Claremont Colleges
Scholarship @ Claremont

Pomona Senior Theses | Pomona Student Scholarship

2008

Acoustical Measurement of the Human Vocal Tract: Quantifying Speech & Throat-Singing

Bryant R. Foresman
Pomona College

Follow this and additional works at: https://scholarship.claremont.edu/pomona_theses

 Part of the Astrophysics and Astronomy Commons

Recommended Citation

Foresman, Bryant R., "Acoustical Measurement of the Human Vocal Tract: Quantifying Speech & Throat-Singing" (2008). *Pomona Senior Theses*. 25.
https://scholarship.claremont.edu/pomona_theses/25

This Open Access Senior Thesis is brought to you for free and open access by the Pomona Student Scholarship at Scholarship @ Claremont. It has been accepted for inclusion in Pomona Senior Theses by an authorized administrator of Scholarship @ Claremont. For more information, please contact scholarship@claremont.edu.

Acoustical Measurement of the Human Vocal Tract: Quantifying Speech & Throat-Singing

Bryant Foresman
Senior Thesis, Pomona College Department of Physics
Spring 2008

Abstract:

The field of biological acoustics has witnessed a steady increase in the research into overtone singing, or "throat-singing," in which a singer utilizes resonance throughout the vocal tract to sing melodies with the overtones created by a vocal drone. Recent research has explored both how a singer vocalizes in order to obtain rich harmonics from a vocal drone, as well as how further manipulations of the vocal apparatus function to filter and amplify selected harmonics. In the field of phonetics, vowel production is quantified by measuring the frequencies of vocal tract resonances, or formants, which a speaker manipulates to voice a particular vowel. Thus, an investigation of throat singing is closely linked to human speech production. Formants are usually detected in vowel spectra obtained using Fast Fourier Transform algorithms (FFTs). An alternative method that provides much higher frequency resolution is external excitation of the vocal tract and measurement of the pressure response signal at the mouth's opening, which can be used to calculate the acoustic impedance spectrum. We demonstrate the use of such an "acoustic impedance meter" to measure the formant frequencies of common vowels as well as the oscillatory modes of simple resonant pipe systems. The impedance meter accurately measures fundamental pipe modes and a variety of formant frequencies with an uncertainty of 1 Hz. Finally, we assess how the impedance meter may be used to measure the unique resonances achieved by qualified throat singers.

Table of Contents

Chapter 1: Introduction to Throat Singing, Human Speech, and Acoustic Impedance Measurement Techniques

1.1: Introduction & Motivation

Throat singing, also known as overtone singing or biphonic singing, has aroused scientific curiosity since the 70s and 80s, during which period Mongolian and Tuvan music reached an international audience. [1] In essence, throat singing allows one person to create multiple notes simultaneously by manipulating the resonant qualities of the vocal tract. Tuva, a small province of Russia in Central Asia, is home to musicians who practice five sub-styles of throat singing, or *khoomei*, which means "throat" in the local language. *Khoomei* has origins in the animistic religion of Tuva and in this context allows individuals to commune with nature via sonic imitation of their surroundings. Overtone singing is practiced by other cultures in Central Asia, as well as by Tibetan monks, the Xhosa people of Africa, and select individuals in musical history. [2] Overtone singing is closely related to vowel production, and research in acoustical phonetics applies directly to a study of throat singing.

The fascinating physiology of throat singing provides many opportunities in acoustic analysis, modeling and measurement techniques. The physical complexity of the vocal tract and vocal chords has prompted a variety of approaches. Some researchers decompose the system into simpler parts to allow analytical or nearly analytical explanations, whereas others propose numerical techniques. Often, researchers employ techniques normally used to study vowel production and phonetics. Common to all models is the necessity of verifying results via experiment upon actual throat singers.

Spectral analysis of digital sound samples via Fourier transform algorithms is the most generally accepted method of data analysis. In this technique, resonant frequencies

are easily quantified and compared to theoretical results. [3] A recent adjunct of general spectral analysis is measurement of the acoustic impedance (essentially a sonic analogue of electrical impedance) at the opening of a throat singer's mouth. This technique, which involves directing an external sound source into the vocal tract and measuring a response signal, removes the spectral content of the vocal chords (which are not in operation during measurement) and replaces it with a predictable source that can be controlled by the researcher.

Exciting the vocal tract with an external source provides a higher degree of resolution in the frequency domain than vocal chord excitation. Whereas vocal chord excitation limits resonances to the harmonics of the speech fundamental, which will be at least 100 Hz (the speaking voice of a male in the bass range), external excitation can apply any audio frequency desired. [4] Further, in some cases a subject can morph between different vocal configurations during measurement, which provides dynamic information useful in understanding why *khoomei* sounds so vastly different than "normal" singing. A device capable of measuring acoustic impedance is applicable to a host of other scenarios, such as musical instrument analysis and predicting the resonances of complex physical systems. [3,5]

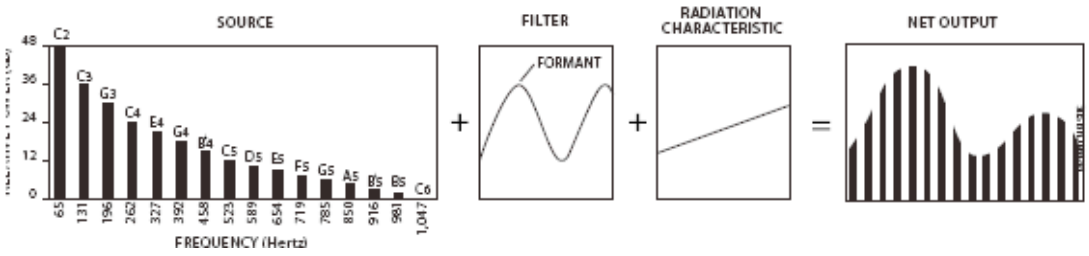
1.2: Understanding Speech Production and Throat Singing

A qualitative understanding of the principle behind human speech as well as throat singing is very useful in assessing acoustic impedance as an experimental quantity. When research into *khoomei* began, the origin of the additional pitches was unknown, which led to the "double source" versus "resonance" debate. The former theory held that

a secondary physiological source was responsible for generating the additional pitches, whereas the latter asserted that the pitches arise from a highly amplified resonance somewhere in the vocal tract. Recent research supports the resonance theory, as does the fact that the notes available to a throat singer are limited to the overtone series, suggesting a single source whose multiple modes of oscillation are exploited. Some forms of throat singing do employ a second sound source, namely the laryngeal folds, which usually oscillate one octave below the vocal chords. Consequently, their harmonics coincide with those of the vocal chords. [6] The affirmation of the resonance theory underpins a qualitative understanding of *khoomei.*

A basic three-part model of the human voice illustrates conceptually how both *khoomei* and human speech in general are achieved. The production of sound begins with the source (the vocal chords), which act as a transducer converting mechanical energy into sonic energy. Because the motion of the vocal chords is not sinusoidal, harmonics are generated along with the fundamental frequency. The power in these harmonics, which are integer multiples of the fundamental frequency, falls off nearly exponentially as frequency increases. Next, as the pressure waveform from the vocal chords propagates through the vocal tract, it is filtered due to the natural resonances of the tract dictated by its boundary conditions. This will be discussed in more detail shortly. Finally, a frequency dependent radiation characteristic is imposed on the waveform as it travels through the air outside the vocal tract. This final effect tends to attenuate lower frequencies and acts as a high-pass filter of sorts. The three-part process is summarized in Figure 1 below:

Figure 1: Three-Part Model of the Human Voice [2]



A throat singer controls parts 1 (source) and 2 (vocal tract filter) of this model to achieve simultaneous pitches as follows. In order to boost the power in the overtones generated by the non-sinusoidal motion of the vocal chords, a throat singer constricts his or her voice such that the chords burst open very quickly and remain shut for a longer period of time, effectively modifying the duty cycle of the waveform. This motion deviates *very* strongly from a sinusoid, and thus more power is allotted to the overtones. Thus, a throat singer begins the process mechanically with muscles in the throat, and generates the set of overtones that will later be perceived as separate, "additional" pitches. [2] All further manipulation of the additional pitches (once again, these are actually amplified overtones of the vocal drone) is achieved without modifying this basic sound produced by the vocal chords: a throat singer holds a vocal drone of constant pitch determined by the length and tension in the chords, and instead modifies the shape of the vocal tract (part 2, Figure) to choose which overtones to amplify and express as "separate" pitches.

The filtering characteristics of the vocal tract figure prominently in the fields of linguistics and phonetics. The inherent resonances of the vocal tract are called "formants," and the frequencies of these resonances are called "formant frequencies."

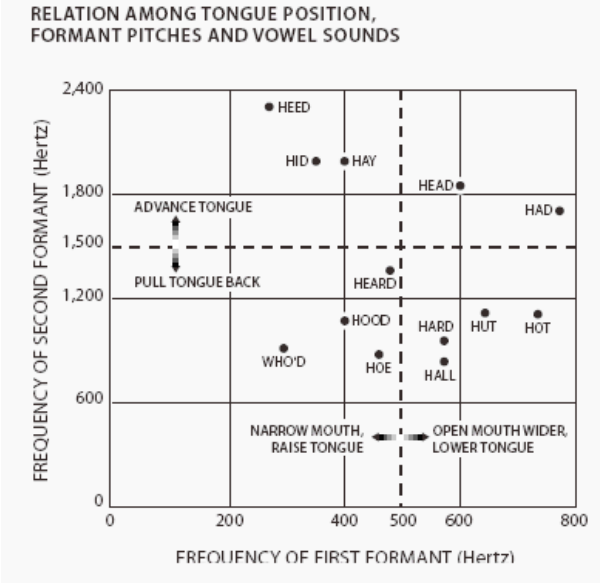
There are multiple formants for a given configuration of the vocal tract, and they are numbered with increasing integers as they increase in frequency (e.g. “first formant,” or F1). It is critical to note that, for a given vocal tract configuration (for example, the overall position assumed when enunciating the first vowel in the phrase “excellent thesis, man!”), the formant frequencies are fixed and do not depend on the fundamental pitch or overtones produced by the vocal chords. This is so because formant frequencies result from the *shape* of the vocal tract, and only when the shape is changed (e.g. when changing vowels) can the resonant frequencies change.

While the vocal tract is very complex, it can be approximated by a pipe with a closed end at the vocal chords and an open end at the mouth. The closed end (vocal chords) forms a pressure antinode, where air molecules collide with the chords and increase in density and thus pressure, and the open end (mouth) forms a pressure node, where air molecules are free to move and thus incapable of exhibiting a local increase of density, yielding a constant, minimal pressure region. In this model, formants correspond to distinct standing waves in the “pipe”, which are achieved when an odd-integer-number of quarter-wavelengths fit along the total path length of the pipe:

$$L = n\frac{\lambda}{4} \qquad (1)$$

where L is the length of the pipe, $n = 1,3,5\ldots$ and λ is the wavelength. In fact, the locations of the formants in the frequency domain are what dictate our perception of vowels. Figure 2 shows a formant map that plots the second formant frequency versus the first formant frequency for a variety of vowel sounds.

Figure 2: Formant Map for Vowel Sounds [2]



By changing the shape of the vocal tract, throat singers manipulate formant locations such that they coincide with the overtones of the constricted drone. This amplifies the drone overtones to an extreme extent. The perceived function of the formants is no longer merely to color the drone sound and create different vowels, but form separate, audible pitches from the drone's harmonics.

Constricting and widening various regions of the vocal tract is the key to changing formant frequencies. If the region around a pressure node is constricted, the local minimal pressure will take longer to force air molecules through the narrowed region, and the wave will slow down and decrease in frequency (remember that the sound wave is actually air molecules sloshing back and forth longitudinally). If, on the other hand, a singer constricts the region around a pressure antinode, where air molecules have roughly zero velocity, the density, and thus the pressure, will vary more quickly because the volume has decreased (by constriction). Thus, the wave speeds up and increases in

frequency. By a similar argument, widening a region has the opposite effect in both cases. Not only do throat singers match formant frequencies with overtones, they effectively merge formants together to create "double resonances." This may not seem intuitively obvious, but formant merging is demonstrated very effectively by measuring acoustic impedance at the mouth. [2]

1.3: Styles of *Khoomei*

Having developed a qualitative understanding of how throat singers achieve their unique sound, we proceed with an explanation of the sub-styles that comprise Tuvan *khoomei.* Within *khoomei,* there are three main styles that describe fundamentally different methods of both producing a rich drone as well as filtering this sound with the vocal tract. To any of these three styles may be added a number of ornaments or embellishments that modify the sound in some manner.

The first of the three styles is actually called *khoomei*, even though this is a blanket term referring to all styles of throat singing in general. The *khoomei* sub-style is performed in a singer's mid-range with moderate tension in the throat. The tongue sits in between the teeth in the bottom jaw and is raised or lowered to filter the drone. Movement of the lips and inner regions of the throat is also common. *Khoomei* produces a wide range of mild, simultaneous harmonics, in which a particular harmonic is stronger than the others and holds the melody.

The second style is called *sygyt*, which is performed high in the singer's range with a high level of tension in the throat. The tongue is cupped on the roof of the mouth, and air is allowed to flow out around the back upper molars and through the mouth.

Sygyt features a very strongly filtered sound, in which a single, whistle-like, melody-carrying harmonic is perceived. The drone sound is highly suppressed and is sometimes nearly inaudible. The whistle-like tone of *sygyt*, which sounds very similar to a sinusoid, suggests the presence of a very strongly peaked resonance in the sound spectrum and illustrates the extreme degree of filtering achieved by the singer.

The third style is called *kargyraa*, in which the ventricular folds (fleshy flaps above the vocal folds that are not normally employed in phonation) oscillate at half the frequency of the vocal chords. This produces a low drone one octave below the vocal drone. The numerous *kargyraa* harmonics (up to 3 or 4 may be perceived depending on the experience of the listener), are amplified and filtered by changing the shape of the mouth in a fashion similar to vowel production. *Borbangnadyr and ezengileer* refer to embellishments upon any of these three styles and do not alter the operation of either the sound source or the filtering mechanism. [1] The different techniques underlying the three styles of *khoomei* produce an acoustic palette that allows a wide range of musical expression.

1.4: Analytical Modeling of Throat Singing

Researchers have developed a variety of physical models to explain the different styles of *khoomei*. In agreement with the "resonance" model of throat singing, models emphasize how the vocal tract filters a sound source from the vocal chords. In the explanation of the *sygyt* style of throat singing, some research suggests that decomposing the vocal tract into a longitudinal (pipe-like) resonator and a Helmholtz resonator is successful in predicting characteristic resonances. [3] More detailed models decompose

the vocal tract into a series of truncated cones and account for visco-thermal energy loss, the yielding walls of the throat, and radiation into a partially open glottis and the environment outside the mouth. These in-depth models derive theoretical vocal tract transfer functions (VTTFs) that characterize how well the tract passes sound waves as a function of frequency. [6]

1.5: Measurement of Acoustic Impedance

Regardless of the theoretical model used to explain the physics of throat singing and human speech, experimental verification is required in all cases. For models that focus on the vocal tract's filtering effects, an experimental method that treats the vocal tract separately from the sound source (vocal chords) is preferred. Further, a method that is capable of measuring parameters of the vocal tract in a dynamic situation (e.g. when a throat singer morphs between configurations) is very helpful in understanding the resonant qualities of the vocal tract.

Acoustic impedance, defined as the pressure divided by the volume velocity[*] at a given location, is useful for quantifying vocal tract resonances. A device capable of measuring acoustic impedance (hereafter referred to simply as an "impedance meter") can be constructed in a variety of ways, but most apparatuses have similar features. Common to all impedance meters is the need to direct a synthesized sound source from a speaker/driver into the mouth normal to its opening. In addition, a microphone must be placed very close to the mouth's opening without significantly interfering with sound production. These issues are usually resolved by directing the sound source through an

[*] linear particle velocity times cross-sectional area (technically the integral of linear particle velocity dotted into the normal vector over the surface in question).

acoustic coupling such as an impedance matching horn, to the end of which a small microphone may be attached. [4]

Impedance meters differ from one another in two primary ways. The first is the method of exciting the vocal tract with the frequency range of interest. In one approach, the frequencies of interest may be rendered as discrete sinusoids that are slowly ramped upwards. [7] In the second approach, the frequencies of interest are synthesized into a single broadband source. Non-ideal transduction on the part of amplifiers, speakers and microphones results in the need to normalize the measured acoustic impedance by a reference impedance. [4] As we will see, implementing normalization is vastly different for the swept sinusoidal source versus the broadband source. In our case, this will result in a preference for the swept sinusoid source (see Chapter 2), but a convincing argument supports the use of broadband excitation as well.

Impedance meters also differ with respect to the experimental quantities they measure at the mouth's opening. Some impedance meters assume an ideal velocity-current source from the end of the impedance matching horn and consequently measure only pressure. [4] Others measure pressure and velocity simultaneously with two different transducers (a microphone and a velocity sensor). [7] The former of the two techniques can be justified theoretically and is not only more cost-effective, but also interferes less with the subject's sound production due to fewer components near the opening of the mouth. Finally, impedance meters may offer real time display of measured spectra, which can be useful in providing feedback during experimentation. [4]

1.6: Design Goals & Analytical Incentives

My goal is to design a functional impedance meter using readily available equipment and software. I plan to approach the construction of an acoustic impedance measurement device from the following perspectives. The device should feature inexpensive hardware that interfaces easily with PC soundcards. I will synthesize sound sources and perform spectral analysis using *MATLAB* and will make my code available for further study.

My project is above all one in design and construction, including proper choice of components and developing robust code in *MATLAB*. Consequently, I will devote the majority of my time to addressing these issues. In order to verify that my impedance meter functions properly, I will first measure resonant pipes, for which oscillatory modes may be determined analytically. Second, I will measure the formant frequencies of a variety of vowels and compare my results to accepted values. I will give an indication of how my apparatus may be used to investigate the acoustics of throat singing without making an effort to do so. The reasons for this are twofold. First, as previously mentioned, design issues are of primary concern in this project. An effective apparatus based on solid principles drawn from the literature is of the highest priority. The second reason is that I will be the experimental subject and no not claim to produce genuine throat singing. My apparatus will function best with the participation of a qualified throat singer.

The generosity of the Pomona College Physics Department has allowed me to build a relationship with *Chirgilchin*: *Master Throat Singers from Tuva*, from whom I obtained a number of recordings. By constructing an impedance meter, I hope to open

the possibility of making measurements with *Chirgilchin* at some point in the future. Finally, it is my hope that any scientific progress in the understanding of human vocal acoustics will contribute to an aesthetic appreciation of throat singing as well as a basic idea of how humans produce and manipulate sound.

Chapter 2: Background and Theory of Impedance Meters

2.1 Introduction to Impedance Meters:

While all impedance meters share certain features, they may function very differently, particularly with respect to signal generation and the experimental quantities they measure. Differing assumptions accompany the various models and are worthy of review. Further, the various types of impedance meters serve differing analytical goals and necessitate particular computational capabilities and monetary budgets. It is my goal to present a broad picture of the issues relevant to construction and to justify my specific approach.

2.2 Basic Construction:

Basic principles of construction apply to all impedance meters. Impedance meters create an acoustic current source by means of driving sound through some sort of impedance matching transmission line. The transmission line is almost always an impedance matching horn, which functions to channel energy with minimal reflections. Often, but not always, this current source can be treated as an ideal velocity source (i.e., one whose velocity flow is not affected significantly by a load, much as an ideal current source in electronics provide a theoretically constant current). [7] The parameters and materials of our current source qualify it as an ideal velocity source. [4] For a constant velocity, pressure will be proportional to the acoustic impedance (recall that acoustic impedance is sound pressure divided by volume velocity). Consequently, we only need

to measure the pressure at the opening of the mouth in order to measure the acoustic impedance. [4]

The exact opening of the mouth is the ideal location for the outlet of the impedance matching horn, for we wish to measure the acoustic impedance at this location. However, this will interfere unduly with the subject's ability to reproduce natural configurations of the lips and mouth. If we place the outlet of the horn directly outside the subject's mouth, we may treat the system comprised of the vocal tract and the half-space around the mouth as a lumped acoustic element driven by the source. In this case, the vocal tract and the half-space will be driven in parallel, and we may apply acoustic circuit analysis to the problem. [4] We will derive theoretical results shortly, but for the meantime we return to construction issues.

An example signal pathway for an impedance meter in the literature is shown in Figure 2.1. [4] This setup dispenses with velocity measurement, and consequently only a pressure transducer (microphone) exists. The low pass filter element is used for speech signal suppression and will not be necessary in my apparatus, which will not record speech simultaneously to excitation. The reason for permitting speech during excitation is to help the subject maintain a constant vowel conformation via aural feedback. The speech signal is later suppressed. For simplicity, I dispense with speech during measurement and require the subject to maintain a constant configuration by sensation alone. The setup in Figure 2.1 requires two computers with interfaces to analogue devices (analogue-to-digital converters (ADCs) and digital-to-analogue converters (DACs)).

This setup can be greatly simplified if we use the soundcards installed on the computers to perform our conversions. Although soundcards installed on typical commercial computers are of relatively low quality and are subject to distortion issues, we will see that the use of a proper amplifier can minimize the distortion of the soundcard's analogue waveform output. Two soundcards installed on one computer would perhaps be more elegant than a single soundcard in two computers, but we choose the latter for simplicity in this experiment. More detail on the setup employed in this project can be found in the procedure section.

Figure 2.1: Example Signal Path

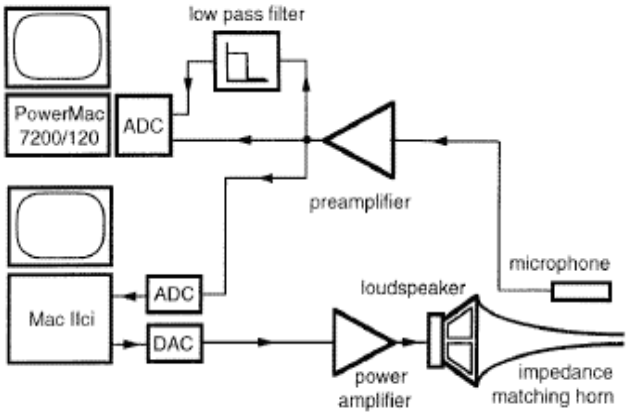


Figure 2. A schematic diagram of the apparatus used for real-time acoustic vocal tract excitation (RAVE) during phonation.

2.3 Signal Generation

An important factor in any impedance meter is the manner in which it will generate an excitation signal. In one type of device, the sound source is a distortion-free, sinusoidal sweep-tone that covers the frequency range of interest. As the frequency is ramped up by discrete intervals, the response of the vocal tract is measured for each frequency, and spectral analysis is performed. The impedances calculated on each

discrete frequency are compiled into a spectrum covering the entire frequency domain. The second method involves generating a broadband source with all desired harmonics synthesized into a single waveform. This broadband source captures the response impedance spectrum in one pass. Each of the two methods has distinct advantages and disadvantages, and our current task is to evaluate these and make an appropriate selection.

2.3.1 Swept Sinusoidal Method

Elimination of transients from the response signal is an important experimental step for the swept-sinusoidal and broadband methods, and will be discussed briefly before we explore the properties of swept-sinusoidal meters. We may calculate how long our signal requires to complete one round trip of the vocal tract and assume a steady state after roughly ten of these round trips. This gives:

$$T_{rt} = 2\frac{d}{c} \qquad (2.1)$$

$$T = 10T_{rt} \qquad (2.2)$$

where T_{rt} is the time required for one round trip of the vocal tract, d is the length of the vocal tract (1 way), c is the speed of sound in air, and T is the time required for a steady state. For a vocal tract d = 17.5 cm long (average male), pressure waves traveling in air at c = 343 m/s will require

$$T_{rt} = 2\frac{d}{c}$$
$$= 2\frac{0.175m}{343ms^{-1}}$$
$$= 0.00102s$$
$$= 1.02ms$$

for a single round trip from equation (2.1) and thus approximately 10 milliseconds from equation (2.2) for a steady state. We will have the opportunity to visually confirm this steady state during the experiment, and *T* may be increased if necessary.

An advantage of the swept-sinusoidal method over the broadband method is apparent when implementing a response normalization algorithm. Before we proceed, a brief discussion of normalization is necessary.

Because components in the apparatus such as amplifiers, speakers and microphones do not represent ideal transducers, and because the apparatus will absorb acoustic energy as well as exhibit its own resonant frequencies, an essential feature in the measurement process is normalizing the measured acoustic impedance by a reference impedance. If we failed to normalize our measurements by this reference, we would actually be detecting peaks in the response spectrum that we due solely to the peculiarities of our apparatus and its preference for transmitting certain frequencies over others. A discussion concerning selection of the reference impedance will follow shortly.

The method of swept sinusoids presents us with a fairly simple solution to normalization. We calculate the power in each discrete sinusoid and store this information for later use. Since we will ultimately be concerned with a ratio of powers for each sinusoid (the measured value divided by the reference value), we need not fuss

over what type of "power" or "amplitude" we are measuring. MATLAB's vector manipulation capabilities provide an ideal setting in which to implement normalization. We may calculate the power in the sinusoid in two ways. The first involves using a Fast Fourier Transform (FFT) algorithm, and selecting the maximum in the absolute-squared-magnitudes of the transform values (to eliminate complex numbers). This method is essential to the broadband excitation method, but the swept sinusoid method offers an alternative.

Since we are theoretically dealing with pure sinusoids in the swept method, the FFT will contain a single value corresponding to the frequency of this sinusoid. Provided we can demonstrate that our signal is sufficiently sinusoidal so as to excite predominately a single resonance in the vocal tract, we may use the root-mean-square (RMS) amplitude of the response signal to calculate the power for each sinusoid. This has a major advantage over the FFT method. We no longer need to window our response to prevent spectral broadening and need not consider artifacts from MATLAB's FFT algorithm.

A distinct disadvantage to the swept sinusoid method is the relatively long time required for data acquisition. We may quantify this time as follows. We will divide our frequency range of interest into intervals with a desired spacing determined by

$$\Delta F = \frac{F_m - F_0}{n} \qquad (2.3)$$

where ΔF is the frequency spacing between sinusoids, F_m is the maximum frequency, F_0 is the minimum frequency, and n is the number of sinusoids in this interval. Once each sinusoid has reached a steady state, we would like to sample each sinusoid for a given

period of time. We choose our steady-state sampling interval for *each* sinusoid to be 50 periods of the sinusoid with the lowest frequency (we must choose the same interval for each sinusoid so as to compare RMS values). Each sinusoid in the series will contain this duration, yielding a total *sampling* time of

$$\tau = 50T_0 = \frac{50}{F_0} \qquad (2.4)$$

where τ is the steady state sampling time, and T_0 is the period corresponding to the sinusoid with lowest frequency F_0. We finally allow a time T seconds from equation (2.2) before each sinusoid to eliminate transients as well as T seconds between sinusoids. Thus, for each sinusoid, we require a time t

$$t = 2T + \tau \qquad (2.5)$$

Multiplying by n from equation (2.3) and rearranging terms, we achieve a total time required for the measurement:

$$T_t = \left(\frac{F_m - F_0}{\Delta F}\right)\left(2T + \frac{50}{F_0}\right) \qquad (2.6)$$

If we were to measure the entire range of audible frequencies from 20 Hz to 20 kHz with a frequency spacing of 20 Hz, this would require

$$T_t = \left(\frac{20000-20}{20}\right)\left(2(0.0102)+\frac{50}{20}\right)\text{sec}$$
$$\approx 2500\,\text{sec}$$
$$\approx 40\,\text{min}$$

This is prohibitively long, especially considering the fact that requiring a subject to assume a configuration for that length of time would surely constitute some form of torture. Further, results would be subject to inevitable shifts in mouth, tongue and lip position.

Luckily, we have no such need to measure the entire audio spectrum. Vowels can be effectively identified by the first two formants alone. Recall that a formant simply designates a resonance of the vocal tract, where the first formant (F1) is the fundamental resonance of the tract. In most situations, however, information about the first three formants is useful. [4] Modeling the throat as a close-open ended pipe yields formant frequencies that agree roughly with experiment. [2] Since an odd-integer number of quarter wavelengths must fit in the length of the throat (modeled as a pipe) due to the boundary conditions, we have

$$d = n\frac{\lambda}{4} \qquad \text{n = 1,3,5…} \qquad (2.7)$$

where d is the length of the throat and λ is the wavelength. Using c=λf, we achieve

$$f = \frac{cn}{4d} \qquad (2.8)$$

where f is the frequency and c is the speed of sound in air. For a typical male tract of 17.5 cm, we achieve for the first three formant frequencies

F1= 490 Hz (first formant)

F2= 1470 Hz (second formant)

F3= 2450 Hz (third formant)

These formants will be shifted slightly in the frequency domain due to the particular vowel configuration adopted, as discussed in Chapter 1. At this point in the analysis, we do not apply the theory of effective length* to the vocal tract length *d,* for we are only attempting to approximate the formant frequencies. Analysis of various throat singing styles has demonstrated that resonant phenomena of interest occur below 2500 Hz as well. [8] Consequently, we only need to measure up to roughly 2500 Hz, which we will extend to 3125 Hz to allow a variation of 25% in the positive direction for F3. In addition, we are not interested in frequencies below F1 and choose 375 Hz as our lowest frequency, which is a variation in F1 of 25% in the negative direction. Using these values with an improved frequency spacing of 10 Hz in equation (2.6) yields

$$T_t = \left(\frac{3125-375}{10}\right)\left(2(0.0102)+\frac{50}{375}\right)\text{sec}$$
$$\approx 43\,\text{sec}$$

* Sound waves in pipes with at least one open end are affected by the nonzero impedance of the half-space around the open end(s), which allows for a nonzero, albeit small, pressure at the physical boundary of the pipe. It is as if the sound waves actually achieve a node slightly beyond the physical boundary of the pipe, hence the concept of an "effective length," which is always slightly larger than the physical pipe length and depends on the geometry of the open end(s).

which is very reasonable, minimizing the likelihood of morphed configurations and excessive drooling. This represents a lower limit on the time required. We will see that lengthening the sinusoids as well as the time between sinusoids leads to cleaner data without drastically increasing data collection time. Finally, when we have identified the rough locations of formants and other resonances via this first order approach, we may resolve them further in a new measurement by centering our frequency range symmetrically about the resonances and choosing a very fine frequency spacing.

To verify that the impedance meter correctly measures the resonances of complex systems such as the vocal tract, it is necessary to measure a simple system for which resonances can be readily calculated from theory. One such system is an actual pipe with one open end and one closed end (while the pipe model approximates the vocal tract, it not fully correct). This pipe will have resonances given very accurately by equation (2.8) upon substitution of an effective length, which will be discussed shortly. This model assumes that we can treat the oscillatory mode within the pipe as one-dimensional. In acoustical physics, this condition is usually satisfied by requiring that a quarter-wavelength is greater in size than any system dimension not parallel to the axis chosen for modeling 1D waves (in this case, the 1D axis chosen lies along the length of the pipe):

$$\frac{\lambda}{4} > d \qquad (2.8.1)$$

where λ is the wavelength and d is a system dimension not parallel to the main 1D system dimension. Rearranging and substituting $\lambda = c/f$, we obtain

$$f < \frac{c}{4d} \qquad (2.8.2)$$

where f is the frequency and c is the speed of sound. For a pipe, we let d equal the diameter of the pipe, which could support oscillatory modes. The frequency obtained upon this substitution will dictate an upper limit on the frequencies at which theory will agree with experiment for the pipe system. This said, we may still attempt to measure resonances above this frequency, but we cannot expect that they will be given accurately by equation (2.8) used with the appropriate effective length.

For a cylindrical pipe with one unflanged (not tapered outwards) open end, theoretical calculations dictate an effective length of

$$L_e = L + 0.61r \qquad (2.8.3)$$

where L_e is the effective length, L is the physical pipe length, and r is the radius of the pipe. [9] If two unflanged open ends exist, the radial term in equation (2.8.3) must be doubled. This effective length is used in place of the physical length d in equation (2.8).

As a further check upon the efficacy of the impedance meter, we measure a double-open ended pipe into which an integer number of half-wavelengths must fit, yielding

$$f = \frac{nc}{2L} \qquad (2.8.4)$$

Where n = 1,2,3… This system will require two end-corrections in creating the effective length, so the radial term in (2.8.3) is doubled.

In measuring an inanimate object, we can obtain a much better signal to noise ratio by increasing the power through the speaker. This will create a sound pressure level (SPL) in the vicinity of the horn that is potentially unsafe for human exposure, in which case the researcher must take the precaution of wearing hearing protection.

2.3.2: Broadband Method

In the broadband excitation method, data is readily collected in one fell swoop. Since measurement time is not an issue, a steady state excitation can be sample for seconds or more, providing excellent spectral resolution via FFT analysis. In this sense, the broadband method is ideal. However, implementing a normalization algorithm is more difficult. An identical number of samples should be selected and windowed properly. Spectral broadening due to discontinuities at the beginning and end of the waveform will be inevitable, and thus it is difficult to determine exactly where peaks exist in the frequency domain. Writing a program in MATLAB to reliably select peaks is thus difficult and subject to error. Because it is relatively straightforward to implement the normalization algorithm for the swept sinusoidal method in MATLAB, we choose to implement this form of excitation.

2.4 Experimental Quantities: Pressure and Velocity

The definition of acoustic impedance, namely pressure divided by volume velocity, implies that an impedance meter must measure these two experimental quantities. However, since most meters employ an impedance matching transmission line such as an exponential horn, which essentially supplies a constant velocity regardless of the load, the acoustic impedance will simply be proportional to the pressure.

Some researchers have detected small variations in the velocity as a function of frequency, and espouse velocity probes that accurately measure particle velocity, which can be converted into volume velocity. [8] However, these variations are small and unlikely to affect strong peaks in the impedance spectrum. Such probes must be calibrated and introduce the possibility of systematic error. In addition, particle velocity may differ across the cross sectional area of a system of interest, especially around edges and corners, where frictional forces exist. Finally, velocity probes are not inexpensive. Thus, we have a strong preference for measuring pressure alone.

The exponential horn used in this project has been demonstrated to be an effective velocity current source that is independent of load for the frequency range of interest. [4] Thus, we have the benefit of a design that requires only one transducer, namely a microphone. This eliminates the difficulty of combining data from two instruments. We need not worry about introducing a phase difference between pressure and velocity data due to differences in processor speeds or data selection techniques. This setup is more cost-effective, which meets a primary design incentive. Finally, past studies demonstrate that a single pressure transducer yields data consistent with theory.

2.5 Theory of the Vocal Tract Impedance Spectrum [4]

Critical to understanding the impedance of the vocal tract is developing a model for the environment directly outside the vocal tract. The half-space outside the subject's mouth presents the vocal tract with an external radiation impedance given by:

$$Z_E = \alpha z \frac{jkr}{1 + jkr} \qquad (2.9)$$

where k is the wavenumber, r is the radial distance of the opening, z is the specific acoustic impedance of the medium in which the sound propagates (air in this case), and α is a geometrical factor that depends on the solid angle into which the sound is allowed to propagate. If $kr << 1$, equation (2.9) simplifies to

$$Z_E \approx jkr\alpha z \qquad (2.10)$$

Thus, the external radiation impedance is linear in frequency under these conditions. We know that

$$k = \frac{2\pi f}{c} \qquad (2.11)$$

Thus $kr << 1$ stipulates that

$$f << \frac{c}{2\pi r} \qquad (2.12)$$

For a radius of 1 cm, equation (2.12) takes the value of roughly 5.5 kHz. Thus, we are fairly safe if we are interested in frequencies up to 3 kHz. Even though this is not much lower than 5.5 kHz, equation (2.9) does not exhibit any local maxima, which will be crucial in our next step of reasoning.

If we consider the vocal tract loaded by the external radiation impedance, basic acoustic circuit theory tells us that a resonance will occur when the complex parts of the vocal tract impedance and the radiation impedance cancel:

$$\mathrm{Im}(Z_{VT}) = -\mathrm{Im}(Z_E) \qquad (2.13)$$

where Im(x) denotes the imaginary part of the complex number x. In order to utilize this condition, we now draw further from acoustic circuit theory. Our excitation source will drive the vocal tract impedance and the external radiation impedance in parallel, for the pressure must be single-valued at the "junction" of these two acoustic circuit elements:

$$Z_p = \frac{1}{1/Z_{VT} + 1/Z_E} \qquad (2.14)$$

where Z_p represents the impedance of the parallel combination of the vocal tract impedance Z_{VT} and the external radiation impedance Z_E. Equation (2.10) implies that

Z_E is linear in frequency, whereas Z_{VT} will have strong peaks due to resonances. Thus, equation (2.14) indicates that Z_p will mirror these extrema. Finally, we see that the vocal tract resonance condition (2.13) will lead to a real-valued maximum for Z_p in equation (2.13). Thus, Z_p has maxima at the resonances of the vocal tract, and we may detect these resonances experimentally by measuring the impedance of the parallel combination of the vocal tract and radiation impedances. Further, these maxima will be followed by steep drops in the spectrum, for the reactance of the vocal tract changes sign very quickly at a resonance, leading immediately to an anti-resonance. This provides a reliable criterion for detecting resonances.

Chapter 3: Apparatus, Construction, Setup and Procedure

3.1 Overview of Apparatus and Materials

3.1.1 The Impedance Meter: Introduction to Design and Construction

The impedance meter in this project measures the pressure response at the mouth when the vocal tract is excited by a series of swept sinusoids. The excitation is synthesized in MATLAB on one computer, amplified, and sent to the speaker and horn. The pressure response is recorded and processed in MATLAB on a second computer. In detailing construction issues, we will follow the signal path and thus illustrate the apparatus's chronology.

3.1.2 Equipment

What follows is a comprehensive list of equipment used to construct and assess the impedance meter.

1. 2 *Dell Optiplex GX620* Desktop Computers with *Microsoft Windows XP Pro* Service Pack 2, 3.00 GHz Pentium 4 CPUs, 1.00GB RAM, and *SoundMAX Integrated Digital Audio* soundcards
2. *MATLAB* 7.0.4.365 (R14) Service Pack 2 (type >>version in MATLAB to determine your version)
3. *Pasco Scientific* PI-9587C Digital Function Generator-Amplifier
4. *VIFA* K10MD-19 4 Ohm, 3" speaker

5. Exponential Horn Construction: Pine stock, 3.5" OD Schedule 40 ABS, Plaster of Paris, paraffin wax, wood filler, Silicone lubricant
6. *Realistic* Electret Tie Clip Microphone Cat. No. 33-1052 (Radio Shack)
7. *CircuitSpecialists.com* MS8209 Auto Ranging Multimeter with Sound Pressure Level (SPL) measurement capability
8. *Tektronix* TDS 1002 Digital Oscilloscope

3.1.3 Signal Path

Figure 3.1 illustrates the signal path for the impedance meter used in this experiment. Computer 1 is used for signal generation, and Computer 2 for recording and post-processing. Figure 3.2 shows a picture of the experimental setup in the basement of the Physics Department at Pomona College. Computer 1 is on the left, and the amplifier and speaker/horn components are between the two computers. A laboratory stand for positioning objects in front of the horn sits to the left of the speaker/horn.

Figure 3.1: Signal Path

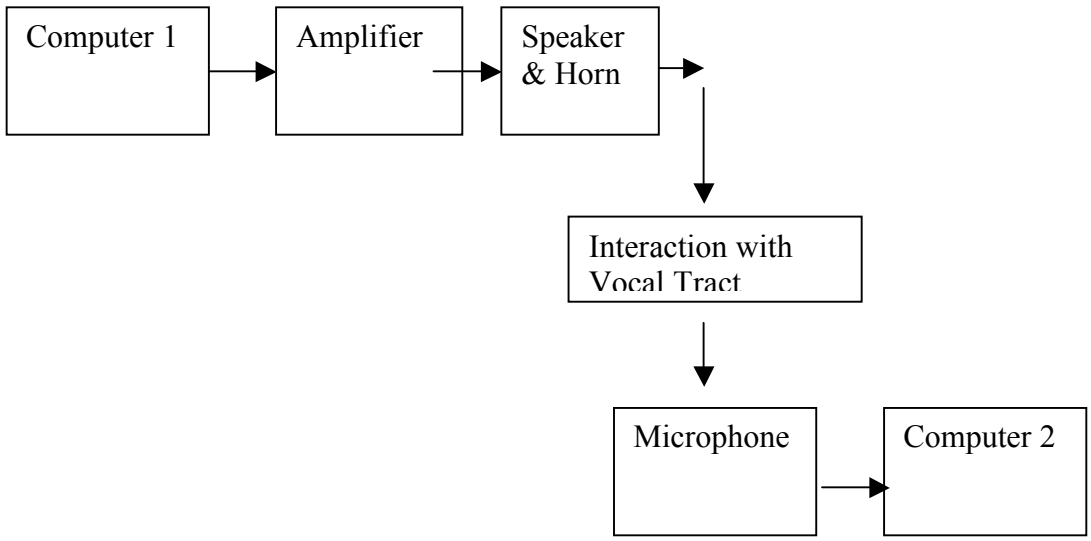


Figure 3.2: Experimental Setup Picture

3.2 Coding and Construction

3.2.1 Code

The code for this apparatus was written in MATLAB 7.0.4.365 (R14) Service Pack 2. The two computers (Figure 3.1) were used because MATLAB does not support simultaneous recording and playback, even if one's sound card is full-duplex (i.e. supports simultaneous recording and playback). This introduces a need to eliminate from all samples the time between the start of recording on one computer and the initiation of a series of sinusoids on the other. This priority dictated code that allowed the user to interact graphically with raw sound data, in order to both eliminate periods of silence as well as unwanted transients in the response. It should be brought to attention that when

recording and playback is performed in MATLAB, conversion of files into .wav format is not necessary. Waveforms are stored directly in MATLAB directories and are referred to as vectors within MATLAB. This greatly facilitates manipulation of the recordings.

A series of sinusoids of duration, amplitude and frequency spacing is created by the user on computer 1. Waveforms are stored as vectors, and MATLAB's *audioplayer* object allows convenient playback from the main console window. Computer 2 contains a program that records these sinusoids upon prompting the user to start recording. Recording was automated so that a user could record a series of vocal configurations one after the other.

Post-processing was achieved on computer 2. A program allowed the user to graphically select the beginning and end of a steady-state region in the first sinusoid generated in the series. Since all subsequent sinusoids were spaced evenly, these beginning and end values were used to evaluate the RMS amplitude of every sinusoid. Two different programs of this sort existed: one for free-field measurements (i.e. mouth closed in correct position) and one for vocal tract measurements. The former generated a free-field impedance spectrum, and the latter generated a vocal-tract impedance spectrum and normalized it by the free-field spectrum. Both of these programs allow the user to eliminate any DC component in the recorded waveforms by graphically identifying the DC offset.

When measuring resonances in regions of low signal-to-noise ratios, identical response signals were taken many times and averaged to eliminate the zero-mean noise in the response waveforms. New programs were written to automate this process.

3.2.2 Exponential Horn and Mounting of Hardware

The exponential horn used in this experiment is shown in Figure 3.3 along with the aluminum speaker mount and the speaker. Two pine molds (Figure 3.4) were created on a wood lathe (courtesy Glenn Flohr, Physics Department machinist) and coated in melted paraffin wax and Silicone lubricant to prevent bonding between the wood and Plaster of Paris that would be used to cast the horn. The shape of the horns was as exponential as could be achieved on the lathe. Only the horn on the right was used, for we were not able to extract the left horn from the cast. The wooden mold was placed vertically (large opening down) in the ABS during casting and hammered out after 10 minutes, before the plaster had completely hardened. The horn was 11 13/16 inches long, the large opening was 3 inches in diameter, and the small opening was 11/16 inches in diameter.

Figure 3.5 shows the hardware mounted on the outlet of the exponential horn. The semicircular cowl was 30 mm in depth (along horn axis) and was mounted so as to achieve adjustable height and depth. The microphone was mounted at roughly 40 degrees from the plane of the horn opening to insure reception of the response signal. The microphone was coincident with the bottom surface of the 13/16” horn outlet hole so as to prevent reflection back into the horn.

Figure 3.3: Exponential Horn with Speaker Mount and Speaker

Figure 3.4: Pine Horn Molds (smaller mold on right was used)

Figure 3.5: Mounted Hardware

3.3 Setup, Calibration, and Safety

3.3.1 MATLAB & Soundcard Output

Due to voltage limitations on certain hardware devices, it is useful to track the signal voltage as it propagates from the computer to the speaker. Figure 3.6 shows a plot of the soundcard output voltage versus sinusoid amplitude (0 through 1) in MATLAB for maximum soundcard volume. The linear relationship with slope 1.4992 V was used to protect the amplifier. A sinusoid amplitude of 0.75 in MATLAB was used to create a 1.12 V signal into the amplifier.

3.3.2 Safety: SPL and Speaker Wattage

The most important factor in determining how strongly we will amplify the soundcard signal is auditory safety for the experimental subject (me). The SPL meter was used to limit the SPL to a maximum of 80 dB at the outlet of the horn for all frequencies from 375 to 3125 Hz. [4] The natural resonances of the horn greatly amplified the signal and were the limiting factor in determining the maximum power through the speaker.

Figure 3.6: Soundcard Voltage Test

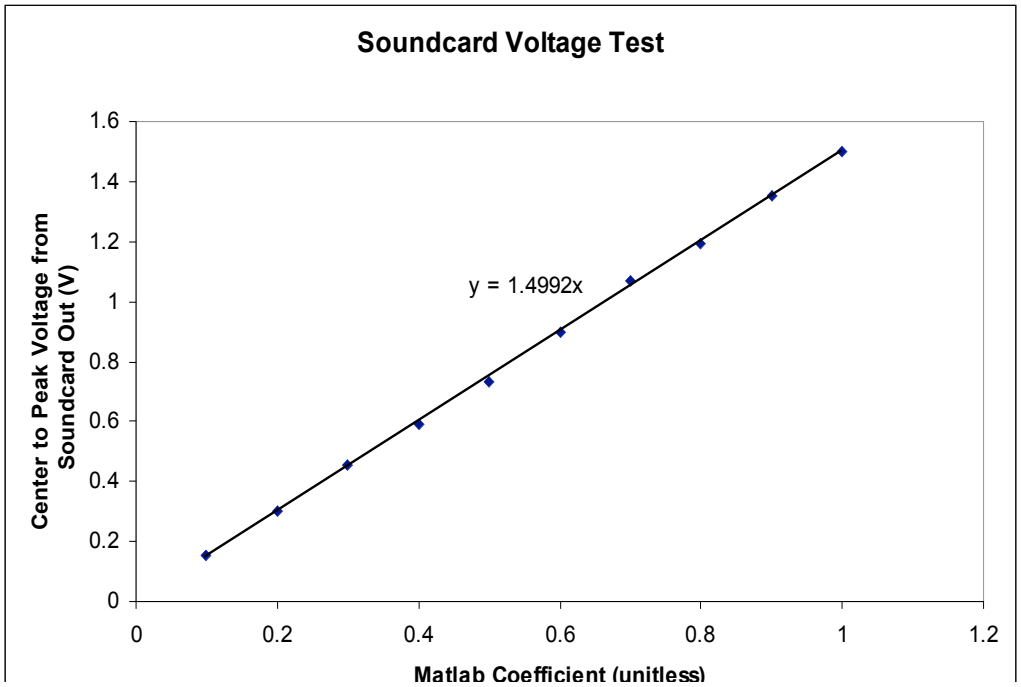
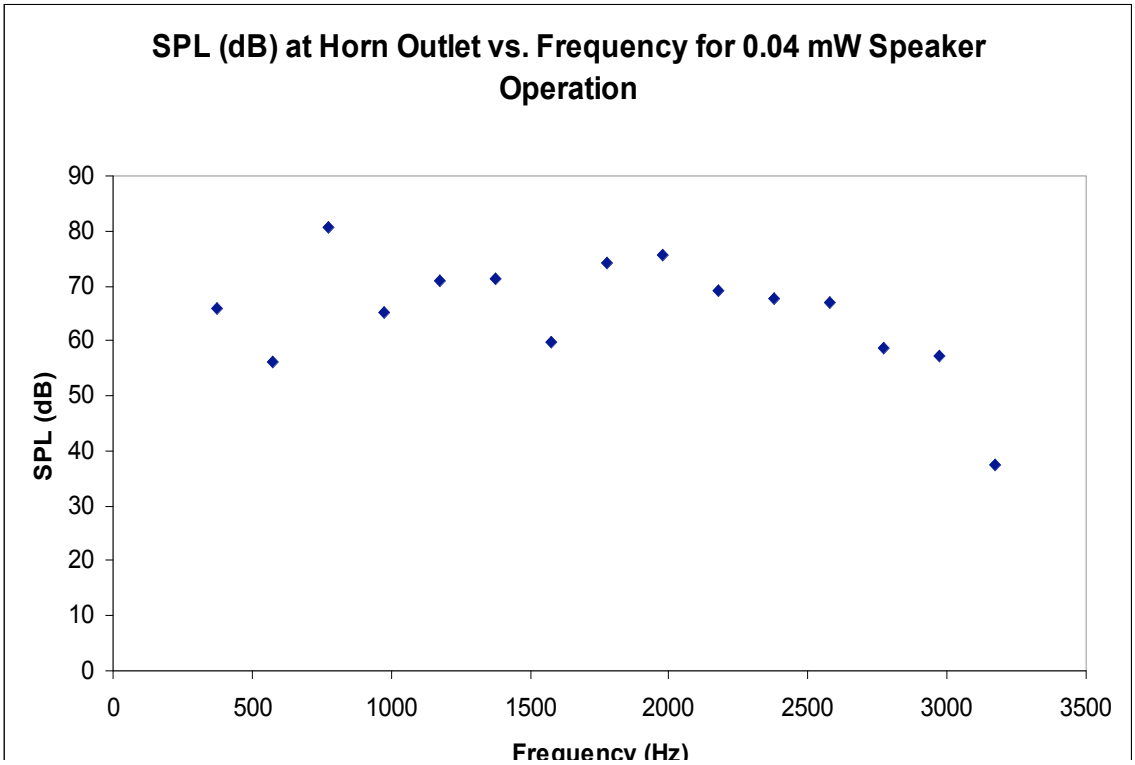


Figure 3.7 shows a plot of the SPL at the horn outlet as a function of frequency for 0.04 mW speaker operation (0.009 V_{rms} across 4 Ω). The SPL safety limit of 80 dB was reached well before the generally accepted 1 W limit for driving speakers with pure sinusoids (sinusoids are particularly hard on speakers, especially at lower frequencies). Thus, 0.04 mW was chosen as the operating power for vocal tract measurements. For measurements on non-human objects such as pipes, the operating power was chosen to be 0.5 W (1 V RMS) to improve the signal-to-noise ratio. I wore earphones during all measurements.

Figure 3.7: SPL vs. Frequency for 0.04 mW Speaker Operation



3.3.3: Speaker Harmonics and RMS Amplitude

In order to argue that RMS amplitude can be used to determine the response to each sinusoid, we must show that the harmonics resulting from non-ideal transduction in the system hardware are negligible. To illustrate this, we measure the output spectrum at the horn outlet for 4 frequencies spanning our frequency range of interest, namely 375 to 3125 Hz. Figure 3.8 shows these four spectra, with letters a) through d) corresponding to 375, 1300, 2240, and 3175 Hz, respectively. These plots were generated in the free audio analysis package *Audacity*, which uses FFT analysis to calculate the power spectrum. The y-axis unit (1 box) is 10 dB, and the x-axis unit is 1 kHz. In all cases, the difference between any harmonic and the fundamental (denoted by the large spike with a vertical

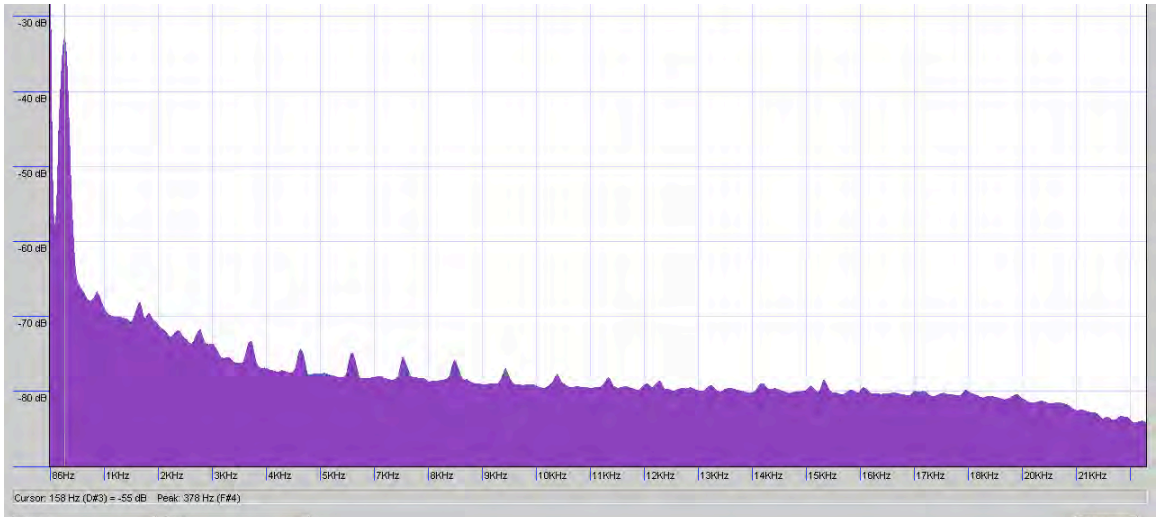
cursor) is at most 30 dB. Since we are dealing with power, we have the following expression for dB:

$$L_{DB} = 10\log\left(\frac{P_1}{P_0}\right) \qquad (3.1)$$

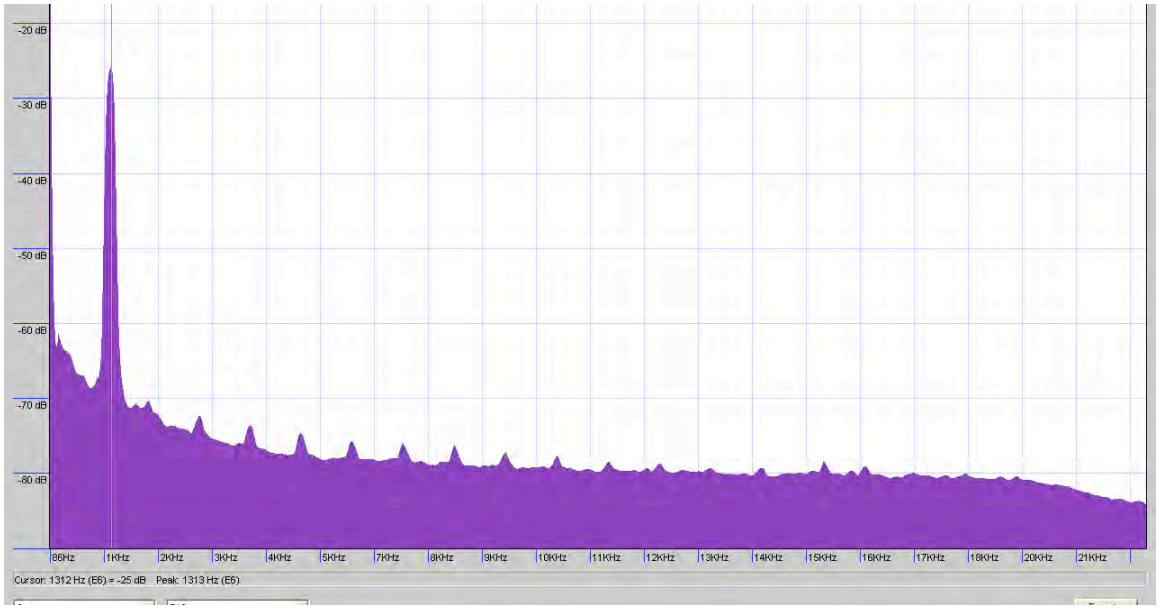
where L_{DB} is the ratio in dB, P_0 is the reference value (the fundamental in this case) and P_1 is the value to which the reference is compared (in this case, any harmonic). Thus, all harmonics are at most 3 orders of magnitude lower than the fundamental.

Figure 3.8: Output Spectra for 375, 1300, 2240, and 3175 Hz

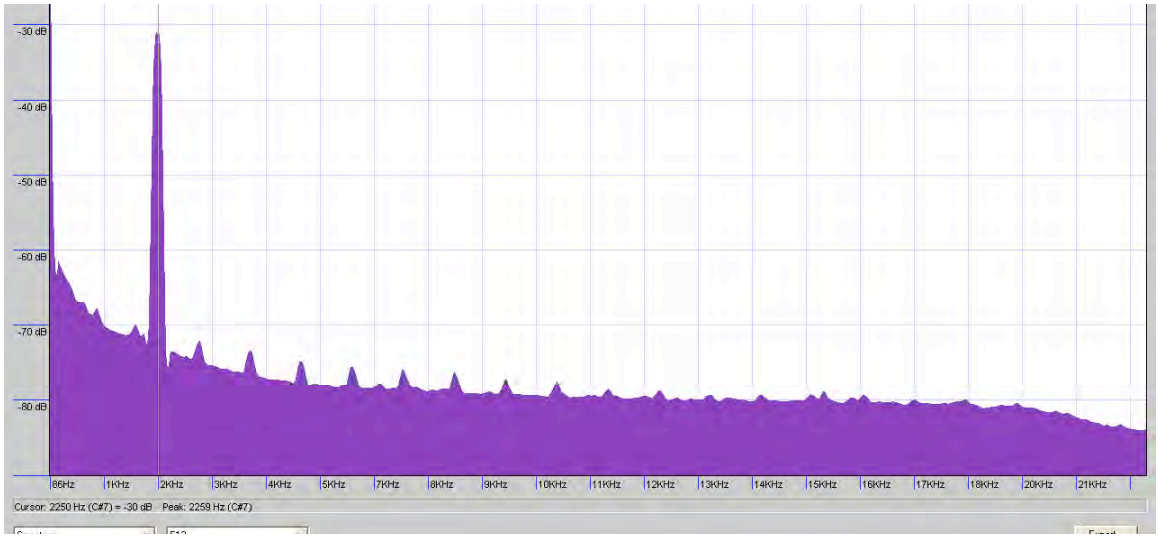
a) 375 Hz



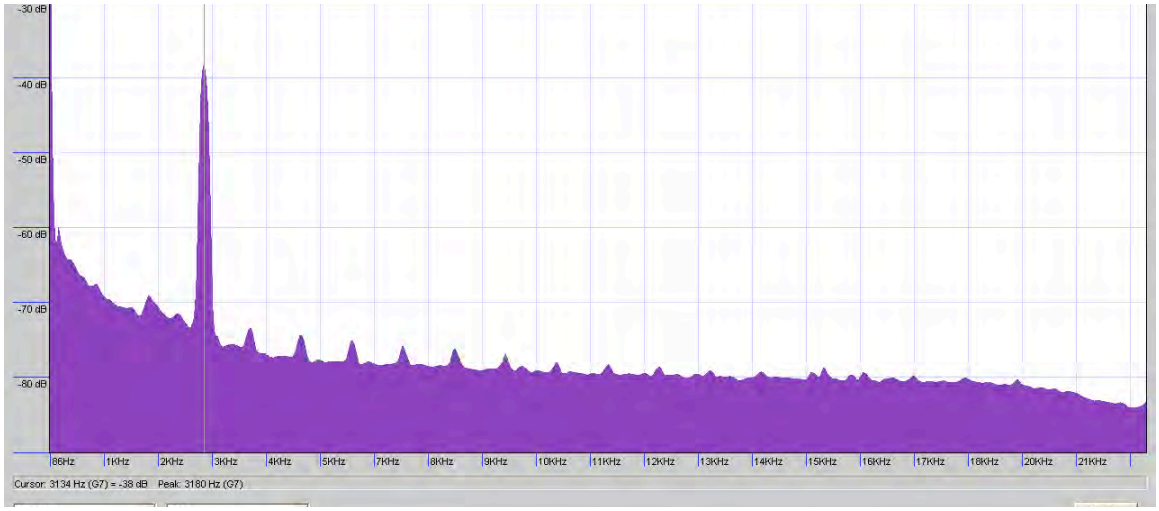
b) 1300 Hz



c) 2240 Hz



d) 3175 Hz



We now attempt to justify the use of RMS amplitude on a semi-theoretical basis, pointing out that the technique has already been effectively demonstrated in the literature. [7] RMS amplitude for a continuous function is given by:

$$f_{RMS} = \sqrt{\frac{1}{(b-a)}\int_{a}^{b}\left[f(x)\right]^{2} dx} \qquad (3.2)$$

where $f(x)$ is a continuous function, and f_{RMS} is the RMS amplitude of the function over the interval $x = a$ to $x = b$. In our case, $f(x)$ is a periodic function of time and is actually converted into a discrete time series upon audio sampling at 44.1 kHz. Further, $f(x)$ consists of a fundamental with harmonics that are always less than 3 orders of magnitude lower in power than the fundamental. Since power is proportional to the square of a waveform, and we are dealing with the squared waveform in equation (3.2) when calculating RMS, we may expect that any harmonic component in $f(x)$ does not contribute

significantly to the overall integral and thus to the RMS amplitude. Once again, we are qualitatively legitimizing RMS amplitude, emphasizing that it has already been proven effective in the literature.

3.4 Measurement Procedure

3.4.1 Pipe Measurement Procedure

A 14.35 ± 0.10 cm pipe with a diameter of 2.70 ± 0.10 cm was used for the closed-open ended pipe measurements. Since this pipe has one unflanged open end, equation (2.8.3) dictates an effective length of 15.17 ± 0.16 cm. Equation (2.8) then yields 565.1 ± 5.9 Hz for the fundamental and 2825 ± 30. Hz for the 2nd harmonic, so the frequency range of 375 to 3125 Hz used for vocal tract measurements applies. The double open-ended pipe used was 34.00 ± 0.01 cm in length and 2.50 ± 0.10 cm in diameter, which yielded an effective length of 37.05 ± 0.22 cm (two length corrections). A frequency range of 250 to 2250 Hz at a spacing of 10 Hz was used for the double open-ended pipe.

Resonances were obtained by first measuring the free-field impedance spectrum. This was achieved by presenting the horn outlet with the closed end of the pipe, as shown in Figure 3.9. This configuration effectively removes the pipe from the system and presents the horn with the correct free-field that is applicable when driving the open end of the pipe. The pipe was mounted symmetrically about the center of the horn outlet, 3 mm away from the microphone. After free-field measurement, the pipe was reversed, and the response spectrum was measured. An adjustable laboratory stand greatly facilitated this process.

Figure 3.9: Measurement of Pipe Free-field Impedance Spectrum

The speaker operating power was initially chosen to be 0.5 W (instead of the 0.04 mW used for vocal tract measurements) to improve the signal-to-noise ratio. This created SPLs at the horn outlet in excess of 100 dB, which would certainly cause hearing damage without proper hearing protection. In order to verify that 0.04 mW speaker operation effectively identifies resonances, measurements were retaken at this wattage.

Initial measurements were taken for a frequency range of 375 to 3125 Hz, divided into 275 intervals for a frequency spacing of 10 Hz. While the theoretical value of the length of each sinusoid τ is 0.13 seconds from equation (2.4), graphical investigation of the steady-state regions of the sinusoids dictated a value of 0.25 seconds. Similarly, equation (2.2) dictates a value of 10 milliseconds for *T,* the time between sinusoids, whereas practice required 50 milliseconds. This required a total time of 83 seconds, which is still very reasonable.

Once the fundamental and 1st two harmonics were located in the frequency domain, they were resolved further by exciting the pipe with a frequency range of

approximately 15 Hz above and below the resonance, with a total of 30 sinusoids for a frequency spacing of 1 Hz. This created an effective "zoom" on the desired resonance. I retained the values for τ and T used in the first series of measurements, to create a total measurement time of roughly 9 seconds. Finally, the harmonic corresponding to n = 25 in equation (2.8) near 14,130 Hz was identified similarly to the resonance zooms. As discussed in Chapter 2, equation (2.8.2) dictates the upper limit on how accurately equation (2.8) models pipe resonances. For our pipe, which has a diameter of 2.70 cm, equation (2.8.2) yields an upper frequency limit of 3175 Hz, so we cannot expect that the frequency of the harmonic n = 25 will agree with 14,940 Hz. However, we may still measure this resonance and assess to what degree theory differs from experiment.

3.4.2 Vocal Tract Procedure

To measure vocal tract resonances, the adjustable cowl in the horn was positioned so that my mouth was centered on the horn outlet, roughly 1.5 cm from the microphone to eliminate the effect of breathing on the response signal. This said, extreme care was taken to breath through the nose, and the system was isolated from inhalation and exhalation by means of duct tape placed around the cowl. Care was taken to keep the configuration fixed throughout all measurements.

A speaker power of 0.04 mW was used for auditory safety. My vocal configurations were measured from 375 to 3125 Hz, with a frequency spacing of 10 Hz. When measuring vowels with particularly low first formants (i.e. the "i" vowel in "tree," which has F1 ~ 250 Hz), I used a range of 175 to 2925 Hz with the same spacing. The

vowels measured are shown in Table 3.1. I retained the values for τ and *T* used in the first series of pipe measurements.

Table 3.1: Vowels Measured

Vowel symbol	Vowel Sound
A	"father"
E	"said"
I	"tree"
O	"hoe"
U	"spoon"
æ	"had"

When regions of the vowel spectra could not be resolved due to a low signal-to-noise ratio, I averaged 10 measurements of identical responses to eliminate zero-mean noise. Because data collection takes much longer in this manner, I chose to zoom in on regions where formants were expected so as to minimize the number of swept sinusoids in the measurement.

Chapter 4: Results and Analysis

4.1 Pipe Measurements

Figure 4.1 displays impedance spectra for the 14.35 cm closed-open ended pipe at 0.5 W speaker operation. The y-axis displays RMS amplitude normalized to unity. The x-axis displays frequency in Hz from 375 to 3125 Hz. The upper graph shows a plot of the free-field impedance spectrum (dashed blue line) and the un-normalized response spectrum (solid red line). The horn shows strong resonances at the peaks in the spectrum. The response spectrum mirrors these maxima, underlining the need to normalize response spectra by the free-field spectrum.

The lower graph in Figure 4.1 shows the normalized response spectrum. Three resonances can be identified as strong maxima followed by sudden plunges in the amplitude at 565 ± 10 Hz, 1655 ± 10 Hz, and 2885 ± 10 Hz, which represent the fundamental and the next two harmonics (the fundamental is called the 1st harmonic when we are dealing with true harmonics, i.e. integer multiples of the lowest resonance). The uncertainty in these frequency values is limited to10 Hz, which is the frequency spacing of the sinusoids in the excitation. While the third harmonic does not have an overwhelmingly sharp peak, the inflection of the spectrum changes sign directly before the minimum at 3000 Hz, and this location was chosen to identify the resonance. The condition for a resonance is a maximum follow by a *steep* descent to a minimum, which was the case at 2885 Hz. The maximum before that would have passed through the inflection point on its way to the minimum, which is not characteristic of a resonance.

Figure 4.2 shows the same series of measurements for 0.04 mW speaker operation. In this case, the fundamental and next two harmonics are identified at 565 ± 10 Hz, 1675 ± 10 Hz, and 2885 ± 10 Hz.

Figure 4.1: Pipe Impedance Spectra for 0.5 W Speaker Operation

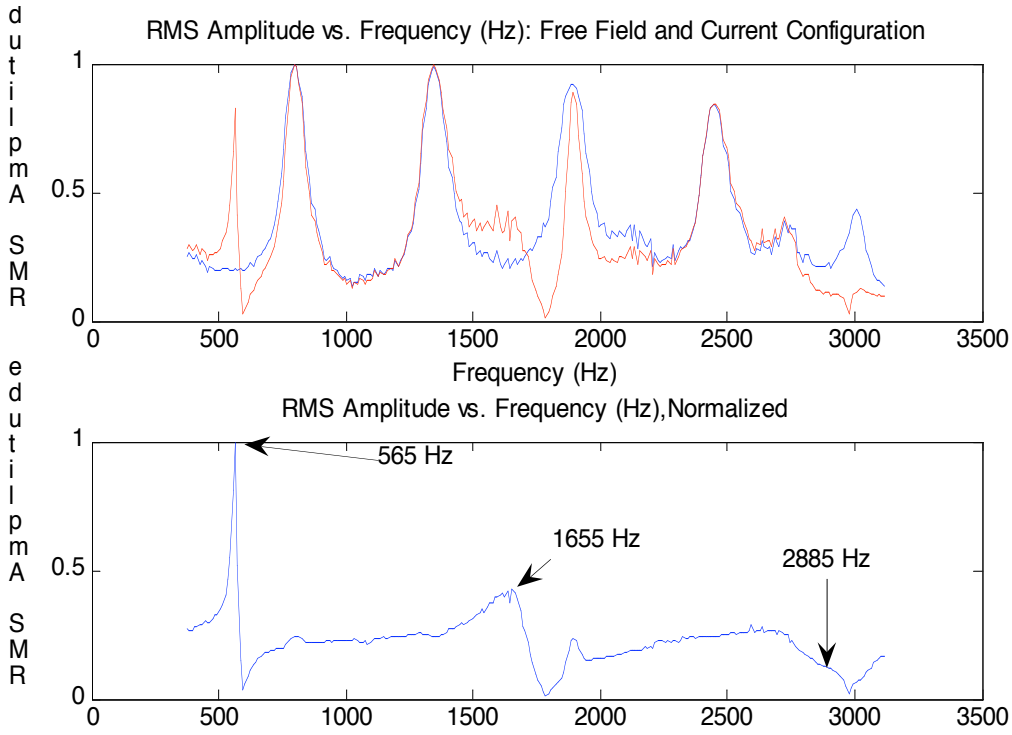
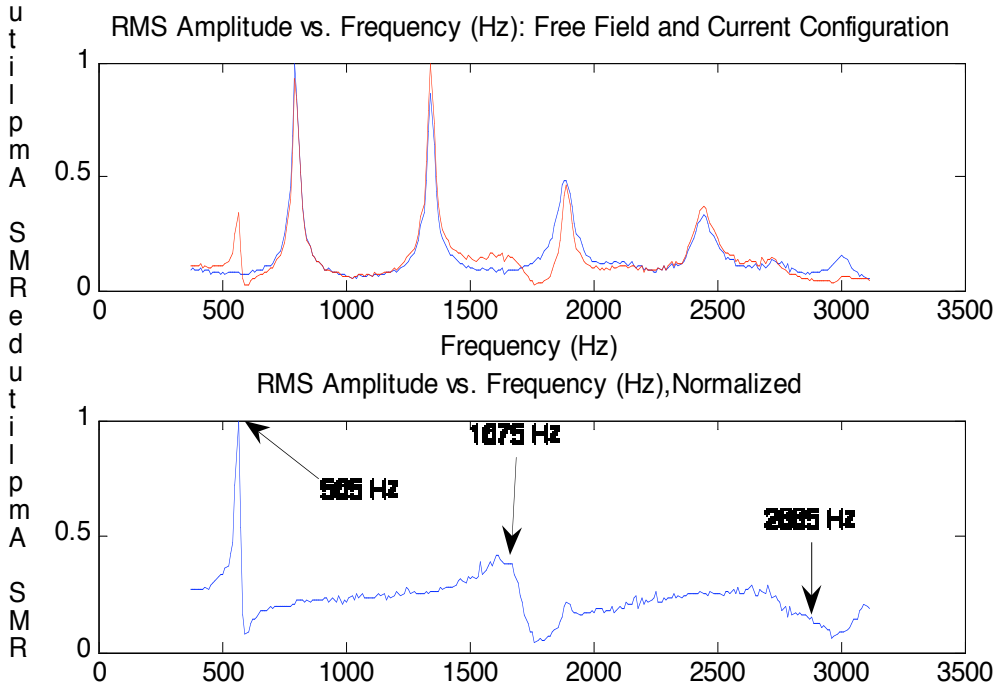


Figure 4.2: Pipe Impedance Spectra for 0.04 mW Speaker Operation



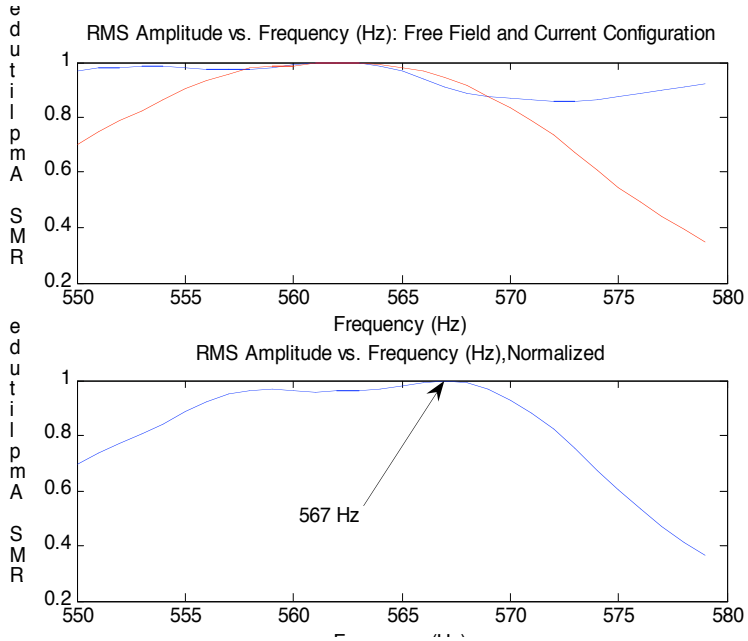
In order to resolve these minima further and improve our uncertainty, we zoom onto each peak with a series of sinusoids spaced at 1 Hz and distributed approximately 15 Hz above and below each peak identified in Figure 4.1. Figures 4.3 a)-c) show the free-fields, un-normalized impedance spectra, and normalized impedance spectra for each resonance. The fundamental was resolved to 567 ± 1 Hz, the second harmonic to 1658 ± 1 Hz, and the third harmonic to 2867 ± 2 Hz (or, depending on interpretation, 2965 ± 2 Hz; see plot for these 2 peaks). Zooms were not acquired at 0.04 mW speaker operation because the 375 to 3125 Hz measurements at 0.04 mW indicate that results will be consistent with the 0.5 W measurements.

Finally, to determine whether or not the impedance meter can successfully identify resonances near the limit of human hearing, the resonance corresponding to n = 25 (13^{th} harmonic) with a theoretical value of 14130 ± 150 Hz was resolved. Technically, 20 kHz is the limit of human hearing, but most individuals can only hear up

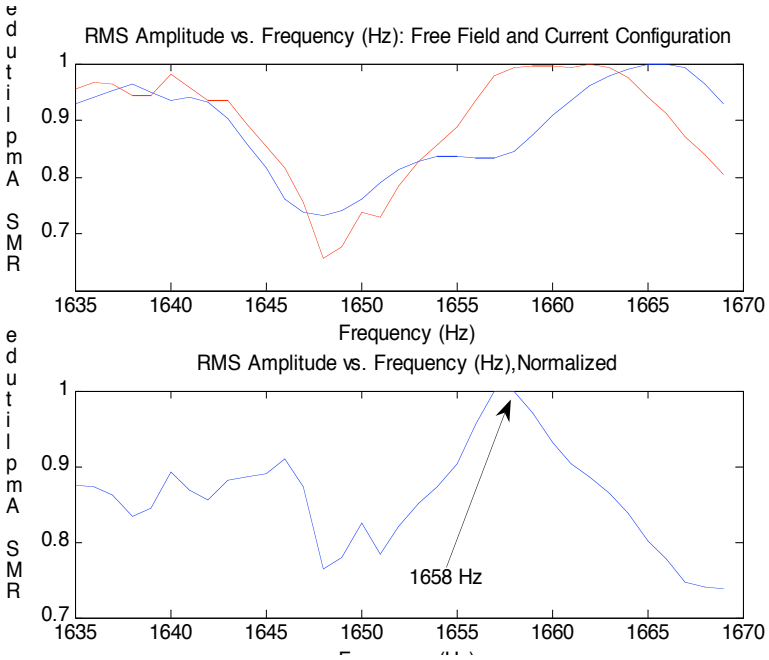
to 15 kHz. As discussed in Chapter 3, we do not expect the measured resonant frequency to agree with 14940 Hz, for the quarter-wavelength assumption (2.8.1) breaks down due to potential oscillatory modes along the diameter of the pipe. A frequency range of 13,975 to 14,275 Hz, with a frequency spacing of 2 Hz was chosen to span the theoretical uncertainty of 150 Hz quoted above. Figure 4.4 shows the resulting spectra, with a maximum clearly visible at 14,051 ± 5 Hz.

Figure 4.3: Zooms of Pipe Fundamental and Next Two Harmonics at 0.5 W Speaker Operation

a) Fundamental



b) 2nd harmonic



c) 3rd harmonic

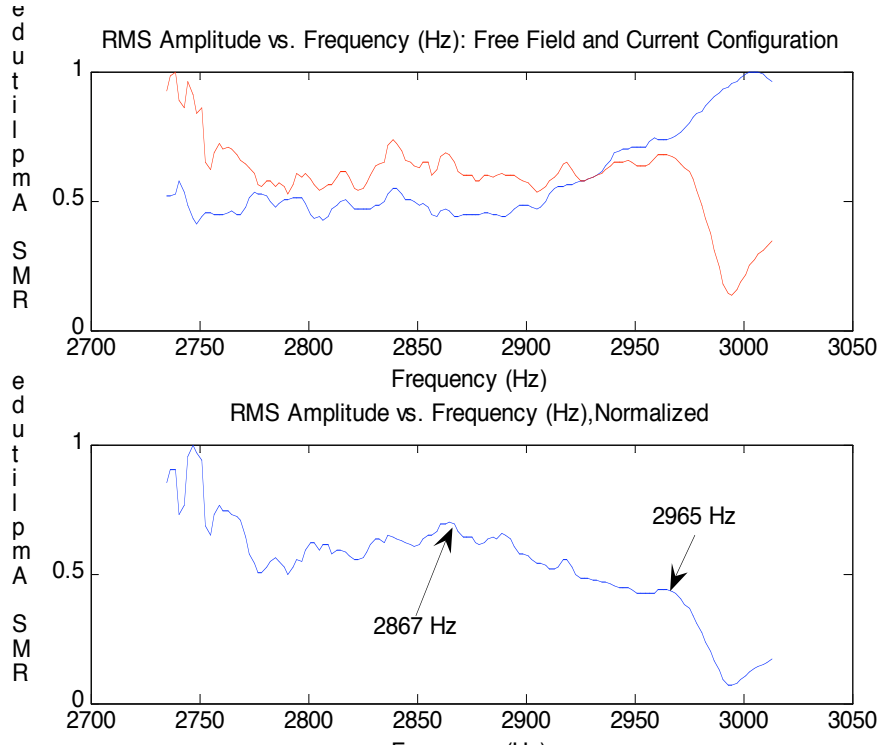


Figure 4.4: 13th Harmonic Zoom

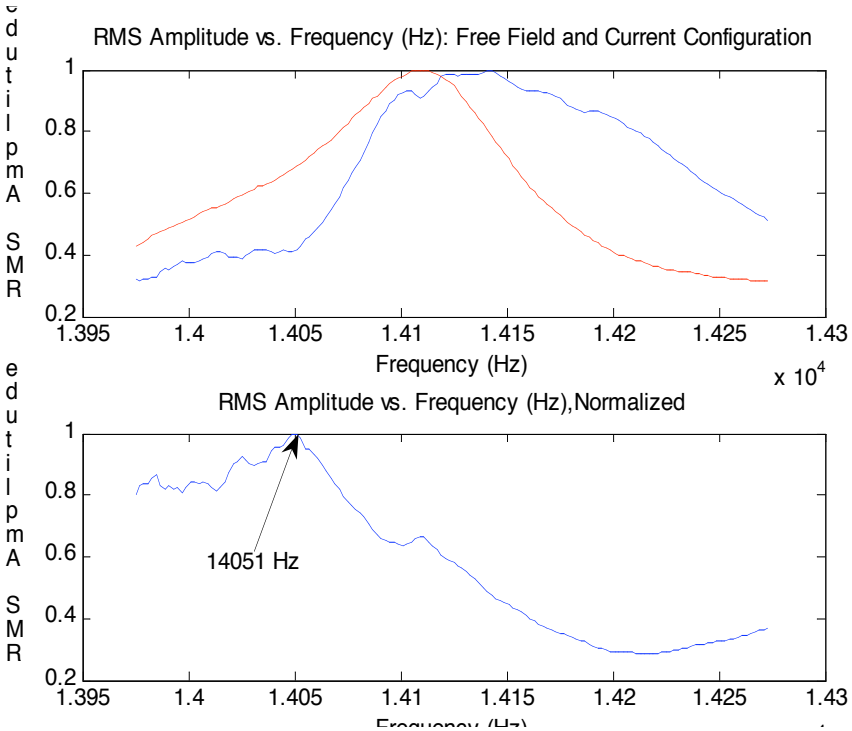


Table 4.1 summarizes the resonances identified by the impedance meter zooms and compares these values to theoretical values from equation (2.8).

Table 4.1: Comparison of Values with Theory

n (eqn. 2.8)	**Harmonic #**	**Theoretical Frequency (Hz)**	**Measured Frequency (Hz)**	**Agree?**
1	fundamental	565.1 ± 6.0	567.0 ± 1.0	Yes
3	2	1695 ± 18	1658 ± 1	No
5	3	2825 ± 30.	2867 ± 2 (2965 ± 2)	No (No)
25	13	14130 ± 150	14051 ± 5	Yes

Experimental values capture the theoretical values within the uncertainties for the fundamental, but not for the 2nd and 3rd harmonics. This may be due to the hardware interfering with the pipe system and shifting resonant frequencies by providing a different effective length. However, the accuracy in determining the fundamental does not seem consistent with this interpretation. It should be noted that the 2nd harmonic measurement errs in the negative direction, whereas that for the 3rd harmonic in the positive direction. Thus, the impedance meter accurately predicts resonances in the vicinity of 500 to 600 Hz at 0.5 W (this is well below the frequency limit of 3175 Hz dictated by equation (2.8.2)) but disagrees with theory for higher modes. Due to the similarity of the 0.5 W and 0.04 W plots with a frequency range of 375 to 3125 Hz, we have reason to believe that the results will be similar for the latter operating power.

As previously mentioned, we suspect that actual resonances in the vicinity of the 13th harmonic are not accurately given by equation (2.8), for at this point, the wavelength is small enough that reflections between the walls may have an effect on the overall oscillatory modes. Further, edge-effects near the opening of the pipe will be more apparent at higher frequencies due to increased reflections, and our microphone measures pressure at the center of the pipe only. Thus, we should be skeptical of the experimental resonant frequency obtained for this mode, for it may represent an entirely different resonance. Nevertheless, the experimental value captures the theoretical value.

Figure 4.4.1 shows a plot of the impedance spectrum for the double open-ended pipe. Resonances are identified at 470 ± 10 Hz, 950 ± 10 Hz, 1430 ± 10 Hz, and 1940 ± 10 Hz. Table 4.1.1 compares the measured values to the theoretical values from equation (2.8.4) with the effective length of 37.05 ± 0.22 cm.

Figure 4.4.1: Impedance Spectrum for Double Open-Ended Pipe

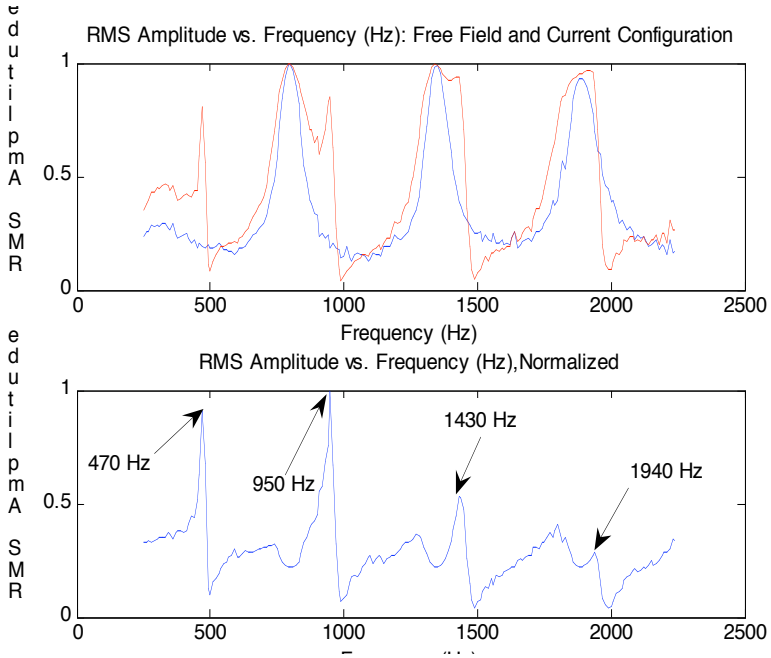


Table 4.4.1: Summary of Values for Double Open-Ended Pipe

n eqn (2.8.4)	**Theoretical Frequency (Hz)**	**Measured Frequency (Hz)**	**Capture?**
1	462.9 ± 2.8	470. ± 10.	yes
2	925.8 ± 5.5	950. ± 10.	no
3	1389 ± 8	1430 ± 10	no
4	1852 ± 11	1940 ± 10	no

Once again, we see that the impedance meter correctly predicts the frequency of the fundamental resonance, which in this case lies in the vicinity of 500 Hz. The measured values for the 2nd, 3rd, and 4th harmonics are systematically larger than those predicted by equation (2.8.4). As in the previous case, it is possible that interaction

between the meter and the pipe's open end leads to a shorter effective length due to reflections off the wall of the horn (the area of Plaster of Paris between the horn outlet and the ABS pipe). A shorter effective length would systematically increase the predicted resonant frequencies, and might explain why our measurements err systematically in the positive direction.

4.2 Vocal Tract Measurements

Figures 4.5 a-f show plots of the impedance spectra for the vowels "a," "e," "i," "o," "u," and "æ" (see table 3.1 for pronunciation). In all cases, I produced the vowels. The graphs have the same format as the impedance spectra plots for the pipe measurements. Resonances (maxima followed by abrupt minima) are labeled with the appropriate frequency. The labels also show a formant identification (1st formant = F1, 2nd formant = F2, 3rd formant = F3) based on well-accepted formant frequency values reported in Table 4.2. The values for "e," "i," "u," and "æ" were taken from a phonetics textbook that reported averages of a collection of authorities' data (sources marked "1" in the last column). [10] The values for "a" and "o" were taken from the data archives of the Eastman Computer Music Center at the University of Rochester (sources marked "2" in the last column).* When there is a high degree of uncertainty as to whether a formant identification in Figure 4.5 is correct, the formant label is followed by a question-mark (?). We will discuss this uncertainty shortly and will present a method that involves calculating formant ratios.

Many of the formants identified on Figure 4.5 are subject to serious concern due to the signal-to-noise ratio. This is particularly apparent for the "e," "i," and "u" vowels,

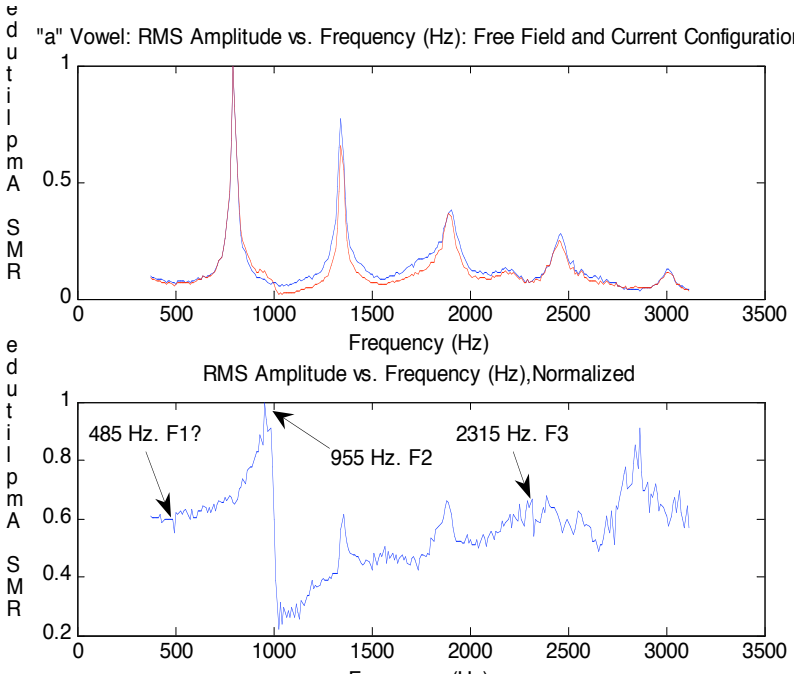
* http://ecmc.rochester.edu/onlinedocs/Csound/Appendices/table3.html

especially in the regions where we expect to identify the 1st formant. This leads to ambiguity in identifying peaks. The preferred peak in such a region was chosen to be the one with the steepest falloff to a minimum, which should be a characteristic of resonant frequencies (see Chapter 2 for a discussion of impedance spectra).

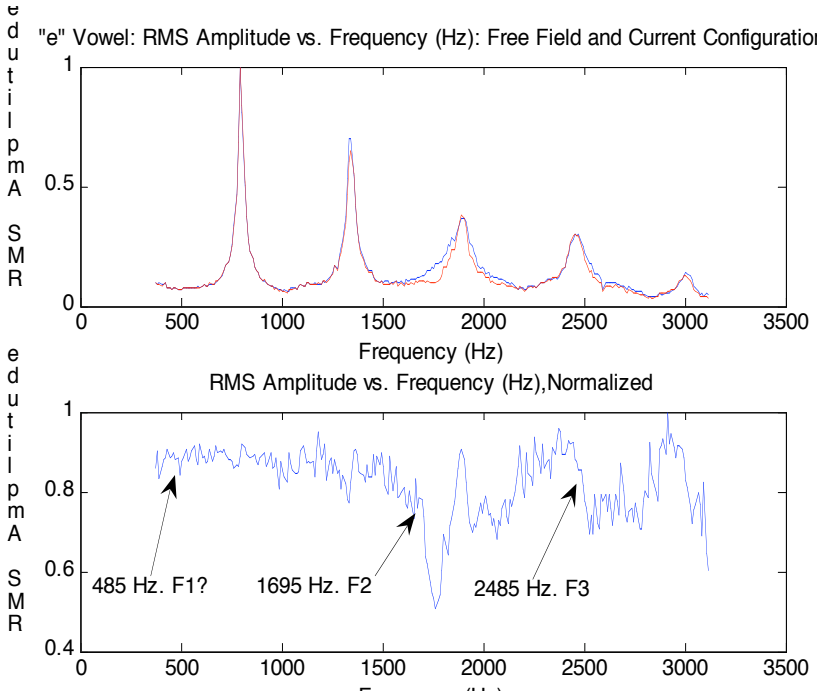
Often, normalized spectra (the lower graphs in each vowel plot) will mirror the characteristic resonances of the horn seen in the free-field and un-normalized vowel spectra, which could lead to misidentification of formants. An example is the "a" vowel in Figure 4.5 a), which shows peaks at roughly 1350 and 1900 Hz that clearly originate from the horn resonances. Juxtaposition of the free-field with the normalized spectrum is thus crucial in order to avoid formant misidentification.

Figure 4.5: Impedance Spectra for "a," "e," "i," "o," "u," and "æ" Vowels

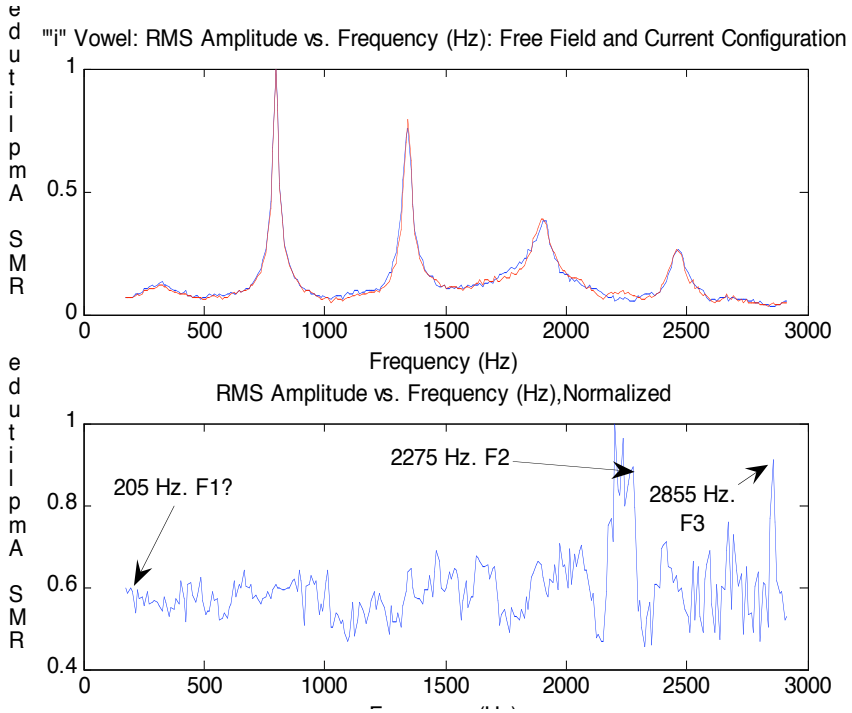
a) "a" Vowel



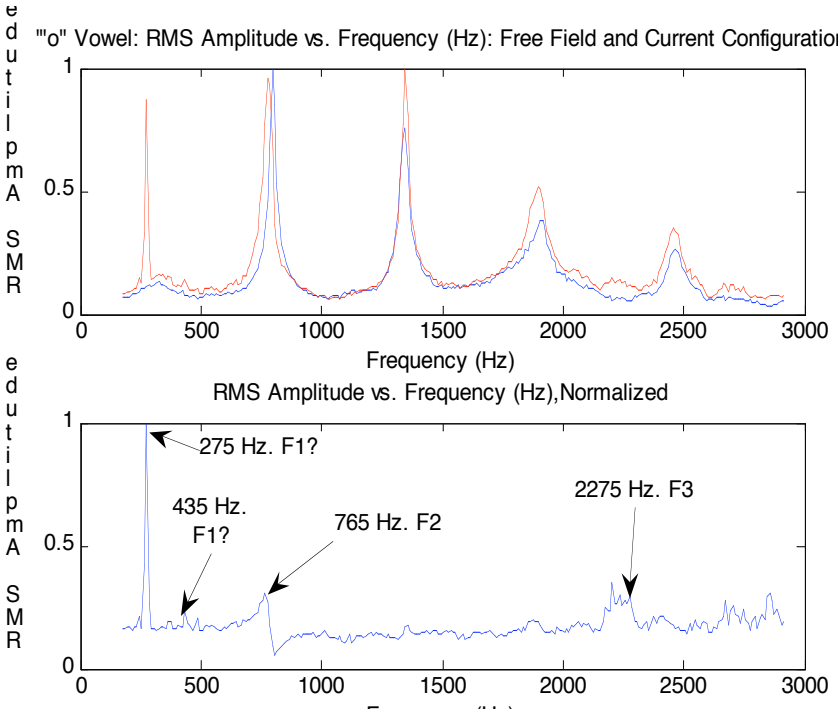
b) “e” Vowel



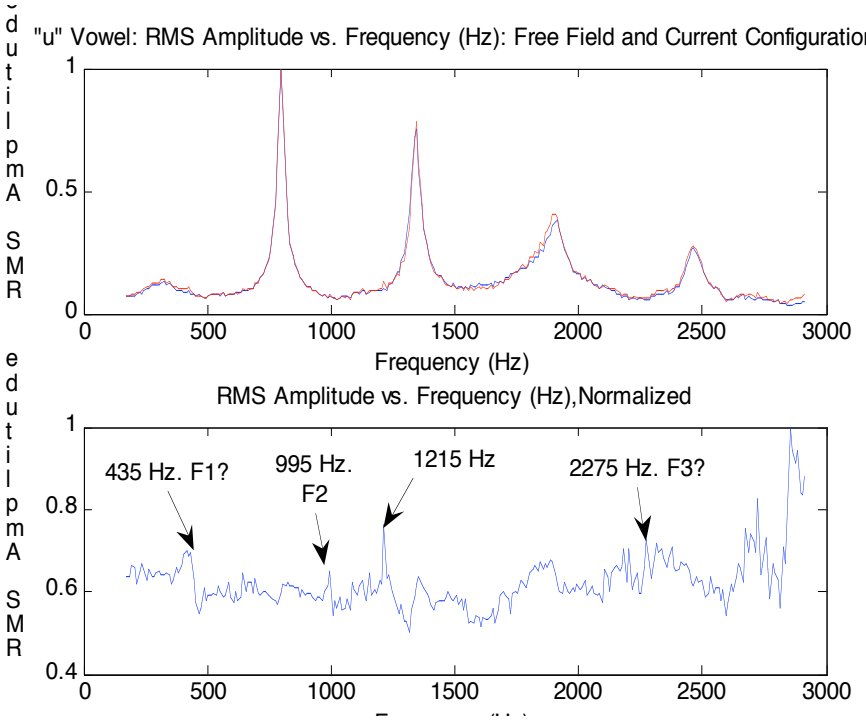
c) “i” Vowel



d) “o” Vowel



e) “u” Vowel



f) “æ” Vowel

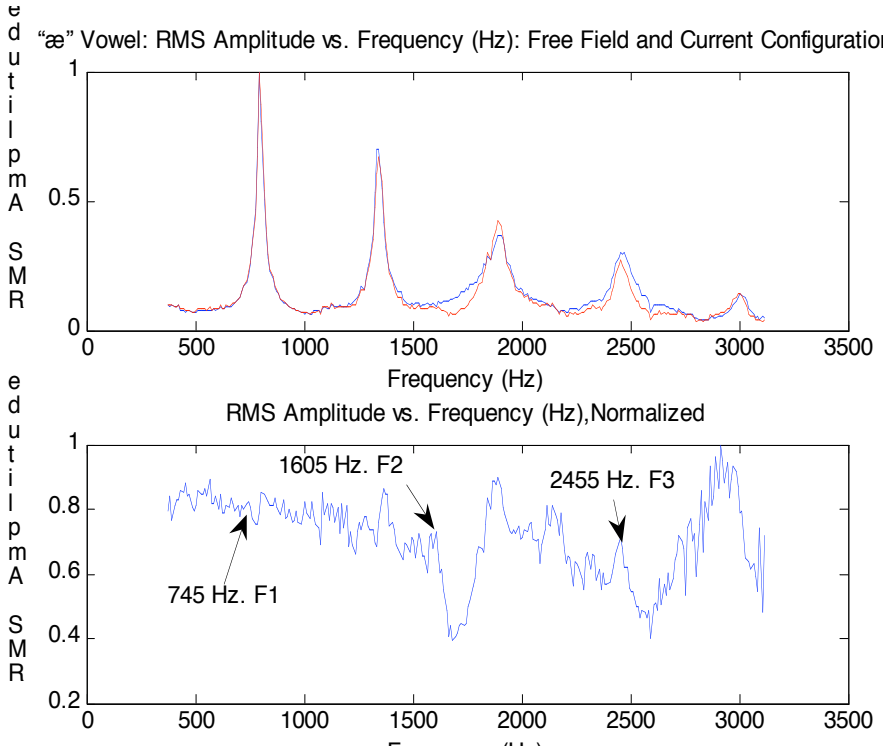


Table 4.2: Formant Frequencies (Literature Values)

Vowel	pronunciation	F1 (Hz)	F2 (Hz)	F3 (Hz)	source
a	“father”	600	1040	2250	2
e	“said”	550	1770	2490	1
i	“tree”	280	2250	2890	1
o	“hoe”	400	750	2400	2
u	“spoon”	310	870	2250	1
æ	“had”	690	1660	2490	1

Sources:

1 = Ladefoged, Peter. *A Course in Phonetics*

2 = Eastman Computer Music Center at the University of Rochester

Table 4.3 juxtaposes the experimental formant frequency values with those in Table 4.2 and displays the percent error between values. For each formant, literature values from Table 4.2 are displayed in boldface to the left of the experimental values from Figure 4.5. The uncertainty for all experimental values is 10 Hz.

Table 4.3: Percent Error Between Experimental and Literature Formant Frequencies

Vowel	F1 (Hz)		% Error	F2 (Hz)		% Error	F3 (Hz)		% Error
a	**600**	485	19.2	**1040**	955	8.2	**2250**	2315	2.9
e	**550**	485	11.8	**1770**	1695	4.2	**2490**	2485	0.2
i	**280**	205	26.8	**2250**	2275	1.1	**2890**	2855	1.2
o	**400**	435	8.8	**750**	765	2.0	**2400**	2275	5.2
u	**310**	435	40.3	**870**	995	14.4	**2250**	2275	1.1
æ	**690**	745	8.0	**1660**	1605	3.3	**2490**	2455	1.4

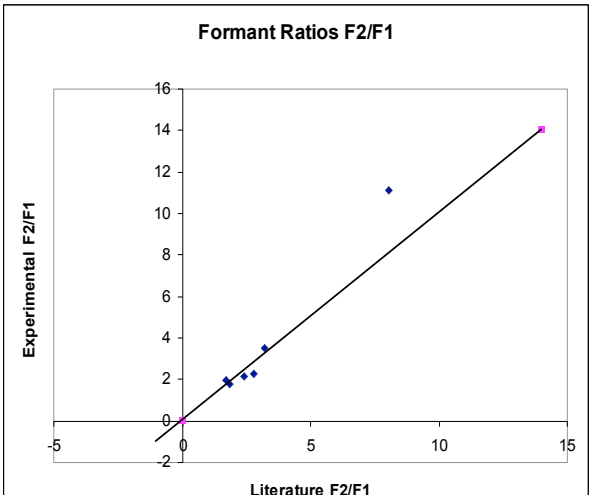
The impedance meter is most accurate in identifying the 2nd and 3rd formant frequencies, whereas it registers values for the 1st formant that differ significantly from expected values. There is no way to verify that I was producing vowels accurately by phonetic standards, and in fact my only method of assessing vowel accuracy is by identifying formant frequencies. Thus, because my vowels might not represent the vowels quoted in the literature (even though they are averages of multiple speakers), deviations from literature formant frequencies are not unexpected.

Perhaps more informative than the actual values of F1, F2 and F3 are the ratios F2/F1 and F3/F1. Literature values represent averages amongst different kinds of speakers (men, women, etc.) and thus may exhibit large differences from purely male

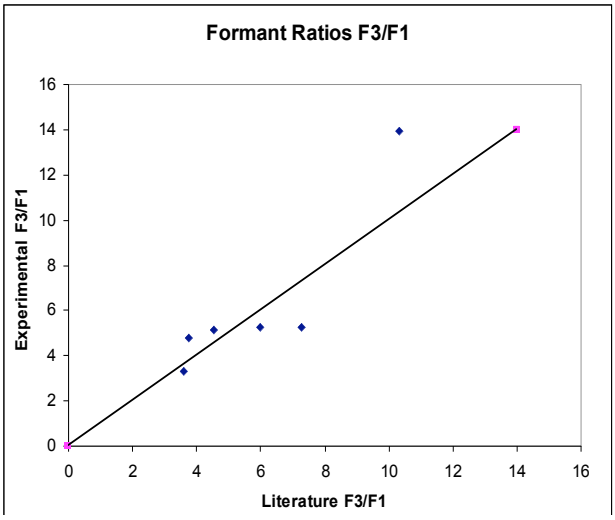
formants (mine). The ratios amongst the formants, however, should be relatively consistent from person to person since we are dealing with the human vocal tract in all cases. Figure 4.5.1 a) plots the experimental value of F2/F1 on the y-axis versus the literature value of F2/F1 on the x-axis, and includes a line with slope 1 denoting equality between literature and experiment. We see a slope of roughly 1 for most vowels. The data point at roughly (8,11) is the "i" vowel, which has a very low 1st formant. Figure 4.5.1 b) plots the same data for F3/F1. The uppermost-right data point is again the "i" vowel. For F3/F1, the relationship deviates more strongly from the line of slope 1, and it appears that the experimental value of F3/F1 is nearly constant for four data points with different literature values.

Figure 4.5.1: Formant Ratio Plots

a) F2/F1



b) F3/F1

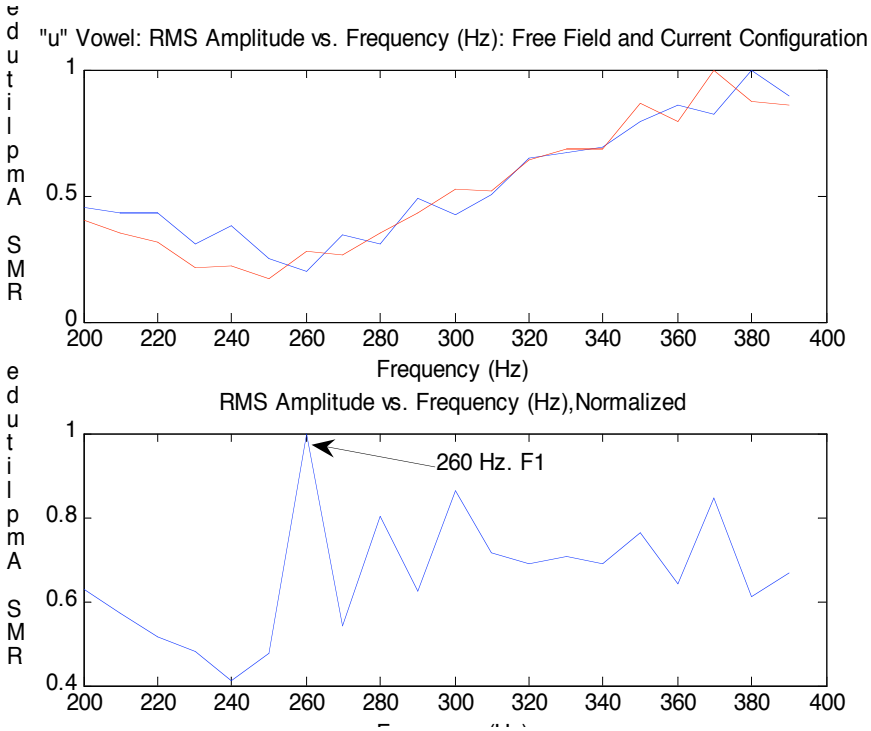


We have already noticed that the signal-to-noise ratio is especially low in the region of the fundamental for particular vowels. In order to address this issue, we may measure the response signal multiple times for a given vowel or configuration, average the results, and proceed with a calculation of the RMS amplitudes. Since noise is random and has an average of zero, whereas our signal is systematic, this will greatly increase the signal-to-noise ratio.

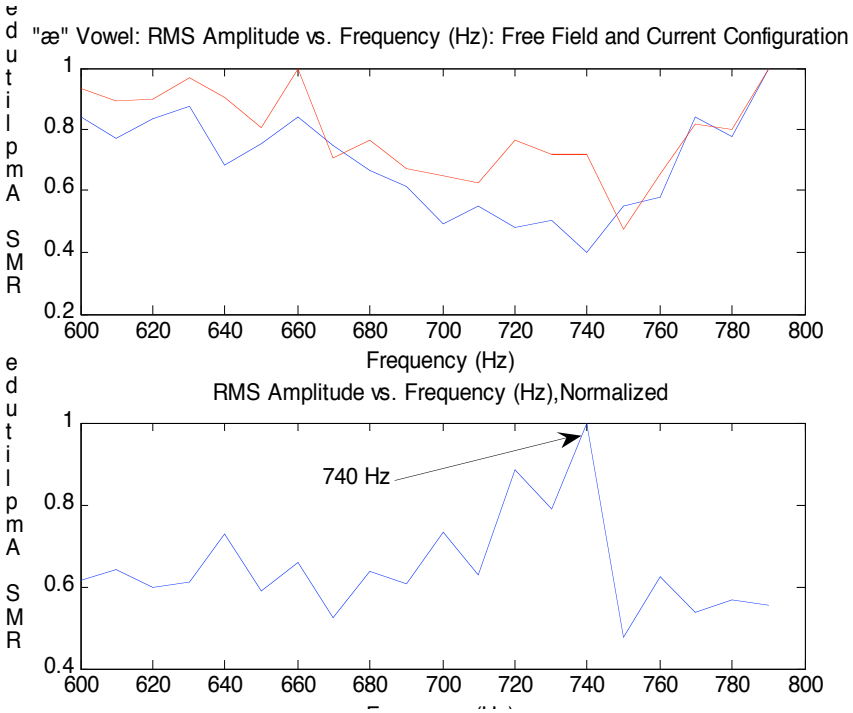
Figures 4.6 a) and b) show plots of the spectra for the "u" and "æ" vowels. In order to resolve F1, the "u" plot ranges from 200 to 400 Hz and the "æ" plot from 600 to 800 Hz, both at a frequency spacing of 10 Hz. These spectra were calculated from a signal that was the average of 10 separate sinusoidal excitations. The specific vowels suffered from noise in this frequency domain (see Figure 4.5), which theoretically contains the 1st formant.

Figure 4.6: Impedance Spectra for “u” and “æ” Via 10 Averaged Response Signals

a) “u” Vowel



b) “æ” Vowel



F1 is identified for the “u” vowel at 260 ± 10 Hz, which deviates from the 310 Hz reported in the literature by 12.9 %, as opposed to the 40.3 % difference observed when only one response signal was used to calculate the spectrum. Similarly, F1 is identified for the “æ” vowel at 740 ± 10 Hz, which deviates from the 690 Hz reported in the literature by 7.2 %, as opposed to the 8.0 % difference observed when only one response signal was used to calculate the spectrum. For the “u” vowel, this technique greatly improved the accuracy of the formant identification, whereas accuracy in the “æ” vowel case did not change significantly. This may be due to the fact that the signal-to-noise ratio, although small in the range of F1 for “æ” (~700 Hz), was nonetheless sufficient to make a reliable formant prediction. We conclude that a series of swept sinusoids from 375 to 3125 Hz is capable of identifying F2 and F3 with much higher accuracy than F1. F1 may be identified by a combination of zooming and signal averaging.

Chapter 5: Conclusions

This project details the construction and operation of an acoustic impedance meter capable of quantifying the resonances of physical systems. This device allows an uncertainty of 1 Hz in the frequency domain during peak (resonance) identification. This is especially useful when resonant frequencies cannot be calculated from theory, as is the case with the human vocal tract. A scientific understanding of human speech and the art of throat singing relies on accurate measurement of resonant frequencies, and this impedance meter was developed with the goal of analyzing various form of Tuvan throat singing.

I satisfied a host of construction priorities that were essential to making this project feasible and successful within the timeframe allowed for an undergraduate thesis. Inexpensive and readily available materials were used, which make this device reproducible and accessible. Further, the MATLAB source code used for measurement automation and post-processing is available for further experimentation or improvement, and is easily exported to other platforms.

The impedance meter in this experiment accurately predicts resonant frequencies in two scenarios. The first is measurement of the fundamental frequencies of two kinds of pipes: one with a single open end, and the other with two open ends. Experimental values captured theoretical values and allowed an uncertainty of only 1 Hz. The second situation in which my impedance meter accurately measures resonances is the identification of the 2^{nd} and 3^{rd} formants of a variety of vowels, including “a,” “e,” “i,” “o,” and “u.” The exact theoretical locations of these formants are not well established in the literature, for they differ from person to person based on physiology, accent and

inflection. In the field of phonetics, researchers often work with averages of formant frequencies gathered from a population of speakers. Thus, formant identification in this project was not based on exact agreement between experimental and literature values[*]. Rather, it was based on reasonable proximity between literature values and the frequencies determined by the graphical "footprints" of formants in the impedance spectrum (maxima followed by sharp descents to minima).

Perhaps the best method to assess how well the meter predicts formant frequencies is to compare experimental and literature values of the formant ratios F2/F1 and F3/F1, which should be similar from speaker to speaker. Data indicates that the impedance meter effectively measures F2/F1, while F3/F1 appears to be constant for number of vowels that should have different values of F3/F1. Superimposing voiced vowel spectra (obtained by FFT analysis) with impedance spectra could aid in identifying a particular speaker's formants, and represents a future development for this project.

While my impedance meter correctly measures certain resonances, it is highly inaccurate in identifying others. Experimental values for the 2^{nd} and 3^{rd} harmonics of a pipe with one closed end failed to capture theoretical values. This error did not appear to be systematic, for the experimental value for the 2^{nd} harmonic erred in the negative direction, while that for the 3^{rd} harmonic erred in the positive direction. That said, two data points are not overwhelmingly solid ground for judging systematic error, and additional successive harmonics (4^{th}, 5^{th}, 6^{th}) should be measured. The experimental value for the 13^{th} harmonic captured the theoretical value. We note that the uncertainty in the theoretical value due to measuring the pipe was 150 Hz in this case. This provides us with a very large interval in the frequency domain that may capture other system

[*] Further, literature values to not cite uncertainties.

resonances, which we would misidentify as the 13th harmonic. Indeed, at the frequency of the 13th harmonic (14,130 ± 150 Hz), we expect that complex resonances (not purely longitudinal) may exist due to the small wavelength and consequent reflections throughout the cylindrical pipe.

The impedance meter systematically overestimates harmonics 2, 3, and 4 of a double open-ended pipe. We hypothesize that the meter might interfere with the pipe system and create a shorter effective length by reflecting sound back into the pipe. This would certainly cause the meter to measure resonant frequencies higher than those predicted by theory.

When the 1st formant (F1) for any vowel is measured, a low signal-to-noise ratio exists, which required averaging of multiple sinusoidal series to eliminate zero-mean noise. This procedure led to the identification of F1 for the "u" and "æ" vowels. This greatly increased measurement time, during which it was difficult to maintain a constant vowel configuration. The low signal-to-noise ratio in the vicinity of F1 resulted from the safety stipulation that the SPL not exceed 80 dB at the horn outlet. It is possible that the SPL could be increased, but further research into the effects of sound on the Eustachian tubes and ears is advised.

The swept sinusoidal method of excitation chosen for this impedance meter lends itself naturally to measurement and post-processing of data. For each sinusoid in the excitation series, the root-mean-square (RMS) amplitude of the response signal was calculated to indicate the power with which the system responded to the excitation. RMS amplitudes are readily compiled into a spectrum using MATLAB's vector

manipulation capabilities. Further, recording and playback in MATLAB is very simple, for no file conversion is necessary.

While the swept sinusoidal method has distinct advantages, the broadband method could solve a host of issues confronted in this experiment. Broadband measurement time would be on the order of 1 second, which would allow stable vowel configurations. In the event that signal-averaging were necessary*, this would only increase to 10 seconds. This is the timescale of some of the shortest measurements made with the swept sinusoidal method. The broadband method would also simplify SPL measurements, for 1 reading would suffice to guarantee auditory safety. This said, the researcher must find a way to reliably and consistently identify peaks in the FFTs of multiple windowed signals.

Included in this project is a considerable amount of detail regarding issues such as vocal physiology, human speech production, auditory safety, and signal processing. These figured prominently in construction and operation decisions. The motivation for designing this impedance meter is to measure the acoustics of throat singing as performed by masters of the art form. The Pomona College Department of Physics graciously contributed to funding a concert given by *Chirgilchin: Master Throat Singers from Tuva* in September 2007 on the college's campus. This provided me with the opportunity to record the different styles of *khoomei* (throat singing) and to create a relationship with the artists. The data I gathered will be crucial in the event that I make further measurements with the artists using my impedance meter. Comparing FFTs of the data with the impedance spectra will further validate the external excitation technique used by this impedance meter. The meter offers a much higher frequency resolution than

* an alternative would be to sample for a longer period of time, which would further resolve peaks in the spectrum after the FFT analysis necessary for the broadband method.

conventional FFT analysis, and this should help clarify the resonances that give throat singing its unique sound. Ultimately, I hope that such an understanding will elucidate throat singing technique both acoustically and physiologically, as well as contribute to a musical appreciation of the art form.

Acknowledgements

I would like to thank Alma Zook, my thesis advisor, for her hours of guidance, as well for her shared interest in acoustics and the physics of music. Thanks to Bryan Penprase, the current department chair, who brought before the department my funding requests, which led to a joint concert and data-gathering session with *Chirgilchin: Master Throat Singers from Tuva.* I look forward to a continued relationship with *Chirgilchin* after graduation and a lifetime of interest in the acoustics of the human voice. I would like to thank Glenn Flohr, the department machinist, for his hours of assistance in the machine shop, for turning one of the best exponential horns I could have imagined, and for joining me numerous times at Haldeman Pool to swim laps. Thank you to Dave Haley, the department master lab-technician, for helping me navigate the department's vast store of equipment, as well as for filling the halls (even in the basement) with echoes of enthusiasm, even if they interfered with measurements from time to time... Thank you to Michael Guerre, the department electrician, for informing my decisions about speaker purchase and signal pathways, and for his shared interest in all forms of overtone singing. Thank you to Mary Paster, professor of linguistics, for her help in deciphering the IPA (International Phonetic Alphabet), for charts of formant frequencies, and for her shared interest in phonetics and sound production. Thank you to *Chirgilchin* for their outstanding concert in September 2007, their participation in data collection, and their warm spirit. Thank you to John Lopes, SCC coordinator, and Sandra Fenton, Grants Administrator/Assistant to the Associate Dean, for their invaluable assistance in coordinating *Chirgilchin's* concert, from green-room towels to artist contracts. Thank

you to my hall mates Michel Grosz, Zak Silverman, and Lucas Allen-Williams, who have endured my attempts at throat singing.

References

[1] Tran Quang Hai, 8th International Congress on Acoustics, 4 pp.; 4. (2004).

[2] T. C. Levin and M. E. Edgerton, Sci. Am. 281, 80 (1999).

[3] M. Kob, Appl. Acoust. **65**, 1249 (2004).

[4] J. Epps, J. R. Smith, and J. Wolfe, Meas Sci Technol **8**, 1112 (1997).

[5] O. K. Mawardi, J. Acoust. Soc. Am. **21**, 84 (1949).

[6] S. Adachi and M. Yamada, J. Acoust. Soc. Am. **105**, 2920 (1999).

[7] M. Kob and C. Neuschaefer-Rube, Med. Eng. Phys. **24**, 467 (2002).

[8] M. Kob and C. Neuschaefer-Rube, 3rd International Workshop on Models and Analysis of Vocal Emissions for Biomedical Applications, 187; 187 (2003).

[9] D. R. Raichel, *The science and applications of acoustics* (Springer, New York, 2000).

[10] P. Ladefoged, *A course in phonetics* (Harcourt Brace Jovanovich, New York, 1975).