is that the means are the same (i.e. statistically not different). If the alpha value is below 0.05 then it can be said that there is a statistical difference in the percent correct means. The summarized results are shown in Table 4-2. The percent correct value was not statistically different when listening to TradUP versus DC. Interestingly, this was also true for AM vs DC, but not true for TradUP vs AM (i.e. there was a statistical difference in the mean percent correct values). In a similar way, CntUP = SED but CntUP > PAM. SED and PAM percent correct means were not statistically different.

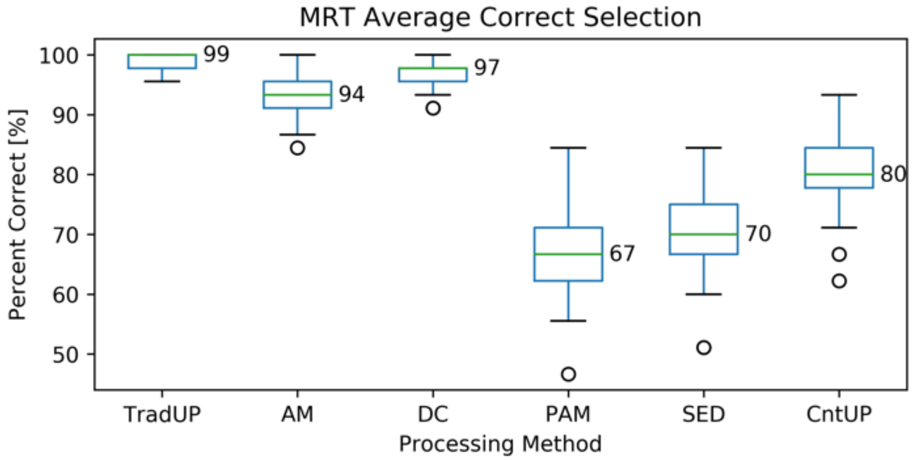


Figure 4-9: A boxplot of the MRT correct selection data by drive signal processing method. Higher is better.

Table 4-2: Method rank from MRT selection accuracy data

Rank	Method
1	TradUP = DC
2	AM = DC
4	CntUP = SED
5	PAM = SED

Note: An equals sign represents no significant difference (i.e. not rejecting the null hypothesis). Moving down a row represents statistical difference (α=0.05)

The time it took a participant to select an answer for the MRT portion of the study was also investigated. This data is summarized in Table 4-3. Using a Dunn's test again, TradUP and DC were not statistically different in selection time while the other methods followed a similar trend to the selection accuracy rank data (Table 4-4).

Table 4-3: MRT time to select by method in seconds. Lower is better.

	TradUP	DC	AM	PAM	SED	CnUP
Avg	1.15	1.18	1.28	2.10	1.71	1.89
Stdev	0.86	0.82	0.94	1.71	1.30	1.48

Table 4-4: Method rank from MRT time to select data.

Rank	Method
1	TradUP = DC
3	AM
4	SED
5	PAM = CntUP

Note: An equals sign represents no significant difference (i.e. not rejecting the null hypothesis). Moving down a row represents statistical difference (α=0.05)

4.5.3 Paired comparison

The first step in processing the PC datasets was to determine the number of wins each method had versus all other methods for every subject. From this, that subject's preference rank was determined. For example, Table 4-5 shows the compiled wins for participant 1. To illustrate how to read the table, start at column "PAM" and then move down to row "SED" and note the cell value is 1. This means that PAM won versus SED. If the cell was 0 it would mean that PAM did not win versus SED. This table was made for each subject by looking at how many times each method won versus another method. If the method won a majority of the time in the five meetings, then a 1 was placed in Table 4-5 in that corresponding cell. Compiling the win table for each participant creates the subject's rank. If there was a tie, then both methods shared the higher rank. For example, if PAM and SED were tied for 4th, then they were both recorded as placing 4th when computing the average ranking.

The overall average ranking is shown in Figure 4-10. TradUP was always ranked best at 5 because the average ranking was 5 with a standard deviation of 0. In order to determine the statistical rank, the non-parametric Dunn's test had to be use again due to none normally distributed data. The results of this are shown in Table 4-6. The average rank of TradUP was not statistically different than DC, but similar to the MRT data, TradUP was statistically ranked higher on average than AM and AM was not statistically different than DC. PAM, SED, and CntUP were not statistically different from one another, but were statistically ranked lower on average than TradUP, DC, and AM.

Table 4-5: Example wins table for participant 1. The axes representation the different signal processing methods. A value of 1 means that column method beat that row method.

	AM	DC	PAM	SED	TradUP	CntUP
AM	0	1	0	0	1	0
DC	0	0	0	0	1	0
PAM	1	1	0	0	1	1
SED	1	1	1	0	1	1
TradUP	0	0	0	0	0	0
CntUp	1	1	0	0	1	0

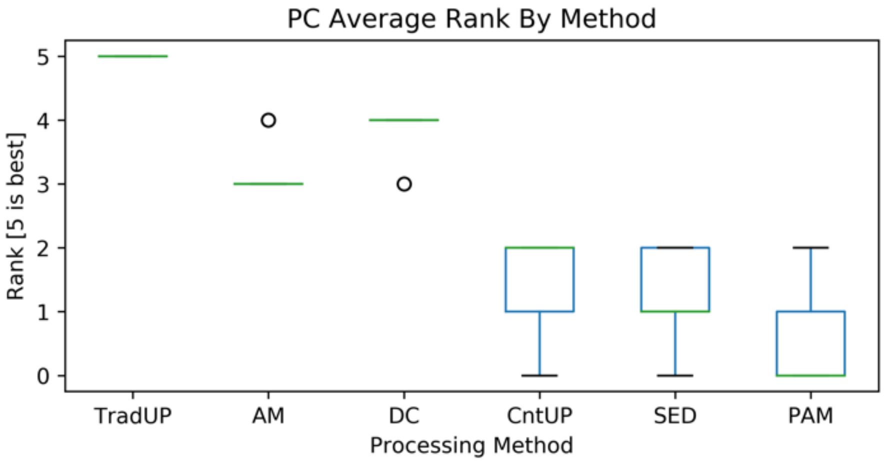


Figure 4-10: A boxplot of the average rank of the different processing methods based on the PC data. Higher is better.

Table 4-6: Method rank from PC data

Rank	Method
1	TradUP = DC
2	AM = DC
4	PAM = SED = CntUP

Note: An equals sign represents no significant difference (i.e. not rejecting the null hypothesis). Moving down a row represents statistical difference (α=0.05)

Additionally, the number of ABBA inconsistent answers by method was tabulated (Table 4-7). This shows the number of times a given method was involved when the subject inconsistently answered their preference. The rank trend in Table 4-7 follows that of the PC rank (Table 4-6). It is interesting to note that TradUP was only part of ABBA inconsistent response 19 times which is much less than the other methods. The minimal ABBA inconsistent response for TradUP is expected because it was always ranked the most intelligible in the PC rank. While the rank data compares the methods relatively, the value of Table 4-7 is the absolute comparison to 0.

Looking at the data a different way, the numbers of wins each method had against other methods was tabulated (Table 4-8). Meaning every time an ABBA question pair was answered consistently the winning method was given a point. The sum of these points is shown in Table 4-8. The rank trend also matched the trend of the PC rank (Table 4-6), but had CntUP and PAM in an alternate positions compared to the ABBA inconsistent answer rank (Table 4-7).

Table 4-7: ABBA inconsistent answers by method. Lower is better.

TradUP	DC	AM	PAM	SED	CntUP
19	159	197	228	231	270

Table 4-8 Method win count. Higher is better.

TradUP	DC	AM	CntUP	SED	PAM
1156	767	630	199	139	82

Note. This data includes all correct ABBA answered from all 47 participants.

The PC portion of the study was also timed. When a new question would start the software would automatically play sample A and then sample B. The time from when sample B finished playing until an answered was selected was recorded. This data was averaged and compiled (Table 4-9). Looking at it statistically again with a Dunn's test, it was found that answers for the traditional speaker (i.e. TradUP) were statistically quicker than all CNT methods ($\alpha=0.05$). CNT methods were all statistically not different from one another.

Table 4-9: PC Time to select by method in seconds. Lower is better.

	TradUP	DC	AM	PAM	SED	CntUP
Avg	1.05	1.72	1.98	2.06	1.84	2.20
Stdev	1.32	2.62	2.86	3.03	2.44	3.21

4.6 Conclusions

In conclusion, it was found that a CNT loudspeaker using DC offset processing was just as intelligible as a traditional moving coil loudspeaker. AM, PAM, SED, and CntUP were statistically less intelligible than TradUP. However, DC offset requires a class A/B amplifier which may not always be an option. Using the PC rank data, it was found that AM was considered equally intelligible as DC and that AM/DC were more intelligible than PAM/SED/CntUP. Therefore, AM should be used if a high enough frequency response class D amplifier is an option.

Further, given the circumstances with COVID-19, the execution of a jury study was shown to be possible using participants' automobiles and a drive-up test site. This format led to various random additional variables to account for including rain, wind, and other nearby drivers. The data led to statistically significant results regarding the intelligibility of CNT loudspeakers.

4.7 Acknowledgements

The authors would like to acknowledge all of the participants in the study for taking the time to help contribute to this work.

5 Correlation of jury study results and psychoacoustic metrics

5.1 Abstract

Multiple linear regression was used to determine a weighted set of psychoacoustic metrics that best correlated to the jury study results discussed in Chapter 4. Thirty-three metrics, including a new custom metric, were reduced to a non-colinear set of metrics using factor analysis and principle component analysis. That metric set was iterated over to determine the optimal metric combination producing the best correlation to the study results. The resulting best fit included only one metric: The novel Total Harmonic Distortion for Speech Intelligibility (THDSI). The fit from the final regression was $R^2 =$ 0.12. In addition to the regression analysis, a modern speech recognition machine learning model was compared to the jury study results.

5.2 Introduction

The execution of a jury study to understand the subjective response of humans to a certain stimulus is a laborious process that would ideally be replaced with a psychoacoustic metric i.e. an objective metric proven too closely follow a subjective preference. That way when a new stimulus arises there is no need to execute a full jury study. The stimulus can be processed computationally by the psychoacoustic metric resulting in an estimate of the subjective response as if a jury study were completed.

The use of psychoacoustic metrics is very common in the field of acoustics. Humans psychoacoustic response to pressure waves is drastically different than the objective metric of pressure. On a broad level humans perceive sound differently based on the frequency and level of the sound, but on a very specific level humans hear sound differently based on a variety of factors (e.g. how much background noise there is, what precedes the sound, and even what the human is doing while listening to the sound). Therefore many psychoacoustic metrics have already been developed. The most common metrics are weighted sound pressure level (dBA/dBC) and loudness (sones) which seek to account for the broad frequency and level discrepancies between objective pressure and human perception. There are many more psychoacoustic metrics which tend to lend themselves toward specific usage applications. Some examples include (in alphabetical order):

Cepstrum: The inverse Fourier transform of the logarithm of a Fourier transformed signal. Cepstrum makes periodic events more pronounced. Usage: Gearbox noise analysis

Impulsiveness: This metric attempts to transform the non-linear response humans have to fast and large changes in pressure level to a linear metric. Usage: Gunshot analysis

Kurtosis: A measurement of how quickly the signal changes. Usage: Health monitoring

Prominence ratio: the ratio of a tone level to noise. This metric is similar to the other tonal psychoacoustic metrics Tonality and Tone-To-Noise. Usage: Automotive turbo whine

Roughness: This metric seeks to quantify the psychoacoustic response to signal modulation for signal modulations up to 70Hz. This is very similar to the low frequency modulation psychoacoustic metric Fluctuation Strength (modulations <4Hz). Usage: Electric razor sound quality

Total harmonic distortion (THD): This is the ratio of energy in the fundamental compared to the summation of the energy in all of the harmonics. Usage: 60Hz electrical power quality

Total harmonic distortion + noise (THDN): Similar to THD, this metric is the ratio of the energy in the fundamental to the energy in the entire signal. Usage: Amplifier performance quantification

Sharpness: This metric tries to account for high frequencies being more annoying to humans than low frequencies. It can be thought of as a high frequency weighting function. Usage: Vacuum cleaner sound quality

Speech intelligibility index (SII): The percent of speech that is intelligible given a specific background signal and speaker signal. SII (1997 ANSI S3.5) is an improvement on the 1969 ANSI S3.5 Articulation Index (AI) metric [78]. Usage: Determine how difficult it would be to understand the passenger in your vehicle for difference vehicle conditions

Speech transmission index (STI): The percent of speech that is intelligible given a specific speech signal and a perturbed speech signal [79]–[81]. Usage: telecom industry to understand the effects of transmission across their service.

While use of these metrics in their specific usage application is obvious, they can even be used in combination with other metrics. This is common when attempting to correlate to jury study results. There are many examples from architectural acoustics [82]–[86], the automotive industry [87]–[93], and beyond [94]–[99]. The basic idea is to perform a regression analysis on the metrics (i.e. the independent variables) to see how they correlate to jury study results (i.e. the dependent variables). When performing subjective to objective comparisons historically, coefficient of determinations (R^2) have been found to be as high as 0.5-0.8 in Vardaxid et al.'s literature review of building sound quality [100], 0.5 in Gozalo et al.'s sound scape comparison [98], 0.92 for Moravec et al.'s washing machine comparison [99], and 0.92 for Astolfi et al.'s work in secondary education classrooms [82].

There are many different types of regression analyses that can be done. Typical regression analysis is done when the dependent variable is not categorical (e.g. is can be any value). Logistic regression is used when the dependent variable is categorical (e.g.

Yes/No). Additionally under the regression category is linear and non-linear methods. Both of these methods have many different models that have subtle assumption differences (e.g. the intercept is forced to 0). The standard regression is single linear regression (aka linear regression),

$$y = mx + b \tag{EQ5-1}$$

where y is the dependent variable, x is the independent variable, m is the weighting factor applied to x and b is the intercept. If there were multiple metrics multiple linear regression could use used,

$$y = m_1x_1 + m_2x_2 + m_nx_n + \cdots + b \tag{EQ5-2}$$

where y is the dependent variable, x_n is the nth metric, m_n is the nth weighting factor, and b in the intercept. Multiple regression can also be done with non-linear models as well as logistic regression analysis.

When performing regression analysis it is important that the dependent variables not be colinear [87]. Collinearity is a way to express high correlation ($R^2 > 0.7$) among dependent variables of a model. Ideally there is low correlation among the dependent variables and high correlation between the dependent variables and the independent variables of a model. Collinearity or multicollinearity can result in variations of any individual metrics causing variation in other metrics. The general process flow to develop the non-colinear metric set to include in the regression model is overviewed in Figure 5-1.

In concert with performing regression on common metrics, it is also possible for researchers to develop custom metrics for the specific stimulus they are studying [91]. In this study, the stimulus or audio files listened to by the subject were transient speech. This is a very important caveat because basically all of the psychoacoustic metrics mentioned above are designed for stationary or at least semi stationary data where the level is not changing drastically over, for example, the Fourier Transform (FFT) block. This presents a unique challenge when processing the common metrics appropriately.

A typical way to adapt common psychoacoustic metrics to transient signals is to compute a max or average value of the metrics vs time. For example, a loudness vs time array could be computed for a transient signal. Then the maximum and/or average of that array could be taken. Then that max/average single value would be the metric used in the regression.

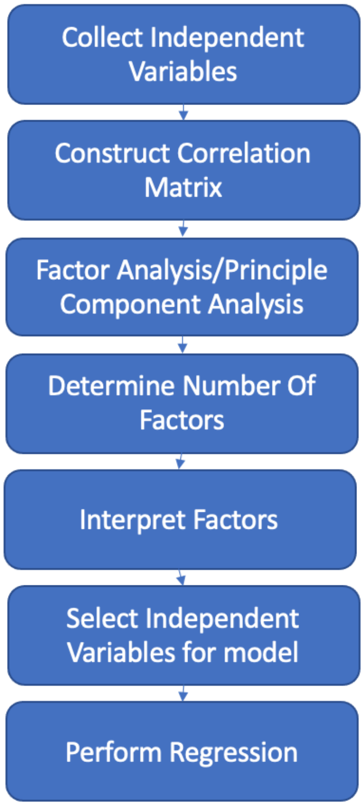

Collect Independent Variables
Construct Correlation Matrix
Factor Analysis/Principle Component Analysis
Determine Number Of Factors
Interpret Factors
Select Independent Variables for model
Perform Regression

Figure 5-1: Flowchart of how to decide what independent variables to use in a regression model [87].

One illustrative example of why transient signals are difficult is THD. It uses the frequency domain values to calculate the ratio of the energy in the fundamental to the sum of the energy in the harmonics, but the metric has no specified frequency resolution (i.e. acquisition time). With a 2 second audio clip, like in this jury study, the frequency resolution could be as low as 0.5Hz (1/2s), but that would take an average of the energy over the whole phrase "The word is X," which really is not expressing the THD level of what is just in the word the subject was to understand, "X." To help solve this problem, the signal could be processed in sections (i.e. FFT blocks) computing a max/average value of the THD vs time array as a custom metric.

For THD specifically, the power industry has developed a custom transient THD metric that utilizes small FFT blocks and overlap [101]–[103]. In development of this transient THD metric the power industry had the helpful advantages of A) knowing where to roughly look for the fundamental (i.e. 60Hz) and B) having very little noise outside of the fundamental and its harmonics. These advantages make it easier to determine the processing parameters and develop a robust algorithm.

As another example of the traditional approach to custom psychoacoustic metric generation, Huang et al. used the Wigner–Ville Transform to generate a metric for shock absorber sound quality [90].

In addition to the more traditional approach of custom metric generation as mentioned above, this work also tried to look at modern speech recognition machine learning algorithms to see if their results could be used as an independent variable. While complete explanation of how the algorithms work is outside of the scope of this effort, the general idea is to compute a cepstrum result every 10ms and input that data into a trained a Hidden Markov Model (HMM) [104], [105]. The output of the model is post-processed to give the estimated spoken word as well as confidence estimate.

Overall, the goal of this effort was to develop a set a psychoacoustic metrics and weights that have the best correlation to the jury study intelligibility results from Chapter 4. Then if a new signal processing method is discovered, its intelligibility can be estimated without having to conduct a completely new jury study. In this effort the common psychoacoustic metrics are used where appropriate, a custom psychoacoustic metric is developed, and a modern speech recognition algorithm is examined.

5.3 Methodology

5.3.1 Dependent variable selection

The MRT section of the jury study from Chapter 4 included 270 questions. 45 questions for each of the 6 processing methods (i.e. AM/DC/PAM/SED/TradUP/CntUP). The TradUP results were not included in this analysis as the main goal was to correlate to CNT results. Therefore, 225 incorrect or correct selection data points were generated from 38 valid jury subjects. The author did not include the 9 subjects that were thrown out from the paired comparison juror quality investigation outlined in Chapter 4. The primary dependent variable this effort tracked to was "Percent Correct" being the percent that the 38 jurors answered correctly when listening to a file.

For example, the first file (of 225) listened to was "SED_F4_b01_w5.wav." This means the processing method was SED, the speaker was female 4 (out of the 4 female and 5 male options), b01 means MRT word list 1 (of 50), and the word spoken was 5 (of 6). Therefore, when the subject listened to the first file they heard this wav file while looking at all six words in list 1. The subject then selected the word they thought they heard. They could have either got it correct or not correct. Therefore, there was correct or not correct (1 or 0) data for all 38 subjects for all 225 files. From this "Percent Correct" was computed for each of the 225 files. This was the primary dependent variable for the regression.

5.3.2 Signal processing

In order to generate psychoacoustic metric statistics, calibrated audio files were needed. These files were recorded from a Larson Davis AEC206 headphone test fixture. The same tablet and headphone combination used in the jury study were placed in the anechoic chamber with the headphones on the headphone test fixture. All 225 files were played through the tablet and headphones then recorded with a calibrated National Instruments 9234 24bit ADC module. This generated the calibrated files that were used when processing metrics.

Some of the speech intelligibility metrics in this study required a “clean” and “noisy” signal that were time synced. The data recorded from the headphone test fixture (i.e. the noisy signal) had to be time synced back to original clean data. Below is an illustration of how that was done for file F1_b01_w3.wav representing the first female speaker speaking word 3 from word list 1.

1) F1_b01_w3.wav was processed with methods AM/DC/PAM/SED/CntUP to generate AM_F1_b01_w3.wav, DC_F1_b01_w2.wav, etc…
2) The generated files were played through the amplifier into the CNT speaker and the response were recorded with an artificial head (aka Aachen head) at a 1m distance (0.5m for PAM) in the anechoic chamber
3) Those files were uploaded to the tablet and used in the jury study
4) After the study, the tablet and headphones were placed in the anechoic chamber and all 225 signals were played into the headphones while the output was recorded with the headphone test fixture.
5) Using cross-correlation and manual visual alignment, the noisy signals were time synced to the clean signal. Additionally, they were down-sampled to the same sample rate (i.e. 51.2kHz→48kHz) and truncated to the exact same length.

The standard files were the “phrase” files. Meaning they contained the phrase “the word is X.” During evaluation of the metrics, it was also decided to try and process just the single word spoken (i.e. just “X”), because the dependent variable Percent Correct was only focused on how well the subject understood the word and the phrase “the word is..” that came before it would likely just cause the metric to compute incorrectly. This was especially concerning when taking a maximum or average value of the metric versus time array. In order to generate the truncated files with just the single word X, Audacity was used to manually select the start and stop signal indices on the clean signals. The noisy signals were then processed with cross-correlation and manual adjustment to get truncated noisy files that were time-synced to the truncated clean file.

5.3.3 Metric development

5.3.3.1 Common Psychoacoustic metrics

As mentioned in the introduction, there are a variety of options for psychoacoustic metrics to include. The metrics selected and processing parameters for this effort are provided in

Table 5-1. Loudness and sound pressure level were chosen because they are the standard broad metrics. Total Harmonic Distortion (THD), THD plus Noise (THDN), and Signal-to-Noise (S/N) were chosen because the different carbon nanotube (CNT) signal processing methods all have to do with harmonic distortion. The blocksize was kept small (1024) relative to the 48kHz sampling rate to help reduce the transient effects. It is not clear how Head Acoustics Artemis software determines the fundamental frequency. The author assumes it takes the bin with the highest value from the FFT results, which with the lack of other information makes sense, but is not correct for this case. Regardless, it was still included. Sharpness was included because the CNT is more efficient at higher frequencies so a metric that looks just at higher frequency content seemed logical. Cepstrum was included even though the author did not feel it would provide much value. There was no technical reason this specific data could not be processed with Cepstrum so the metric was included. Similarly, Kurtosis was also included. Power Spectral Density (PSD) was included as a different time weighting than the Sound Pressure Level (SPL). PSD and SPL would likely give back similar results for a stationary signal, but it was unclear what would happen given the transient input so it was included.

Articulation-Band Correlation Modified Rhyme Test (ABCMRT), Speech intelligibility Index (SII), Speech Transmission Index (STI) and, Short Term Objective Intelligibility (STOI) seemed like the most promising metrics as they all look at intelligibility specifically so they were included. ABCMRT is an objective speech estimator that follows the MRT development logic. SII is used to determine intelligibility estimates for speech levels in noisy environments (i.e. measurements taken synchronously). STI is used to determine intelligibility estimates for perturbation mechanisms like phone calls (i.e. measurements taken asynchronously). STOI is used to determine speech intelligibility for degraded signals in cochlear implant simulations.

Table 5-1: Included standard subjective (i.e. psychoacoustic) metrics included in this analysis

Metric [units]	Processing settings	Software
Loudness [Sones]	DIN 45631 Sound field: Free Single values: Max/Average Averaged left and right ear	Head Acoustics Artemis
Harmonic Distortion [%]	1024 blocksize Overlap 50% THD/THDN/SN Single values: Max/Average Averaged left and right ear	Head Acoustics Artemis
Sharpness [acum]	Method: Aures Loudness - DIN 45631 Sound field: Free Single values: Max/Average Averaged left and right ear	Head Acoustics Artemis
Cepstrum [dB]	1024 Blocksize Window: Hanning Overlap: 50% Single values: Max/Average Averaged left and right ear	Head Acoustics Artemis
Kurtosis [none]	Overlap: 50% Integration time: 125ms Single values: Max/Average Averaged left and right ear	Head Acoustics Artemis
Sound pressure level (SPL) [dB re 2e-5 Pa]	Time weighting: Fast (125ms) Spectral weighting: Z Single values: Average Averaged left and right ear	Head Acoustics Artemis
Sound pressure level (SPL) [dBA re 2e-5 Pa]	Time weighting: Fast (125ms) Spectral weighting: A L5/L10/L25 (= statistics p95/p10/p75, respectively) Single values: Average Averaged left and right ear	Head Acoustics Artemis

Power spectral density (PSD) [dBA re 2e-5Pa]	1024 Blocksize Window: Hanning Overlap: 50% Spectral weighting: Z Single values: Average/Peak hold Averaged left and right ear	Head Acoustics Artemis
Articulation-Band Correlation Modified Rhyme Test (ABCMRT) [%] [106]	Used standard settings to map to all 21 critical bands Input both the original “clean” file and the “noisy” file the subjects listened to in the study Used left ear data only	MATLAB
Speech Intelligibility Index (SII) [%]	ANSI S3.5 Background signal was a 40dBA overall level pink noise Single values: Max/Average Input both the original “clean” file and the “noisy” file the subjects listened to in the study Used left ear data only	MATLAB
Speech Transmission Index (STI) [%] [79]–[81]	Delta frequency: 2 Hz for phrase files and 8Hz for single word files due to their short time duration. Default = 0.06Hz Input both the original “clean” file and the “noisy” file the subjects listened to in the study Used left ear data only	Python
Short Term Objective Intelligibility (STOI) [%] [107], [108]	Input both the original “clean” file and the “noisy” file the subjects listened to in the study Used left ear data only	Python

5.3.3.2 THDSI

Based on the common psychoacoustic metric investigation and the understanding that the different CNT drive signal processing methods are designed to cope with the frequency doubling issue, it became unfortunately obvious that the THD/THDN metrics would likely not work well. There were two main issues:

1) With the Artemis software, there was no way to enter a fundamental frequency or somehow teach the software where to look. This meant that it would likely not compute THD correctly. The signal to noise (S/N) metric would still be correct, but ideally THD could be calculated.
2) THD has to be computed in the frequency domain. This means that some block of time has to be averaged. In Artemis a blocksize as low as 256 can be selected (~5ms at 48kHz) so this would likely help, but would create large frequency resolution in the frequency domain encompassing the energy from the neighboring spectral lines.

To solve issue #1, it was proposed that the ideal metric could be fed both the clean and noisy signals, like ABCMRT, SII, STI, and STOI require and the algorithm could learn the correct fundamental frequency from the clean signal. For example, using short time Fourier transforms the time synced clean and noisy signals would be input and processed with the same blocksize. The algorithm would determine which bin in the clean signal had the highest level and call that the fundamental bin. It would then use that bin index and its harmonics on the noisy signal to compute THD.

To solve issue #2, three common time-frequency analyses were investigated. Short time Fourier transforms (STFTs), Morlet wavelets, and Wigner-Ville transforms were studied for an example file. The wavelet processing had worse frequency resolution at low frequency versus the STFTs which was a huge issue because most of the fundamental frequencies were low (<200Hz) where the wavelet frequency resolution was the highest. While the short time resolution at high frequencies was a bonus for wavelets, the very high frequency resolution at low frequencies meant that wavelets would not work. The Wigner-Ville transform was a concern for three reasons i) the frequency domain artifacts from the harmonics ii)) the significant computation time with 2s of data at a sampling rate of 48kHz and iii) The signals included multiple “events” especially in the phrase “the word is X” signals. Even the single word “X” signals still had multiple events in them for words like “pu-ck”. The author understands that different windowing methods could have been used to reduce the artifacts, but STFT appeared to work sufficiently so a more in depth Wigner-Ville investigation seemed unnecessary. Therefore, it was decided STFT was the best path forward with the main concern being that frequency resolution increased with decreased time between STFT blocks, but a variety of blocksizes could be computed quickly to determine an optimal blocksize setting.

In summary, the Total Harmonic Distortion for Speech Intelligibility (THDSI) metric requires inputs of time synced, same duration, clean and noisy signals. The signals are then STFT processed. In each spectra, the fundamental frequency bin index is determined from the highest value in the clean spectra. The fundamental level is set to the noisy signal value at that fundamental bin index. Then the harmonics are summed from integer multiples of the fundamental index in the noisy signal. THDSI is then computed as the energy in the harmonics divided by the energy in the fundamental. THDNSI can also be computed as the energy in all frequency bins except the fundamental divided by the energy in the fundamental.

To illustrate THDSI with an example, imagine there are two signals, an original and modified signal. The original is 5 seconds long, but has a 50Hz sine wave at an amplitude of 10 for the time period 1-5s. Then there is a modified signal which is the original signal after going through some perturbation. It has an amplitude of 0 for the first second. For time 1-2 seconds it has a 50Hz sine wave at amplitude 10 just like the original. For time 2-3 it has two sine waves, one at 50Hz 10 amplitude, but a second at 100Hz and 2.5 amplitude. For time 3-4 the first harmonic increases from 2.5 to 5. Then for the last second, time 4-5, the fundamental goes to 0 and the 100Hz sine wave at amplitude 5 still exists (Table 5-2). The spectrogram of the modified signal is shown in Figure 5-2.

If the common THD metric was calculated over the whole 0-5s period, the result would be 25%. Artemis would improve on this assuming a small enough blocksize was used. Where the need for a different metric is more obvious is in the last second, time 4-5s. Here the original signal has energy at the fundamental, but it was, for some reason, reduced to 0 in the modified signal. The Artemis algorithm would falsely call 100Hz the new fundamental and compute a THD of 0%, because there is no energy at 200Hz, when it actually should be infinite. Figure 5-3 shows the result from the THDSI computation for the modified signal.

Table 5-2: Example signals for THDSI

Time Period (Seconds)	Original Signal	Modified Signal (Fundamental/Harmonic)	Common THD computed over 0-5s*	Artemis THD*	THDSI*
0-1	0	0/0	25%	0%	0%
1-2	10	10/0	25%	0%	0%
2-3	10	10/2.5	25%	25%	25%
3-4	10	10/5	25%	50%	50%
4-5	10	0/5	25%	0%**	Inf

*Values computed on the modified signal
**Artemis computes 0% even though it should be infinite, because in the original signal there was energy at the fundamental.

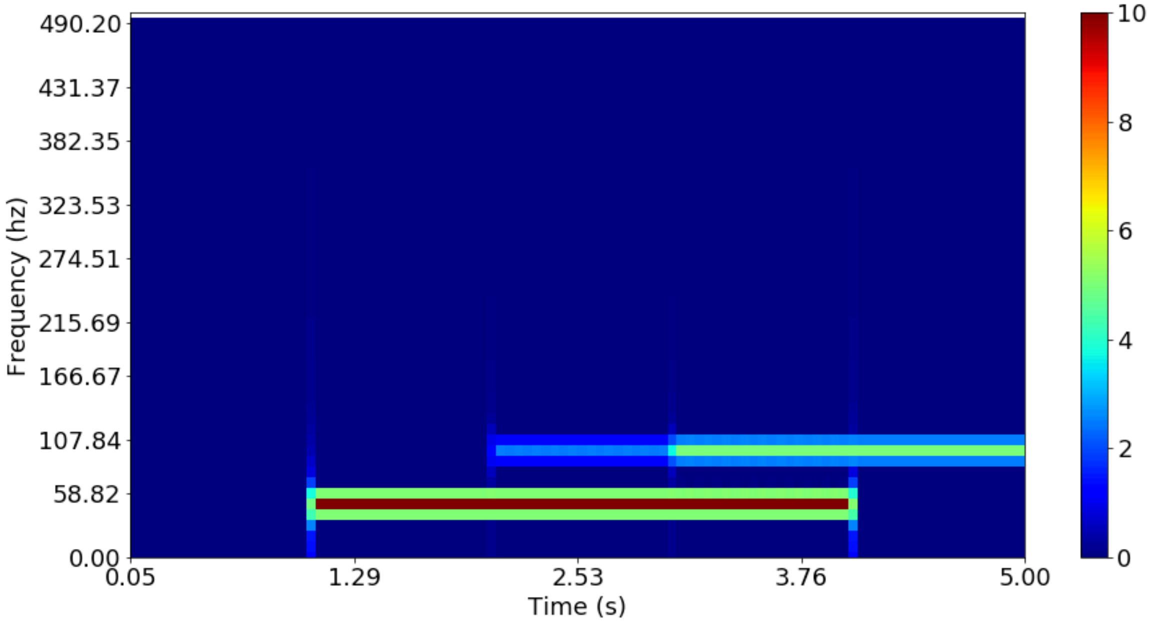


Figure 5-2: STFT analysis of the example modified signal.

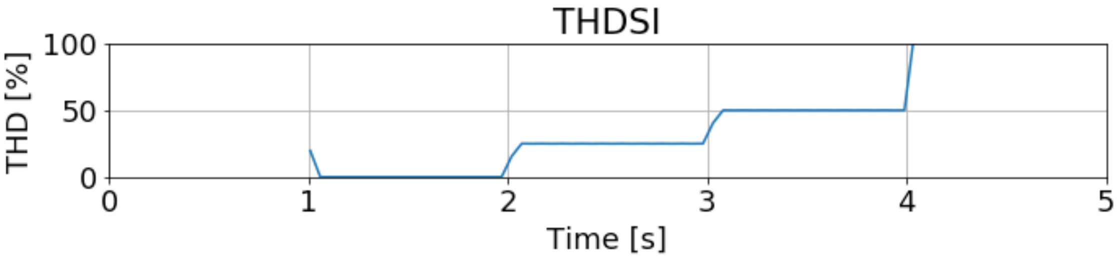


Figure 5-3: The THDSI result from the simulated modified signal.

Initially there are two obvious parameters for THDSI, the STFT blocksize and amount of overlap. The author also added three more parameters: down sample factor, A-weight, and a required threshold above the noise to detect a valid fundamental frequency. The down sample factor was incorporated to help increase algorithm performance since the data were acquired at 48kHz and the frequencies of interest were in the primary human speech range of less than 10kHz. The A-weight parameter was important so that higher frequency harmonics would not increase the THD unnecessarily as the subjects would not have perceived them. The threshold parameter requirement becomes obvious when there is silence in the file. For example, imagine a signal where the speaker says “the word is dent.” At the time between the words there is just noise in the signal so selecting the fundamental from the clean signal as the max bin value would just be selecting the bin with the highest noise. Therefore, a level and frequency threshold for the fundamental bin selection were added. The threshold level criteria,

$$|noisySignal(\max value\ index)| > threshold * average(|noisySignal|) \quad \text{EQ5-3}$$

where the noisySignal(max value index) is the value of the noisy signal at the index where the STFT spectra is maximum in the clean signal and average(noisySignal) is the average of all noisy signal STFT spectra values. The threshold frequency criteria required the fundamental to be at 20Hz or greater. If either the threshold level or frequency criteria were not met, then the algorithm would output a Not A Number (i.e. NaN) for THDSI and THDNSI for that spectral line.

With the above listed parameters the THDSI algorithm worked as expected with simulated signals, but had to be altered after investigation with CNT specific signals. Due to CNT's frequency response being logarithmic with respect to frequency (Figure 4-2), the low frequency values, where the fundamentals were in the clean signal, were significantly lower in output value and resulted in the THDSI calculation outputting very high levels (~10,000%). To help correct for this, the noisy signal was multiplied by the frequency response function (FRF) of the CNT loudspeaker. Figure 5-4 shows an example spectra where the STFT original result (blue) was modified to the corrected result (orange) to increase the lower frequencies values. This helped reduce THDSI levels that would be unrealistically high due to dividing by a very low fundamental level. It is important to note, as was done in Chapter 4, that the FRF for the CNT loudspeaker was not a true FRF since a true FRF cannot be computed, because the input power is at half the frequency of the output pressure. The FRF used here is more accurately called the linear autopower for a constant voltage input.

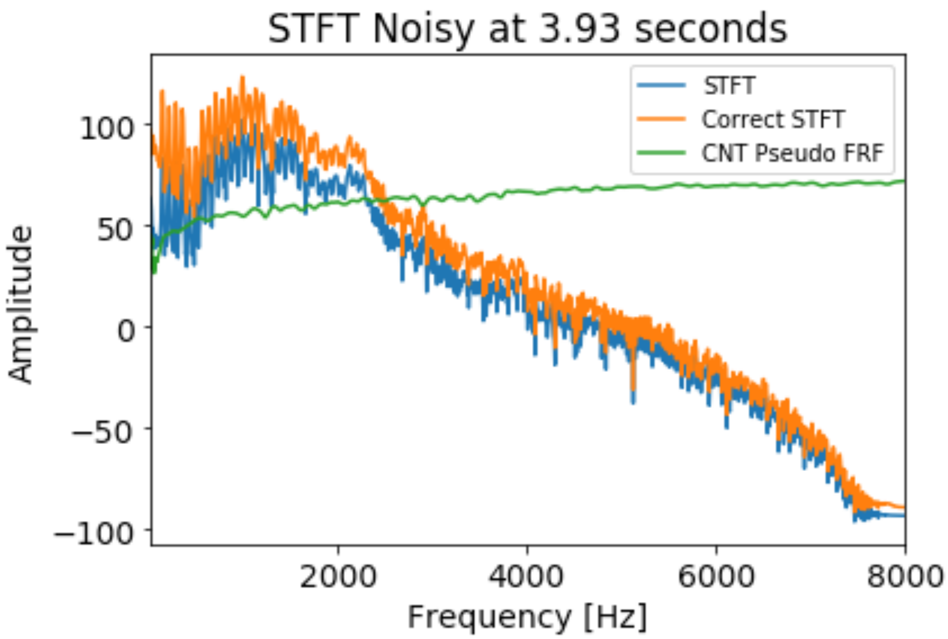


Figure 5-4: Example STFT spectra showing how the noisy signal was modified by the pseudo FRF of the CNT loudspeaker.

An example THDIS processing is shown in Figure 5-5. Figure 5-5 a shows the detected fundamental frequencies in the clean audio file versus time. Note: the drop outs in the data are when the threshold criteria were not met and the THDSI algorithm output NaN. Figure 5-5 b and c show the THDSI and THDNSI results noting that the values are very high relative to the traditional THD metric at upwards of 3000% and 4000%+ for THDSI and THDNSI, respectively. These levels were significantly worse prior to adjusting the noisy signal by the CNT FRF. The main reason these percentages are so elevated relative to the traditional THD metric is that the CNT output at the fundamental frequency (~100Hz in this example) is very low. Therefore, when computing THDSI the denominator is small. With that said, the main use of this metric will be to make *relative* comparisons of files, so the *absolute* high levels are assumed to be tolerable. From these data, the max and average over the whole time period can be computed as the single metrics to use in the regression. Note: Only the left ear data were processed with THDSI.

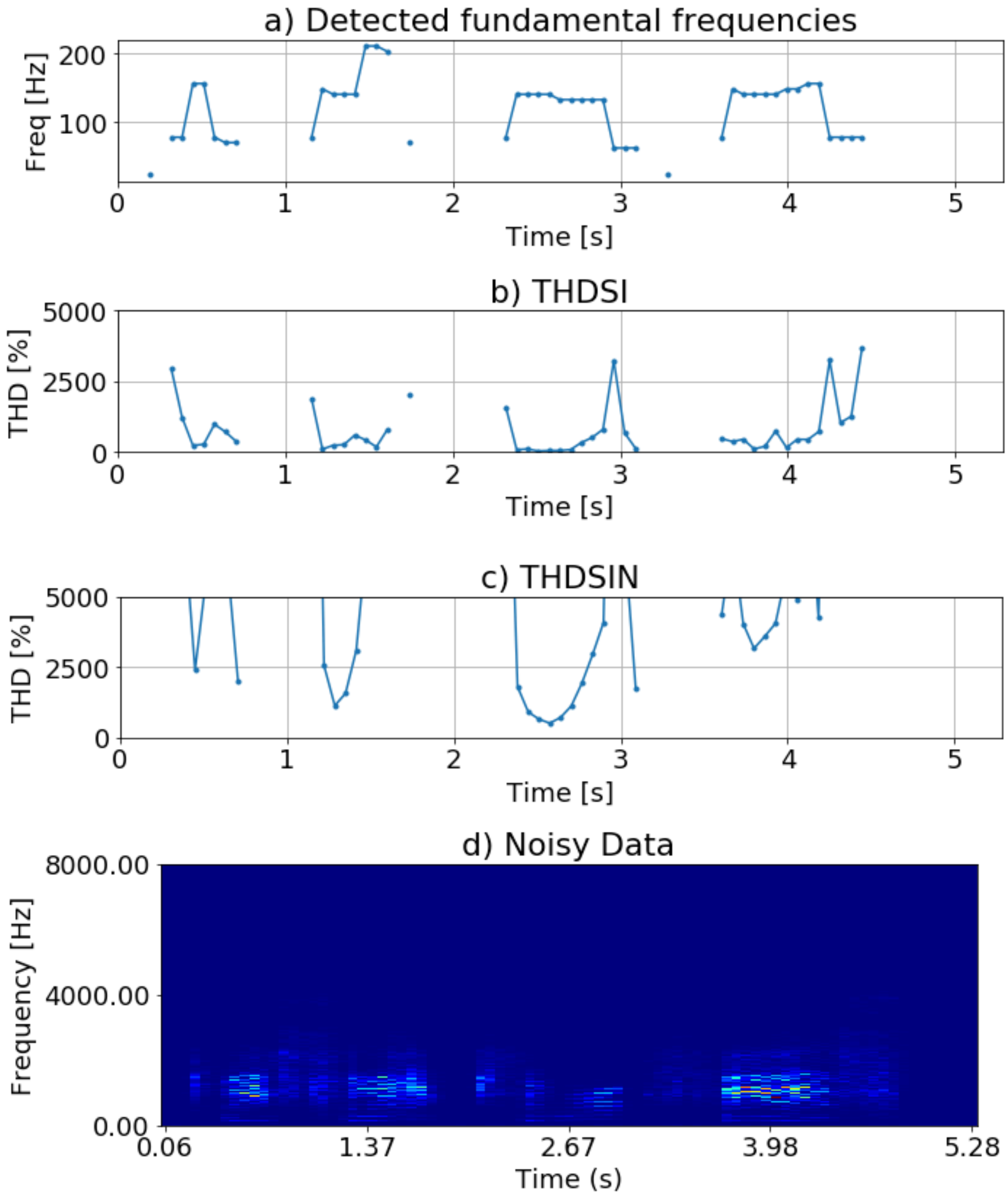


Figure 5-5: Example THDSI output for file AM_F1_b25_w3.wav (The word is fizz). a) shows the detected fundamental frequency in the clean signal (F1_b25_w3.wav) b) shows the computed THDSI result c) shows the computed THDNSI result and d) shows the spectrogram of the noisy signal.

5.3.3.3 Logistic Regression models of jurors

As mentioned in the introduction, linear regression is used when the dependent variable is not categorical (i.e. it can be any number). Logistic regression is used when the dependent variable is categorical (e.g. True/False, 5 point scale, etc..). With the percent correct main dependent variable requiring linear regression, the expected data flow after metric computation is shown in Figure 5-6.

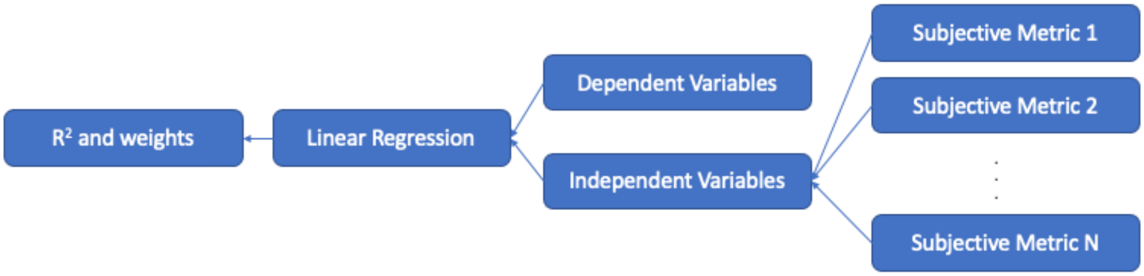


Figure 5-6: General data flow where each "psychoacoustic metric" is a common psychoacoustic metric or a custom metric like THDSI.

Another custom metric idea the author had was to use 80% of the subject results (i.e. the training data) to develop separate logistic regression models for each juror. Then use the remaining 20% (i.e. the test data) to determine the performance of the models and compute a "Predicted Percent Correct" metric. The general idea is laid out in Figure 5-7. The predicted percent correct metric would be another independent variable going into the final regression (Figure 5-6).

The parameters used for the logistic regression were the "liblinear" solver, an inverse of regulation strength of 10,000,000, and the intercept was forced to zero. These parameters were chosen by trial and error.

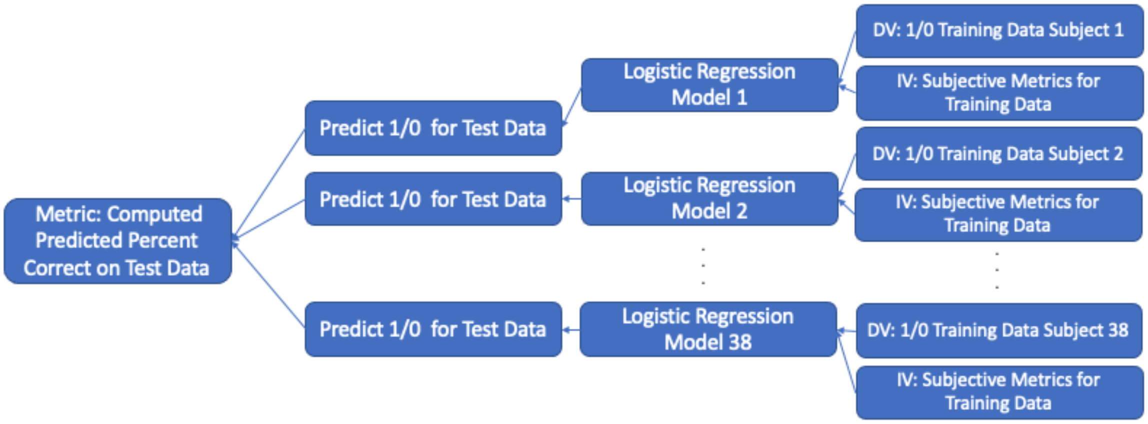

DV: 1/0 Training Data Subject 1
IV: Subjective Metrics for Training Data
Logistic Regression Model 1
Predict 1/0 for Test Data
DV: 1/0 Training Data Subject 2
IV: Subjective Metrics for Training Data
Metric: Computed Predicted Percent Correct on Test Data
Predict 1/0 for Test Data
Logistic Regression Model 2
Predict 1/0 for Test Data
Logistic Regression Model 38
DV: 1/0 Training Data Subject 38
IV: Subjective Metrics for Training Data

Figure 5-7: General flow of data to create the predicted percent correct metric that would then be fed in as one of the independent variables in Figure 5-6. DV and IV represent the Dependent Variables and the Independent Variables used in the logistic regression models.

5.3.3.4 Google Speech Recognizer

The final idea the author had for a custom metric was to use the output of a modern speech recognition machine learning algorithm as an independent variable. The 225 complete phrase files were down sampled to the recommended 16kHz sample rate and converted to linear 16 WAV format. They were then uploaded to the speech recognition algorithm. In this case it was Google's "command_and_search" model in March of 2021. The output was the predicted phrase spoken and a confidence percent. The predicted word in the phrase was then matched to the actual word spoken to determine a percent correct metric that would be used as an independent variable in the final regression.

5.3.4 Linear regression

Following the methodology shown in Figure 5-1, the first step before performing the regression is to remove the highly correlated independent variables (i.e. psychoacoustic metrics) using the correlation matrix, factor analysis (FA), and principle component analysis (PCA). In summary, the difference between FA and PCA is that FA forms a model of theoretical latent "factors" that predict the independent variables and PCA reduces the independent variables to a smaller set of orthogonal "components". Putting that another way, PCA assumes no other information exists that could cause variation within the independent variables while FA does not. Typically, both methods show similar results. For this effort, both were used together and both were used with the criteria of explaining 70% of the variation when determining the number of factors and components from FA and PCA, respectively.

Once the correlation matrix was computed, the absolute value was plotted in a colormap (Figure 5-8). From there, a new colormap was generated with a threshold level of 0.5 where values less than 0.5 where set to 0 and values greater than 0.5 were set to 1 (Figure 5-9). From this colormap, the list of which variables needed to be reduced was decided. Putting it another way, if there was a white cell in Figure 5-9 then a decision had to be made on which metric to not include in the regression analysis.

In order to help determine which correlated metric should be kept for the model, the FA factors and the PCA components were examined and the metric with the highest factor loading or component weighting was used. For example, in Figure 5-9 it is shown that Max THD, Avg THD, Max THDN, Avg THDN, Max S/N, Avg S/N, and Avg H2 were all correlated. This makes intuitive sense because they are all trying to represent the amount of harmonic distortion and noise. Looking at the FA loadings from these metrics in Table 5-3 factor "1" and Table 5-4 component "0," Avg THDN had the highest loadings/weighting at 0.94 and 0.21, respectively. If there was not a tie in the FA loadings, like this case, the metric with the highest value in the FA loadings was kept and the other correlated metrics were removed from the metric set. If there was a tie for the highest loading then the author would use the highest component weighting to determine which metric to keep. If there was a tie in both the FA loadings and PCA components the author would randomly select one of those two as the metric to keep and all others would be removed. Completing this exercise on all of the metrics would result in a correlation matrix as shown in Figure 5-10, where the remaining metric set is uncorrelated and can be used in regression.

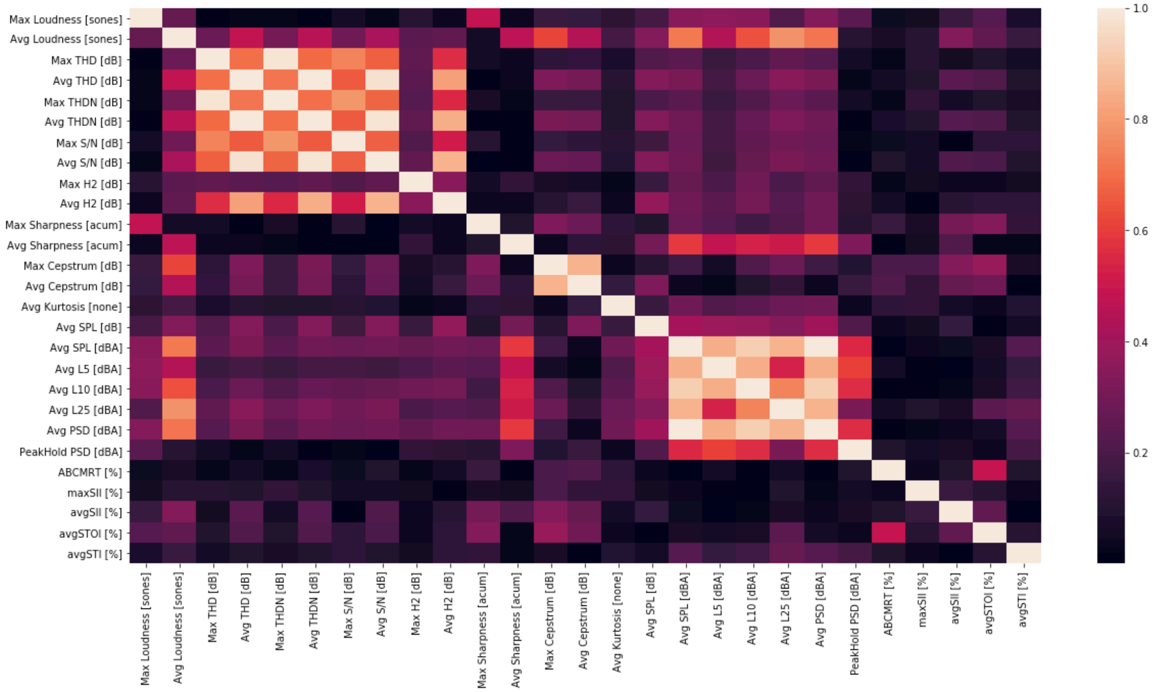


Figure 5-8: An example correlation matrix for the phrase files (I.e. the word is X) against the common metrics.

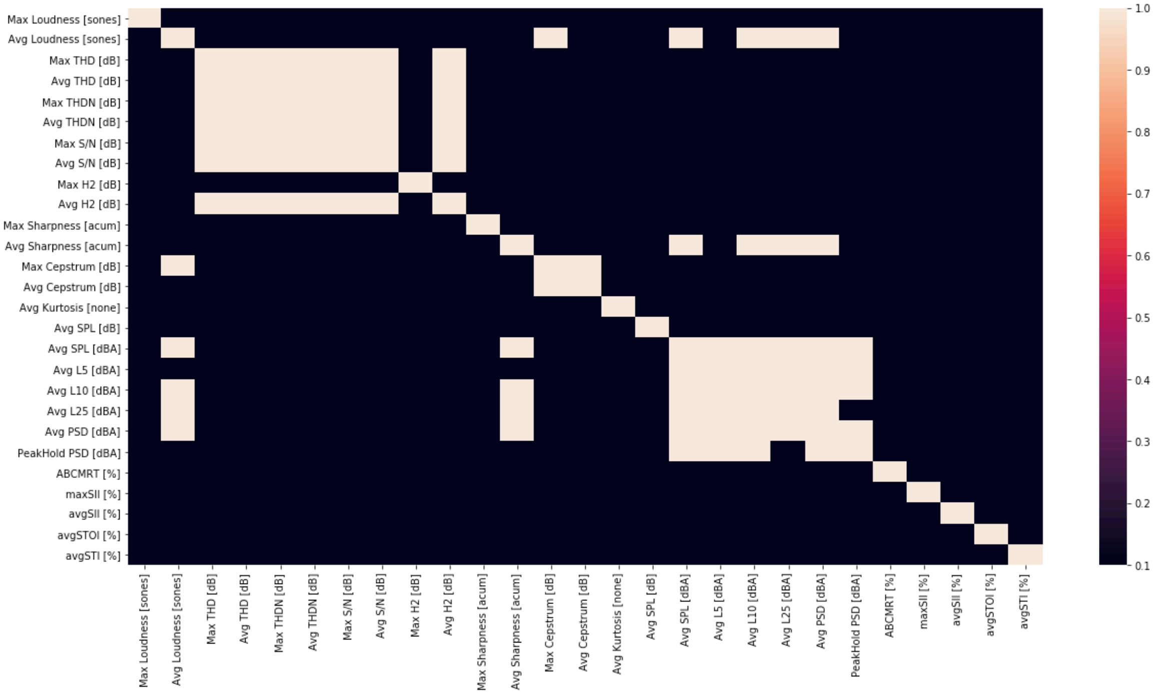
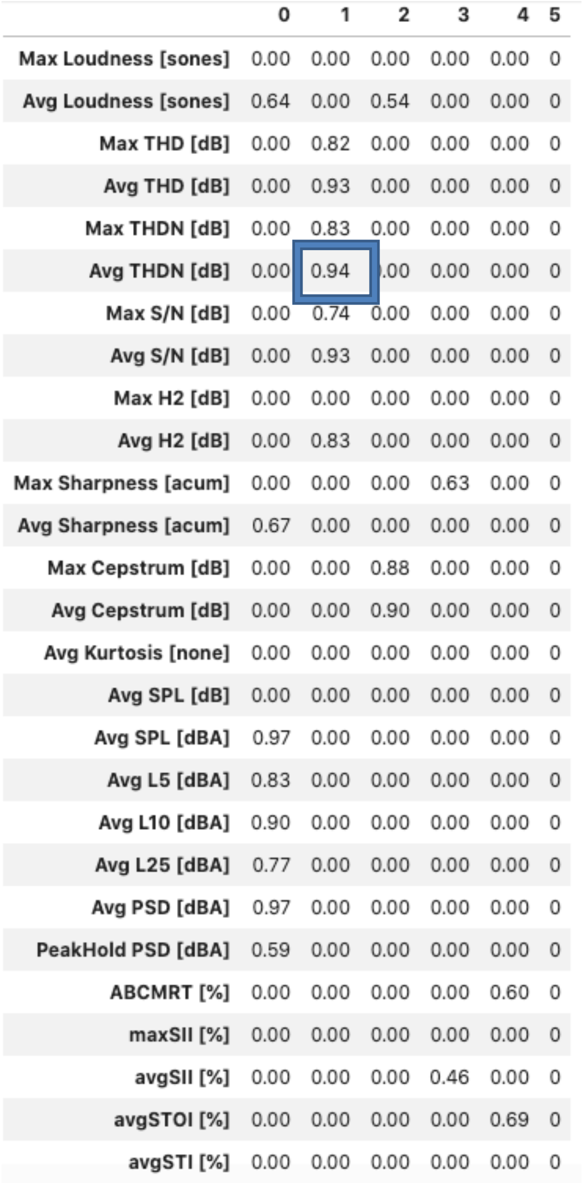


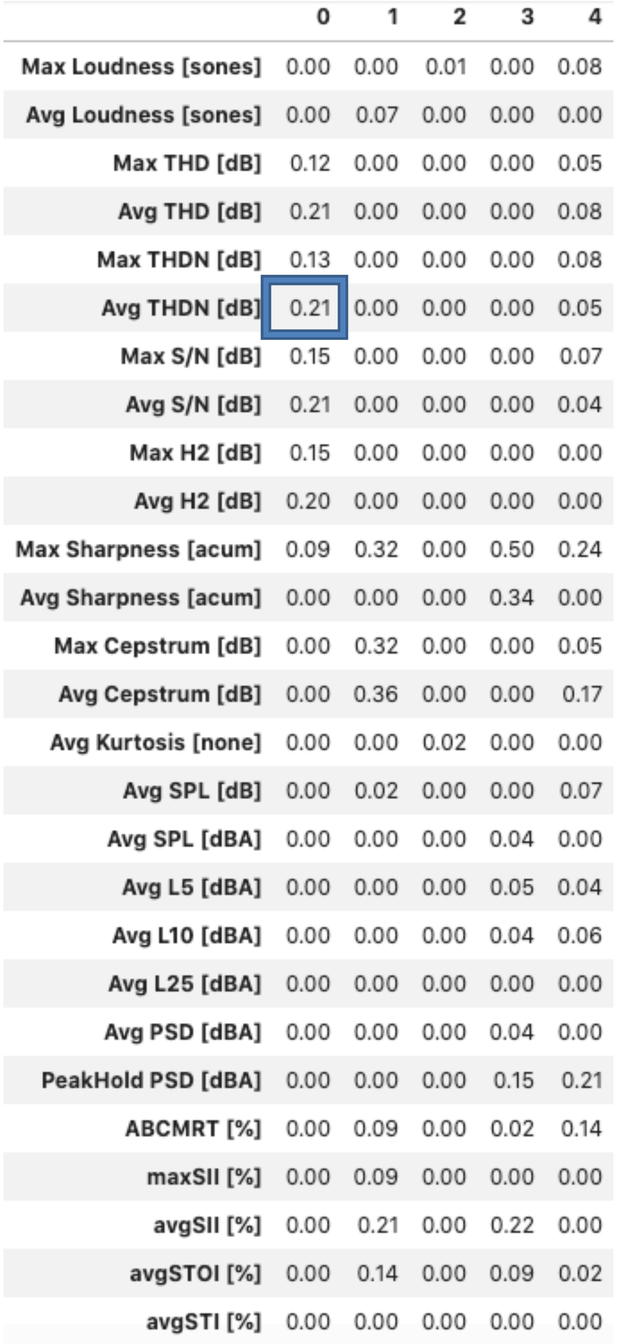
Figure 5-9: An example showing how the threshold of 0.5 was applied to the correlation matrix in Figure 5-8. Any cells where there is white means that metric was correlated with another metric and a decision had to be made about which metric to not include moving forward. Note: The colormap is symmetric and could be plotted as an upper or lower triangular matrix.

Table 5-3: Example factor analysis loadings table for the phrase files against the common metrics. The blue box is highlighting the cell referenced in the text example. Note: All values less than 0.4 were set to 0 to make the table more readable.



	0	1	2	3	4	5
Max Loudness [sones]	0.00	0.00	0.00	0.00	0.00	0
Avg Loudness [sones]	0.64	0.00	0.54	0.00	0.00	0
Max THD [dB]	0.00	0.82	0.00	0.00	0.00	0
Avg THD [dB]	0.00	0.93	0.00	0.00	0.00	0
Max THDN [dB]	0.00	0.83	0.00	0.00	0.00	0
Avg THDN [dB]	0.00	0.94	.00	0.00	0.00	0
Max S/N [dB]	0.00	0.74	0.00	0.00	0.00	0
Avg S/N [dB]	0.00	0.93	0.00	0.00	0.00	0
Max H2 [dB]	0.00	0.00	0.00	0.00	0.00	0
Avg H2 [dB]	0.00	0.83	0.00	0.00	0.00	0
Max Sharpness [acum]	0.00	0.00	0.00	0.63	0.00	0
Avg Sharpness [acum]	0.67	0.00	0.00	0.00	0.00	0
Max Cepstrum [dB]	0.00	0.00	0.88	0.00	0.00	0
Avg Cepstrum [dB]	0.00	0.00	0.90	0.00	0.00	0
Avg Kurtosis [none]	0.00	0.00	0.00	0.00	0.00	0
Avg SPL [dB]	0.00	0.00	0.00	0.00	0.00	0
Avg SPL [dBA]	0.97	0.00	0.00	0.00	0.00	0
Avg L5 [dBA]	0.83	0.00	0.00	0.00	0.00	0
Avg L10 [dBA]	0.90	0.00	0.00	0.00	0.00	0
Avg L25 [dBA]	0.77	0.00	0.00	0.00	0.00	0
Avg PSD [dBA]	0.97	0.00	0.00	0.00	0.00	0
PeakHold PSD [dBA]	0.59	0.00	0.00	0.00	0.00	0
ABCMRT [%]	0.00	0.00	0.00	0.00	0.60	0
maxSII [%]	0.00	0.00	0.00	0.00	0.00	0
avgSII [%]	0.00	0.00	0.00	0.46	0.00	0
avgSTOI [%]	0.00	0.00	0.00	0.00	0.69	0
avgSTI [%]	0.00	0.00	0.00	0.00	0.00	0

Table 5-4: An example showing the output from PCA using the same inputs as the FA in Table 5-3. The blue box is highlighting the cell referenced in the text example. Note: PCA only needed 5 components (columns) to represent 70% of the total variation whereas FA needed 6 factors.



	0	1	2	3	4
Max Loudness [sones]	0.00	0.00	0.01	0.00	0.08
Avg Loudness [sones]	0.00	0.07	0.00	0.00	0.00
Max THD [dB]	0.12	0.00	0.00	0.00	0.05
Avg THD [dB]	0.21	0.00	0.00	0.00	0.08
Max THDN [dB]	0.13	0.00	0.00	0.00	0.08
Avg THDN [dB]	0.21	0.00	0.00	0.00	0.05
Max S/N [dB]	0.15	0.00	0.00	0.00	0.07
Avg S/N [dB]	0.21	0.00	0.00	0.00	0.04
Max H2 [dB]	0.15	0.00	0.00	0.00	0.00
Avg H2 [dB]	0.20	0.00	0.00	0.00	0.00
Max Sharpness [acum]	0.09	0.32	0.00	0.50	0.24
Avg Sharpness [acum]	0.00	0.00	0.00	0.34	0.00
Max Cepstrum [dB]	0.00	0.32	0.00	0.00	0.05
Avg Cepstrum [dB]	0.00	0.36	0.00	0.00	0.17
Avg Kurtosis [none]	0.00	0.00	0.02	0.00	0.00
Avg SPL [dB]	0.00	0.02	0.00	0.00	0.07
Avg SPL [dBA]	0.00	0.00	0.00	0.04	0.00
Avg L5 [dBA]	0.00	0.00	0.00	0.05	0.04
Avg L10 [dBA]	0.00	0.00	0.00	0.04	0.06
Avg L25 [dBA]	0.00	0.00	0.00	0.00	0.00
Avg PSD [dBA]	0.00	0.00	0.00	0.04	0.00
PeakHold PSD [dBA]	0.00	0.00	0.00	0.15	0.21
ABCMRT [%]	0.00	0.09	0.00	0.02	0.14
maxSII [%]	0.00	0.09	0.00	0.00	0.00
avgSII [%]	0.00	0.21	0.00	0.22	0.00
avgSTOI [%]	0.00	0.14	0.00	0.09	0.02
avgSTI [%]	0.00	0.00	0.00	0.00	0.00

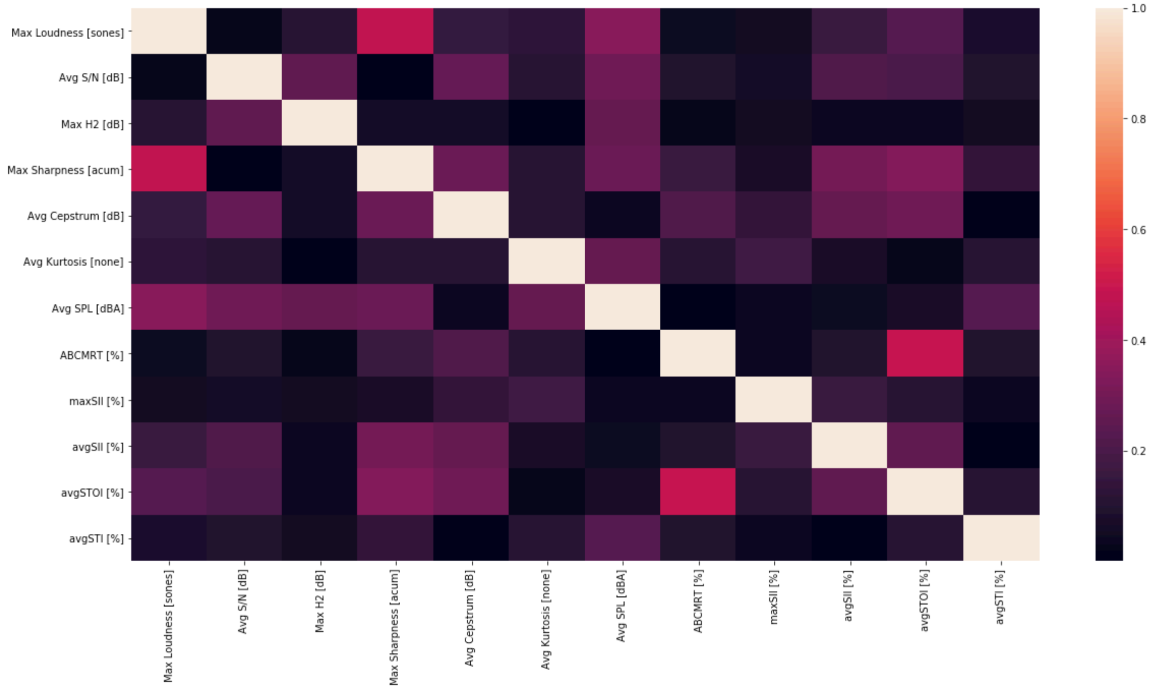


Figure 5-10: An example correlation matrix after the correlated metrics were removed using FA and PCA.

In this effort both single and multiple linear regression were used. Single ordinary least squares linear regression was used initially to understand how each individual metric correlated to the dependent variable prior to FA/PCA analysis. While performing multiple linear regression, 14 different variations of linear regression models were tested (

Table 5-5). The standard ordinary least squares (OLS) regression resulted in the highest correlation so was therefore used to determine the optimal set of metrics. Nonlinear regression was investigated, but found to perform worse than linear so it was not used.

To determine the optimal set of metrics from all of the non-colinear independent variables, all possible permutations of the metric set were regressed against the percent correct dependent variable and the metric set that had the lowest adjusted R^2 was selected. Adjusted R^2 is defined as,

$$adjR^2 = 1 - \frac{(1 - R^2)(n - 1)}{n - k - 1} \qquad \text{EQ5-4}$$

where R^2 (%) is the coefficient of determination, n is the number of points in the data, and K is the number of independent variables. Adjusted R^2 was used because all permutations were compared from k = 2 to k = total number of non-collinear independent variables.

Table 5-5: Linear models used in the study

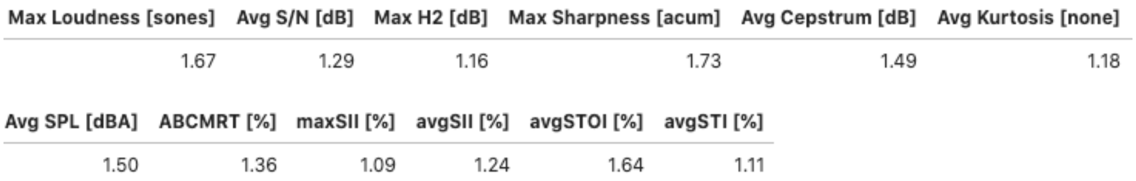
Regression Model Name	**Comments**
Ordinary Least Squares	Minimize residual sum of squares
Ridge Regression	With built in leave one out cross validation
SGD Regressor	Minimize a regularized empirical loss with stochastic gradient descent
Elastic Net Model	Iterative fitting along a regularization path
Lars	Least angle regression
Lasso	Linear model with L1 prior as regularizer
Lasso-Lars	Lasso model fit with lars
Orthogonal Matching Pursuit	
Bayesian ARD Regression	
Bayesian Ridge Regression	
RANSAC Regression	Random sample consensus
Theil-Sen Regressor	Robust multivariate regression model
Passive Aggressive Regressor	
Epsilon-Support Vector Regression	With free parameters C and epsilon
Partial Least Squares	Transformer and regressor

As a quality check step Variation inflation factor (VIF) analysis was performed to confirm no multicollinearity existed for the metrics entering regression. VIF is,

$$VIF = \frac{1}{1 - R^2} \qquad \text{EQ5-5}$$

where R^2 is the coefficient of determination between the single metric in question and all other metrics in the analysis. An example is shown in Table 5-6. VIF values were required to be below 3, but typically were below 2. Putting the 3 requirement another way, the R^2 between each independent variable and all other independent variables going into regression had to be less than 2/3 (~0.66).

Table 5-6: An example VIF analysis output.



Max Loudness [sones]	Avg S/N [dB]	Max H2 [dB]	Max Sharpness [acum]	Avg Cepstrum [dB]	Avg Kurtosis [none]
1.67	1.29	1.16	1.73	1.49	1.18

Avg SPL [dBA]	ABCMRT [%]	maxSII [%]	avgSII [%]	avgSTOI [%]	avgSTI [%]
1.50	1.36	1.09	1.24	1.64	1.11

Once the optimal metric set was regressed, the T-Statistic p-value was computed. If the p-value was less than $\alpha = 0.05$ that meant the independent variable was worth keeping (i.e. that null hypothesis that the independent variable does not correlate to the dependent variable could be rejected). The metric set was then further refined to only include metrics that had a small p-value. That metric set became the final model.

Using the final metric set, the coefficient of determination R^2 and the adjusted R^2 (EQ5-4) were computed as the final fit of the independent variables to the dependent. The weights for each metric were also computed. The weights could be applied in EQ5-2 when testing future drive signal processing methods.

5.4 Results

5.4.1 Logistic regression

The analysis described in section 5.3.3.3 unfortunately resulted in a low coefficient of determination, $R^2 = 0.023$ (Figure 5-11). Additionally, the result seemed to predict many values of ~84 percent correct (i.e. the vertical line of dots in Figure 5-11). The cause of this was unknown. Therefore, the predicted percent correct metric was not used as an independent variable moving forward.

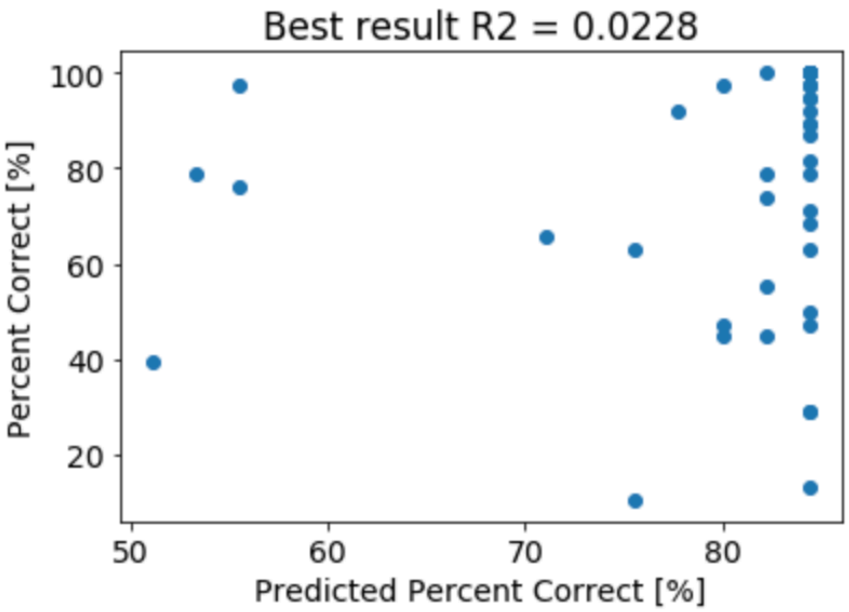
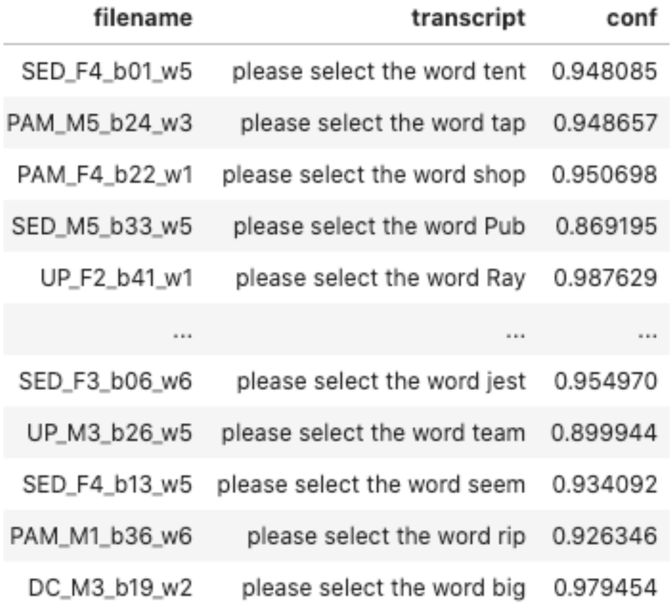


Figure 5-11: Scatter plot showing the lack of correlation for the computed independent variable "Predicted Percent Correct" to the dependent variable Percent Correct.

5.4.2 Google speech recognizer

The analysis described in 5.3.3.4 was initially conducted on the clean original phrase files. With these files, the algorithm correctly guessed the word 80% of the time. An example of the output is shown in Table 5-7.

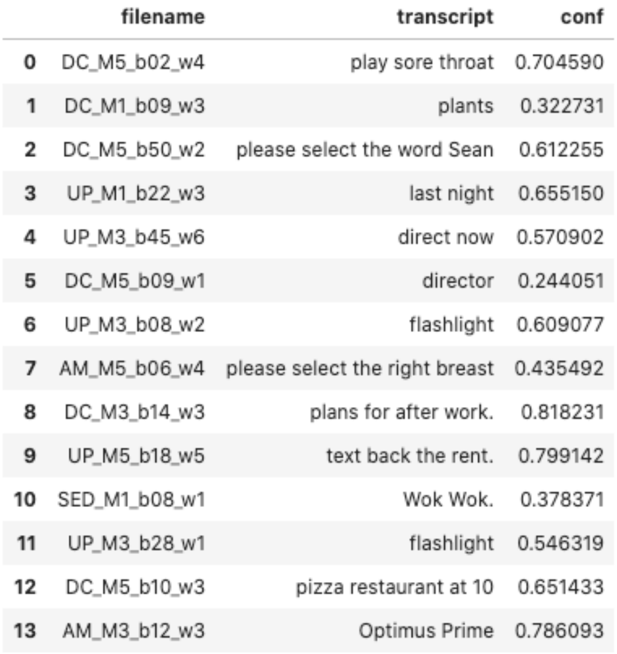
Table 5-7: Example output from the Google speech recognition model using the clean phrase files.



filename	transcript	conf
SED_F4_b01_w5	please select the word tent	0.948085
PAM_M5_b24_w3	please select the word tap	0.948657
PAM_F4_b22_w1	please select the word shop	0.950698
SED_M5_b33_w5	please select the word Pub	0.869195
UP_F2_b41_w1	please select the word Ray	0.987629
...	...	...
SED_F3_b06_w6	please select the word jest	0.954970
UP_M3_b26_w5	please select the word team	0.899944
SED_F4_b13_w5	please select the word seem	0.934092
PAM_M1_b36_w6	please select the word rip	0.926346
DC_M3_b19_w2	please select the word big	0.979454

The same analysis was then performed on phrase and single word files. An example output is shown in Table 5-8. There were only 13 of the 225 files that even resulted in a predicted transcript for the phrase files. Of those 13 only two had the phrase "Please select the..", but unfortunately neither of those two were correct resulting in a 0% correct score. Since the model input is 10ms cepstrum values, it is hypothesized that the noise in the signals would drastically effect the performance. Recall from Chapter 4, that the DC offset method (45 of 225 files) had a 97.2% correct score. So while the noise did not affect the jury subjects, it was too much for the speech recognition algorithm at this time. Therefore it was not used as an independent variable in the regression.

Table 5-8: The complete output from the Google speech recognition model using the noisy phrase files.



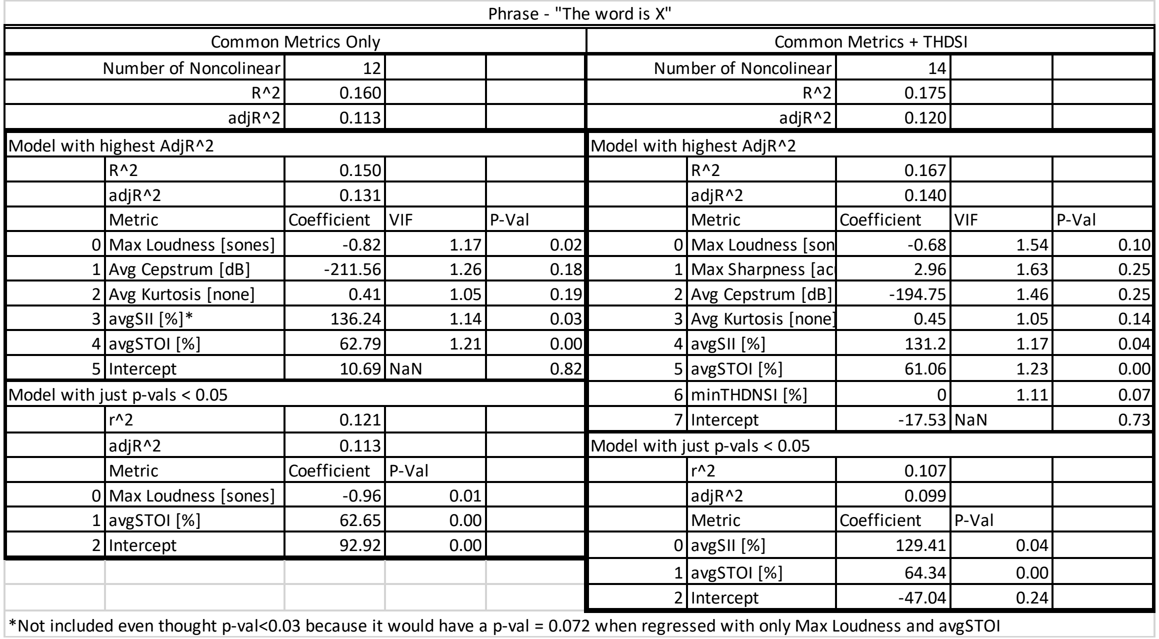
	filename	transcript	conf
0	DC_M5_b02_w4	play sore throat	0.704590
1	DC_M1_b09_w3	plants	0.322731
2	DC_M5_b50_w2	please select the word Sean	0.612255
3	UP_M1_b22_w3	last night	0.655150
4	UP_M3_b45_w6	direct now	0.570902
5	DC_M5_b09_w1	director	0.244051
6	UP_M3_b08_w2	flashlight	0.609077
7	AM_M5_b06_w4	please select the right breast	0.435492
8	DC_M3_b14_w3	plans for after work.	0.818231
9	UP_M5_b18_w5	text back the rent.	0.799142
10	SED_M1_b08_w1	Wok Wok.	0.378371
11	UP_M3_b28_w1	flashlight	0.546319
12	DC_M5_b10_w3	pizza restaurant at 10	0.651433
13	AM_M3_b12_w3	Optimus Prime	0.786093

5.4.3 Multiple linear regression

Initially the phrase files (i.e. the word is “X”) were processed. The analysis was done including THDSI metrics and not including them (Table 5-9). From the complete original metric set, the common metrics were reduced to 12. The results are shown in the left half of Table 5-9. The R^2 and adjR^2 using all 12 of these metrics was 0.16 and 0.11, respectively. As described in section 5.3.4, all permutations of the 12 metrics were tested to determine the metric set (i.e. model) with the highest adjR^2. This resulted in an R^2 and adjR^2 of 0.15 and 0.13, respectively for a model including max loudness, average cepstrum, average kurtosis, average SII, and STOI. A t-test was computed for all of these independent variables and the resulting p-value was used to determine if the metric should be kept (i.e. if there was significant correlation between the metric and the dependent variable). This analysis showed that average cepstrum and average kurtosis had to be removed. A new model was fit using max loudness, average SII, and average STOI. The p-values were again computed and average SII was now 0.07 which was up from 0.03 in the initial model. Therefore, average SII was removed and the metric set for the final model was determined to be max loudness and average STOI. The resulting R^2 was 0.12, the weights were -0.96 and 62.65 for max loudness and STOI, respectively. The intercept was 92.92.

The same process was repeated including using the THDSI metrics and is shown on the right half of Table 5-9. The resulting model did not contain THDSI, but contained average STOI and average SII.

Table 5-9: Results from processing the phrase files



Phrase - "The word is X"									
Common Metrics Only					Common Metrics + THDSI				
	Number of Noncolinear	12				Number of Noncolinear	14		
	R^2	0.160				R^2	0.175		
	adjR^2	0.113				adjR^2	0.120		
Model with highest AdjR^2					Model with highest AdjR^2				
	R^2	0.150				R^2	0.167		
	adjR^2	0.131				adjR^2	0.140		
	Metric	Coefficient	VIF	P-Val		Metric	Coefficient	VIF	P-Val
0	Max Loudness [sones]	-0.82	1.17	0.02	0	Max Loudness [son	-0.68	1.54	0.10
1	Avg Cepstrum [dB]	-211.56	1.26	0.18	1	Max Sharpness [ac	2.96	1.63	0.25
2	Avg Kurtosis [none]	0.41	1.05	0.19	2	Avg Cepstrum [dB]	-194.75	1.46	0.25
3	avgSII [%]*	136.24	1.14	0.03	3	Avg Kurtosis [none]	0.45	1.05	0.14
4	avgSTOI [%]	62.79	1.21	0.00	4	avgSII [%]	131.2	1.17	0.04
5	Intercept	10.69	NaN	0.82	5	avgSTOI [%]	61.06	1.23	0.00
Model with just p-vals < 0.05					6	minTHDNSI [%]	0	1.11	0.07
	r^2	0.121			7	Intercept	-17.53	NaN	0.73
	adjR^2	0.113			Model with just p-vals < 0.05				
	Metric	Coefficient	P-Val			r^2	0.107		
0	Max Loudness [sones]	-0.96	0.01			adjR^2	0.099		
1	avgSTOI [%]	62.65	0.00			Metric	Coefficient	P-Val	
2	Intercept	92.92	0.00		0	avgSII [%]	129.41	0.04	
					1	avgSTOI [%]	64.34	0.00	
					2	Intercept	-47.04	0.24	
*Not included even thought p-val<0.03 because it would have a p-val = 0.072 when regressed with only Max Loudness and avgSTOI									

The process was then computed for the single word files with the thought being that the jury study subject was only tasked with understanding the word “X” so any metric data computed for the beginning of the phrase “the word is…” would likely not correlate. For example, when max loudness was computed the loudness versus time array may have had a peak at a time not when the word was said. Therefore, all of the metrics were computed over the very short duration single word clips. The results are shown in Table 5-10.

The common metrics, not including THDSI metrics, resulted in a final model containing max SII and average STOI with an $R^2 = 0.082$. Including THDSI the resulting model had only average THDNSI as a metric with an $R^2 = 0.122$. The compiled adjusted R^2 values for all of the final models is shown in Table 5-11. The conclusion from this is that the best model to predict percent correct is the single word model which contains the single psychoacoustic metric average THDNSI.

Table 5-10: Results from processing the single word files.

Single Word- "X"									
Common Metrics Only					Common Metrics + THDSI				
Number of Noncolinear		9			Number of Noncolinear		11		
R^2		0.107			R^2		0.179		
adjR^2		0.695			adjR^2		0.136		
Model with highest AdjR^2					Model with highest AdjR^2				
	R^2	0.100				R^2	0.177		
	adjR^2	0.080				adjR^2	0.155		
	Metric	Coefficient	VIF	P-Val		Metric	Coefficient	VIF	P-Val
0	Avg L5 [dBA]	-0.37	1.14	0.28	0	Max H2 [dB]	0.61	1.02	0.25
1	maxSII [%]	455.04	1.05	0.047	1	maxSII [%]	355.30	1.03	0.10
2	avgSII [%]	56.09	1.13	0.10	2	avgSTOI [%]	23.41	1.14	0.01
3	avgSTOI [%]	36.56	1.05	0.00	3	avgSTI [%]	-19.50	1.03	0.09
4	avgSTI [%]	-13.41	1.02	0.26	4	minTHDSI [%]	0.02	1.21	0.09
5	Intercept	-355.66	NaN	0.09	5	avgTHDNSI [%]	-4.02E-04	1.33	0.00
Model with just p-vals < 0.05					6	Intercept	-228.10	NaN	0.25
	r^2	0.082			Model with just p-vals < 0.05				
	adjR^2	0.073				r^2	0.122		
	Metric	Coefficient	P-Val			adjR^2	0.118		
0	maxSII [%]	545.57	0.03			Metric	Coefficient	P-Val	
1	avgSTOI [%]	33.87	0.02		0	avgTHDNSI [%]	-4.09E-04	0.00	
2	Intercept	-437.01	0.00		1	Intercept	95.85	0.00	

Table 5-11: Comparison between the different file lengths showing the best performing model was the single word model that included average THDNSI.

Phrase	AdjR^2	Single Word	AdjR^2
No THDSI	0.113	No THDSI	0.073
THDSI	0.099	THDSI	0.118

The resulting best fit model equation is

$$Estimated\ Percent\ Correct = -0.000409 * avgTHDNSI + 95.85 \qquad \text{EQ 5-6}$$

where the Estimated Percent Correct is the model's prediction of the percent correct and avgTHDNSI is the computed average THDNSI (%) over the duration of only the single word (i.e. not the whole phrase). The resulting fit from this model is shown in Figure 5-12.

While the final coefficient of determination is low, there are many aspects of the final model that make intuitive sense. For example, the slope of the model line (Figure 5-12) is negative. This means that as THDNSI increases, the estimated percent correct decreases, which follows intuition. Additionally, the y intercept is 95.85 which is close to 100%. It follows intuition that the model would have a y intercept of near 100% meaning that when the model is input 0% THDNSI the estimated percent correct should be near 100%.

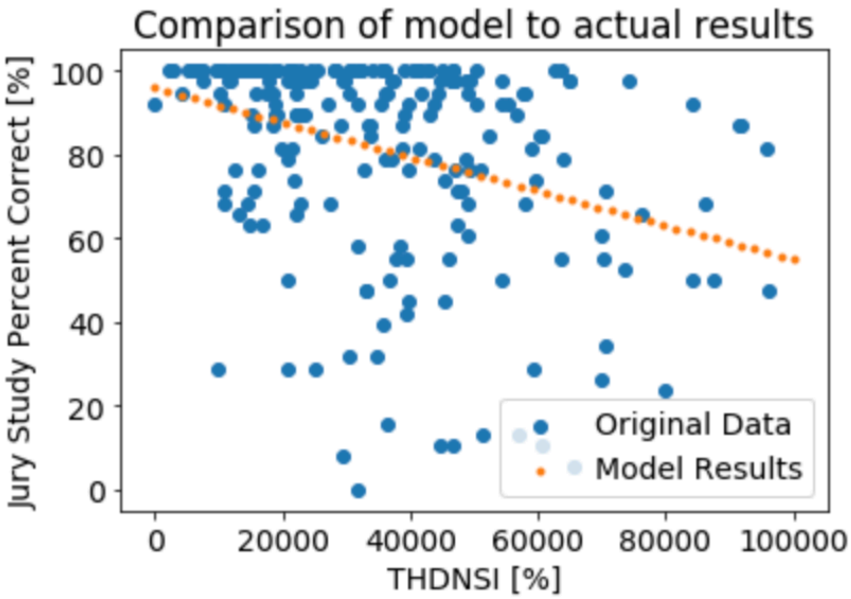


Figure 5-12: Demonstration of the final model fit over the jury study data.

5.5 Conclusions

In conclusion, a correlation between the subjective results of the jury study in Chapter 4 and a variety of psychoacoustic metrics was undertaken. The psychoacoustic metrics included common sound quality metrics like loudness, but also included a new custom metric titled Total Harmonic Distortion for Speech Intelligibility THDSI. Additionally, unique novel metrics like developing logistic regression models for each jury study subject and using modern speech recognition algorithms were investigated. Unfortunately, all of this effort only led to a model with a coefficient of determination (R^2) of 0.122. With jury studies it can be difficult to get correlation and that was definitely the case in this effort. Moving forward, alternative machine learning algorithms could be utilized to help provide additional independent variables to investigate.

5.6 Acknowledgements

The author would like to thank Head Acoustics for use of their Artemis software to calculate many of the metrics used in this effort.

6 Conclusions and recommended future work

6.1 Conclusions

In order to determine if a new technology is applicable for any given application it requires certain knowledge about that technology. For example, one can't determine if the hammer will work for nailing without knowing if the hammer is too heavy to lift. Whenever a new technology is discovered a myriad of questions arise about it. The development for CNT loudspeakers is no different.

A lot has changed in the community's understanding of carbon nanotube loudspeakers since the start of this effort in fall of 2014. In addition to the results presented in this document, significant improvements have been made by others in the durability [71], modeling [12], [40]–[47], and application spaces for these transducers [30]–[36]. The main contributions of this document were quantification of the true efficiency for various signal processing methods, discovery of new signal processing methods that can be used with class D amplifiers, and subjective data outlining which signal processing methods are the most intelligible.

Combing the results from Chapters 2 & 3, a summary of the efficiency and THD findings are presented in Table 6-1. From this data the main conclusion that can be made is that regardless of drive signal processing method, the efficiency is on the same order of magnitude (i.e. no drive signal method significantly increases the efficiency). This efficiency follows the analytic prediction for an open CNT thermophone (Figure 2-5) Putting it another way, for an open CNT loudspeaker, the efficiency will always be on the E-6% efficiency order of magnitude (frequencies < 300Hz). This is drastically different than the ~E-2% efficiency of a moving coil loudspeaker (Table 2-3), which is roughly four orders of magnitude more efficient.

While these transducers are, by their physical nature, less efficient that does not mean their efficiency cannot be improved. For example, tuning the resonance of a CNT loudspeaker in an enclosure would cause that system to have a higher output than an open CNT loudspeaker therefore improving the efficiency.

It is also interesting to note from Table 6-1 that the THD of the FCAC (aka SED) method is significantly lower than the other methods, but as mentioned in Chapter 3 this is largely due to that method, as well as TCAC, being able to be optimized for single sine wave signals and not working well on complex signals. With that knowledge it was not surprising to then see SED be one of the worst performers in the jury study in Chapter 4. Aside from CntUP, DC offset had the highest THDs in Chapter 3, but performed the best in the jury study in Chapter 4. This suggests that THD is not a great metric to track sound quality for CNT drive signals for stationary signals. For the transient signals in Chapter 5, a similar conclusion was made that THD was not the ideal psychoacoustic metric which led to the development of THDSI as a better, but still not perfect, metric.

One important conclusion from the efficiency work is that there are multiple different drive signal processing methods that work to linearize the thermoacoustic frequency doubling. Each of these methods allow for use with different amplification hardware. CntUP/SED/TCAC can all be used with class D amplifiers. AM can be used with some class D amplifiers as well as radio frequency amplifiers, and DC can be used with class A/B amplifiers.

One surprising conclusion from the efficiency effort was that efficiency was not a function of carrier frequency (Figure 2-10). This was largely surprising because according to the analytical model [13] efficiency should increase with frequency. Therefore, if a higher carrier frequency is used the higher the efficiency was expected to be. This did not turn out to be the case. Other important conclusions were that the optimal B/A ratio for DC offset was 0.62 (Figure 2-7) and the optimal modulation index for AM was 1.5 (Figure 2-11).

Table 6-1: Summary of efficiency and THD results from Chapters 2 & 3. All data was for input power of ~72 W_{rms}

	Efficiency (μ%)	THD (%)
AC/CntUP	4.3 - 319	≈ ∞
DCAC/DC (B/A=0.62)	1.69 - 308	43 - 93
AMAC/AM	1.24 - 228	22 – 95
FCAC/SED	1.01 - 1083	0.68 - 59
TCAC	1.26 - 388	1.7 - 11

Summarizing the results from the jury study in Chapter 4, DC offset was the most intelligible drive signal processing method followed by amplitude modulation (AM) (Table 4-2). Additionally, it was found that executing a jury study outside the lab by way of a drive up jury study can lead to statistically significant results without the health risk of bringing tens of subjects into the lab during a pandemic.

Recalling the findings from the psychoacoustic metric correlation in Chapter 5, the main conclusion reached was that there are not any combination or common psychoacoustic metrics that correlated well to the jury study results in Chapter 4. A new metric, Total Harmonic Distortion for Speech Intelligibility (THDSI) was developed, but only resulted in a coefficient of determination (R^2) equal to 0.12. That model is shown in EQ 5-6.

6.2 Future work

6.2.1 Efficiency

Looking back at what has been done in regards to efficiency, there are two main areas to focus on moving forward. The first is a new method that was published by Torraca et al. [109]. The new method, adaptive predistortion (AP), uses a sliding FIFO buffer of historical data to compute a dynamic DC offset. It then adds that DC offset to the signal and modulates the summation at the nyquist. This allows for a drive signal processing method that can be used on a class D amplifier. In subjective testing in the lab, the sound quality was vastly superior to TCAC and FCAC (aka SED). Interestingly, the authors claim this method to also be significantly more efficient than other methods. However, they only use SPL measurements to confirm this and they also don't acquire data at frequencies up to the nyquist (i.e. where the modulation occurs). So this method should be investigated as the others were done in Chapters 2 & 3.

The second area to focus on is enclosure design. Different application opportunities are pushing this development, but at this time it is just that, application specific. A much broader study of enclosure efficiency should be undertaken so that the applications can start their designs based on what is learn from the optimal enclosure effort.

6.2.2 Drive signal processing

At this point, especially with the addition of AP as described in 6.2.1, there is not a significant need for drive signal development. Gains in the efficiency and durability arenas warrant more attention.

6.2.3 Sound quality

The intelligibility results of the jury study in Chapter 4 were a big step forward in discerning which drive signal processing method was most intelligible. Unfortunately, the resulting psychoacoustic metric correlation in Chapter 5 did not leave much promise that any future metrics, like AP, could be compared without executing a new jury study. Therefore, it is suggested that additional machine learning algorithms be tested to determine if a new model (i.e. independent variable) can be found that better correlates to the jury study results.

The main difficulty with developing modern machine learning algorithms is determining what data to feed it. Preliminary investigations show Convolutional Neural Network (CNN) models seem to perform the best for audio file correlation. Unfortunately it is not as easy as inputting the wav files into the models. While that can be done, the results are not as correlated as when meta data are used [110]. Audio recognition models (e.g. models that estimate genre) seem to be typically fed short time Fourier transform (aka spectrogram) data with varying construction of the various layers within the CNN. It is recommended to look into these methods, possibly feeding the model other metrics like THDSI and the others that showed some correlation in Chapter 5.

6.3 Recommendation applications

After spending nearly six and a half years with the CNT technology, I believe it is unlikely that CNT loudspeakers will become more popular than traditional moving coil loudspeakers primarily due to their inefficiency. The ideal application for CNT loudspeakers are applications that require low weight, small size, or custom directivity and have access to ample power.

7 Reference List

[1] F. Braun, "Notiz über Thermophonie," *Ann. Phys.*, vol. 301, no. 6, pp. 358–360, 1898, doi: 10.1002/andp.18983010609.

[2] H. D. Arnold and I. B. Crandall., "The Thermophone as a Precision Source of Sound," *Phys. Rev.*, vol. 10, no. 1, pp. 22–38, Jul. 1917, doi: 10.1103/PhysRev.10.22.

[3] X. Yu, R. Rajamani, K. A. Stelson, and T. Cui, "Carbon nanotube-based transparent thin film acoustic actuators and sensors," *Sensors and Actuators A: Physical*, vol. 132, no. 2, pp. 626–631, Nov. 2006, doi: 10.1016/j.sna.2006.02.045.

[4] A. R. Barnard *et al.*, "Advancements toward a high-power, carbon nanotube, thin-film loudspeaker," *Noise Control Engineering Journal*, vol. 62, no. 5, pp. 360–367, Sep. 2014, doi: 10.3397/1/376235.

[5] T. M. Bouman, A. R. Barnard, and M. Asgarisabet, "Experimental quantification of the true efficiency of carbon nanotube thin-film thermophones," *The Journal of the Acoustical Society of America*, vol. 139, no. 3, pp. 1353–1363, Mar. 2016, doi: 10.1121/1.4944688.

[6] T. Bouman, A. Barnard, and J. Alexander, "Continued Drive Signal Development for the Carbon Nanotube Thermoacoustic Loudspeaker Using Techniques Derived from the Hearing Aid Industry," Jun. 2017, pp. 2017-01–1895, doi: 10.4271/2017-01-1895.

[7] S. Iijima, "Helical microtubules of graphitic carbon," *Nature*, vol. 354, no. 6348, pp. 56–58, Nov. 1991, doi: 10.1038/354056a0.

[8] A. E. Aliev, M. D. Lima, S. Fang, and R. H. Baughman, "Underwater Sound Generation Using Carbon Nanotube Projectors," *Nano Lett.*, vol. 10, no. 7, pp. 2374–2380, Jul. 2010, doi: 10.1021/nl100235n.

[9] A. R. Barnard, D. M. Jenkins, T. A. Brungart, T. M. McDevitt, and B. L. Kline, "Feasibility of a high-powered carbon nanotube thin-film loudspeaker," *The Journal of the Acoustical Society of America*, vol. 134, no. 3, pp. EL276–EL281, Sep. 2013, doi: 10.1121/1.4817261.

[10] A. R. Barnard, T. A. Brungart, T. M. McDevitt, and D. M. Jenkins, "Background and development of a high-powered carbon nanotube thin-film loudspeaker," presented at the Inter Noise, New York City, New York, USA, Aug. 2014, Accessed: Sep. 29, 2014. [Online].

[11] L. Xiao *et al.*, “Flexible, Stretchable, Transparent Carbon Nanotube Thin Film Loudspeakers,” *Nano Lett.*, vol. 8, no. 12, pp. 4539–4545, Dec. 2008, doi: 10.1021/nl802750z.

[12] S. S. Asadzadeh, A. Moosavi, C. Huynh, and O. Saleki, “Thermo acoustic study of carbon nanotubes in near and far field: Theory, simulation, and experiment,” *Journal of Applied Physics*, vol. 117, no. 9, p. 095101, Mar. 2015, doi: 10.1063/1.4914049.

[13] L. Xiao *et al.*, “High frequency response of carbon nanotube thin film speaker in gases,” *Journal of Applied Physics*, vol. 110, no. 8, p. 084311, 2011, doi: 10.1063/1.3651374.

[14] A. E. Aliev, Y. N. Gartstein, and R. H. Baughman, “Increasing the efficiency of thermoacoustic carbon nanotube sound projectors,” *Nanotechnology*, vol. 24, no. 23, p. 235501, Jun. 2013, doi: 10.1088/0957-4484/24/23/235501.

[15] B. J. Mason, S.-W. Chang, J. Chen, S. B. Cronin, and A. W. Bushmaker, “Thermoacoustic Transduction in Individual Suspended Carbon Nanotubes,” *ACS Nano*, vol. 9, no. 5, pp. 5372–5376, May 2015, doi: 10.1021/acsnano.5b01119.

[16] M. E. Kozlov, C. S. Haines, J. Oh, M. D. Lima, and S. Fang, “Sound of carbon nanotube assemblies,” *Journal of Applied Physics*, vol. 106, no. 12, p. 124311, 2009, doi: 10.1063/1.3272691.

[17] J.-H. Pöhls *et al.*, “Physical properties of carbon nanotube sheets drawn from nanotube arrays,” *Carbon*, vol. 50, no. 11, pp. 4175–4183, Sep. 2012, doi: 10.1016/j.carbon.2012.04.067.

[18] M. B. Jakubinek *et al.*, “Thermal and electrical conductivity of tall, vertically aligned carbon nanotube arrays,” *Carbon*, vol. 48, no. 13, pp. 3947–3952, Nov. 2010, doi: 10.1016/j.carbon.2010.06.063.

[19] “Acoustics - Determination of sound power levels and sound power levels of noise sources using sound pressure - Engineering methods for an essentially free field over a reflecting plane,”,” American National Standards Institute, Standard ANSI S12.54, 2011.

[20] S. Garrett, “Two-Way Loudspeaker Enclosure Assembly and Testing as A Freshmen Seminar,” presented at the Proceedings of Congress On Sound & Vibration 17, Cairo, Egypt, 2010.

[21] L. H. Tong, C. W. Lim, and Y. C. Li, “Gas-Filled Encapsulated Thermal-Acoustic Transducer,” *Journal of Vibration and Acoustics*, vol. 135, no. 5, p. 051033, Oct. 2013, doi: 10.1115/1.4024765.

[22] P. Jia and C. W. Lim, “Thermal-Acoustic Wave Generation and Propagation Using Suspended Carbon Nanotube Thin Film in Fluidic Environments,” *Journal of Applied Mechanics*, vol. 83, no. 9, p. 091007, Sep. 2016, doi: 10.1115/1.4033893.

[23] V. Leibman, “Frequency transposing hearing aid,” U.S. Patent 5,014,319, May 07, 1991.

[24] J. M. Alexander, J. G. Kopun, and P. G. Stelmachowicz, “Effects of Frequency Compression and Frequency Transposition on Fricative and Affricate Perception in Listeners With Normal Hearing and Mild to Moderate Hearing Loss,” *Ear & Hearing*, vol. 35, no. 5, pp. 519–532, Sep. 2014, doi: 10.1097/AUD.0000000000000040.

[25] A. E. Aliev *et al.*, “Alternative Nanostructures for Thermophones,” *ACS Nano*, vol. 9, no. 5, pp. 4743–4756, May 2015, doi: 10.1021/nn507117a.

[26] M. Daschewski, M. Kreutzbruck, and J. Prager, “Influence of thermodynamic properties of a thermo-acoustic emitter on the efficiency of thermal airborne ultrasound generation,” *Ultrasonics*, vol. 63, pp. 16–22, Dec. 2015, doi: 10.1016/j.ultras.2015.06.008.

[27] T. Bouman, “Drive signal development for the thermacoustic loudspeaker,” Master of Science in Mechanical Engineering, Michigan Technological University, Houghton, Michigan, 2016.

[28] S. A. Romanov, A. E. Aliev, B. V. Fine, A. S. Anisimov, and A. G. Nasibulin, “Highly efficient thermophones based on freestanding single-walled carbon nanotube films,” *Nanoscale Horiz.*, vol. 4, no. 5, pp. 1158–1163, 2019, doi: 10.1039/C9NH00164F.

[29] H. Hu, K. Zhang, and D. Wang, “Frequency response calculations of carbon nanotube based nanothermophones,” *J. Phys.: Conf. Ser.*, vol. 1633, p. 012008, Sep. 2020, doi: 10.1088/1742-6596/1633/1/012008.

[30] R. H. Baughman, “Carbon Nanotubes--the Route Toward Applications,” *Science*, vol. 297, no. 5582, pp. 787–792, Aug. 2002, doi: 10.1126/science.1060928.

[31] X. Yu, R. Rajamani, K. A. Stelson, and T. Cui, “Active Control of Sound Transmission Through Windows With Carbon Nanotube-Based Transparent Actuators,” *IEEE Trans. Contr. Syst. Technol.*, vol. 15, no. 4, pp. 704–714, Jul. 2007, doi: 10.1109/TCST.2006.890277.

[32] M. M. Thiery, “Advanced Uses for Carbon Nanotubes: A Spherical Sound Source and Hot-films as Microphones,” Master of Science in Mechanical Engineering, Michigan Technological University, Houghton, Michigan, 2017.

[33] S. A. Senczyszyn, “Commercialization of the Carbon Nanotube Thermophone for Active Noise Control Applications,” Master of Science in Mechanical Engineering, Michigan Technological University, Houghton, Michigan, 2018.

[34] A. R. Barnard, “Solid state transducer, system, and method,” U.S. Patent 20200410973, Dec. 31, 2020.

[35] S. Prabhu and A. Barnard, “Design and characterization of an enclosed coaxial carbon nanotube speaker,” *The Journal of the Acoustical Society of America*, vol. 147, no. 4, pp. EL333–EL338, Apr. 2020, doi: 10.1121/10.0001029.

[36] M. Zhang *et al.*, “Self-Powered, Electrochemical Carbon Nanotube Pressure Sensors for Wave Monitoring,” *Adv. Funct. Mater.*, vol. 30, no. 42, p. 2004564, Oct. 2020, doi: 10.1002/adfm.202004564.

[37] K. Suzuki *et al.*, “Study of Carbon-Nanotube Web Thermoacoustic Loud Speakers,” *Jpn. J. Appl. Phys.*, vol. 50, p. 01BJ10, Jan. 2011, doi: 10.1143/JJAP.50.01BJ10.

[38] D. Passeri *et al.*, “Thermoacoustic Emission from Carbon Nanotubes Imaged by Atomic Force Microscopy,” *Adv. Funct. Mater.*, vol. 22, no. 14, pp. 2956–2963, Jul. 2012, doi: 10.1002/adfm.201200435.

[39] D. Ahn and S.-E. Ahn, “Thermoacoustic properties of multi-wall carbon nanotube sheet for loudspeaker application,” *Materials Letters*, vol. 263, p. 127242, Mar. 2020, doi: 10.1016/j.matlet.2019.127242.

[40] M. Daschewski, R. Boehm, J. Prager, M. Kreutzbruck, and A. Harrer, “Physics of thermo-acoustic sound generation,” *Journal of Applied Physics*, vol. 114, no. 11, p. 114903, Sep. 2013, doi: 10.1063/1.4821121.

[41] C. W. Lim, L. H. Tong, and Y. C. Li, “Theory of suspended carbon nanotube thinfilm as a thermal-acoustic source,” *Journal of Sound and Vibration*, vol. 332, no. 21, pp. 5451–5461, Oct. 2013, doi: 10.1016/j.jsv.2013.05.020.

[42] Y. Yang and J. Liu, “Computational characterization on the thermoacoustic thermophone effects induced by micro/nano-heating elements,” *Microfluid Nanofluid*, vol. 14, no. 5, pp. 873–884, May 2013, doi: 10.1007/s10404-012-1121-5.

[43] M. Asgarisabet, “Multi physics modeling and validation of carbon nanotube loudspeakers,” Master of Science in Mechanical Engineering, Michigan Technological University, Houghton, Michigan, 2016.

[44] M. Asgarisabet, A. R. Barnard, and T. M. Bouman, “Near field acoustic holography measurements of carbon nanotube thin film speakers,” *The Journal of the Acoustical Society of America*, vol. 140, no. 6, pp. 4237–4245, Dec. 2016, doi: 10.1121/1.4971328.

[45] P. Guiraud, S. Giordano, O. Bou Matar, P. Pernod, and R. Lardat, “Two temperature model for thermoacoustic sound generation in thick porous thermophones,” *Journal of Applied Physics*, vol. 126, no. 16, p. 165111, Oct. 2019, doi: 10.1063/1.5121395.

[46] P. Guiraud, S. Giordano, O. Bou-Matar, P. Pernod, and R. Lardat, “Multilayer modeling of thermoacoustic sound generation for thermophone analysis and design,” *Journal of Sound and Vibration*, vol. 455, pp. 275–298, Sep. 2019, doi: 10.1016/j.jsv.2019.05.001.

[47] P. Kumar *et al.*, “Understanding the Low Frequency Response of Carbon Nanotube Thermoacoustic Projectors,” *Journal of Sound and Vibration*, p. 115940, Jan. 2021, doi: 10.1016/j.jsv.2021.115940.

[48] G. Chitnis, A. Kim, S. H. Song, A. M. Jessop, J. S. Bolton, and B. Ziaie, “A thermophone on porous polymeric substrate,” *Appl. Phys. Lett.*, vol. 101, no. 2, p. 021911, Jul. 2012, doi: 10.1063/1.4737005.

[49] L. H. Tong, C. W. Lim, S. K. Lai, and Y. C. Li, “Gap separation effect on thermoacoustic wave generation by heated suspended CNT nano-thinfilm,” *Applied Thermal Engineering*, vol. 86, pp. 135–142, Jul. 2015, doi: 10.1016/j.applthermaleng.2015.04.031.

[50] W. Yi, L. Lu, Z. Dian-lin, Z. W. Pan, and S. S. Xie, “Linear specific heat of carbon nanotubes,” *Phys. Rev. B*, vol. 59, no. 14, pp. R9015–R9018, Apr. 1999, doi: 10.1103/PhysRevB.59.R9015.

[51] K. Jiang, Q. Li, and S. Fan, “Spinning continuous carbon nanotube yarns,” *Nature*, vol. 419, no. 6909, pp. 801–801, Oct. 2002, doi: 10.1038/419801a.

[52] M. Zhang, “Strong, Transparent, Multifunctional, Carbon Nanotube Sheets,” *Science*, vol. 309, no. 5738, pp. 1215–1219, Aug. 2005, doi: 10.1126/science.1115311.

[53] Y. Wei, L. Liu, P. Liu, L. Xiao, K. Jiang, and S. Fan, “Scaled fabrication of single-nanotube-tipped ends from carbon nanotube micro-yarns and their field emission applications,” *Nanotechnology*, vol. 19, no. 47, p. 475707, Nov. 2008, doi: 10.1088/0957-4484/19/47/475707.

[54] A. E. Aliev, M. H. Lima, E. M. Silverman, and R. H. Baughman, “Thermal conductivity of multi-walled carbon nanotube sheets: radiation losses and quenching of phonon modes,” *Nanotechnology*, vol. 21, no. 3, p. 035709, Jan. 2010, doi: 10.1088/0957-4484/21/3/035709.

[55] L. Hu, D. S. Hecht, and G. Grüner, “Carbon Nanotube Thin Films: Fabrication, Properties, and Applications,” *Chem. Rev.*, vol. 110, no. 10, pp. 5790–5844, Oct. 2010, doi: 10.1021/cr9002962.

[56] K. Jiang, J. Wang, Q. Li, L. Liu, C. Liu, and S. Fan, “Superaligned Carbon Nanotube Arrays, Films, and Yarns: A Road to Applications,” *Adv. Mater.*, vol. 23, no. 9, pp. 1154–1161, Mar. 2011, doi: 10.1002/adma.201003989.

[57] P. Liu, Y. Wei, L. Liu, K. Jiang, and S. Fan, “Formation of free-standing carbon nanotube array on super-aligned carbon nanotube film and its field emission properties,” *Nano Res.*, vol. 5, no. 6, pp. 421–426, Jun. 2012, doi: 10.1007/s12274-012-0224-3.

[58] H. Tian *et al.*, “Graphene-on-Paper Sound Source Devices,” *ACS Nano*, vol. 5, no. 6, pp. 4878–4885, Jun. 2011, doi: 10.1021/nn2009535.

[59] H. Tian *et al.*, “Single-layer graphene sound-emitting devices: experiments and modeling,” *Nanoscale*, vol. 4, no. 7, p. 2272, 2012, doi: 10.1039/c2nr11572g.

[60] T. Sugimoto and Y. Nakajima, “Acoustic characteristics of a flexible sound generator based on thermoacoustic effect,” Montreal, Canada, 2013, pp. 030004–030004, doi: 10.1121/1.4799163.

[61] Y. Wei, X. Lin, K. Jiang, P. Liu, Q. Li, and S. Fan, “Thermoacoustic Chips with Carbon Nanotube Thin Yarn Arrays,” *Nano Lett.*, vol. 13, no. 10, pp. 4795–4801, Oct. 2013, doi: 10.1021/nl402408j.

[62] R. Dutta *et al.*, “Gold Nanowire Thermophones,” *J. Phys. Chem. C*, vol. 118, no. 50, pp. 29101–29107, Dec. 2014, doi: 10.1021/jp504195v.

[63] W. Fei, J. Zhou, and W. Guo, “Low-voltage Driven Graphene Foam Thermoacoustic Speaker,” *Small*, vol. 11, no. 19, pp. 2252–2256, May 2015, doi: 10.1002/smll.201402982.

[64] A. E. Aliev, S. Perananthan, and J. P. Ferraris, “Carbonized Electrospun Nanofiber Sheets for Thermophones,” *ACS Appl. Mater. Interfaces*, vol. 8, no. 45, pp. 31192–31201, Nov. 2016, doi: 10.1021/acsami.6b08717.

[65] C. S. Kim, S. K. Hong, J.-M. Lee, D.-S. Kang, B. J. Cho, and J.-W. Choi, “Free-Standing Graphene Thermophone on a Polymer-Mesh Substrate,” *Small*, vol. 12, no. 2, pp. 185–189, Jan. 2016, doi: 10.1002/smll.201501673.

[66] M. Zhang, P. Wolmarans, and A. E. Aliev, “The fabrication and characterization of nanocarbon foams for their utilization in thermoacoustic device,” *The Journal of the Acoustical Society of America*, vol. 142, no. 4, pp. 2538–2538, Oct. 2017, doi: 10.1121/1.5014274.

[67] A. Ghasemi Yeklangi, S. E. Khadem, and S. Darbari, “Fabrication and investigation of a thermoacoustic loudspeaker based on carbon nanotube coated laser-scribed

graphene," *Journal of Applied Physics*, vol. 124, no. 22, p. 224501, Dec. 2018, doi: 10.1063/1.5038729.

[68] Z. L. Ngoh *et al.*, "Experimental characterization of three-dimensional Graphene's thermoacoustic response and its theoretical modelling," *Carbon*, vol. 169, pp. 382–394, Nov. 2020, doi: 10.1016/j.carbon.2020.06.045.

[69] V. Vesterinen, A. O. Niskanen, J. Hassel, and P. Helistö, "Fundamental Efficiency of Nanothermophones: Modeling and Experiments," *Nano Lett.*, vol. 10, no. 12, pp. 5020–5024, Dec. 2010, doi: 10.1021/nl1031869.

[70] A. E. Aliev *et al.*, "Thermal management of thermoacoustic sound projectors using a free-standing carbon nanotube aerogel sheet as a heat source," *Nanotechnology*, vol. 25, no. 40, p. 405704, Oct. 2014, doi: 10.1088/0957-4484/25/40/405704.

[71] W. H. Nelson, "Active noise control using carbon nanotube thermophones: Case study for an automotive HVAC application," M.Sc. Thesis, Michigan Technological University, Houghton, Michigan, USA, 2020.

[72] A. Hall *et al.*, "Signal conditioning of carbon nanotube thin film loudspeakers," in *14th IEEE International Conference on Nanotechnology*, Toronto, ON, Canada, Aug. 2014, pp. 668–671, doi: 10.1109/NANO.2014.6968142.

[73] A. E. Aliev *et al.*, "Thermoacoustic sound projector: exceeding the fundamental efficiency of carbon nanotubes," *Nanotechnology*, vol. 29, no. 32, p. 325704, Aug. 2018, doi: 10.1088/1361-6528/aac509.

[74] G. Fairbanks, "Test of Phonemic Differentiation: The Rhyme Test," *The Journal of the Acoustical Society of America*, vol. 30, no. 7, pp. 596–600, Jul. 1958, doi: 10.1121/1.1909702.

[75] P. W. Nye and J. H. Gaitenby, "Consonant intelligibility in synthetic speech and in a natural speech control (modified rhyme test results)," Haskins Laboratories, SR-33, 1973.

[76] H. A. David, *The method of paired comparisons*, 2. ed., Rev. London: Griffin, 1988.

[77] S. Voran, "Using articulation index band correlations to objectively estimate speech intelligibility consistent with the modified rhyme test," in *2013 IEEE Workshop on Applications of Signal Processing to Audio and Acoustics*, New Paltz, NY, USA, Oct. 2013, pp. 1–4, doi: 10.1109/WASPAA.2013.6701826.

[78] "Methods for Calculation of the Speech Intelligibility Index," American National Standards Institute, Standard ANSI S3.5, 1997.

[79] T. Houtgast and H. J. M. Steeneken, “Predicting and measuring speech intelligibility in rooms,” *The Journal of the Acoustical Society of America*, vol. 71, no. S1, pp. S19–S19, Apr. 1982, doi: 10.1121/1.2019264.

[80] K. L. Payton and L. D. Braida, “A method to determine the speech transmission index from speech waveforms,” *The Journal of the Acoustical Society of America*, vol. 106, no. 6, pp. 3637–3648, Dec. 1999, doi: 10.1121/1.428216.

[81] “Sound system equipment - Part 16: Objective rating of speech intelligibility by speech transmission index,” International Electrotechnical Commission, Standard IEC 60268-16, 2020.

[82] A. Astolfi and F. Pellerey, “Subjective and objective assessment of acoustical and overall environmental quality in secondary school classrooms,” *The Journal of the Acoustical Society of America*, vol. 123, no. 1, pp. 163–173, Jan. 2008, doi: 10.1121/1.2816563.

[83] J. E. Dockrell and B. Shield, “Children’s perceptions of their acoustic environment at school and at home,” *The Journal of the Acoustical Society of America*, vol. 115, no. 6, pp. 2964–2973, Jun. 2004, doi: 10.1121/1.1652610.

[84] M. Hagen, J. Kahlert, C. Hemmer-Schanze, L. Huber, and M. Meis, “Developing an Acoustic School Design: Steps to Improve Hearing and Listening at Schools,” *Building Acoustics*, vol. 11, no. 4, pp. 293–307, Dec. 2004, doi: 10.1260/1351010042900086.

[85] S. M. Kennedy, M. Hodgson, L. D. Edgett, N. Lamb, and R. Rempel, “Subjective assessment of listening environments in university classrooms: Perceptions of students,” *The Journal of the Acoustical Society of America*, vol. 119, no. 1, pp. 299–309, Jan. 2006, doi: 10.1121/1.2139629.

[86] J. K. Ryu and J. Y. Jeon, “Subjective and objective evaluations of a scattered sound field in a scale model opera house,” *The Journal of the Acoustical Society of America*, vol. 124, no. 3, pp. 1538–1549, Sep. 2008, doi: 10.1121/1.2956474.

[87] N. Otto, S. Amman, C. Eaton, and S. Lake, “Guidelines for Jury Evaluations of Automotive Sounds,” Sound and vibration, 2001.

[88] M. J. M. Nor, M. H. Fouladi, H. Nahvi, and A. K. Ariffin, “Index for vehicle acoustical comfort inside a passenger car,” *Applied Acoustics*, vol. 69, no. 4, pp. 343–353, Apr. 2008, doi: 10.1016/j.apacoust.2006.11.001.

[89] T. J. Shin, D. C. Park, and S. K. Lee, “Objective evaluation of door-closing sound quality based on physiological acoustics,” *Int.J Automot. Technol.*, vol. 14, no. 1, pp. 133–141, Feb. 2013, doi: 10.1007/s12239-013-0015-1.

[90] H. B. Huang, R. X. Li, X. R. Huang, M. L. Yang, and W. P. Ding, “Sound quality evaluation of vehicle suspension shock absorber rattling noise based on the Wigner–Ville distribution,” *Applied Acoustics*, vol. 100, pp. 18–25, Dec. 2015, doi: 10.1016/j.apacoust.2015.06.018.

[91] Y. Fang and T. Zhang, “Sound Quality of the Acoustic Noise Radiated by PWM-Fed Electric Powertrain,” *IEEE Trans. Ind. Electron.*, vol. 65, no. 6, pp. 4534–4541, Jun. 2018, doi: 10.1109/TIE.2017.2767558.

[92] G. Volandri, F. Di Puccio, P. Forte, and L. Mattei, “Psychoacoustic analysis of power windows sounds: Correlation between subjective and objective evaluations,” *Applied Acoustics*, vol. 134, pp. 160–170, May 2018, doi: 10.1016/j.apacoust.2017.11.020.

[93] A. Gonzalez, M. Ferrer, M. de Diego, G. Piñero, and J. J. Garcia-Bonito, “Sound quality of low-frequency and car engine noises after active noise control,” *Journal of Sound and Vibration*, vol. 265, no. 3, pp. 663–679, Aug. 2003, doi: 10.1016/S0022-460X(02)01462-1.

[94] L. Lehto, L. Laaksonen, E. Vilkman, and P. Alku, “Occupational voice complaints and objective acoustic measurements—do they correlate?,” *Logopedics Phoniatrics Vocology*, vol. 31, no. 4, pp. 147–152, Jan. 2006, doi: 10.1080/14015430600654654.

[95] T. Lokki, J. Pätynen, A. Kuusinen, and S. Tervo, “Disentangling preference ratings of concert hall acoustics using subjective sensory profiles,” *The Journal of the Acoustical Society of America*, vol. 132, no. 5, pp. 3148–3161, Nov. 2012, doi: 10.1121/1.4756826.

[96] G. Pietila and T. C. Lim, “Intelligent systems approaches to product sound quality evaluations – A review,” *Applied Acoustics*, vol. 73, no. 10, pp. 987–1002, Oct. 2012, doi: 10.1016/j.apacoust.2012.04.012.

[97] J. F. Santos, S. Cosentino, O. Hazrati, P. C. Loizou, and T. H. Falk, “Objective speech intelligibility measurement for cochlear implant users in complex listening environments,” *Speech Communication*, vol. 55, no. 7–8, pp. 815–824, Sep. 2013, doi: 10.1016/j.specom.2013.04.001.

[98] G. Rey Gozalo, J. Trujillo Carmona, J. M. Barrigón Morillas, R. Vílchez-Gómez, and V. Gómez Escobar, “Relationship between objective acoustic indices and subjective assessments for the quality of soundscapes,” *Applied Acoustics*, vol. 97, pp. 1–10, Oct. 2015, doi: 10.1016/j.apacoust.2015.03.020.

[99] M. Moravec, G. Ižaríková, P. Liptai, M. Badida, and A. Badidová, “Development of psychoacoustic model based on the correlation of the subjective and objective sound

quality assessment of automatic washing machines," *Applied Acoustics*, vol. 140, pp. 178–182, Nov. 2018, doi: 10.1016/j.apacoust.2018.05.025.

[100] N.-G. Vardaxis, D. Bard, and K. Persson Waye, "Review of acoustic comfort evaluation in dwellings—part I: Associations of acoustic field data to subjective responses from building surveys," *Building Acoustics*, vol. 25, no. 2, pp. 151–170, Jun. 2018, doi: 10.1177/1351010X18762687.

[101] C. Moo, Y. Chang, and P. Mok, "A digital measurement scheme for time-varying transient harmonics," *IEEE*, vol. 10, pp. 588–594, 1995.

[102] Y.-J. Shin, E. J. Powers, M. Grady, and A. Arapostathis, "Power Quality Indices for Transient Disturbances," *IEEE Trans. Power Delivery*, vol. 21, no. 1, pp. 253–261, Jan. 2006, doi: 10.1109/TPWRD.2005.855444.

[103] J.-C. Montano, M.-D. Borras, M. Castilla, A. Lopez, J. Gutierrez, and J.-C. Bravo, "Harmonic distortion index for stationary and transient states," in *2009 Compatability and Power Electronics*, Badajoz, Spain, May 2009, pp. 123–128, doi: 10.1109/CPE.2009.5156023.

[104] T. Sainath and C. Parada, "Convolutional Neural Networks for Small-footprint Keyword Spotting," presented at the Sixteenth Annual Conference of the International Speech Communication Association, 2015.

[105] G. Chen, C. Parada, and G. Heigold, "Small-footprint keyword spotting using deep neural networks," in *2014 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Florence, Italy, May 2014, pp. 4087–4091, doi: 10.1109/ICASSP.2014.6854370.

[106] S. D. Voran, "A multiple bandwidth objective speech intelligibility estimator based on articulation index band correlations and attention," in *2017 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, New Orleans, LA, Mar. 2017, pp. 5100–5104, doi: 10.1109/ICASSP.2017.7953128.

[107] C. H. Taal, R. C. Hendriks, R. Heusdens, and J. Jensen, "A short-time objective intelligibility measure for time-frequency weighted noisy speech," in *2010 IEEE International Conference on Acoustics, Speech and Signal Processing*, Dallas, TX, USA, 2010, pp. 4214–4217, doi: 10.1109/ICASSP.2010.5495701.

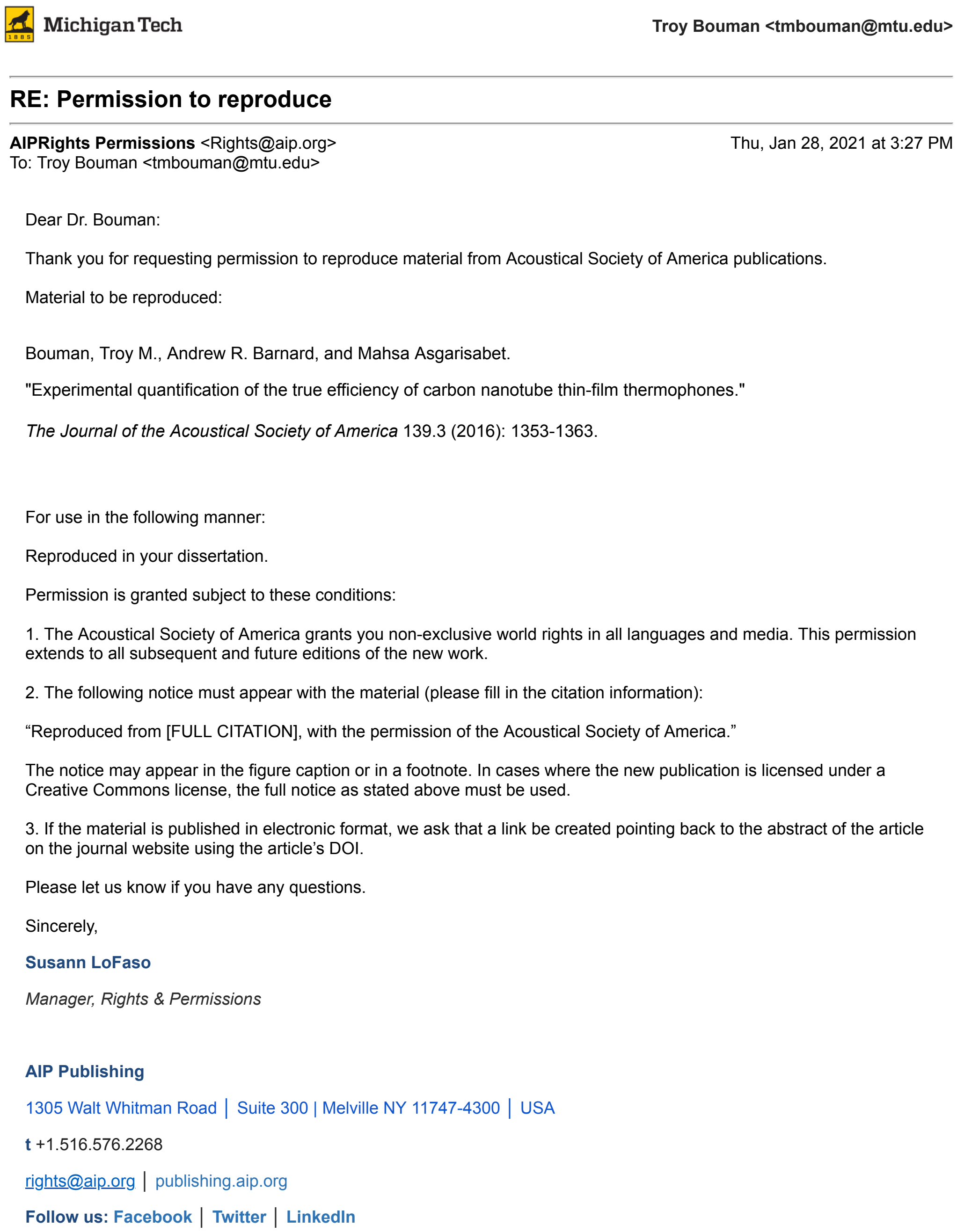
[108] C. H. Taal, R. C. Hendriks, R. Heusdens, and J. Jensen, "An Algorithm for Intelligibility Prediction of Time–Frequency Weighted Noisy Speech," *IEEE Trans. Audio Speech Lang. Process.*, vol. 19, no. 7, pp. 2125–2136, Sep. 2011, doi: 10.1109/TASL.2011.2114881.

[109] P. La Torraca, Y. Ricci, M. Bobinger, P. Pavan, and L. Larcher, “Linearization of thermoacoustic loudspeakers by adaptive predistortion,” *Sensors and Actuators A: Physical*, vol. 297, p. 111551, Oct. 2019, doi: 10.1016/j.sna.2019.111551.

[110] S. Dieleman and B. Schrauwen, “End-to-end learning for music audio,” in *2014 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Florence, Italy, May 2014, pp. 6964–6968, doi: 10.1109/ICASSP.2014.6854950.

A Copyright documentation

A.1 JASA Permission



Michigan Tech

Troy Bouman <tmbouman@mtu.edu>

RE: Permission to reproduce

AIPRights Permissions <Rights@aip.org>
To: Troy Bouman <tmbouman@mtu.edu>

Thu, Jan 28, 2021 at 3:27 PM

Dear Dr. Bouman:

Thank you for requesting permission to reproduce material from Acoustical Society of America publications.

Material to be reproduced:

Bouman, Troy M., Andrew R. Barnard, and Mahsa Asgarisabet.

"Experimental quantification of the true efficiency of carbon nanotube thin-film thermophones."

The Journal of the Acoustical Society of America 139.3 (2016): 1353-1363.

For use in the following manner:

Reproduced in your dissertation.

Permission is granted subject to these conditions:

1. The Acoustical Society of America grants you non-exclusive world rights in all languages and media. This permission extends to all subsequent and future editions of the new work.

2. The following notice must appear with the material (please fill in the citation information):

“Reproduced from [FULL CITATION], with the permission of the Acoustical Society of America.”

The notice may appear in the figure caption or in a footnote. In cases where the new publication is licensed under a Creative Commons license, the full notice as stated above must be used.

3. If the material is published in electronic format, we ask that a link be created pointing back to the abstract of the article on the journal website using the article’s DOI.

Please let us know if you have any questions.

Sincerely,

Susann LoFaso

Manager, Rights & Permissions

AIP Publishing

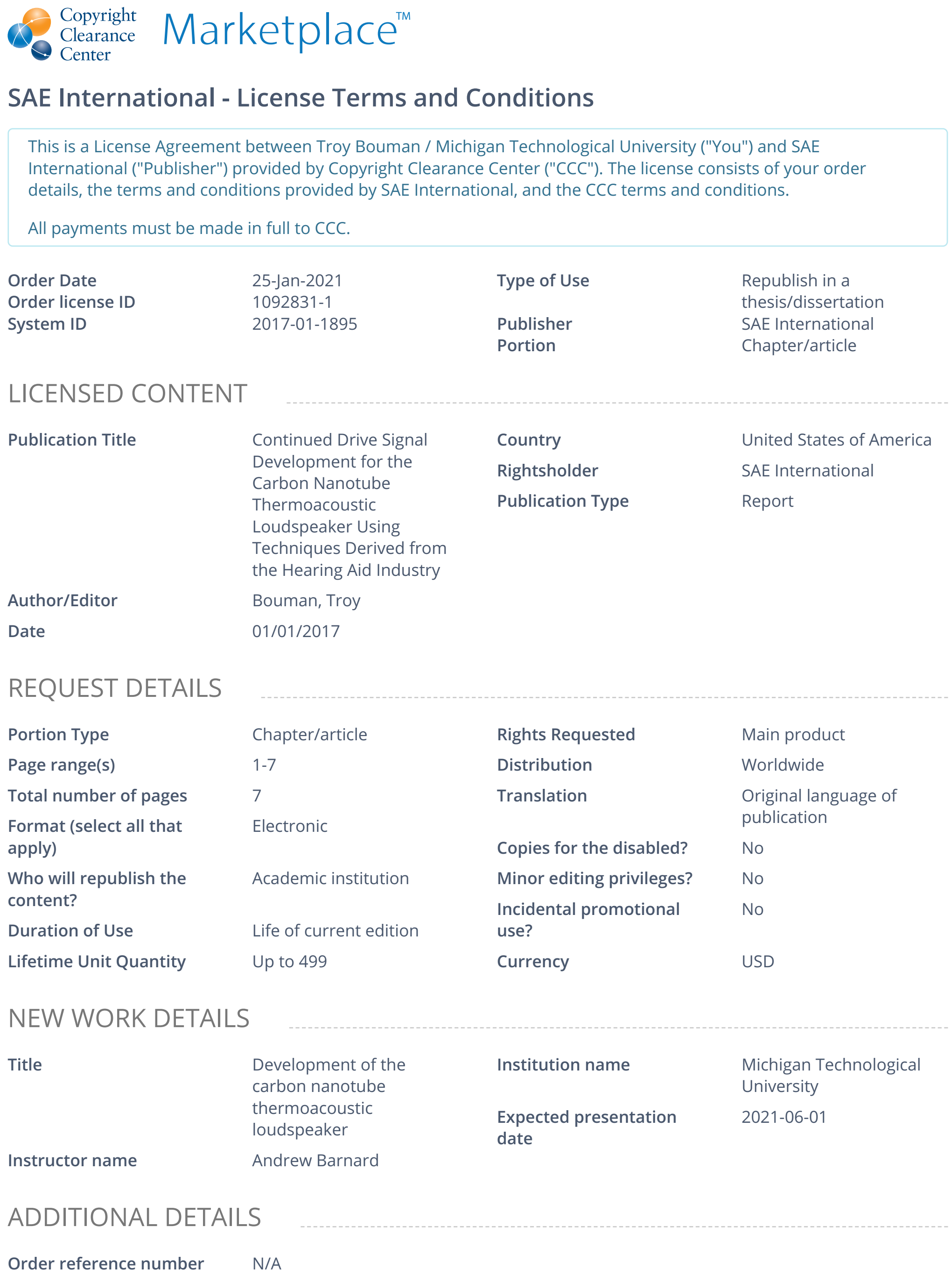
1305 Walt Whitman Road │ Suite 300 | Melville NY 11747-4300 │ USA

t +1.516.576.2268

rights@aip.org │ publishing.aip.org

Follow us: Facebook │ **Twitter** │ **LinkedIn**

A.2 SAE Permission



Copyright Clearance Center Marketplace™

SAE International - License Terms and Conditions

This is a License Agreement between Troy Bouman / Michigan Technological University ("You") and SAE International ("Publisher") provided by Copyright Clearance Center ("CCC"). The license consists of your order details, the terms and conditions provided by SAE International, and the CCC terms and conditions.

All payments must be made in full to CCC.

Order Date	25-Jan-2021	**Type of Use**	Republish in a thesis/dissertation
Order license ID	1092831-1		
System ID	2017-01-1895	**Publisher**	SAE International
		Portion	Chapter/article

LICENSED CONTENT

Publication Title	Continued Drive Signal Development for the Carbon Nanotube Thermoacoustic Loudspeaker Using Techniques Derived from the Hearing Aid Industry	**Country**	United States of America
		Rightsholder	SAE International
		Publication Type	Report
Author/Editor	Bouman, Troy		
Date	01/01/2017		

REQUEST DETAILS

Portion Type	Chapter/article	**Rights Requested**	Main product
Page range(s)	1-7	**Distribution**	Worldwide
Total number of pages	7	**Translation**	Original language of publication
Format (select all that apply)	Electronic	**Copies for the disabled?**	No
Who will republish the content?	Academic institution	**Minor editing privileges?**	No
		Incidental promotional use?	No
Duration of Use	Life of current edition		
Lifetime Unit Quantity	Up to 499	**Currency**	USD

NEW WORK DETAILS

Title	Development of the carbon nanotube thermoacoustic loudspeaker	**Institution name**	Michigan Technological University
		Expected presentation date	2021-06-01
Instructor name	Andrew Barnard		

ADDITIONAL DETAILS

Order reference number	N/A

The requesting person / organization to appear on the license	Troy Bouman / Michigan Technological University

REUSE CONTENT DETAILS

Title, description or numeric reference of the portion(s)	Continued Drive Signal Development for the Carbon Nanotube Thermoacoustic Loudspeaker Using Techniques Derived from the Hearing Aid Industry	**Title of the article/chapter the portion is from**	N/A
		Author of portion(s)	Bouman, Troy
		Issue, if republishing an article from a serial	2017-01-1895
Editor of portion(s)	N/A	**Publication date of portion**	2017-06-05
Volume of serial or monograph	N/A		
Page or page range of portion	1-7		

CCC Republication Terms and Conditions

1. Description of Service; Defined Terms. This Republication License enables the User to obtain licenses for republication of one or more copyrighted works as described in detail on the relevant Order Confirmation (the "Work(s)"). Copyright Clearance Center, Inc. ("CCC") grants licenses through the Service on behalf of the rightsholder identified on the Order Confirmation (the "Rightsholder"). "Republication", as used herein, generally means the inclusion of a Work, in whole or in part, in a new work or works, also as described on the Order Confirmation. "User", as used herein, means the person or entity making such republication.

2. The terms set forth in the relevant Order Confirmation, and any terms set by the Rightsholder with respect to a particular Work, govern the terms of use of Works in connection with the Service. By using the Service, the person transacting for a republication license on behalf of the User represents and warrants that he/she/it (a) has been duly authorized by the User to accept, and hereby does accept, all such terms and conditions on behalf of User, and (b) shall inform User of all such terms and conditions. In the event such person is a "freelancer" or other third party independent of User and CCC, such party shall be deemed jointly a "User" for purposes of these terms and conditions. In any event, User shall be deemed to have accepted and agreed to all such terms and conditions if User republishes the Work in any fashion.

3. Scope of License; Limitations and Obligations.

    3.1. All Works and all rights therein, including copyright rights, remain the sole and exclusive property of the Rightsholder. The license created by the exchange of an Order Confirmation (and/or any invoice) and payment by User of the full amount set forth on that document includes only those rights expressly set forth in the Order Confirmation and in these terms and conditions, and conveys no other rights in the Work(s) to User. All rights not expressly granted are hereby reserved.

    3.2.
    General Payment Terms: You may pay by credit card or through an account with us payable at the end of the month. If you and we agree that you may establish a standing account with CCC, then the following terms apply: Remit Payment to: Copyright Clearance Center, 29118 Network Place, Chicago, IL 60673-1291. Payments Due: Invoices are payable upon their delivery to you (or upon our notice to you that they are available to you for downloading). After 30 days, outstanding amounts will be subject to a service charge of 1-1/2% per month or, if less, the maximum rate allowed by applicable law. Unless otherwise specifically set forth in the Order Confirmation or in a separate written agreement signed by CCC, invoices are due and payable on "net 30" terms. While User may exercise the rights licensed immediately upon issuance of the Order Confirmation, the license is automatically revoked and is null and void, as if it had never been

issued, if complete payment for the license is not received on a timely basis either from User directly or through a payment agent, such as a credit card company.

3.3. Unless otherwise provided in the Order Confirmation, any grant of rights to User (i) is "one-time" (including the editions and product family specified in the license), (ii) is non-exclusive and non-transferable and (iii) is subject to any and all limitations and restrictions (such as, but not limited to, limitations on duration of use or circulation) included in the Order Confirmation or invoice and/or in these terms and conditions. Upon completion of the licensed use, User shall either secure a new permission for further use of the Work(s) or immediately cease any new use of the Work(s) and shall render inaccessible (such as by deleting or by removing or severing links or other locators) any further copies of the Work (except for copies printed on paper in accordance with this license and still in User's stock at the end of such period).

3.4. In the event that the material for which a republication license is sought includes third party materials (such as photographs, illustrations, graphs, inserts and similar materials) which are identified in such material as having been used by permission, User is responsible for identifying, and seeking separate licenses (under this Service or otherwise) for, any of such third party materials; without a separate license, such third party materials may not be used.

3.5. Use of proper copyright notice for a Work is required as a condition of any license granted under the Service. Unless otherwise provided in the Order Confirmation, a proper copyright notice will read substantially as follows: "Republished with permission of [Rightsholder's name], from [Work's title, author, volume, edition number and year of copyright]; permission conveyed through Copyright Clearance Center, Inc. " Such notice must be provided in a reasonably legible font size and must be placed either immediately adjacent to the Work as used (for example, as part of a by-line or footnote but not as a separate electronic link) or in the place where substantially all other credits or notices for the new work containing the republished Work are located. Failure to include the required notice results in loss to the Rightsholder and CCC, and the User shall be liable to pay liquidated damages for each such failure equal to twice the use fee specified in the Order Confirmation, in addition to the use fee itself and any other fees and charges specified.

3.6. User may only make alterations to the Work if and as expressly set forth in the Order Confirmation. No Work may be used in any way that is defamatory, violates the rights of third parties (including such third parties' rights of copyright, privacy, publicity, or other tangible or intangible property), or is otherwise illegal, sexually explicit or obscene. In addition, User may not conjoin a Work with any other material that may result in damage to the reputation of the Rightsholder. User agrees to inform CCC if it becomes aware of any infringement of any rights in a Work and to cooperate with any reasonable request of CCC or the Rightsholder in connection therewith.

4. Indemnity. User hereby indemnifies and agrees to defend the Rightsholder and CCC, and their respective employees and directors, against all claims, liability, damages, costs and expenses, including legal fees and expenses, arising out of any use of a Work beyond the scope of the rights granted herein, or any use of a Work which has been altered in any unauthorized way by User, including claims of defamation or infringement of rights of copyright, publicity, privacy or other tangible or intangible property.

5. Limitation of Liability. UNDER NO CIRCUMSTANCES WILL CCC OR THE RIGHTSHOLDER BE LIABLE FOR ANY DIRECT, INDIRECT, CONSEQUENTIAL OR INCIDENTAL DAMAGES (INCLUDING WITHOUT LIMITATION DAMAGES FOR LOSS OF BUSINESS PROFITS OR INFORMATION, OR FOR BUSINESS INTERRUPTION) ARISING OUT OF THE USE OR INABILITY TO USE A WORK, EVEN IF ONE OF THEM HAS BEEN ADVISED OF THE POSSIBILITY OF SUCH DAMAGES. In any event, the total liability of the Rightsholder and CCC (including their respective employees and directors) shall not exceed the total amount actually paid by User for this license. User assumes full liability for the actions and omissions of its principals, employees, agents, affiliates, successors and assigns.

6.

Limited Warranties. THE WORK(S) AND RIGHT(S) ARE PROVIDED "AS IS". CCC HAS THE RIGHT TO GRANT TO USER THE RIGHTS GRANTED IN THE ORDER CONFIRMATION DOCUMENT. CCC AND THE RIGHTSHOLDER DISCLAIM ALL OTHER WARRANTIES RELATING TO THE WORK(S) AND RIGHT(S), EITHER EXPRESS OR IMPLIED, INCLUDING

WITHOUT LIMITATION IMPLIED WARRANTIES OF MERCHANTABILITY OR FITNESS FOR A PARTICULAR PURPOSE. ADDITIONAL RIGHTS MAY BE REQUIRED TO USE ILLUSTRATIONS, GRAPHS, PHOTOGRAPHS, ABSTRACTS, INSERTS OR OTHER PORTIONS OF THE WORK (AS OPPOSED TO THE ENTIRE WORK) IN A MANNER CONTEMPLATED BY USER; USER UNDERSTANDS AND AGREES THAT NEITHER CCC NOR THE RIGHTSHOLDER MAY HAVE SUCH ADDITIONAL RIGHTS TO GRANT.

7. Effect of Breach. Any failure by User to pay any amount when due, or any use by User of a Work beyond the scope of the license set forth in the Order Confirmation and/or these terms and conditions, shall be a material breach of the license created by the Order Confirmation and these terms and conditions. Any breach not cured within 30 days of written notice thereof shall result in immediate termination of such license without further notice. Any unauthorized (but licensable) use of a Work that is terminated immediately upon notice thereof may be liquidated by payment of the Rightsholder's ordinary license price therefor; any unauthorized (and unlicensable) use that is not terminated immediately for any reason (including, for example, because materials containing the Work cannot reasonably be recalled) will be subject to all remedies available at law or in equity, but in no event to a payment of less than three times the Rightsholder's ordinary license price for the most closely analogous licensable use plus Rightsholder's and/or CCC's costs and expenses incurred in collecting such payment.

8. Miscellaneous.

   8.1. User acknowledges that CCC may, from time to time, make changes or additions to the Service or to these terms and conditions, and CCC reserves the right to send notice to the User by electronic mail or otherwise for the purposes of notifying User of such changes or additions; provided that any such changes or additions shall not apply to permissions already secured and paid for.

   8.2. Use of User-related information collected through the Service is governed by CCC's privacy policy, available online here:https://marketplace.copyright.com/rs-ui-web/mp/privacy-policy

   8.3. The licensing transaction described in the Order Confirmation is personal to User. Therefore, User may not assign or transfer to any other person (whether a natural person or an organization of any kind) the license created by the Order Confirmation and these terms and conditions or any rights granted hereunder; provided, however, that User may assign such license in its entirety on written notice to CCC in the event of a transfer of all or substantially all of User's rights in the new material which includes the Work(s) licensed under this Service.

   8.4. No amendment or waiver of any terms is binding unless set forth in writing and signed by the parties. The Rightsholder and CCC hereby object to any terms contained in any writing prepared by the User or its principals, employees, agents or affiliates and purporting to govern or otherwise relate to the licensing transaction described in the Order Confirmation, which terms are in any way inconsistent with any terms set forth in the Order Confirmation and/or in these terms and conditions or CCC's standard operating procedures, whether such writing is prepared prior to, simultaneously with or subsequent to the Order Confirmation, and whether such writing appears on a copy of the Order Confirmation or in a separate instrument.

   8.5. The licensing transaction described in the Order Confirmation document shall be governed by and construed under the law of the State of New York, USA, without regard to the principles thereof of conflicts of law. Any case, controversy, suit, action, or proceeding arising out of, in connection with, or related to such licensing transaction shall be brought, at CCC's sole discretion, in any federal or state court located in the County of New York, State of New York, USA, or in any federal or state court whose geographical jurisdiction covers the location of the Rightsholder set forth in the Order Confirmation. The parties expressly submit to the personal jurisdiction and venue of each such federal or state court.If you have any comments or questions about the Service or Copyright Clearance Center, please contact us at 978-750-8400 or send an e-mail to support@copyright.com.

v 1.1

A.3 THDSI Python 3.7 Function

```
import numpy as np

def THDSI(cleanFFT, NoisyFFT, yAxis, binsize, threshold = 2,
overlapFac=0,fs = None, TF = None):
    """
    Calculates the Total Harmonic Distortion for Speech
Intelligibility (THDSI) value at every time step of the Short Time
Fourier Transform (STFT) spectra

    Input
    -----
    * cleanFFT : 2darray

        Where the each row is the STFT spectra at a center time
(i.e. the typical np.fft.rfft() result that is amplitude corrected)
for the clean signal.

    * NoisyFFT : 2darray

        -Where the each row is the STFT spectra at a center time
(i.e. the typical np.fft.rfft() result that is amplitude corrected)
for the noisy signal.
        -Must be the same shape as cleanFFT and computed from the
same STFT settings (e.g. binsize)

    * yAxis : 1Darray

        Array of the  center bin frequencies resulting from the
cleanSTFT and noisySTFT computations

    * binsize : int

        The binsize used in the STFT calculations for cleanSTFT and
noisy STFT

    * Threshold : float

        (Optional) The amplitude threshold used to determine if a
valid fundamental frequency is found. Default = 2

    * OverlapFac : float

        (Optional) Value 0 - 1 representing the percent of overlap
desired Example: 0.5 means 50% overlap. Default = 0

    * fs : int

        (Optional) Sampling rate of the original data used to make
cleanSTFT and noisySTFT. Including fs will result in metrics being
printed about the processing parameters. Default = None
```

```
    * TF : 1Darray

        (Optional): Transfer Function applied to the noisySTFT data.
Must have length equal to the row size of noisySTFT. Default = None

    Output
    ------
    * THDSIvals : 1darray

        Contains the computed THDSI values at each time step.
Returns NaN if no valid frequency is found or if no harmonics are
found.

    * THDNSIvals : 1darray

        Contains the computed THDNSI values at each time step.
Returns NaN if no valid frequency is found or if no harmonics are
found.

    * harmonicVals : 1darray

        Contains the computed summation of all harmoincs at each
time step. Returns NaN if no valid frequency is found or if no
harmonics are found.

    * fundFreqs : 1darray

        Contains the estimated fundamental frequency determined from
the cleanSTFT spectra. Returns NaN if no valid frequency is found or
if no harmonics are found.

    """

    #Check input data quality
    assert (np.shape(cleanFFT) == np.shape(NoisyFFT)),"Clean and
Noisy signals are not the exact same length"

    #Initialize variables
    THDSIvals = []
    THDNSIvals = []
    harmonicVals = []
    fundFreqs = []

    #Print some metrics if given fs
    if fs:
        print("Freq res: {:.2f} [Hz]".format(fs/binsize))
        print("FFT Frame size: {:.0f} [ms]".format(binsize/fs*1000))
        if overlapFac != 0:
            print("Increment between FFTs is {} [ms] - {}%
overlap".format((binsize/fs*1000)*overlapFac,overlapFac*100))
        else:
            print("Increment between FFTs is {} [ms] - {}%
overlap".format((binsize/fs*1000),overlapFac*100))


    ACF = 1/np.mean(np.hanning(binsize)) #1/mean
    ECF = 1/np.sqrt(np.mean(np.hanning(binsize)**2)) #1/rms
    for currCol, (cleanFFTData, noisyFFTData) in
enumerate(zip(cleanFFT.T, NoisyFFT.T)):

        maxIdx = cleanFFTData.argmax()

        if np.any(TF):
            #Compute corrected STFT from CNT transfer function and
apply it to the noisyData before THDSI calc
            noisyFFTData = noisyFFTData/TF

        fund = noisyFFTData[maxIdx]
        if (fund > (threshold*np.mean(noisyFFTData))) &
(yAxis[maxIdx]>20): #Threshold check
            harmIdx = maxIdx
            harmonics = 0
            harmonicMultiplier = 2
            while (maxIdx * harmonicMultiplier) < len(noisyFFTData):
#Loop through harmonics
                harmIdx = maxIdx * harmonicMultiplier
                harmonics += noisyFFTData[harmIdx]
                harmonicMultiplier += 1
            if (harmIdx != maxIdx): #Harmonics founds, data is good
                fundFreqs.append(yAxis[maxIdx])
                harmonicVals.append(harmonics)
                THDSIvals.append(harmonics/fund*100)
                THDNSIvals.append(((np.sum(noisyFFTData)-
fund)/fund)*(ECF/ACF)*100) #Remember to convert from ACF to ECF when
summing to an energy value
            else: #No harmoincs found so data no good. Write NaN
                fundFreqs.append(np.nan)
                harmonicVals.append(np.nan)
                THDSIvals.append(np.nan)
                THDNSIvals.append(np.nan)
        else: #Threshold not met, write NaN
            fundFreqs.append(np.nan)
            harmonicVals.append(np.nan)
            THDSIvals.append(np.nan)
            THDNSIvals.append(np.nan)

    return THDSIvals, THDNSIvals, harmonicVals, fundFreqs
```