THE MUSIC MUSE

Leslie Willson

Thesis submitted to the Faculty of the
Virginia Polytechnic Institute and State University
in partial fulfillment of the requirements for the degree of

MASTER OF SCIENCE

in

Mechanical Engineering

A.L. Wicks, Chair
William R. Saunders
Ricardo Burdisso

Nov 1, 1996
Blacksburg, Virginia

Keywords: Signal Processing, Speech Analysis

THE MUSIC MUSE

Leslie Willson

(ABSTRACT)

Ever wonder why two people can sing the same note with the same loudness, but sound completely different? Middle C is middle C no matter who sings it, yet for some reason Lucianno Pavarotti's middle C sounds richer and more beautiful than Bob Dylan's middle C, for example. But then again, what is beauty in singing? It is a completely biased and abstract concept. To some, Bob Dylan's voice may epitomize tonal beauty, while to others his voice may be comparable to fingernails on a chalk board. Anyway, differences in tone quality, or timbre, are due to differences in the spectral characteristics in different voices. The Music Muse is a computer program designed to help singers train their voices by showing them the individual components of their voices that combine to produce timbre.

In paintings, many colors are combined to produce different hues and shades of color. The individual colors that make up the hue are difficult to distinguish. Similarly in music, harmonics with varying amplitudes combine to create voice colors, or timbres. These individual harmonics are difficult to distinguish by the ear alone. The Music Muse splits the voice up into its harmonic components by means of a Fourier transform. The transformed data is then plotted on a harmonic spectrum, from which singers can observe the number of harmonics in their tone, and their amplitudes relative to one another. It is these spectral characteristics that are important to voice timbre.

The amplitudes of the harmonics in a voiced tone are determined by the resonant frequencies of the vocal tract. These resonances are called formants. When a harmonic that is produced by the vocal cords has a frequency that is at or near a formant frequency, it is amplified. Formants are determined by the length, size, and shape of the vocal tract. These parameters differ from person to person, and change during articulation. Optimal tonal quality during singing is obtained by placing formants at a desired frequency. The Music Muse calculates the formants of the voice by means of cepstral analysis. The formants are then plotted. With this tool, singers can learn how to place their formants.

One of the difficulties of voice training is that singing is rated on a scale of quality, which is difficult to quantify. Also, feedback tends to be biased, and therefore subjective in nature. The Music Muse provides singers with the technology to quantify quality to a degree that makes it less of an abstract concept, and therefore more attainable.

TABLE OF CONTENTS

LIST OF FIGURES

CHAPTER 1: INTRODUCTION

Unlike science, art is a subjective field. Science is definitive and objective. The beauty of science is that it provides us with unbiased explanations for life. Consequently, it does not allow any room for personal interpretation, and is not considerate of human emotions. Art, however, is the complete opposite. It is based more on emotion and individual expression. It has little room for definition, because its interpretation is dependent on personal experiences. Science is quantitative, while art is purely qualitative. Art is an important factor in human life because it allows us to escape from the strictured harshness of everyday reality. It allows us to focus on beauty, which cannot be defined. This lack of definitiveness makes art difficult to control. However, if art were scientific, it would be much less expressive.

Most art forms do, however, have a bit of a scientific base. For example, visual arts, such as painting, are based on how the eye processes information. Artists can create the illusion of three dimensions by creating certain combinations of dark and light that the eye interprets spatially. Music, which is the art genre that is the subject of this thesis, is based on how the ear processes sound information. It has measurable pitch scales based on the sound resolution of the ear, and even harmonic combinations that have been universally agreed upon to be pleasant or unpleasant. However, music is still evaluated on a scale of quality, which is difficult to quantify. Vocal quality does seem to correlate with skill, which is something that can be taught. Nevertheless, the skill is difficult to teach because the voice instrument is intangible, and the feedback tends to be subjective in nature. Piano teachers can demonstrate the correct fingering necessary for skillful playing, but voice teachers cannot reach inside their students bodies to demonstrate correct articulation, breathing, or vocal dynamics. The Music Muse is a voice training computer program that has been designed to add a measure of science to singing to make it tangible, but not enough to rob it of its human and expressive nature.

When a voiced tone is produced, a single sound is heard by the untrained listener. However, the produced tone is actually a combination of overtones. These overtones, or harmonics, add brilliance and beauty to

the sound. In painting, different colors are combined to create different hues and shades of color. Similarly in singing, overtones with varying amplitudes are combined to create voice colors, or timbres. However, the individual harmonics are even more difficult to distinguish with the ear than the colors that combine to create different hues in painting. The Music Muse can distinguish the harmonics in a sound by calculating the Fourier transform of the tone.

The number of harmonics in a sound and their relative amplitudes are displayed by the Music Muse on a graph called a harmonic spectrum. The Music Muse records a voiced sound, and displays the resulting harmonic spectrum, thereby providing the user with a visual image of tonal quality.

With the Music Muse, singers can watch their overtones change while they sing. This will teach them which types of sounds produce which types of harmonic spectra. The Music Muse also displays a graph of the voice signal in the time domain, so that the singers can observe the rate and extent of their vibrato. One screen on the Music Muse can calculate the formants, or resonant frequencies, of the voice.

One more Music Muse feature is its ability to compare voices. Singers can compare their voices to professional singers, or any other singers, and observe their spectral differences. For example, a student may compare his voice to a voice teacher, and notice that his own voice has traces of nasality or other unpleasantries that cannot be detected in the teacher's voice. Singers may also compare themselves singing in different ways to see if they can make their nasality disappear.

With the Music Muse, voice files can be saved and played again later. With this feature, singers can keep a library of voice files to compare to. Also, singers can keep an archive of their own voice files to document their progress.

The Music Muse would be extremely useful to voice teachers. The role of the voice teacher is to help students master their instruments. It is impossible, however, for voice teachers to reach down their students' throats to manipulate their voice mechanisms in ways that produces the best resonance. Teachers must rely on verbal communication, which

involves individual interpretation, rendering it an unreliable source. Teachers may use abstract phrases such as "float the sound" or "make the tone lighter," or instruct their students to "place" their resonances forward into their faces, higher into their heads, or lower into their chests. But what do these phrases really mean? This type of instruction forces students to decipher word salads of images before they can understand their instructions. The Music Muse eliminates the conundrum of imprecise verbal communication by producing a unique visual image for every type of sound.

Another reason the Music Muse is helpful to singers is that ears are biased indicators of sound. When sound enters the ear, it is transformed into nerve impulses that can be interpreted by the brain. During this transformation, some sounds are masked, attenuated, or otherwise altered. Also, singers do not hear their own voices in the same way as their listeners. People generally do not know exactly what their voices sound like to other people unless they hear themselves on a recording. The Music Muse provides singers with an unbiased image of sound in its raw, unaltered form. This allows singers to really know what tones they are producing, and how the tones affect what is actually heard.

The Music Muse takes voice training to another level by adding an element of tangibility as well as precision. The visual images provide accurate definitions to the endless variances of sounds. The program also gives singers unbiased impressions of their voices. With this program, voice quality still remains a personal interpretation, but it becomes more explicit, and perhaps easier to attain.

1.1 ORGANIZATION OF THESIS

This thesis will give a detailed explanation of how the Music Muse works, and why it is an excellent tool for improving a singing voice. Part I includes Chapters 2 through 5, which give background information necessary for an understanding of how sensational a program the Music Muse is. Chapter 2 will review the basic principles of acoustics and vibrations needed to understand the program. This chapter will define acoustic waves, harmonics and inharmonics, simple versus complex vibration, and the beating phenomenon. This chapter will also explain resonance, pitch theory and formant theory, harmonic spectra, and how each relates to voice timbre.

Chapter 3 will review the basic principles of signal processing that are used by the Music Muse to break the voice signal down into its harmonic spectrum by means of a Fourier Transform. This chapter will also introduce formant frequencies, and explain how the Music Muse can extract them from their harmonic spectrum by means of cepstral analysis.

Chapter 4 will introduce the hearing mechanism. The different parts of the mechanism will be presented with a detailed explanation of how they work together to convert sound waves into information that can be interpreted by the brain. Hearing phenomena, such as masking, and different ways in which the brain interprets this sound information will also be discussed. Chapter 4 will show the reader just how biased a sound indicator the ear really is, as opposed to the impartial Music Muse. The reader will also gain an understanding of why certain sounds are universally pleasing, while others are not, and how the Music Muse can be a visual sound quality indicator.

Chapter 5 is the voice chapter. It will introduce the voice mechanism and how it produces sound. The discussion of formant frequencies will also be expanded on. Chapter 5 will show the reader how the harmonic spectra that they see using the Music Muse are produced physically.

Part II will finally present the program and how it works. Each screen will have its own chapter, which will explain its purpose, give

detailed instructions on its usage. The last chapter of this section will present related software programs.

PART I: BACKGROUND

The Music Muse is a scientific device that can help singers strengthen their techniques. It is geared towards musicians, but is utilizes concepts of engineering. This section of the thesis will present the engineering concepts that went into the design of the Music Muse, including acoustics, vibrations, and signal processing. It will also relate these concepts to both the hearing mechanism and the voice mechanism. The purpose of this section is to show how engineering has been used to create a tool for artists.

CHAPTER 2: FUNDAMENTALS OF ACOUSTICS AND VIBRATIONS

An acoustic wave is a vibrational disturbance that propagates in an elastic medium.[1] This disturbance displaces the atoms or molecules of the medium from their normal configurations, which subsequently causes internal elastic restoring forces to arise. These elastic restoring forces, coupled with the inertia of the system, cause the particles in the medium to oscillate, and a resulting wave to propagate through the medium. If the frequency of the oscillations is between 20 and 20,000 Hz, which is termed the *audible range*, the resulting wave is generally perceived by humans as sound. Acoustic waves with frequencies above and below the audible range are called *ultrasonic* and *infrasonic waves*, respectively, and are not perceived as sound[2]. The following sections will explain some of the fundamentals of simple acoustics and vibrations that will be referred to in the preceding sections of this thesis.

2.1 Harmonics and Inharmonics.

Consider a rope held at one end and fixed at the other. If the rope is forced at the held end, a wave will travel down the rope to the fixed end, and then reflect back. If the rope is forced rapidly enough, there will be waves traveling in both directions, some of which will interfere and cause a jumble. However, if the rope is forced at just the right frequency, a *standing wave* will be produced. A standing wave is a wave that does not appear to move. The lowest possible frequency that will produce a standing wave is called the *fundamental frequency*, or *first harmonic* of the rope, and looks like the one in Fig. 2.1a. The wavelength of the fundamental frequency is equal to twice the length of the rope. The next highest frequency that will produce a standing wave is twice the fundamental frequency, and the resulting standing wave is shown in Fig. 2.1b. Similarly, the next highest frequency to produce a standing wave is three times the fundamental, shown in Fig. 2.1c. Figs 2.1b and 2.1c show the *second and third harmonics* of the rope, respectively.[3]

---

[1] Beranek and Ver, pp.1; Coppens et. al. pp.1

[2] Coppens et. al. pp 1

[3] Giancoli pp. 372-373

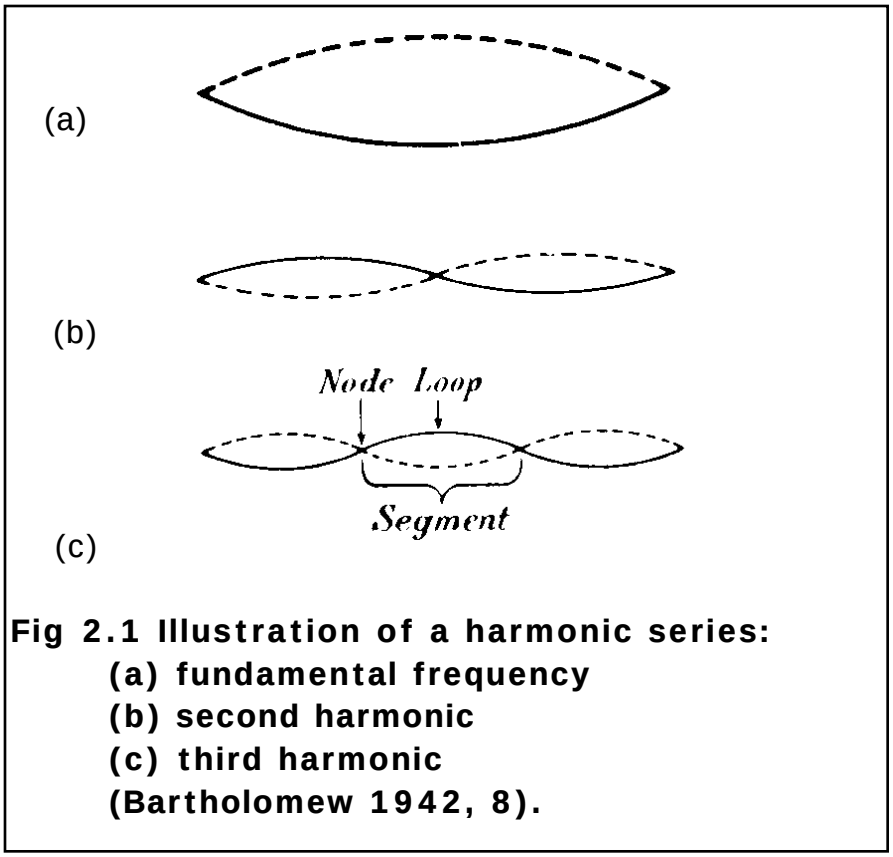


Fig 2.1 Illustration of a harmonic series:
(a) fundamental frequency
(b) second harmonic
(c) third harmonic
(Bartholomew 1942, 8).

Harmonics, or *overtones*, are waves with frequencies that are integer multiples of the fundamental frequency. The places where the wave is at equilibrium are called *nodes.* The fundamental frequency together with its harmonics make up what is known as the *harmonic series*. Frequencies that are not integer multiples of the fundamental are called *inharmonics.*[4]

2.2 Simple and Complex Vibration

The simplest type of vibration is called *simple harmonic motion*. It is characterized by a single frequency, and is best illustrated by a sine

[4] Bartholomew pp. 7-11

wave. If the frequency of simple harmonic motion of a body is within the audible range, it produces a *pure tone*.

Complex vibration occurs when a body vibrates at more than one frequency at the same time. For example, the soundboard of a piano vibrates in a complex fashion when multiple notes are played simultaneously. Also, the eardrum vibrates in a complex fashion, which is why people can hear more than one sound at a time.[5] In music, pure tones are nearly impossible. "Pure tones have to be generated with electronic oscillators; there is no musical instrument that produces them (and even for electronically generated pure tones, there is no guarantee that they will be 'pure' when they actually reach our ear)."[6] In music, the closest approximations to pure sounds are those produced by a tuning fork, a flute, or by a falsetto "oo."[7]

2.3 Beating.

When multiple pure tones come in contact, they interfere with one another, and their amplitudes add together. They interfere constructively when positive parts come in contact, and destructively if positive meets negative. When sound waves with frequencies that are close to each other come in contact, they alternately interfere constructively and destructively, causing periodic fluctuations in the amplitude of the tone produced. These fluctuations are referred to as *beats*. In music, if the two frequencies are close enough in value, the characteristic beating causes an unpleasant rough sensation, that tells the listener the sounds are out of tune. This type of beating will be discussed in further detail in a later chapter.

---

[5] Bartholomew pp. 10
[6] Roederer pp. 19
[7] Bartholomew pp. 15

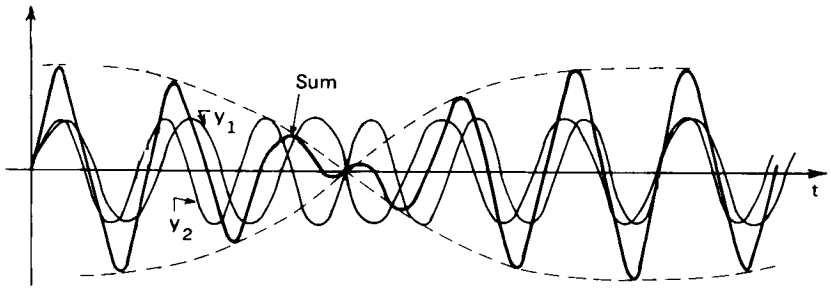


Fig 2.2 Beating phenomenon that occurs when two waves out of phase come in contact. The superposition of two waves slightly out of phase produces a resultant wave, in bold print, that modulates in amplitude (Roederer 1975, 27). (BMP, 99KB)

Beating also occurs when tones with the same frequency come in contact, but they are slightly out of phase. Fig. 2.2 illustrates this effect. The bold dark line represents the waves that results from the superposition of two waves slightly out of phase. The amplitude of the resultant wave modulates. This modulation of amplitude is the beating sensation that is heard. Beats can also occur when the overtones of complex tones come in contact.[8]

2.4 Resonance.

"Everything that possesses the properties of mass and compliance is a resonator."[9] An enclosed volume of air fits this description, and is used most often as a sound resonator. In music, examples of such resonators

[8] Roederer pp. 26-27
[9] Sundberg pp. 11

are the bodies of violins, the soundboards of pianos, and the vocal tracts of humans.

Sound usually dissipates rapidly as it travels away from its source. For this reason, resonators are used to reinforce the sounds produced by musical instruments. For example, when a violin is played, the sounds heard are not those from the vibrating strings, but those from the vibrating volume of air in the body.[10] "The strings by themselves would merely cut through the air, back and forth, without generating much sound."[11] The resonators are what make the instruments.

Resonators are characterized by their natural, or *resonant frequencies*. These are the frequencies at which the resonators vibrate optimally, and they depend on the volume, size, and shape of the resonators. In music, resonant frequencies are called *formants*. The subject of formants will be expanded on later in this chapter.

The type of resonance described above is called *sympathetic resonance*. The bodies of instruments vibrate sympathetically with their sources to reinforce their sounds. To further illustrate this concept, hold a vibrating tuning fork up to a flute. Without blowing into the opening, finger the tuning fork tone on the flute. The flute will produce a faint sound with a pitch equal to that of the tuning fork. If a different note is fingered on the flute, the effect will disappear. Sympathetic resonance occurs when one body vibrates at or near a resonant frequency of another. *Forced resonance* occurs when a body is forced to vibrate at its natural frequency by a body with a different natural frequency, such as when a plate is struck by a fork.[12]

2.5 Relative Pitch vs Formant Theory

The *relative pitch theory* states that no matter what the fundamental frequency is, the harmonic spectrum produced by the tone will remain the same relative to the fundamental for any given register of

---

[10] Bartholomew pp. 37
[11] ibid.
[12] Bartholomew pp. 27-28

the instrument,[13] where a register is defined as a "phonation frequency range in which all tones are perceived as being produced in a similar way, and which possess a similar voice timbre."[14] Fig 2.3 illustrates this concept. The fundamental frequencies of the tones are represented by lines extending from the first harmonics to the corresponding piano keys. No matter which fundamentals are played, the relative strengths of the harmonics in the spectra do not change.

The relative pitch theory, however, is only satisfied eighty or ninety percent of the time. The rest of the time, what is known as the *formant theory* comes into effect. The formant theory states that, due to sympathetic resonance, the partials that are near a formant frequency will be augmented.[15] This theory is illustrated in Fig 2.4. This figure suggests that a formant exists somewhere near the sixth, fifth, and fourth partials of the first, second, and third spectra, respectively. In the first spectrum, the sixth partial is the strongest. In the second spectrum, the fundamental is raised so that the sixth partial moves away from the formant, so its magnitude decreases, while the fifth partial moves closer to the formant and gets stronger. This process is repeated in the third spectrum.

The two theories of harmonic structure do peacefully coexist.[16] Even though Fig 2.3 appears to obey the theory of relative pitch, it does suggest the existence of a formant somewhere between the second and third partials of the first spectrum. As the fundamental rises, its amplitude along with the amplitude of the second partial appear to increase as though they are moving towards the formant, while the third partial decreases as it moves away. Therefore, the formant theory appears to have the most accurate explanation of the harmonic structure of sounds, and will be assumed to be true throughout this thesis.

---

[13] Vennard pp. 125
[14] Sundberg pp. 49
[15] Vennard pp. 125
[16] ibid.

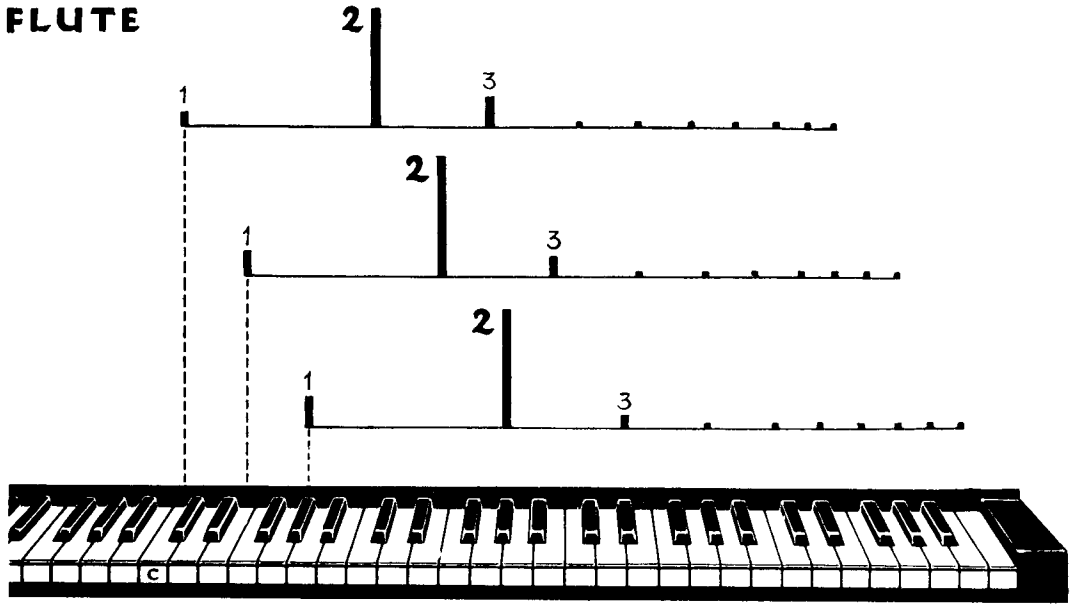


Fig. 2.3 The relative pitch theory states that for a given instrument, the amplitudes of the harmonics in any tone will stay the same relative to the fundamental. In this figure, the amplitudes of all the harmonics appear to remain constant relative to the changing fundamentals (Vennard 1968, 126). (BMP, 223 KB)

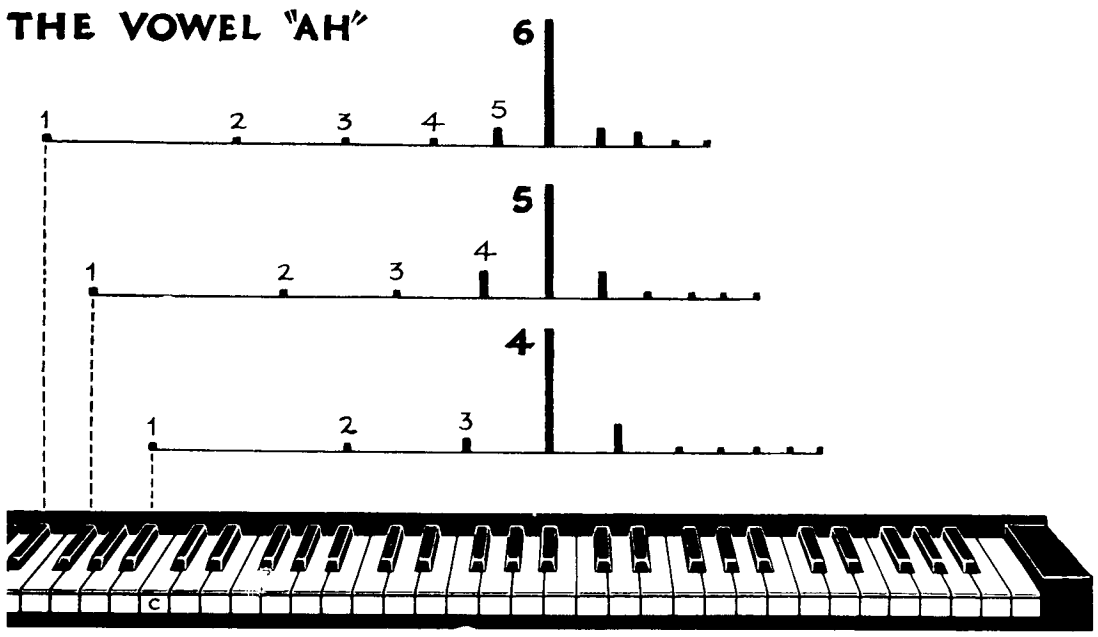


Fig. 2.4 The formant theory states that no matter what the fundamental frequency is, the harmonics that are closest to a formant frequency will always be amplified. This figure suggests the presence of a formant frequency somewhere around the 6th, 5th, and 4th harmonics of the 1st, 2nd, and 3rd spectra, respectively (Vennard 1968, 126). (BMP, 238 KB)

CHAPTER 3: BASIC CONCEPTS IN SIGNAL PROCESSING

The following sections will review the mathematics and concepts of signal processing that are used in the creation, and are necessary for the comprehension of, the Labview® program.

3.1 Conversion of a Signal From Time to Frequency Domain

Any periodic function can be expressed as a Fourier series, which is a sum of weighted sines and cosines. If x(t) is a periodic function of time t, with period T, the Fourier series form of x(t) is:[17]

(3.1) $$x(t) = a_\circ + \sum_{k=1}^{\infty}\left[a_k \cos\frac{2\pi kt}{T} + b_k \sin\frac{2\pi kt}{T} \right]$$

where a_o, a_k, and b_k are the constant Fourier coefficients given by:

(3.2a) $$a_\circ = \frac{1}{T}\int_{-\frac{T}{2}}^{\frac{T}{2}} x(t)dt$$

(3.2b) $$a_k = \frac{2}{T}\int_{-\frac{T}{2}}^{\frac{T}{2}} x(t)\cos\frac{2\pi kt}{T}\,dt \qquad k \geq 1$$

(3.2c) $$b_k = \frac{2}{T}\int_{-\frac{T}{2}}^{\frac{T}{2}} x(t)\sin\frac{2\pi kt}{T}\,dt \qquad k \geq 1$$

"The mathematical conditions for the convergence of equation 3.2 are extremely general and cover practically every conceivable engineering situation. The only important restriction is that, when x(t) is discontinuous, the series gives the average value of x(t) at the discontinuity."[18]

---

[17] Newland pp. 34

[18] ibid.

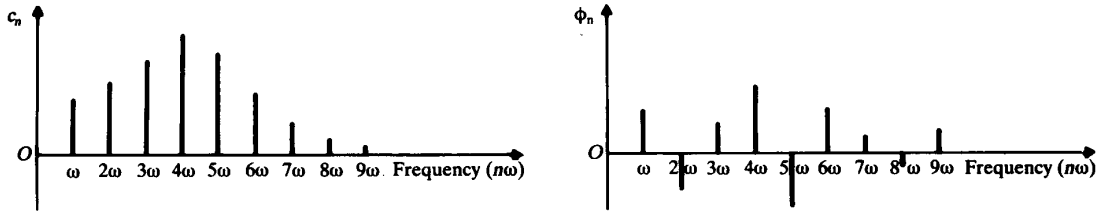


Fig 3.1 Example of a harmonic spectrum showing (a) magnitude and (b) phase. The lengths of the spectral lines represent the amplitudes of the harmonics at each frequency, ω (Rao 1990, 39). (BMP, 485KB)

Substituting $\omega_k = \frac{2\pi k}{T}$ into equations 3.2b and 3.2c yields the following:

(3.3a) $$a_k = \frac{2}{T}\int_{-\frac{T}{2}}^{\frac{T}{2}} x(t)\cos\omega_k t\,dt \qquad k \geq 1$$

(3.3b) $$b_k = \frac{2}{T}\int_{\frac{-T}{2}}^{\frac{T}{2}} x(t)\sin\omega_k t\,dt \qquad k \geq 1$$

The Fourier representation of x(t) now simplifies to:

(3.4) $$x(t) = a_\circ + \sum_{k=1}^{\infty}\left[a_k \cos\omega_k t + b_k \sin\omega_k t\right]$$

where $a_k \cos\omega_k t$ and $b_k \sin\omega_k t$ are the harmonic functions of the series. The harmonics can be plotted as vertical lines on a diagram of amplitude vs. frequency, where a_k and b_k are the amplitudes, and the ω_k's are the

frequencies.[19] The ω_k's are spaced $\Delta\omega$ apart, where $\Delta\omega = \frac{2\pi}{T}$. Fig 3.1 is an example of this type of diagram, called a *harmonic spectrum.*

The concept that all periodic functions can be expressed as the sum of simple sine and cosine functions is fundamental to the Music Muse. The Music Muse acquires voice signals, breaks them up into harmonics, and plots their harmonic spectra. From the spectra, users can observe the timbral characteristics of their voices. The next chapter will go into more detail about timbral characteristics.

3.2 Fourier Transform Integrals

The Fourier series can also be expressed as an integral. If the t-axis is adjusted to that the mean value of x(t) is zero, then a_o = 0. Now, substituting (3.4a), (3.4b), ω_k and $\Delta\omega$ into (3.4), we get:[20]

$$(3.5)\quad x(t) = \sum_{k=1}^{\infty}\left\{\left[\frac{\Delta\omega}{\pi}\int_{\frac{-T}{2}}^{\frac{T}{2}} x(t)\cos\omega_k t\,dt\right]\cos\omega_k t\right\} + \sum_{k=1}^{\infty}\left\{\left[\frac{\Delta\omega}{\pi}\int_{\frac{T}{2}}^{\frac{T}{2}} x(t)\sin\omega_k t\,dt\right]\sin\omega_k t\right\}$$

When the period T $\to \infty$, $\Delta\omega \to d\omega$, and $\sum \to \int$. Now (3.5) becomes

$$(3.6)\quad x(t) = \int_{\omega=0}^{\infty}\frac{d\omega}{\pi}\left[\int_{-\infty}^{\infty} x(t)\cos\omega t\,dt\right]\cos\omega t + \int_{\omega=0}^{\infty}\frac{d\omega}{\pi}\left[\int_{-\infty}^{\infty} x(t)\sin\omega t\,dt\right]\sin\omega t$$

Defining the following functions

$$(3.7a)\qquad A(\omega) = \frac{1}{2\pi}\int_{-\infty}^{\infty} x(t)\cos\omega t\,dt$$

$$(3.7b)\qquad B(\omega) = \frac{1}{2\pi}\int_{-\infty}^{\infty} x(t)\sin\omega t\,dt$$

---

[19] Rao pp. 38-39

[20] Newland pp. 35

and substituting into (3.6) gives

(3.8) $$x(t) = 2\int_0^\infty A(\omega)\cos\omega t d\omega + 2\int_0^\infty B(\omega)\sin\omega t d\omega$$

The terms *A(ω)* and *B(ω)* are the components of the *Fourier Transform* of x(t) into the frequency domain. The final representation of x(t) is known as the *inverse Fourier transform.*[21]

3.3 Complex Fourier Transform

Define X(ω) as

(3.9) $$X(\omega) = A(\omega) - iB(\omega)$$

Now combining (2.8a) and (2.8b) we get

(3.10) $$X(\omega) = \frac{1}{2\pi}\int_{-\infty}^{\infty} x(t)(\cos\omega t - i\sin\omega t)dt$$

Euler's equation says that

(3.11) $$e^{i\theta} = \cos\theta + i\sin\theta$$

Substituting this into (3.11) gives use the *complex Fourier Transform* of x(t):

(3.12) $$X(\omega) = \frac{1}{2\pi}\int_{-\infty}^{\infty} x(t)e^{-i\omega t}dt$$

The inverse Fourier transform for x(t) in terms of X(ω) is

$$x(t) = \int_{-\infty}^{\infty} X(\omega)e^{i\omega t}d\omega$$

---

[21] ibid.

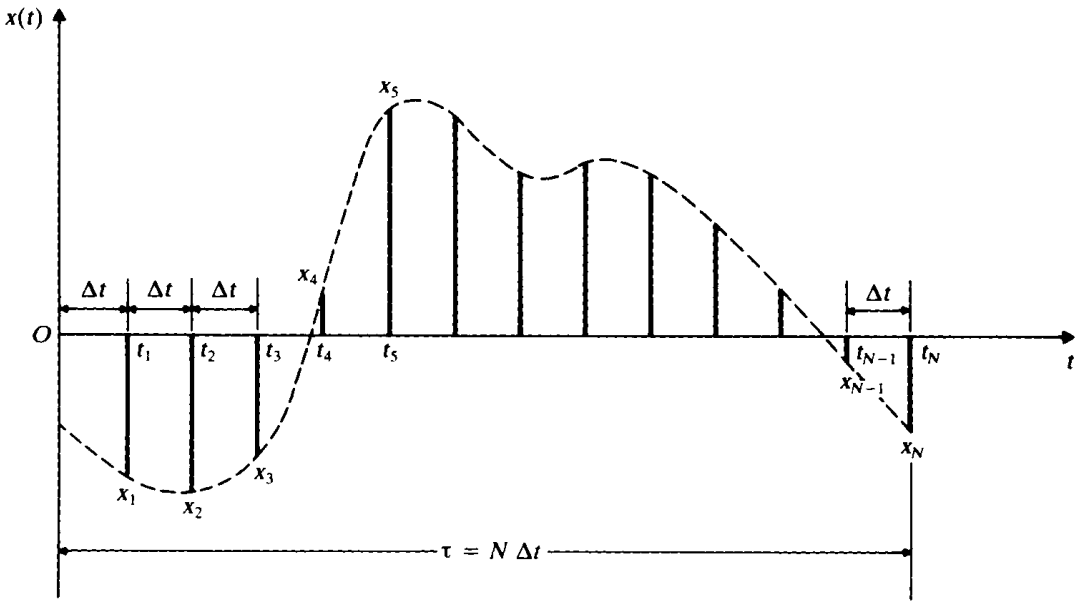


Fig 3.2 Periodic function, x(t), with period τ, discretely sampled at uniform intervals of Δt, for a total of N samples (Rao 1990, 42). BMP, 196KB)

Equations (3.12) and (3.13) together from what is known as the *Fourier transform pair.*[22]

(3.4) Discrete Fourier Transform

The previous expressions for the Fourier transform of a signal all dealt with continuous data. However, in most real situations, such as with the Music Muse, only finite samples of the actual signal are available. Therefore, the Fourier series must have a form that can be analyzed discretely. Consider a uniformly sampled discrete time series, x(t), such as the one in Fig 3.2, with period T, and a total of N samples taken. Modify equations (3.3) and (3.4) by letting the integrals in the

[22] ibid. pp. 38-39

expressions for a_k and b_k range from 0 to T rather than from -T/2 to T/2. Next, combine a_k and b_k into one complex equation by defining:

(3.14) $X_k = a_k - ib_k$

and substituting

(3.15) $e^{-i\omega_k t} = \cos\omega_k t - i\sin\omega_k t$

to give

(3.17) $X_k = \frac{1}{T}\int_0^T x(t)e^{-i\omega_k t}dt$

This equation can be approximated by the *discrete Fourier transform*:

(3.17) $X_k = \frac{1}{N}\sum_{n=0}^{N-1} x(n\Delta t)e^{-i\left(\frac{2\pi k}{N}\right)n\Delta t}$

and x(t) is subsequently approximated by the *inverse discrete Fourier transform*:[23]

(3.18) $$x(t) = \sum_{k=0}^{N-1} X_k e^{i\left(\frac{2\pi k}{N}\right)n\Delta t}$$

3.5 Cepstral Analysis of Speech Signals

A speech signal is commonly modeled as the convolution of two signals: a source signal that is produced by the vocal cords, and the impulse response function of the vocal tract. Voiced sound is produced by the "quasi-periodic pulses of airflow" caused by the vocal cords.[24] The vocal tract is modeled as a linear, time-varying filter. For sustained vowel sounds, such as those produced during singing, the parameters of

[23] ibid. pp. 116
[24] Oppenheim pp. 816

the vocal tract change slowly enough to assume that they are constant over a small interval in time.[25] In speech synthesis, these small intervals are usually around 10 - 30 ms.[26] The resonant frequencies of the filter model of the vocal tract are called *formants*.[27] These formants are significant contributors to voice timbre. In order to calculate these frequencies, however, they must be deconvolved from the source signal produced by the vocal cords. This deconvolution can be accomplished by utilizing a method called *homomorphic deconvolution*, or *cepstral analysis*.

As an example, consider a 20 ms sample of a sung vowel sound. The following mathematical expression illustrates the voice signal model:[28]

(3.19) $x(t) = p(t) * v(t)$

where, $x(t) =$ speech signal
$p(t) =$ source signal
$v(t) =$ vocal tract response

This equation, however, is not valid at the ends of the voice interval, because of discontinuities caused by the airflow pulses that occur before and after the interval. To assuage this problem, the source signal, p(t), is multiplied by a window. The Hamming window is a popular choice because it tapers smoothly at both ends. Equation (3.19) now becomes:[29]

(3.20) $x(t) = p_w(t) * v(t)$

where $p_w(t) = w(t)p(t)$

Since convolution in time is equal to multiplication in the frequency domain, this expression becomes:

---

[25] ibid.
[26] O'Shaughnessy pp. 228
[27] Oppenheim pp. 816
[28] ibid.
[29] ibid. pp. 818

(3.21) $X(\omega) = P_w(\omega)V(\omega)$

where the functions in terms of ω are the complex Fourier transforms of the corresponding time signals.

The spectral contents of P and V are different enough to be separated by linear filtering. In order to separate the two signals, their product must be transformed into a sum of two signals. This transform is logarithmic:

(3.22) $\log[X] = \log[PV] = \log[P(\omega)] + \log[V(\omega)]$

The *complex cepstrum* of a signal X(ω) is defined as the inverse Fourier transform of log[X(ω)]:[30]

(3.23) $\hat{x}(t) = \int_{-\infty}^{\infty} log[X(\omega)]e^{i\omega t}d\omega$

Since X(ω) is complex, it is characterized by magnitude and phase. For speech signals, the phase information is relatively insignificant and may be discarded, and the *real cepstrum* or simply the *cepstrum*, suffices. It is defined as the inverse Fourier transform of the log of the magnitude spectrum:[31]

(3.24) $\hat{c}(t) = \int_{0}^{2\pi} \log[|X(\omega)|]e^{i\omega t}d\omega$

The cepstrum can also be calculated by using the discrete Fourier transform in place of the continuous transform:

(3.25) $\hat{c}_d(t) = \sum_{k=0}^{N-1} \log[|X_k|]e^{i\left(\frac{2\pi k}{N}\right)n}$

[30] O'Shaughnessy pp. 299
[31] ibid. pp. 230

Since the cepstrum is the inverse transform of log[X(ω)], we know from equation (3.22) that it is also the inverse transform of log[P(ω)] + log[V(ω)]. Therefore, it can be seen that

(3.26) $\hat{c} = \hat{p} + \hat{v}$

The vocal tract information can be found in $\hat{v}$, and the excitation information can be found in $\hat{p}$. Since the vocal tract component varies slowly with respect to the excitation, the two components are easy to distinguish. The vocal tract information can generally be found in the first 3 - 4 ms of the cepstrum. It must be noted, however, that equation (3.20) is only valid for $w(t) \approx w(t+M)$, where M is the effective duration of v(t). Since typical cepstral analysis windows violate this rule, is has been found that $\hat{v}$is "repeated every pitch period and is subject to double sinc-like distortion."[32] To compensate, the cepstral windows should be halved to avoid aliasing. As a result, the cepstral windows should be 1 - 2 ms. The formant frequencies are then found by a Fourier transform of this cepstral window.

Fig 3.3 shows an example of cepstral analysis of a female singing an "ah," evaluated using the Music Muse. The signal was sampled at a rate of 10000 samples/sec, and multiplied by a Hamming window. The first 256 points were analyzed, which is equivalent to 25.6 ms. The graph of the log spectrum shows that there is obviously a slowly varying component as well as a rapidly varying component. The rapidly varying component is the excitation information, which shows up in the cepstrum as regularly spaced impulses. The vocal tract information is at the lower end of the cepstrum. A discrete Fourier transform of the first 1.5 ms of the cepstrum yields the formant plot, which follows the envelope of the slowly varying components of the log spectrum. The formant frequencies are the peaks of the plot. Since the formants are resonances, the amplitudes of the frequencies in the harmonic spectrum are highest at or near these frequencies.

---

[32] ibid. 229

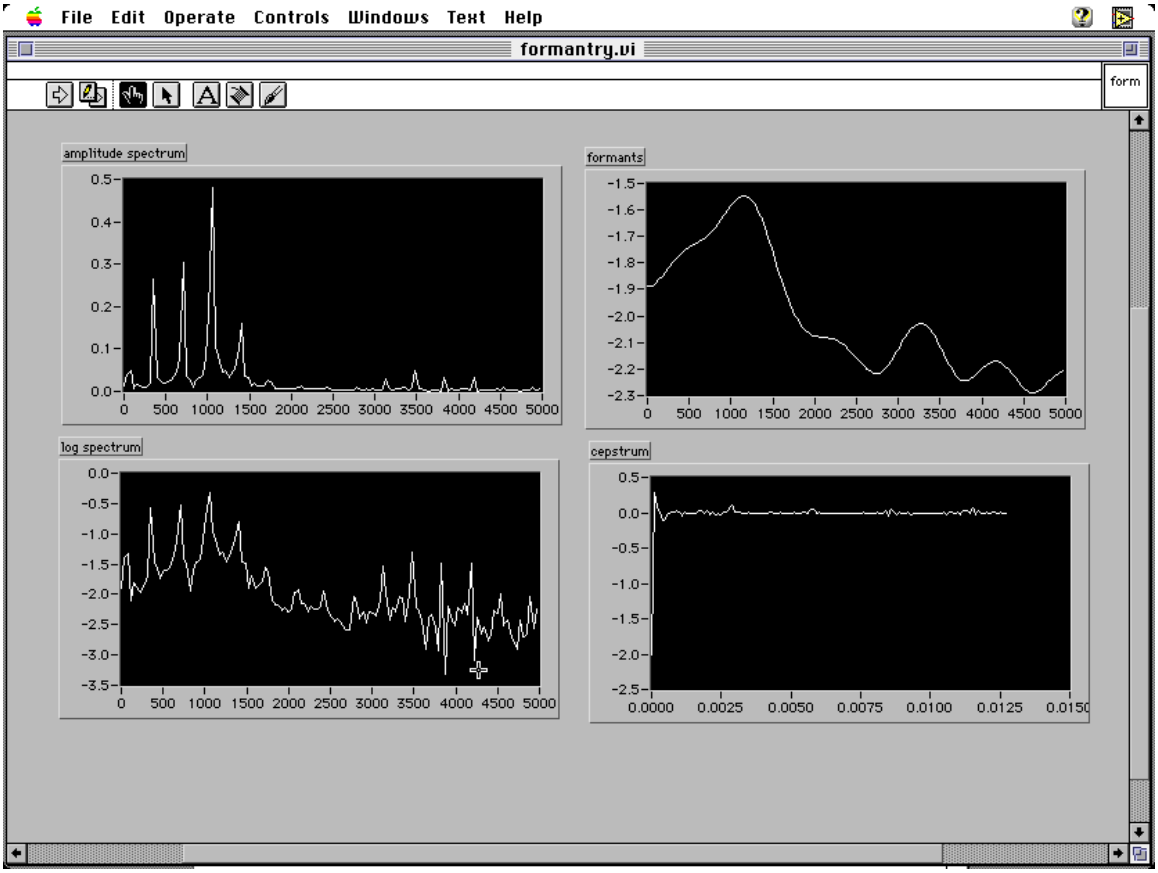


Fig 3.3 Example of cepstral analysis using the Music Muse. The log spectrum is composed of a slowly varying component and a rapidly varying component, which are separated in the cepstrum. A DFT of the slowly varying component, which is located at the low end of the cepstrum, yields the formants plot, the peaks of which determine the formant frequencies. The formant frequencies lie in the range where the amplitudes in the harmonic spectrum are the highest. (BMP, 64KB)

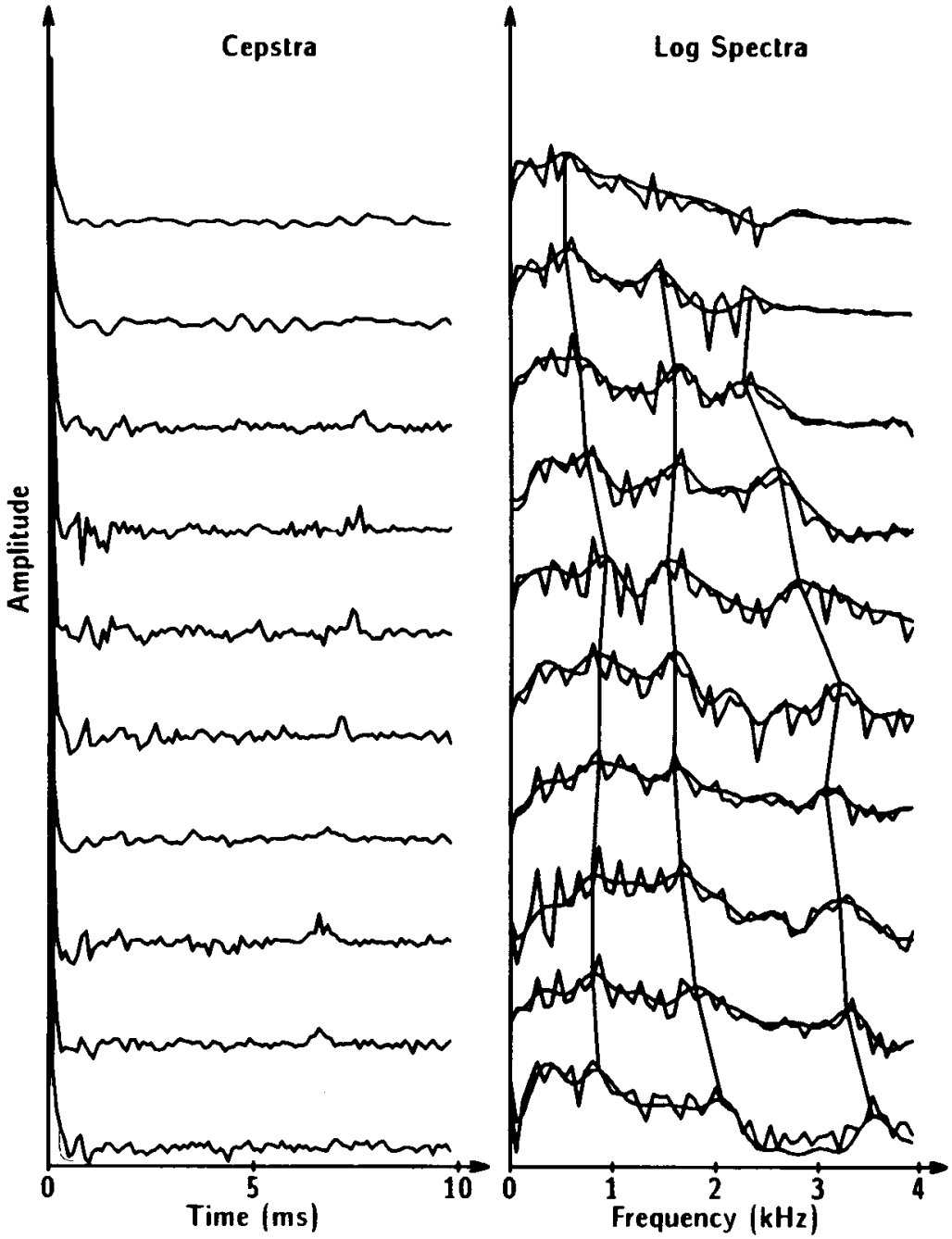


Fig. 3.4 Cepstral analysis of a speech signal divided into ten consecutive segments, each 10 ms in duration. The formant plots are superimposed on top of the corresponding log spectra. The vertical lines connect the

Fig. 3.4 shows another example of the cepstral analysis of a speech signal divided into ten consecutive samples, each with a duration of 10 ms. The formant plots are superimposed on top of the log spectra. The peaks of the formant plots are the formant frequencies, and are all connected to show how the formants vary with time.

CHAPTER 4: THE PSYCHOPHYSICS OF MUSIC

The purpose of the ear is to convert sound waves into nerve impulses that can be interpreted by the brain in terms of pitch. The ear is an incredible device, so ingeniously and meticulously crafted, to some it may suggest the existence of a divine creator. The ear has astounding frequency analyzing and filtering capabilities. It can extract a single pitch from a complex tone. It can also extract multiple pitches from overlapping complex tones, such as those heard in a band or orchestra.[33] The ear is especially responsive to sound frequencies within the voice range (approximately 200 - 5600 Hz), amplifying sound energy within this range, and attenuating energy in other ranges.[34]

Psychophysics is defined by The American Heritage Dictionary as "the psychological study of the relationships between physical stimuli and sensory response.[35] The psychophysics of music deals with sound perception, and how different tones interfere with one another.[36] Learning about psychophysics helps us to understand what sounds are pleasing to the ears, and which are not. The following sections will be an introductory lesson in psychophisics. The first section will explain the anatomy of the ear, and how it works to transform sound waves into a form perceivable by the brain. The remaining sections will divulge more deeply into the brain's interpretation of the sound information, in terms of pitch and timbre. After reading this chapter, it should be clear that the Music Muse is necessary for singers to have unbiased, unaltered representations of their voices. The reader will also understand how vocal quality can be determined by the Music Muse graphs.

4.1 The Ear

The ear has three main sections: the outer ear, the middle ear, and the inner ear.[37] The outer ear collects the sound waves, the middle ear

---

[33] Plomp pp. 2
[34] O'Shaughnessy pp. 128
[35] American Heritage Dictionary, The pp. 1000
[36] O'Shaughnessy pp. 140
[37] Coppens et. al. pp. 257

transforms the sound waves into mechanical motion, and the inner ear converts this motion into electrical signals which are interpreted by the brain.[38]

Sound enters the ear through the *pinna* of the outer ear, which is connected to the *auditory canal*. The sound then travels down the canal to the *tympanic membrane*, or *eardrum*, which is resultingly set into vibration. The eardrum serves as the entrance to the *middle ear*, which contains three *ossicles*, or bones: the *malleus, incus*, and *stapes*, commonly referred to as the hammer, anvil and stirrup respectively. The vibrations of the eardrum are transmitted through the chain of ossicles to the *oval window* , a membrane which marks the beginning of the *inner ear*.[39]

The purpose of the ossicles is to help provide an "approximate impedance match between the air in the auditory canal and the liquid in the inner ear.[40] "The acoustic impedance of the inner ear fluid is about 4000 times that of air."[41] Without the impedance transformation of the ossicles,
all but about 0.1% of the sound waves hitting the eardrum would be reflected back with very little actually entering the inner ear.[42]

The middle ear also helps protect the delicate mechanism of the inner ear from damage caused by excessively intense sounds.[43] The impedance match from the ossicles varies with the intensity of the received sound. "For high intensities, the muscles controlling the motion of the ossicles change their tension to reduce the amplitude of motion of the stapes, thereby protecting the delicate mechanism of the inner ear from damage. This process is known as the *acoustic reflex*. Since it takes about 0.5 ms after perceiving a loud sound for the acoustic reflex to become effective, it offers no protection from sudden impulsive sounds

---

[38] Roederer pp. 18
[39] Coppens et al. pp. 257-258; Roederer pp. 19-20
[40] Coppens et al. pp. 258
[41] O'Shaughnessy pp. 130
[42] ibid.
[43] Coppens et al. pp. 259; O'Shaughness pp. 130

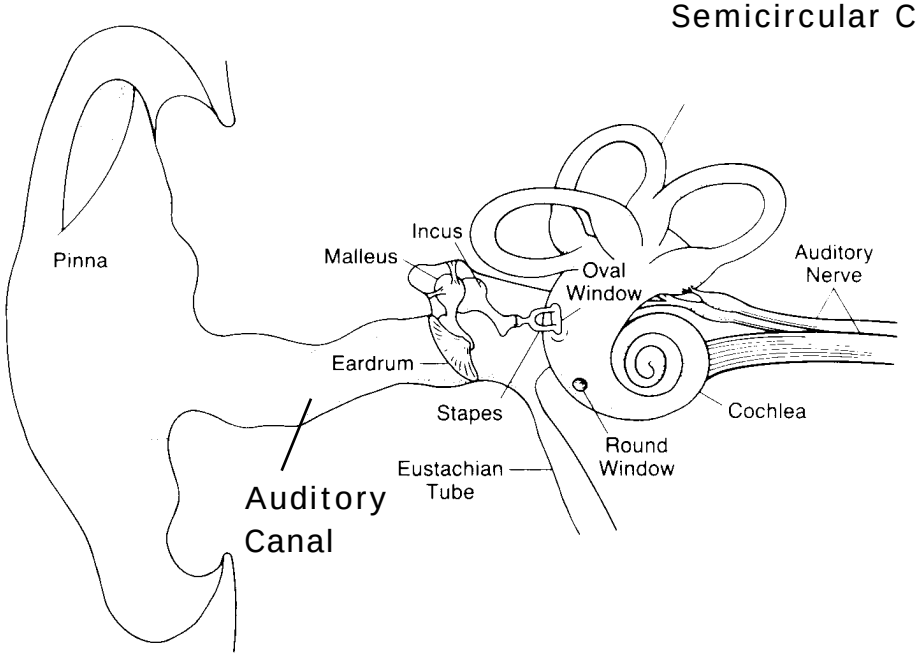


Fig 4.1 Diagram of the hearing mechanism (O'Shaughnessy 1987, 129) (BMP, 213KB)

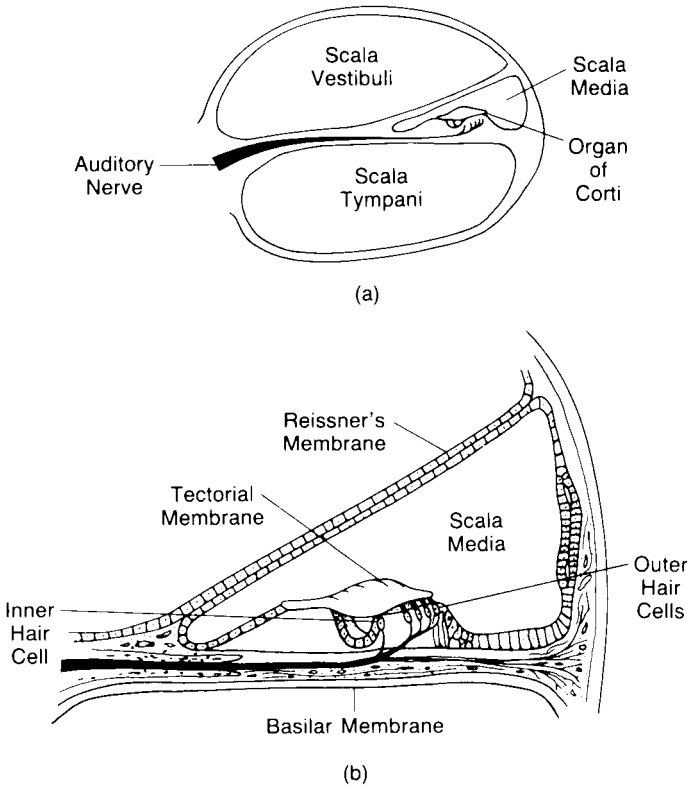


Fig 4.2 Anatomical features of the cochlea in the inner ear:
(a) cross-section of the cochlea
(b) structures within the scala media
(O'Shaughnessy 1987, 131).
(BMP, 179KB)

such as gunshots, explosions, and so forth."[44]

The stapes of the middle ear is attached to the *scala vestibuli* through the oval membrane, which is sealed to keep the inner ear fluid from leaking out.[45] Also attached to the scala vestibuli are the semicircular canals, which do not contribute to hearing, but are there to give us a sense of balance.[46]

The cochlea, located in the inner ear, is a tube of hard bone, wound in the shape of a snail's shell, and partitioned by two membranes into three separate chambers. The largest chamber is the scala vestibuli, and is separated from the smaller middle chamber, the *cochlear duct* or *scala media*, by *Reissner's membrane*. The cochlear duct is then separated from the third chamber, the *scala tympani*, by the *basilar membrane*. The scala vestibuli and the scala tympani are both filled with perilymphatic fluid, which is similar to spinal fluid. Liquid from the scala vestibuli can only enter the scala tympani through a small opening at the apex of the cochlea. The cochlear duct is filled with endolymphatic fluid, which is similar to the intercellular fluid found throughout the body.[47]

The *bony ledge*, which carries the *auditory nerve*, projects from the center of the cochlea into the cochlear liquid. At the end of the bony ledge, the auditory nerve fibers enter the basilar membrane. Above the basilar membrane is the *tectorial membrane* which connects to the bony ledge at one end, and enters the cochlear fluid at the other end.[48]

The *Organ of Corti*, attached to the top of the basilar membrane, contains rows of *hair cells* that span the length of the cochlea. Protruding from each hair cell are several dozen small hairs which extend to the under surface of the tectorial membrane.[49] The vibrations transmitted through the ossicles cause the flexible oval window to vibrate. This vibrating motion creates oscillations in the perilymphatic fluid, which travel down the scala vestibuli, around into the scala tympani, and back to the round

---

[44] Coppens et al. pp. 259
[45] O'Shaughnessy pp. 131
[46] Coppens et al. ibid.
[47] ibid.
[48] ibid.
[49] ibid.

window. The round window serves as a pressure-release termination.[50] This fluid disturbance sets the basilar membrane into motion. "Since the Organ of Corti is attached to the basilar membrane while the tectorial membrane is attached to the bony ledge, the relative motions generated between them flex the hairs, thereby exciting the nerve endings attached to the hair cells into producing electrical impulses" which are sent to the brain to be perceived as sound information (Coppens et al. 1982, 260-61).[51] Wow! How impressive was that? The ear is truly a magnificent organ.

4.2 Pitch

The pitch of a tone is the brain's interpretation of the tone's fundamental frequency. But how does the brain transform frequency information into a pitch? First consider only pure tones. The previous section explained how the oscillations of the cochlear fluid vibrate the basilar membrane, which flexes the hairs and hair cells, which excite nerve endings. The basilar membrane "varies gradually in shape and tautness along its length;" subsequently, "its frequency response varies accordingly.[52]

The basilar membrane is displaced a constant amount of 3.5 - 4 mm, no matter what the octave. Therefore, if a frequency jumps from 40 to 80 Hz or from 5000 to 10000 Hz, the displacement on the basilar membrane is the same. As a result, the relationship between the frequency of the tone and the corresponding distance from the base to the resonance region of the basilar membrane is logarithmic.[53] Fig 4.4 shows this logarithmic relationship.

---

[50] Coppens et al. ibid.; O'Shaughnessy pp. 131-132
[51] Coppens et al. pp. 260-261
[52] O'Shaughnessy pp. 132
[53] Roederer pp. 21-22

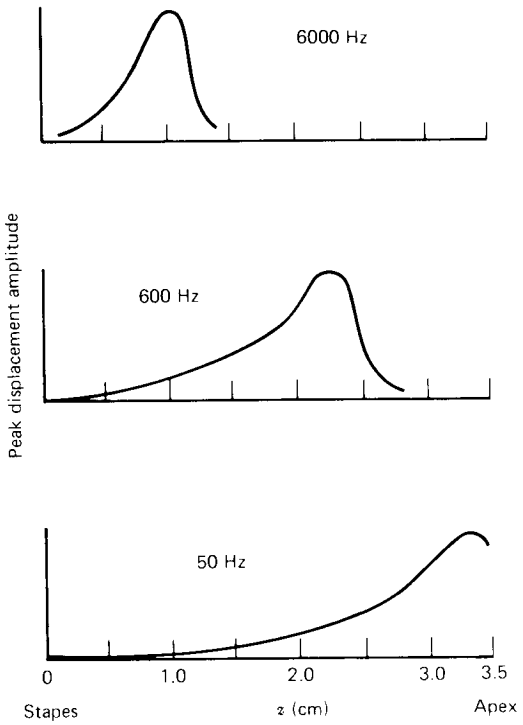


Fig 4.3 Peak displacements amplitudes of the basilar membrane for various pure tones. The lower the frequency, the closer the resonant

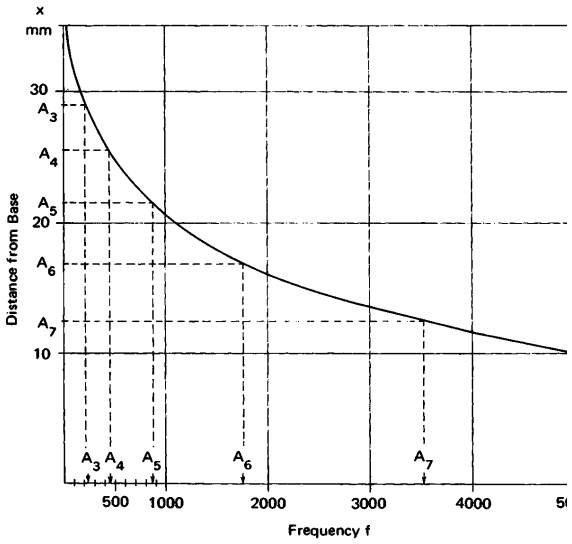


Fig 4.4 Logarithmic relationship between the position of maximum resonance along the basilar membrane and the frequency *f* of a pure tone (Roederer 1975, 22).

After reading about pitch determination, the following question arises: if a given frequency excites the basilar membrane over a finite region, why does the brain interpret a single sound and not a whole smear of sounds? Because the brain also has a "sharpening mechanism in the peripheral neural network.[54] The brain can actually figure out which pitch is right and hone in on it in a short enough time span that only one sound is perceived. This sharpening process "takes place after the signals from the stimulated region on the basilar membrane enter the neural network, in which the activity collected along the whole [maximum resonance] region is 'focused' or 'funneled' into a more limited number of responding neurons while the surrounding neurons have been inhibited (contrast enhancement)."[55]

4.3 The Superposition of Pure Tones

When two pure tones are sounded, their waveforms are linearly superposed, creating a single complex waveform equal to the sum of the individual motions of the waves. The eardrum then vibrates in the form of the resulting complex wave, causing the basilar membrane to be excited maximally in two regions at the same time.[56]

To illustrate the psychoacoustic effects of this linear superposition, consider two pure tones, *f1* and *f2*, that when sounded produce a complex tone *F*. Also, let $f2 = f1 + \Delta f$. If Δf is zero, then *f1* is equal to *f2*, and a single pitch equal to *f1* is produced. If Δf is smaller than a certain limit called the *just noticeable difference*, or *jnd*, then *f1* and *f2* are judged as having the same pitch, and a single pitch is heard. This jnd, or frequency resolution, depends on "the frequency, intensity, and duration of the tone in question-- and on the suddenness of the frequency change. It varies greatly from person to person, is a function of musical training, and unfortunately, depends considerably on the method of measurement employed."[57]

---

[54] ibid. pp. 32
[55] ibid.
[56] ibid. pp. 27
[57] ibid. pp. 23

If Δf is increased slightly still, the resonance regions of the basilar membrane will overlap, and beating will occur. A single frequency equal to $(f1 + f2)/2$ will be heard, whose amplitude will beat at a frequency equal to Δf. As Δf increases, the beats increase as well. When Δf reaches about 15 Hz, the beats disappear and a characteristic unpleasant roughness, or *dissonance*, occurs.[58]

When Δf surpasses what is called the limit of frequency discrimination, Δf_d, the excited regions along the basilar membrane are well separated, and two separate tones are heard. However, the disagreeable sensation of roughness is still apparent until Δf supersedes what is known as the critical band, Δf_{cb}. Now, two separate tones are heard with a pleasurable, or *consonant* sound.[59]

4.4 Central Pitch Processor

One of the most amazing things about the ear is that it can distinguish a single tone from a complex combination of tones. When a complex tone is sounded, the pitch is generally perceived to be that of the fundamental frequency. This is because the brain has a mechanism that learns to recognize certain recurring patterns of harmonics. This mechanism is called the *central pitch processor*. Its main function is to match sounds with patterns that have already been programmed into the brain over time. These patterns correspond to spatial patterns of resonance maxima along the basilar membrane. This recognition process is used by the brain for many other applications. For example, if a person stares at a cloud long enough, the cloud may begin to look like a bear or a tree or some other figure that recurs in nature, because the brain matches it with a recognizable image. This process provides an efficient way for the brain to conserve memory and minimize redundancy.[60]

The unique pitch that is sensed by the central pitch processor is usually the fundamental because it is most often the most prominent.

---

[58] ibid. pp. 28
[59] ibid.
[60] ibid. pp. 54

However, this process also works if some of the elements are missing from a pattern. This is because the brain can perceive "nonexistent--but expected-- contours."[61] This is similar to the way we might recognize the word "*mther*" as a misspelled version of "*mother*." This phenomenon is most prominent with lower frequency tones.[62]

4.5 Thresholds

As was mentioned before, the audible range is about 20 - 20000 Hz. However, sound frequencies below 1 kHz and above 5 kHz require significantly more energy to be heard than those frequencies between 1 and 5 kHz. This is due to both the filtering actions of the outer and middle ear, and also the smaller number of hair cells with very low or very high characteristic frequencies with respect to the mid-frequency range. "The minimum intensity at which sounds can be perceived is known as the *auditory* or *hearing threshold*, which rises sharply with decreasing frequency below 1 kHz and with increasing frequency above 5 kHz."[63]

The audibility threshold rises sharply to a cutoff for frequencies above 5 kHz. The region above 5 kHz is where the greatest amount of variability occurs between different listeners, particularly for listeners above the age of 30. The cutoff frequency for young listeners can be as high as 20 - 25 kHz, whereas people in their 40's and 50's can seldom hear frequencies above 15 kHz. The audibility threshold is for the most part independent of age for the frequency range below 1 kHz.[64]

As sound levels increase, they approach the *threshold of feeling* at about 120 dB, where the sound energy can actually be felt. This threshold is independent of frequency. As sound levels increase past the threshold of feeling, they eventually reach the *threshold of pain* at about 140 dB. This threshold is damaging to the ear, however damage can occur at levels below 100 dB, depending on the duration of the sound and the amount of exposure.[65]

---

[61] ibid. pp. 55
[62] ibid.
[63] O'Shaughnessy pp. 141
[64] Coppens et al. pp. 263
[65] ibid.

As was mentioned earlier, the ear responds to loud sounds by reducing its sensitivity. As a result, the threshold of audibility shifts upward when exposed to loud
sounds, the amount of which depends on the actual intensity and duration of the loud sound. When the sound is removed, the ear gradually recovers, shifting the audibility threshold back down to its original value. If the ear fully recovers, *temporary threshold shift*, or *TTS*, has occurred. If the ear does not recover completely, *PTS* or *permanent threshold shift* has occurred. PTS causes permanent damages to the hair cells in the inner ear.[66]

Another threshold is the *differential threshold*, which is the amount of fluctuation the ear detects during the beat phenomenon. This value depends on the frequencies and intensities of the tones producing the beats, and the number of beats per second. The greatest sensitivity to intensity changes occurs around 3 beats per second. Sensitivity decreases at the frequency extremes, especially for lower frequencies.[67] One other threshold to consider is the *masking threshold*, which will be introduced in the next section.

4.6 Masking

Masking occurs when the "perception of one sound is obscured by the presence of another."[68] When two tones are heard simultaneously, or if one tone is produced a short time after the other, one of the tones can raise the audibility threshold of the other. Simultaneous sounds cause *frequency masking*, where the lower frequency masks the higher one.[69] Frequency masking is illustrated in Fig 4.5. The masking tones are 400 Hz in (a), and 2000 Hz in (b). The masked frequencies are plotted on the x-axes versus their subsequent threshold shifts on the y-axes. The curves are labeled with the intensities of the masking tones. The louder the

---

[66] ibid.
[67] ibid.
[68] O'Shaughnessy pp. 146
[69] ibid.

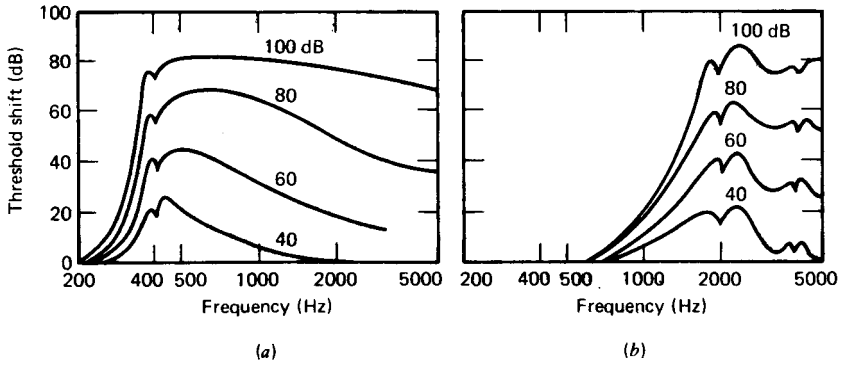


Fig 4.5 Masking of one pure tone by another. The masking tones are 400 Hz in (a) and 2000 Hz in (b). The masked frequencies and their resultant threshold shifts are plotted on the x and y-axes, respectively. The curves are labeled with the intensities, L_I, of the masking tones. The louder the masking tone, the more frequencies it masks. Also, it is evident that lower frequencies mask do more masking than higher frequencies (Coppens et al. 1982, 266). (BMP, 110 KB)

masking tone, the more frequencies it masks. Note that the lower frequency masks more tones than the higher frequency.

Temporal masking occurs when two sounds within each other's critical band are delayed with respect to one another, and one or both sounds are masked. Forward temporal masking is when the masker is sounded first, and backwards temporal masking is when the masker is sounded second.[70]

For forward temporal masking, as the delay between the termination of the masker and the beginning of the masked tone increases, so does the amount of energy required to mask the tone. Forward masking is most effective if the two tones occur within 10 ms. This effect

[70] ibid. pp. 151

decreases with time, until there is not effect beyond about a 200 ms delay.[71]

Backward masking occurs when a short tone burst is not heard because the tone sounded immediately after it has sufficient noise. Backward masking only has an effect with a delay less than 20 ms, and decreases rapidly as the delay increases.[72]

In an earlier section, the jnd of frequency was defined as the minimum difference in frequency required for two tones to be distinguishable. The jnd of intensity is the minimum amount that the intensity of a tone must be changed in order to perceive a difference in loudness. This is similar to the masking threshold, which is the minimum difference that a masked tone must have to be perceived in the presence of the masker.[73]

4.7 Timbre.

Section 3.1 explained how the brain interprets the fundamental frequency of a tone as pitch. What about the rest of the tones in the spectrum? The presence of harmonics in a tone, and their amplitudes relative to each other, are all interpreted as *timbre.* In painting, many shades of color are combined to create desired hue. It is difficult to determine the individual colors that went into the mixture. Only their combined effect can be perceived. Music is similar. It is extremely difficult to distinguish all the different overtones that make up voice color. Their combined effect is interpreted by the brain as timbre.

In music, pure tones sound dull and lifeless. If a pure tone is generated by an oscillator, the sound produced will be "sweet but dull.[74] If the harmonics of the note are added one by one, the sound will gradually become brighter, and often more appealing.[75]

---

[71] ibid. pp. 151-152
[72] ibid.
[73] Roederer pp. 81
[74] Bartholomew pp. 13
[75] ibid.

Timbre is also what distinguishes musical instruments, as well as different singing voices, from one another. Timbre is dictated by the number of harmonics in a tones, and their amplitudes relative to each other. As was previously mentioned, the more harmonics, the brighter the sound. Fig. 4.6 shows the harmonic spectra produced by different instruments, all sounding a concert A (440 Hz). The figure shows that the bright timbre of the horn can be attributed to its relatively large number of harmonics, while the dark sounds of the oboe and flute correlate to their lower numbers of harmonics.

Relative amplitudes of the harmonics in a spectrum also contribute to timbre. Fig 4.7 shows an example of spectral differences calculated on one of the screens of the Music Muse. Voice 1between the harmonic spectra of a male and a female singing a frequency of about 300 Hz on "ah." Voice 1 and Voice 2 are female and male, respectively. The Difference In Overtone data is calculated as Voice 2 - Voice 1. Although the two singers are producing the same note, the resulting sounds have different timbres. These differences are direct results of the spectral differences shown in the graphs. These differences may be difficult to identify with the powerful yet imprecise ear, but with the Music Muse, they can be obtained with relative ease.

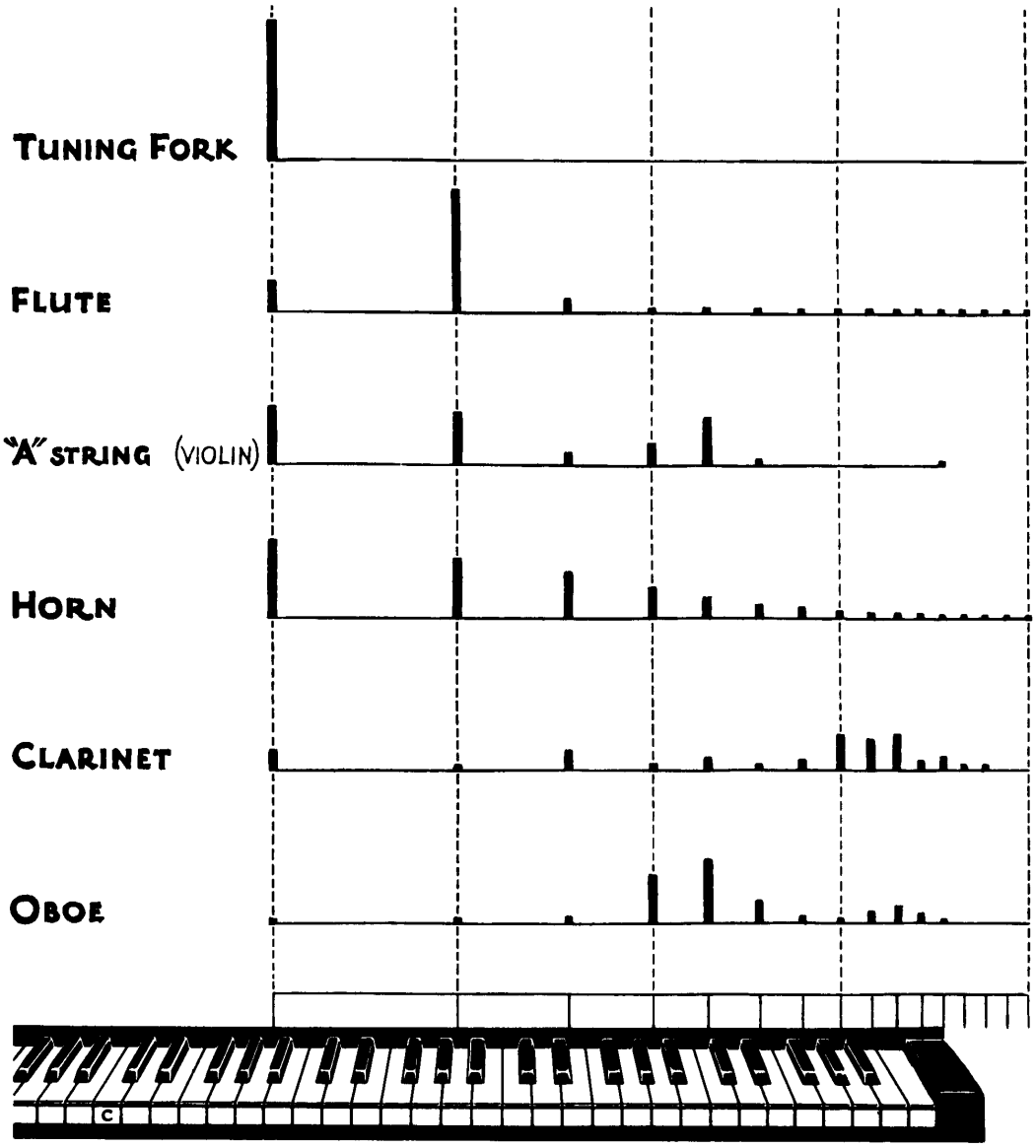


Fig. 4.6 Harmonic spectra of various instruments sounding a 440 Hz tone. These spectral differences are what cause each instrument to produce a unique sound (Vennard 1968, 23). (BMP, 550 KB)

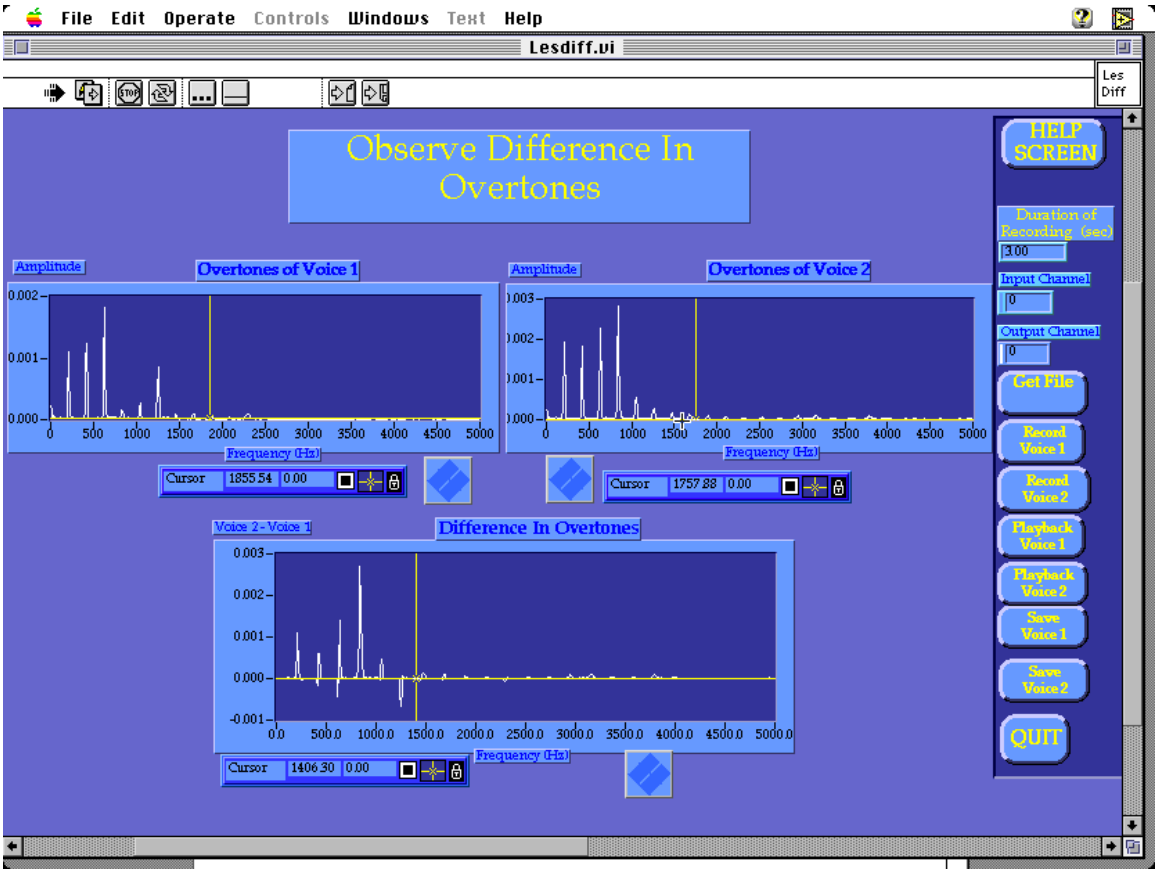


Fig. 4.7 Screen from the Music Muse that shows the spectral differences between two voices. These spectral differences are the direct cause of timbral differences. Voice 1 and Voice 2 are female and male, respectively, producing fundamental frequencies of about 300 Hz. The Difference In Overtones data is calculated as Voice 2 - Voice 1. (BMP, 509KB)

CHAPTER 5: THE VOICE

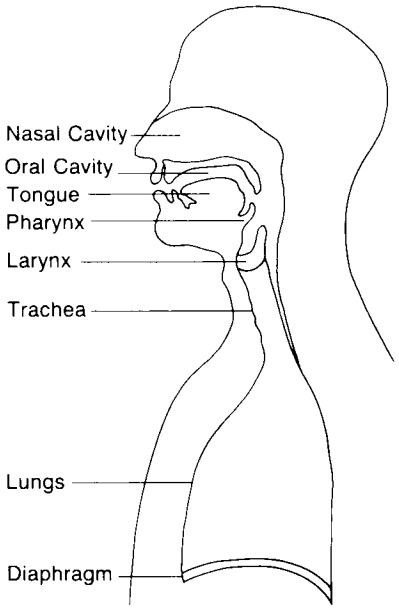


Fig 5.1 A diagram of the voice mechanism (O'Shaughnessy 1987, 40) (BMP, 117KB)

Speech signals are the means by which speakers communicate information to listeners. The resulting sounds are divided into two broad categories: "vowels, which allow unrestricted airflow in the vocal tract; and consonants, which restrict airflow at some point and are weaker than vowels."[76] Singing is generally viewed as the production of sustained vowel sounds. The different organs of the body that are used to produce sound make up what is called the voice organ.[77] A schematic of the voice organ is shown in fig 5.1.

The three major components of the voice organ are the *breathing system*, the *vocal cords*, and the *vocal tract*. Each of these components serves a purpose of its own. The breathing apparatus provides a source of air that passes through the vocal cords and vocal tract.[78] The vocal cords are muscles shaped as folds, covered by a mucous membrane.[79] The folds vibrate as the air from the lungs passes through, creating a sound wave that becomes an excitation signal for the

---

[76] O'Shaughness pp. 39
[77] Sundberg pp. 1
[78] ibid. pp. 9
[79] ibid. pp. 6

vocal tract.[80] The vocal tract is the combination of the pharynx and the mouth and includes the laryngeal chamber, the back of the throat, the mouth, nasal cavity, and to some extent the head cavities.[81] The vocal tract acts as a resonator and filter, amplifying certain excitation frequencies while attenuating others.[82] The acoustic signal that emerges from the vocal cords is the *source signal*. It is comprised of a fundamental frequency and a number of harmonics, which together create the *source spectrum*. As was mentioned in the signal processing chapter, the speech signal that emerges from the lips of the speaker is actually the convolution of the source signal from the vocal cords, and the impulse response characterizing the vocal tract.

The following sections will explain in more detail the physiology of voice production, different singing characteristics, and ways in which singers attempt to produce optimal sounds.

5.1 The Breathing Apparatus.

The purpose of the breathing apparatus is to compress the air in the lungs, so that an airstream is generated past the vocal cords and into the vocal tract. The lungs are located in the chest or thorax cavity.[83] They consist of a spongy structure, and assume the smallest volume possible within the thorax. They expand when air enters, but then push the air out to in attempts to maintain as small a volume as possible. They behave much like a balloon that is filled with air and then allowed to naturally release it. The lungs passively exert forces during respiration.[84]

Two groups of muscles control respiration: the intercostal muscles of the ribs, which are both active and passive contributors, and the abdominal wall and diaphragm, which are active contributors only. The intercostal muscles join the ribs. The inspiratory intercostals contract during inspiration to expand the rib cage. During expiration, they exert a passive force on the lungs in attempts to return to their smaller, relaxed position. The expiratory intercostals do just the opposite. They contract during expiration, and exert a passive force during inspiration. As a consequence, there is a particular lung volume at which the expiratory and inspiratory forces are equal. Whether or not the volume is slightly higher or lower decides which muscles contract or expand, and subsequently, which muscles exert a passive force to restore a relaxed position.[85]

The diaphragm is a large dome-shaped muscle attached to the bottom of the ribcage. It resembles an upside-down salad bowl, under which are the abdominal muscles. When the diaphragm is contracted, it flattens out, pressing the abdominals down, and expanding the volume in the lungs. This lowers the subglottic pressure, and causes air to flow into the lungs. It also causes the abdomen to expand outward, which allows the diaphragmatic activity to be monitored. The downward force on the abdomen is what makes it harder to sing with a full stomach. The

---

[80] O'Shaughnessy pp. 40
[81] Fowler and Willson
[82] O'Shaughnessy pp. 40
[83] ibid. pp.42
[84] Sundberg pp. 27
[85] ibid.

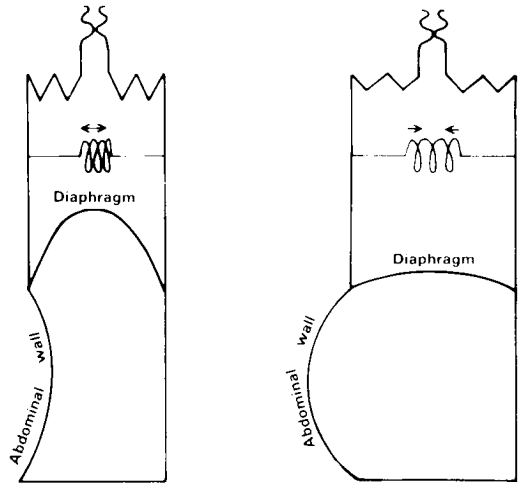


Fig 5.2 Schematic of the diaphragm and abdominals during respiration. The rib cage is modeled as a bellows with a spring to represent the behavior of the intercostal muscles (Sundberg 1987, 29). (BMP, 96KB)

diaphragm is pushed back to its dome shape by the contraction of the abdominal muscles.[86] Fig 5.2 illustrates this effect.

5.2 The Voice Box.

The larynx, or voice box, connects the lungs to the vocal tract through a passage called the trachea. The larynx is a framework of cartilages--the thyroid, cricoid, and arytenoid cartilages, and the epiglottis--joined by ligaments and membranes. The trachea splits into two bronchial tubes before entering the lungs. The epiglottis covers the larynx when food is intended to go down into the stomach.[87]

Inside the larynx are the vocal cords, or vocal folds. The stream of air from the lungs is pressed up through the trachea against the resistance of the vocal cords, which are a pair of

[86] ibid. pp. 28-29
[87] O'Shaughnessy pp. 43

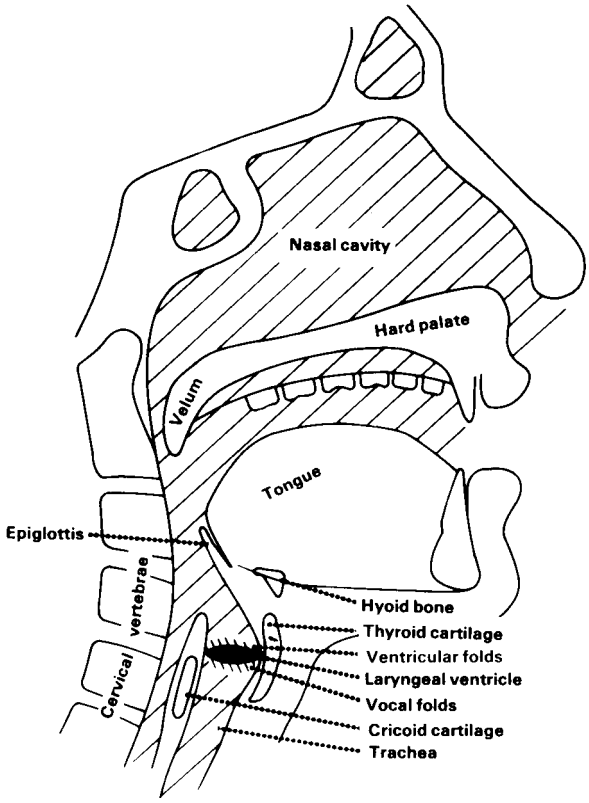


Fig 5.3 Diagram of voice box and vocal tract (Sundberg 1987, 7). (BMP, 195KB)

membranes that are stretched across the top of the trachea, separated by a slit known as the glottis, and act the part of the oscillator in the voice production process. These vocal cords are set into vibration by the air, producing puffs or air that create a sawtooth waveform with many harmonics, at a fundamental frequency referred to as the

phonation frequency.[88] It is this frequency that we hear as a sung tone. As with any fixed-fixed string system, the phonation frequency is determined by the length of the glottis, and the tension of the vocal cords.[89] The vocal cords run front to back in the throat, and are each connected at one

[88] Coppens et al. pp. 275; Fowler and Willson

[89] Sundberg pp. 16

end each connected at one end to the thyroid cartilage, and to two individual arytenoid cartilages at the other.[90] The tension and length in the vocal cords results from contractions in the cricoid muscles, which tilt the cricoid cartilage. Higher sung tones shorten the glottis and tighten the vocal cords.[91] Women generally have naturally shorter, lighter vocal cords than men, which is why they usually have higher voices.[92]

5.3 The Vocal Tract.

After leaving the glottis, the voice source passes through the vocal tract which reinforces the sound and shapes it acoustically. "The vocal tract can be modeled as an acoustic tube with resonances, called formants, and antiresonances."[93] Sound pressure waves that move through the vocal tract with frequencies at or near the formant are amplified, while frequencies at or near the antiresonances are attenuated. The formants depend on the length, size, and shape of the vocal tract, which vary from person to person.[94] This is why two sopranos singing the same note can sound completely different. The frequencies of the formants also determine the timbre of the produced sound, and therefore contribute to voice color.

The vocal tract length is defined as the distance from the glottis to the lip opening. The longer the vocal tract, the lower the first formant frequency. Adult males generally have longer vocal tract lengths than adult females, which is one reason why adult males have deeper voices than adult females.[95] The vocal tract length can be consciously modified by raising or lowering the larynx, or by protruding or otherwise shaping the lips (Sundberg 1987, 22).

The shape of the vocal tract is described by its cross-sectional area; however, this area varies along the length of the tract with the configuration of the articulators (Sundberg 1987, 21). The cross-sectional area therefore must be described by an area function. When the articulators change relative positions, the formant frequencies change as well. As a result, every combination of formant frequencies has its own distinct area function (Fowler and Willson 1995).

5.4 Vowels.

The articulators not only control formant locations, they also control vowel production. As a result, each vowel has its own characteristic set of formants, and its own characteristic area transfer function (Fowler and Willson 1995). The figures at the top of Fig 5.4 show average vocal tract area functions for various English vowel sounds. The figures at the bottom of Fig 5.4 show the corresponding vocal tract configurations.

The articulatory positions used to form vowels differ from person to person. These positions are functions of speech habits and accents, personal physical characteristics of the voice organ, use of the voice organ (which usually depends on training), and even emotions involved.

---

[90] ibid. pp. 6
[91] ibid. pp. 16
[92] Fowler and Willson
[93] O'Shaughnessy pp. 50
[94] Sundberg pp. 20
[95] Fowler and Willson

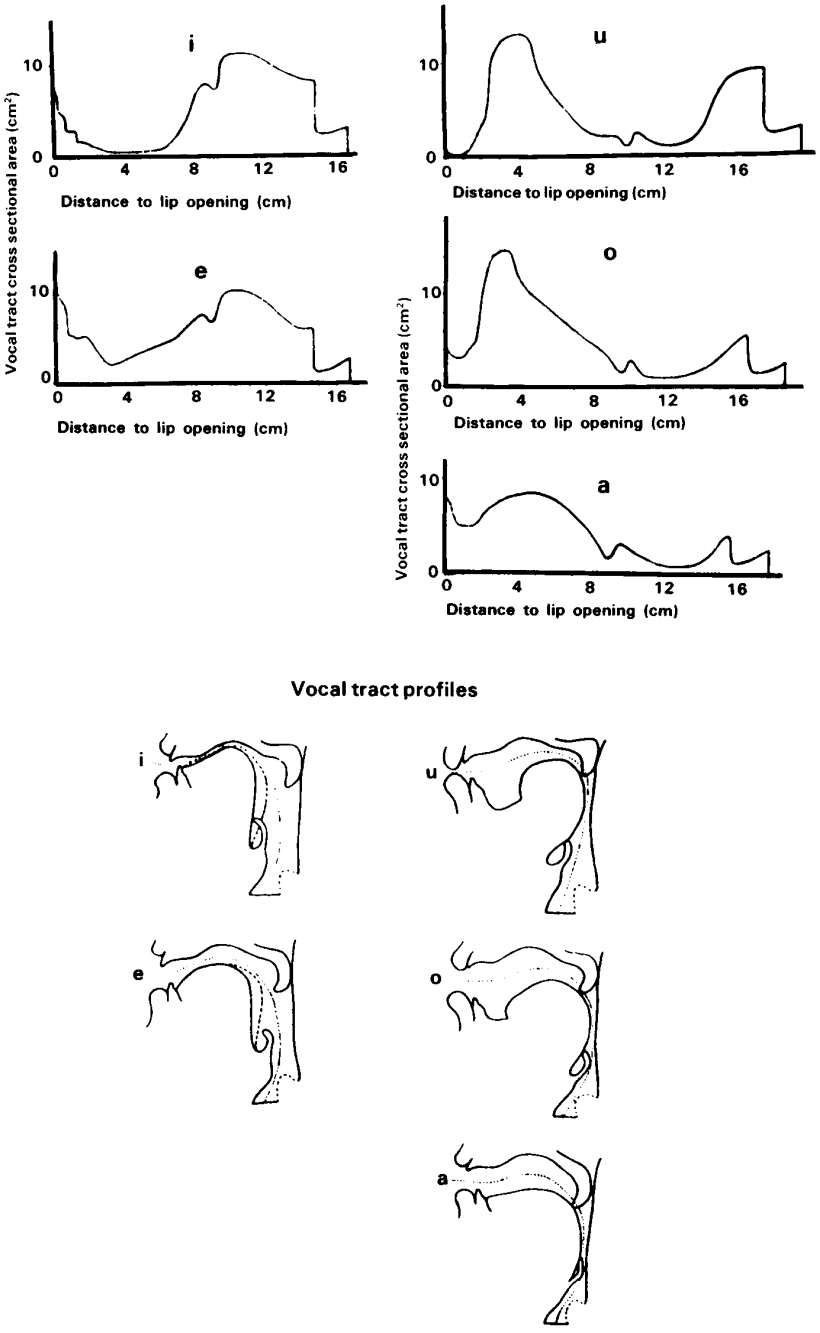


Fig 5.4 The upper graphs show the average area functions for various vowel sounds. These functions are determined by the positioning of the articulators, shown in the vocal tract profiles in the bottom figures (Sundberg 1987, 21). (BMP, 413KB)

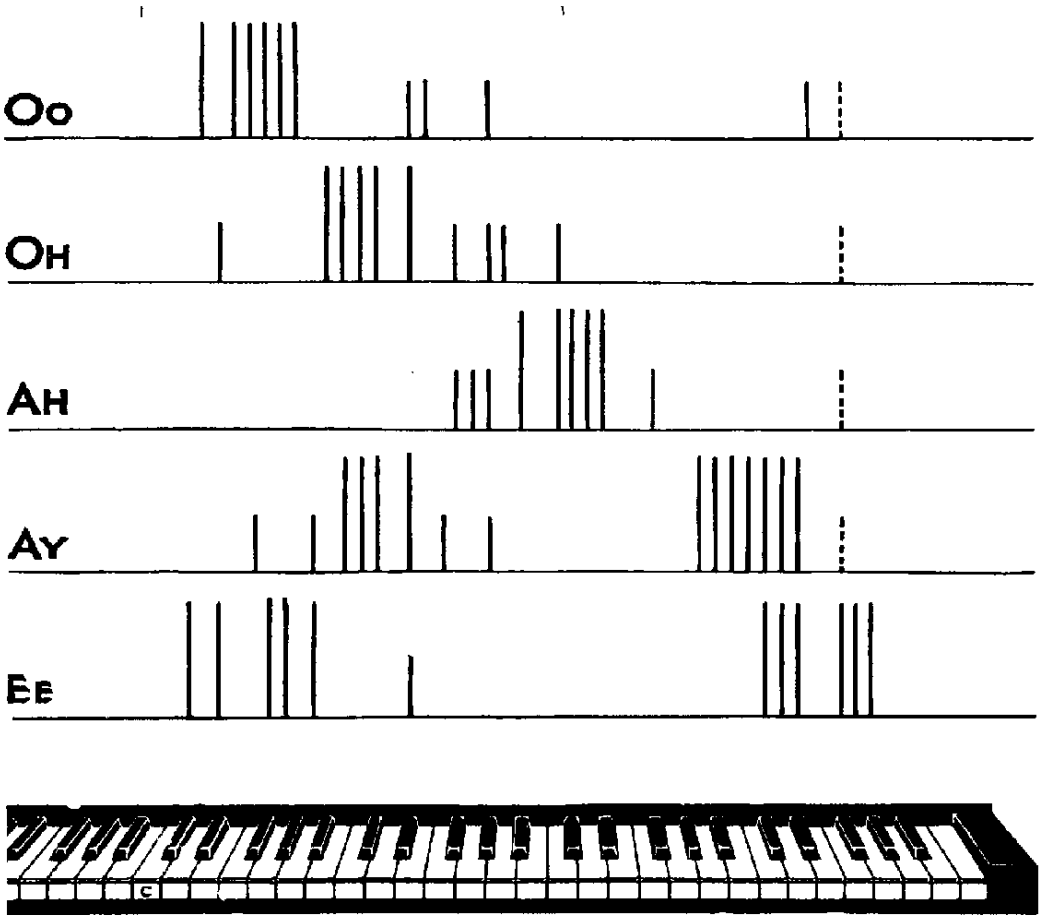


Fig 5.5 Formant distribution for various vowel sounds, plotted against the corresponding piano notes (Vennard 1968, 127).

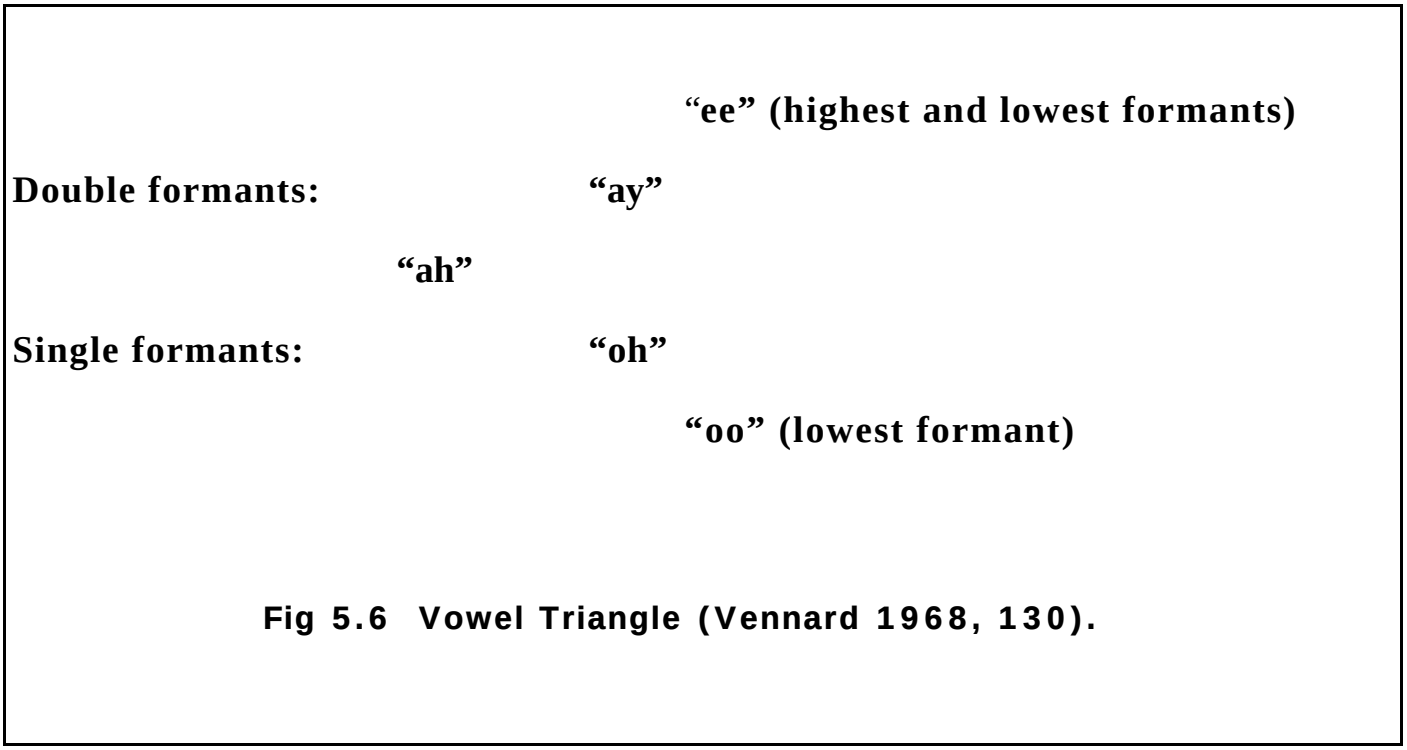


Fig 5.6 Vowel Triangle (Vennard 1968, 130).

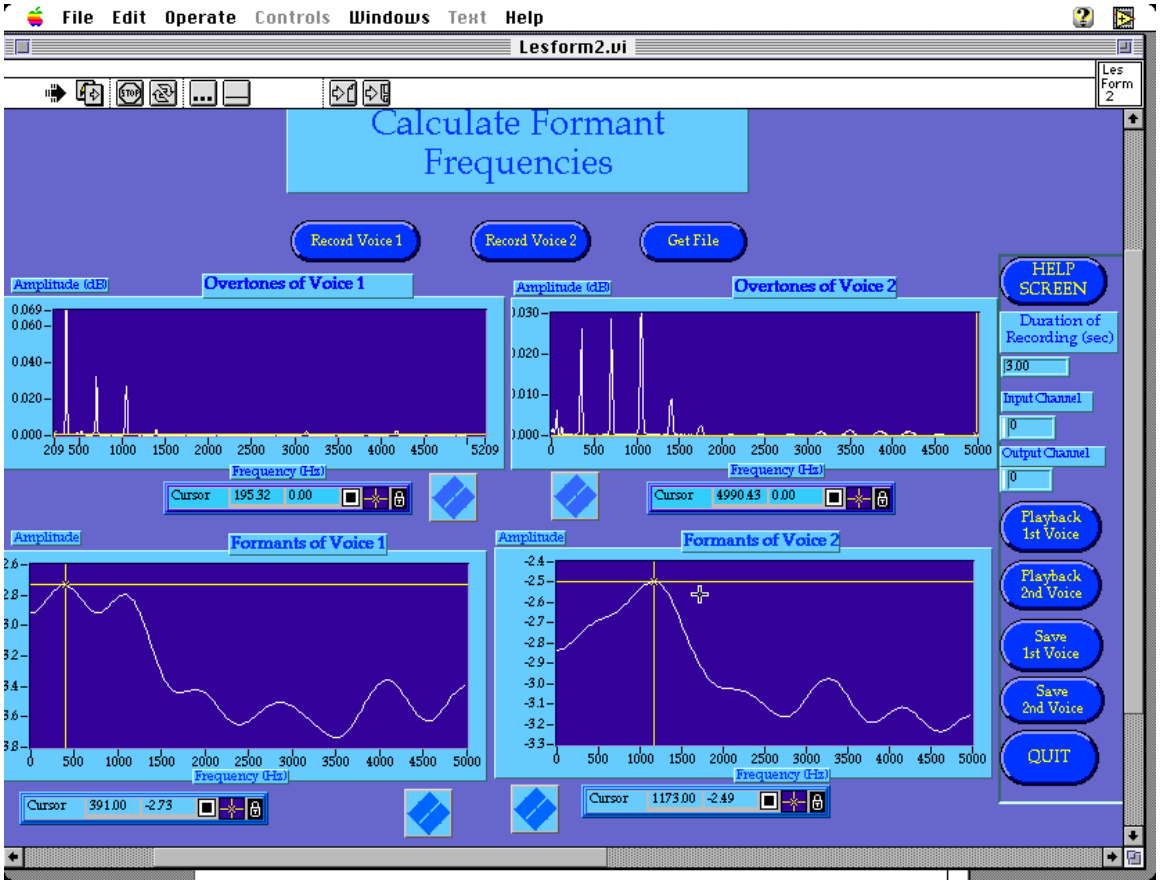


Fig 5.7 The overtones and formants of two vowel sounds sung at the same fundamental frequency. Even though the fundamental frequency remains constant, the spectral components change because of the change in articulatory positions. Voice 1 is "oo" and Voice 2 is "ah." (BMP, 77KB)

	Frequency ranges, Hz		
	Adult males	Adult females	Bandwidths, Hz
F_1	200–800	250–1000	40–70
F_2	600–2800	700–3300	50–90
F_3	1300–3400	1500–4000	60–180

Values are means and are given in hertz. F_1 and F_2 blend in females for some vowels; figures for these vowels are tabulated midway between F_1 and F_2 columns

	Adult males			Adult females		
Vowel	F_1	F_2	F_3	F_1	F_2	F_3
[i]	255	2330	3000	340	2610	3210
[ɪ]	350	1975	2560	425	2170	2900
[ɛ]	560	1875	2550	690	2015	2815
[æ]	735	1625	2465	950	1955	2900
[ɑ]	760	1065	2550		1085	2810
[ʌ]	640	1250	2610	750	1300	2610
[ɔ]	610	865	2540		785	2565
[ʊ]	475	1070	2410	515	1070	2280
[u]	290	940	2180	390	995	2585

Fig 5.8 Exact values of formants and their ranges for adult males and females singing various vowel sound (Thomas 1987, 104). (BMP, 118KB)

However, articulatory positions do have some characteristics that appear to be uniform among many people (Fowler and Willson 1995). Fig 5.5 shows the most commonly agreed upon distribution of formants for the five "pure vowels."

Fig 5.5 brings us to what is known as the vowel triangle, which is shown in Fig 5.6. From Fig 5.5 it can be seen that "oo" and "oh" each have one formant, or two clustered formants, because there is one position where the harmonics are the strongest. Similarly, "ay" and "ee" appear to have two well separated formants each. "Ah" is generally described as having the nature of both types of vowels. "Ah" could have one formant that descends in pitch with the series "ah," "oh," "oo." It also could have two clustered formants that separate in the series "ah," "ay," "oo." Also, note that "ay" is really "oh" plus a higher partial, while "ee" is really "oo" plus a high partial.

To test this theory, if you sing "ee" loudly and then put your hand over your mouth, the resulting sound will be "oo." The same procedure can be used to transform "ay" into "oh."[96]

Vowel sounds can actually be described as different timbres produced by the same instrument. By this definition, changing vowels sounds can be compared to changing the timbre of a trumpet by putting a mute inside the bell.[97] Fig 5.7 shows the formants of an adult female singing two different vowel sounds at the same fundamental frequency. Voice 1 is "oo" and Voice 2 is "ah." The changes in harmonic spectra and formant plots from "oo" to "ah" are quite noticeable. "Oo" appears to have lower formants than "ah." Also note that "oo" has fewer harmonics than "ah," which is why "oo" sounds more dark and hooty while "ah" produces a brighter sound. Fig 5.8 verifies the Music Muse calculations. It shows tabulated average vowel formants for males and females. It can be seen that the formant values that were calculated by the Music Muse lie within the tabulated ranges.

5.5 Placing Formants

There are certain known ways in which people can consciously adjust their formants to obtain a more desirable sound. A theoretical model of the vocal tract constructed by Lindblom and Sundbergh[98] shows a few ways in which placing formants can be accomplished. The model was based on measurements taken from lateral x-ray pictures of a subject pronouncing various long vowels. The articulators in the model included the jaw, tongue body, tongue tip, lips, and larynx. The most significant contributors were the tongue body contour, the jaw opening, and the configuration of the lips.[99]

The tongue body can assume many different shapes. Also, many different sounds are produced with similar tongue body contours. For this reason, its shape must be described as precisely as possible by the direction and quantity of bulging. In the model, the direction -1 means "toward the hard palate, as in the vowel /I/;" direction 0 means "toward the velum, as in the vowel /u/;" and the direction +1 means "toward the lower part of the posterior pharyngeal wall, as in the vowel /α/."[100] A tongue bulging quantity of 1 corresponds to the bulging during the regular pronunciation of the vowels, and a bulging quantity of 0 corresponds to no bulging at all.[101]

When the jaw opening is increased, the distance from the jaw to the cervical vertebrae is reduced. This means that the mouth is widened, and the pharynx is narrowed. The lip opening is mainly dependent on the jaw opening. However, the lip opening can also be rounded or spread as an independent action.[102]

The results of the formant placements of various articulatory configurations is shown in Fig 5.9. Fig 5.9a shows the influence of the jaw opening on the three lowest formants, where c and d

---

[96] Vennard pp. 130
[97] ibid.
[98] Sundberg pp. 96
[99] ibid. pp. 96-97
[100] ibid.
[101] ibid. pp. 97
[102] ibid. pp. 96-97

(a)

(b)

(c)

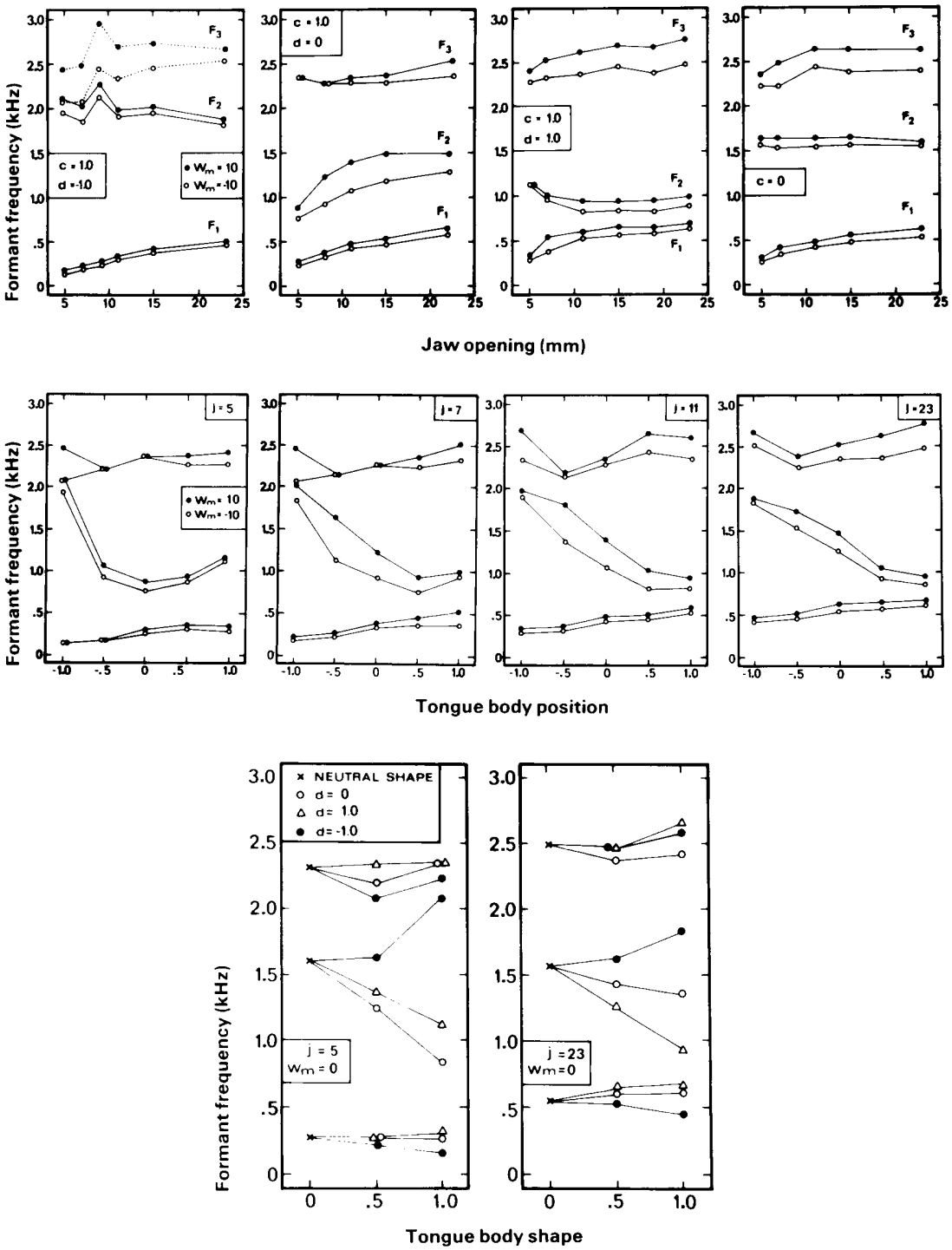


Fig 5.9 (BMP, 431KB)
(a) Influence of jaw opening on three lowest formants (Sundberg 1987, 100).
(b) Influence of tongue shape on three lowest formants for various jaw opening (Sundberg 1987, 100).
(c) Influence of tongue bulging on three lowest formants, for three tongue shapes and two jaw openings (Sundberg 1987, 101).

specify the quantity and direction of tongue bulging, respectively. Fig 5.9b shows the influence of tongue shape on formants for various degrees of jaw openings, symbolized by "j." The filled and open circles in both graphs correspond to spread and rounded lip openings, respectively. Fig 5.9c shows the influence of the degree of tongue bulging, symbolized by "d," on formants for two degrees of jaw openings and three tongue shapes. In Fig 5.9a, dropping the jaw appears to raise the first formant. Also, for a tongue bulging quantity of 1, an increase in the value of d raises all formants as the jaw is opened. For a value of c equal to zero, the formants do not appear to change with the amount of jaw opening. Both Figs 5.9a and b show that rounding the lips lowers all formants, as does lowering the larynx. This is because rounding the lips lengthens thevocal tract, and shaping it much like a long pipe, and long pipes have low resonances.[103] Fig 5.9c shows that for a given tongue body shape, the degree of tongue bulging has the most significant effect on the second formant.

5.6 Vibrato

Ordinarily, if muscles are put under a tolerable amount of stress, they can handle it smoothly for a period of time. After a while, however, fatigue sets in, and the muscles begin to shake. This happens during voice production. Small muscles are used, and the fluctuation in their energy can be heard in a form referred to as vibrato. It is a completely natural phenomenon. Its presence adds quality and richness to vocal tones, and its development is often a sign of vocal skill.[104]

Vibrato is the periodic, sinusoidal modulation of the amplitude and frequencies of a sung tone. It is characterized by its rate and its extent. The rate is the number of cycles that occur per second, or frequency, which is usually in the range of 5.5 to 7.5 Hz.[105] The fact that most spasmodic movements such as stammering occur in this same frequency range shows a relation between vibrato and enervation.[106] The extent of the vibrato is how much the frequency rises and falls during a cycle. The extent is usually in the range of $\pm$ 1 to 2 semitones, where a semitone is defined as a frequency difference of almost 6%.[107]

Fig 5.10 shows an example of one cycle of the vibrato of a baritone voice singing "ah" at middle C (262 Hz). The figure shoes the time and frequency spectrum sampled four times over a 0.6 second interval. Both the time and frequency domain plots appear to change from (a) to (c), while (d) looks identical to (a), as if it is the beginning of a new identical cycle.[108]

The modulation in frequency during vibrato causes the modulation in amplitude. The perceived amplitude of a tone sung with vibrato corresponds to the amplitude of the strongest overtone in the spectrum, which is that overtone located closest to the formant. When the frequency modulates, it

---

[103] ibid. pp. 100
[104] ibid. pp. 163
[105] ibid. pp. 163-164
[106] Vennard pp. 193
[107] Fowler and Willson
[108] Bartholomew pp. 23

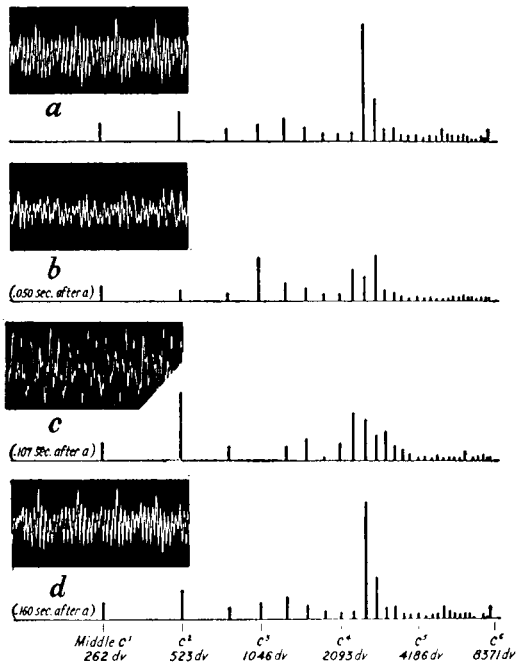


Fig 5.10 Samples of one cycle of vibrato taken at four chronological instants in time (Bartholomew 1942, 22).

moves the strongest overtone closer and farther from the formant, causing the amplitude to simultaneously rise and fall.[109]

The actual changes in frequency take place entirely too fast to be sensed by the ear, and are usually interpreted as variations in intensity. However, their effects can easily be heard by listening to a recorded voice at half speed on a record player. The changes in pitch become quite pronounced.[110]

When vibrato is too slow, or much less than 5.5 Hz, it tends to sound bad. If it is too fast, or much over 7.5 Hz, it tends to sound objectionably nervous. This faster rate of frequency

[109] Fowler and Willson
[110] Bartholomew pp. 23

modulation is called tremolo. If the extent of vibrato exceeds $\pm$ 1 or 2 semitones, it tends to sound wobbly, and becomes a trillo.[111]

Vibrato can be exaggerating by overtaxing the muscles involved. The more the muscles are taxed, the more they tremble. This is why opera singers generally have more rapid vibrati than concert singers. Opera singers are required to overtax their muscles by singing over large orchestras.[112] The emotional involvement of the singer also determines the amount of vibrato produced.[113]

Vibrato in musical sounds is very pleasing to our ears. It is imitated in many other instruments. Stringed interment players can contrive a vibrato by rocking their fingers back and forth on the string. Woodwind players imitate it by biting their reeds at the desired frequencies. These maneuvers are all used to make the instruments sound more "human" or "emotional."[114]

Fig 5.11 is a screen from the Music Muse that can be used to monitor vibrato. The amplitude-time plot shows the vibrato of an untrained female singer. As was previously stated, the rate is the number of undulations, or humps in the graph, per second. The extent is the size of the interval over which the fundamental frequency is spread in the overtone plot.

---

[111] Sundberg pp. 164
[112] Vennard pp. 195-196
[113] Sundberg pp. 164
[114] Vennard pp. 194

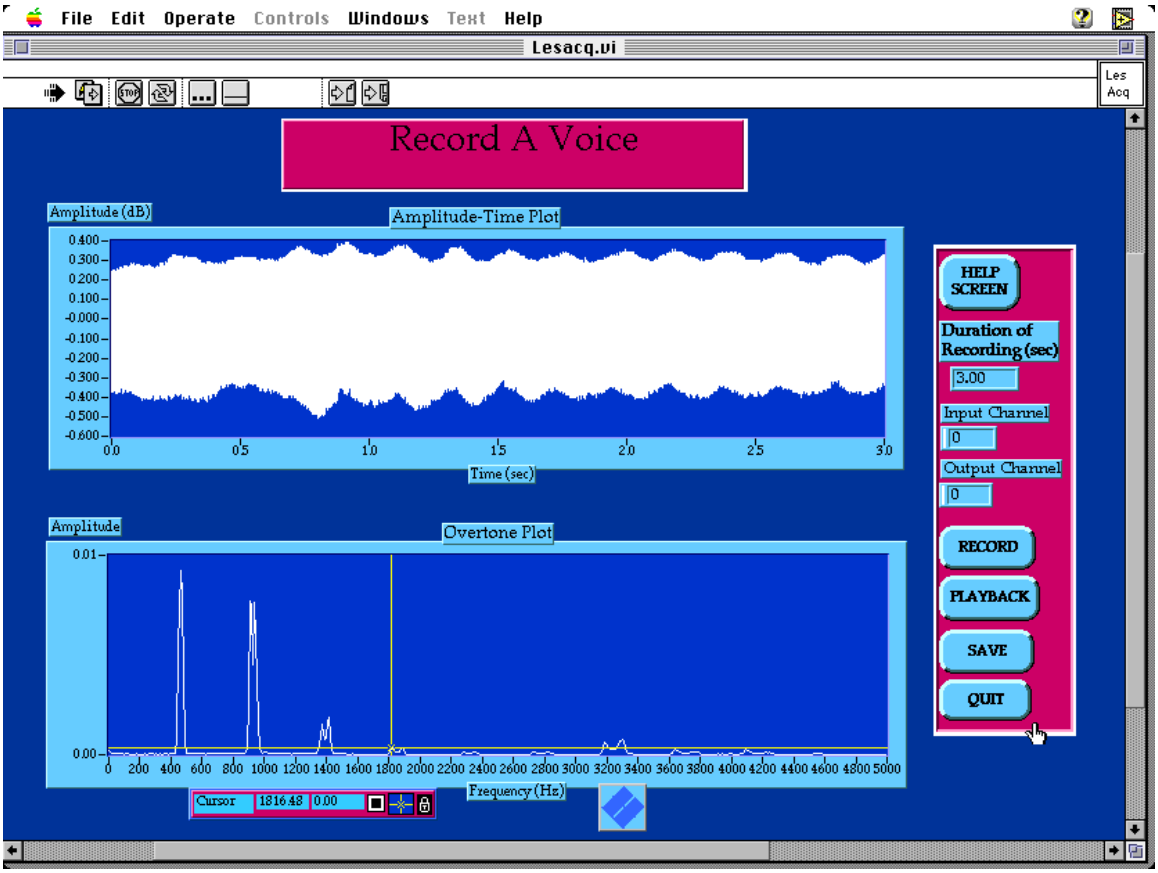


Fig 5.11 Music Muse screen that can be used to monitor vibrato. The rate and extent are observed to be the number of humps in the graph and their amplitudes, respectively. This data was taken from an untrained female singer. (BMP, 509KB)

5.7 Registers.

The most common definition of a singing register is a "phonation frequency range in which all tones are perceived as being produced in a similar way and which possess a similar voice timbre."[115] The transition from one register to another is characterized by a concrete change in timbre. Males usually have two distinguishable registers: the modal or normal register for lower tones, and the falsetto register for higher tones. The transition from one to the other is often accompanied by short breaks into the falsetto range.[116]

Females generally have three registers: the chest register, the middle register, and the head register. The transitions from chest to middle register and from middle to head register usually occur around 400 Hz (G4) and 600 Hz (E5) respectively. The transition frequencies for males is in the vicinity of 200 to 350 Hz (G3 - F4) for modal to falsetto. These frequencies, however, vary substantially from person to person.[117]
The aim in correct singing is to be able to eliminate timbral variations between registers. This can happen through training. The Music Muse can display timbral variations that occur, and can help singers to identify them and to correct them. The differences in registers will still occur on a laryngeal level, but will be imperceivable to the listener.[118] No matter how imperceivable they are to the listener, however, they can always be detected by the Music Muse.

---

[115] Sundberg pp. 49
[116] Sundberg pp. 50
[117] ibid. pp. 50-51
[118] Sundberg pp. 51

PART II: THE PROGRAM

Part I of this thesis explained the science behind the singing voice, and voice quality. It also presented the information that led to the creation of the Music Muse. Part II will now show how these concepts were brought together to create this innovative new program, and will also mention related software that is currently on the market. The final chapter will discuss the conclusion.

CHAPTER 6: THE DESIGN OF THE MUSIC MUSE

The Music Muse is a virtual instrument, or *vi,* that was created using a National Instruments programming system called Labview®. Unlike other programming systems, such as BASIC or C, which use text-based languages to create lines of code, Labview® uses a graphical programming language called G, which creates code in a block diagram form. Labview® programs are called virtual instruments, because they appear and behave just like real instruments. They have front panels which can include graphs, controls and indicators, as well as user interfaces such as knobs and push buttons. Labview® also has a library of functions and vi's that have already been created, that can be used for data aquisition, analysis, storage, and many other applications, and can be incorporated into new programs to make programming easier.[119]

When the program is run, the first screen that is seen is the one in Fig 6.1 It simply gives a brief summary of what the program does, its purpose, how the data is presented, and what the data means. At the bottom of this screen are three option buttons. The Quit button ends the program. The How Does It Work? button calls up a series of screens that give more information on what the program does, and the Main Menu button brings up the main menu of program options.

When the How Does It Work? button is pushed, a series of screens like those in Appendix A are displayed. These screens are for first time users of the program. They define overtones, harmonic spectra, and formants, and explain how each is important to voice timbre. All of this information is presented in language that is meant to be comprehended by people who have no technical background. It is also crucial that the non-technical audience view these screens, or else they will have a small chance of being able to understand the data.

The Music Muse is comprised of five main vi's, all of which have front panels that are accessed from the main menu. The main menu is shown in Fig 6.2. The Single Voice Options vi's collect and process data from a single voice. The Two Voice Comparison Options vi's are used to

---

[119] Labview® for Macintosh User Manual 1994, pp. 1-1 to 1-2

compare the characteristics of two voices. The front panels of the vi's are accessed by clicking on the corresponding buttons. The quit button brings the user back to the first screen.

When a vi button is clicked, before the front panel appears on the screen, an introduction screen comes up. Fig 6.3 shows the introduction screen that appears when the Continuously Record A Voice button is clicked. These screens are there to give the user a description of what the corresponding front panel does, and brief user instructions. The Continue button brings up the front panel. The Return button takes the user back to the main menu. The Example button is for new users. It brings up a series of help screens that go into greater detail on how the front panel is to be used, with more explicit instructions. It also shows a sample screen, and tells how to read the data. All of the vi screens are in Appendix A.

Careful consideration was taken in the design of the Music Muse to make as user friendly and non-intimidating as possible for non-technical users. Each front panel comes equipped with a Help Screen button, which calls up a list of user instructions. The Help Screen window from the Continuously Record A Voice And Compare vi is shown is Fig 6.4 As an added help measure, information on any front panel object can be obtained by pressing the apple key and clicking on the desired item. A pop-up menu will then appear, from which the Descriptions option must be chosen. A written description of the item will then appear. Finally, the Quit and Stop buttons on each of the front panels bring the user back to the previous screens, from which the Example buttons can be accessed for a review of how the front panel screens are used.

In order to run the Music Muse, the user must have either the PC or Macintosh versions of Labview®, a microphone, a speaker, and a data acquisition board. If the microphone or speaker being used are not powerful enough to drive the signal, an amplifier must be used. Also, depending on the frequency response of the microphone, a low pass filter set to a cut off frequency of 2 kHz for the Continuously Record vi's and 5 kHz for the rest of the vi's, should be used. The Music Muse was designed on a Macintosh Quadra 950. During its creation, a Realistic microphone, a Hi-Tex CP-18 speaker, and an NB-A2000 DAQ board were used.

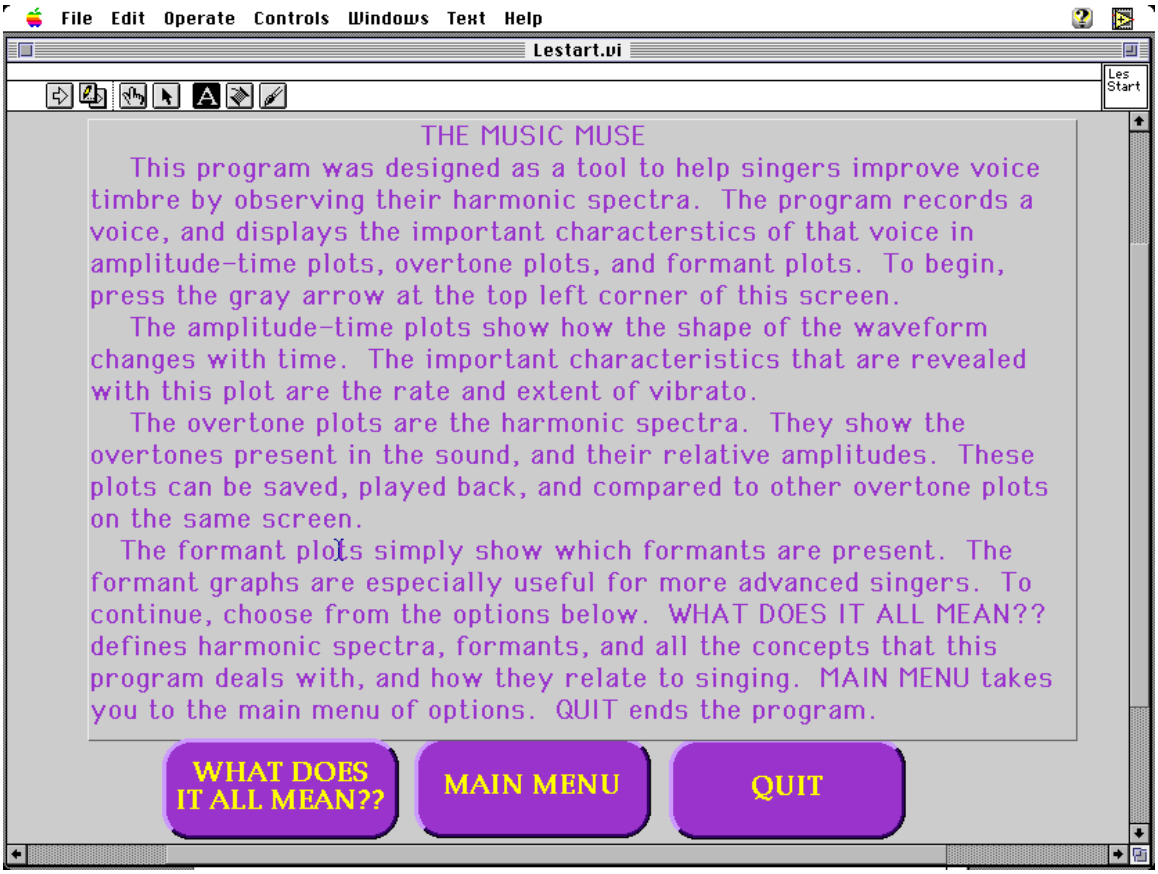


Fig 6.1 First screen of the Music Muse. (BMP, 509KB)

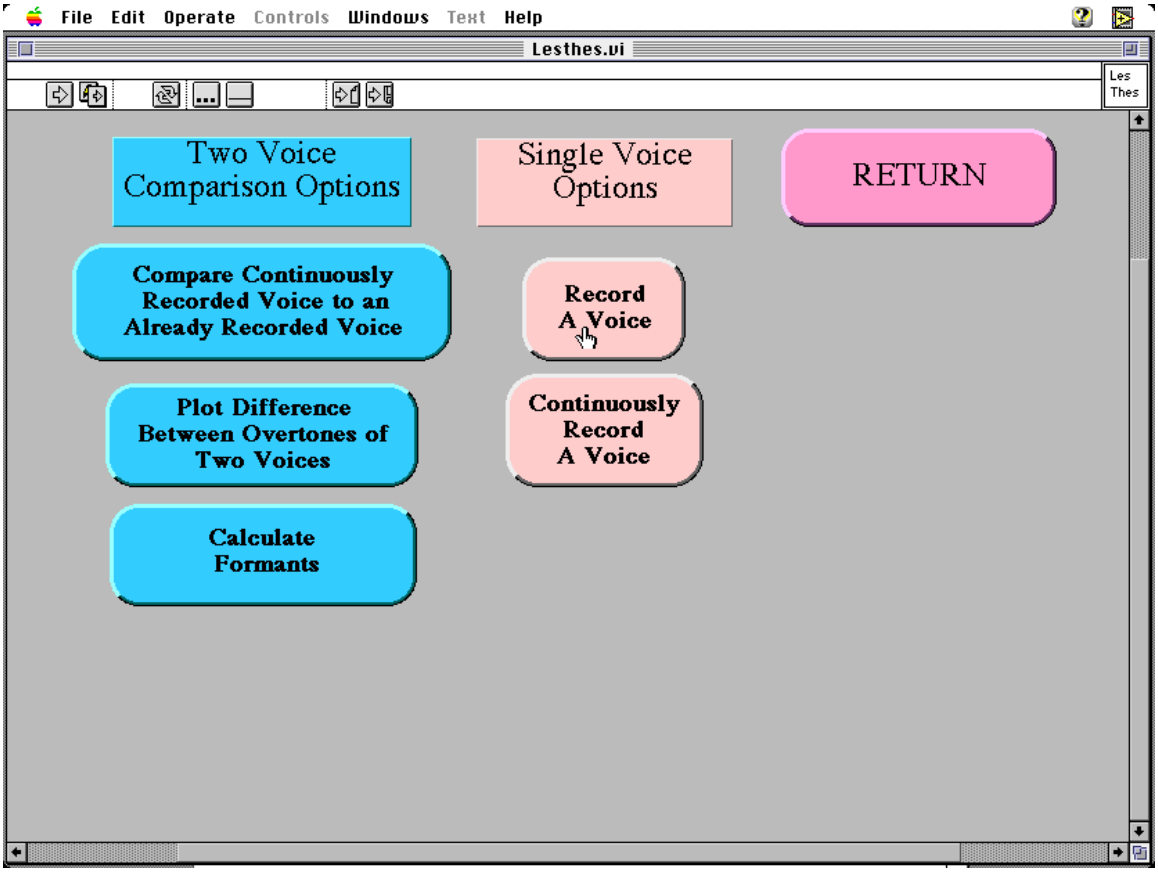


Fig 6.2 Music Muse main menu (BMP, 509KB)

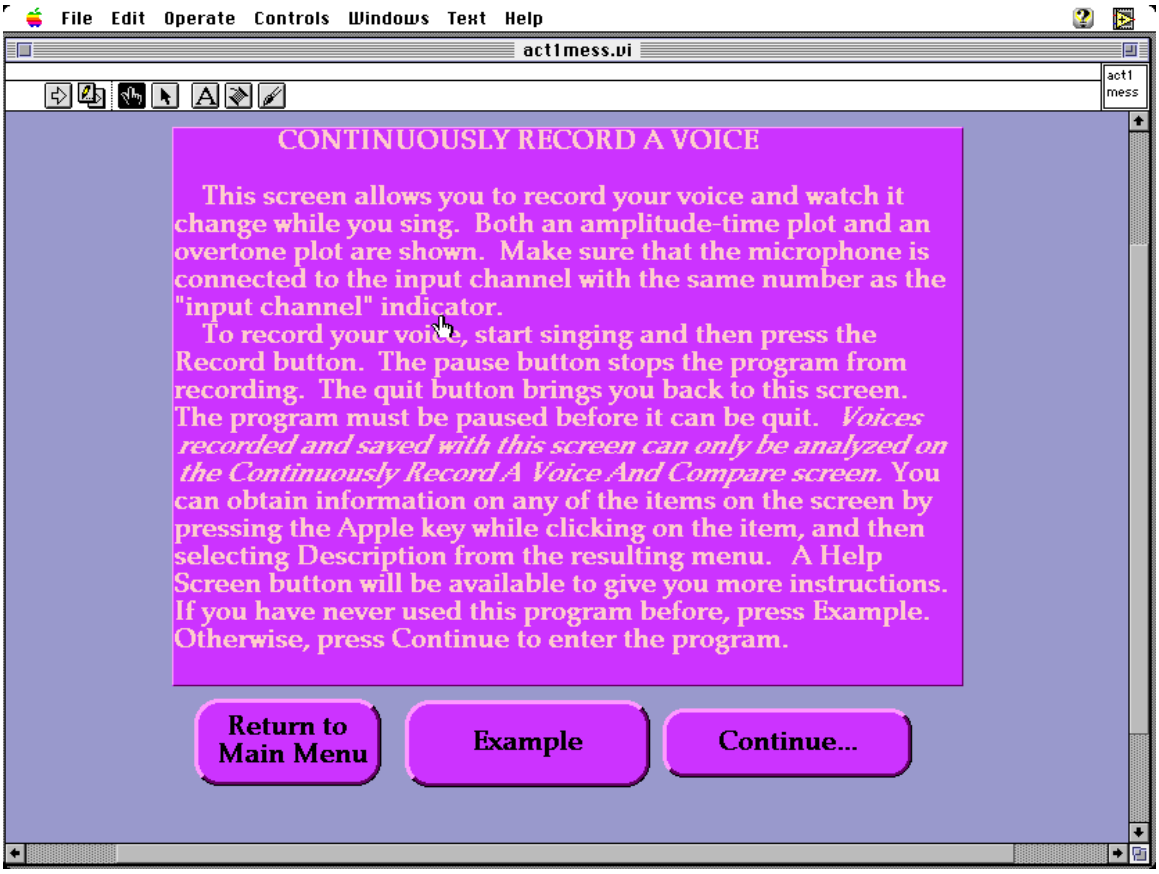


Fig 6.3 Introduction screen for Continuously Record A Voice (BMP, 509KB)

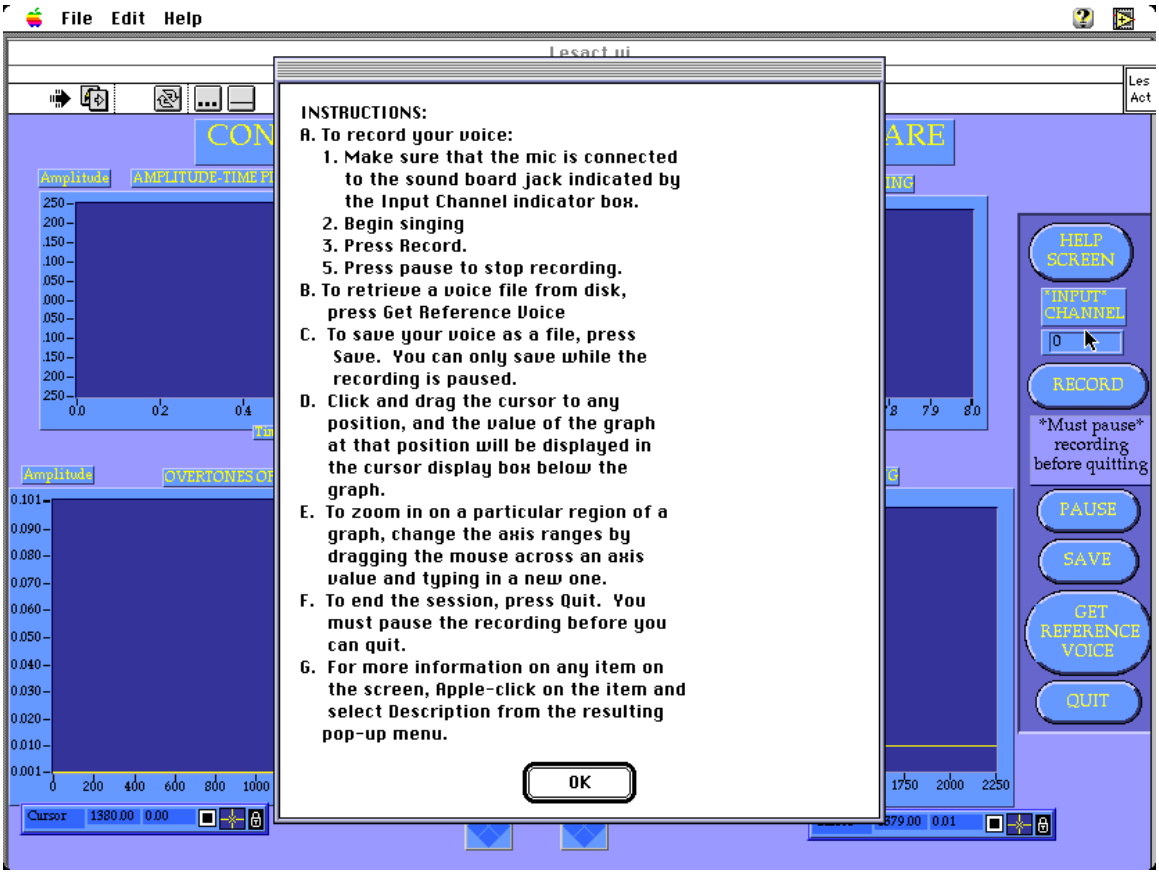


Fig 6.4 Help Screen window from the Continuously Record A Voice And Compare vi. (BMP, 509KB)

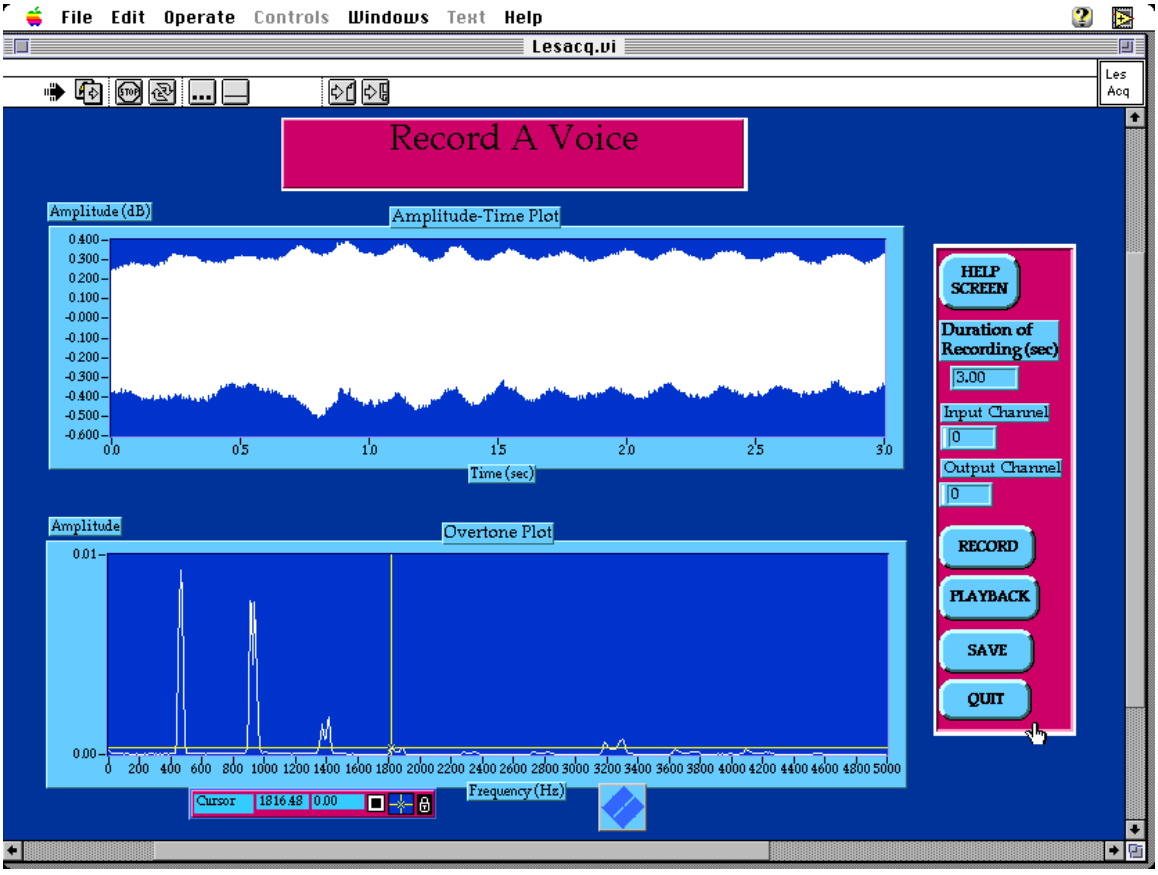


Fig 6.5 Record A Voice vi front panel. (BMP, 509KB)

6.1 Record A Voice

The Record A Voice screen is the simplest Music Muse screen, and is shown in Fig 6.5. It simply records a 3 second sample of a voice, and displays the signal in the time and frequency domains. The time domain plot is used to monitor the amount of vibrato in the signal. The vibrato in Fig 6.5 is from the voice of an untrained female. The harmonics are displayed in the Overtone plot. This overtone plot, as well as all of the other overtone plots in the Music Muse, actually shows the amplitude

spectrum of the data. Each overtone is spread over a small frequency range as a result of leakage, but primarily as a result of the frequency modulation caused by the vibrato.

The voice data that is acquired with this is sampled at 10000 samples/second. This sampling rate was chosen because its bandwidth is high enough to capture most of the relevant frequency data that can be found in the harmonic spectra of voices, and it is low enough to cut down on processing time. The figure in Appendix B shows the comparison plots of data collected and sampled at 10 kHz, and data sampled at 20 kHz. In both cases, a female is singing a fundamental around 900 Hz, which is in the upper vocal range. It can be seen that even for a fundamental in this range, the amplitudes of the harmonics are insignificant beyond the 5000 Hz bandwidth. Therefore, it can be assumed that most human voices will not produce significant harmonic information above this frequency range.

Once the data is collected, it is divided into blocks of 1024 samples. An FFT of each block of data is calculated, and the results are averaged together and plotted on the Overtones plot on the screen. The calculation of the FFT is based on the assumption that the data is ergotic. Therefore, this vi is used to analyze one vowel sound at a time. All of the vi's except for the two Continuously Record screens acquire data the same way.

Before recording a voice on this screen or any other, the microphone and speaker must be connected to the correct jacks on the data acquisition board. These jacks are indicated by the Input Channel and Output Channel indicators on the screens, respectively. Then user must start singing first, and then press the Record button. By singing first and then acquiring the data, the transients in the voice that are associated with the vocal attack are neglected, and the signal is that much closer to being ergotic. Once the voice signal has been acquired, it can be played back through the speakers. When the user presses Playback on this screen or any other, the vi simply converts the time signal back to analog, and plays it back through the speakers.

The yellow perpendicular lines in the Overtone graph of Fig 6.5 indicate the cursor position. The value of the cursor position is indicated in the cursor display box below the graph. The cursor can be dragged to

different positions so that the values of the harmonics can be observed. For a finer resolution, the ranges of the graphs can be manually altered. The user must simply drag the mouse across the axis value to be changed, and type in a new one. These two features are available to all of the Music Muse graphs.

To save the voice data, the user must press save. A dialogue box will appear with instruction. These data files can be retrieved later and analyzed on other screens. The Help Screen button can be pressed for a simple list of user instructions. This button will appear on all of the screens. The black arrow at the top left corner of the screen indicates that the vi is running. If for any reason this button changes from black to gray, the user must click on the arrow to start the vi back up. These instruction all appear in the help and introduction screens of the vi, which are shown in Appendix A.

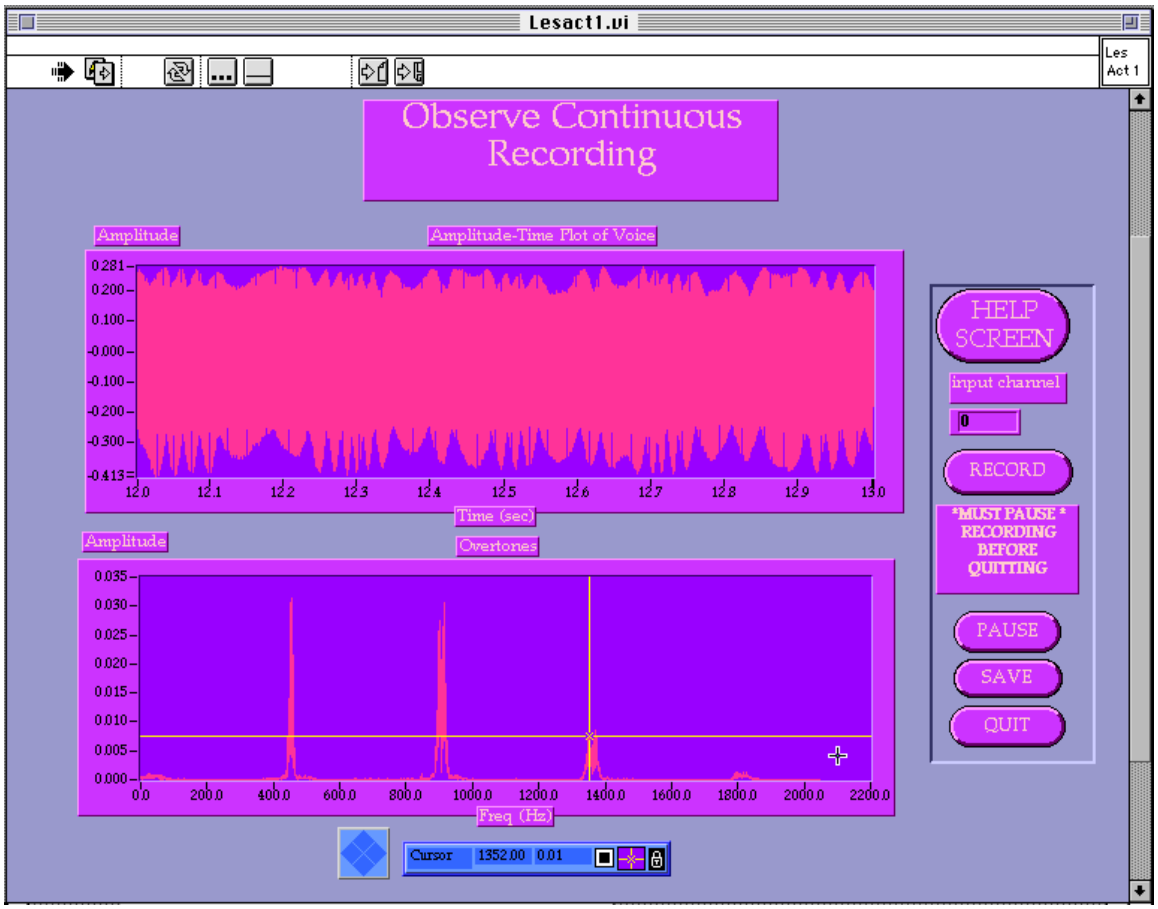


Fig 6.6 Continuously Record A Voice vi front panel. (BMP, 454KB)

6.2 Continuously Record A Voice

The Continuously Record A Voice vi works the same way as the Record A Voice vi, except it continuously acquires data and updates the graphs. Fig 6.6 Shows the front panel of this vi. The Amplitude-Time graph updates itself by adding on new data points at the right hand side, so that the graph appears to move to the left with as time progresses. The Overtone graph updates itself by changing the amplitudes and frequencies of the harmonics as needed.

Because the vi is constantly reading in new data, processing it, and updating the graphs on the front panel, the data is collected with a

sampling rate of only 4000 samples/sec, and the graphs plot 4000 data points at a time. The low sampling rate only provides the vi with a bandwidth of 2000 Hz. Consequently, only a few harmonics can be observed. Also, higher harmonics produced by first soprano singers will be neglected. Also as a consequence of the low sampling rate, these voice signals cannot be played back because the lack of frequency information would distort the sound. This vi calculates the FFT of the entire block of data at once, instead of dividing it into blocks of 1024, as another measure to save computing time. The objective of this vi was to produce the highest quality of results as close to real-time as possible.

The usage instructions are similar to the instructions for the Record A Voice vi. First, the microphone must be connected to the jack indicated by the Input Channel indicator box. Next, the user must first begin singing, and then press Record. To stop the data acquisition process, the user must press Pause. When the acquisition is paused, the last 4000 data points will remain on the graphs. This data can be saved via the Save button. The data saved on this screen can only be analyzed on the Continuously Record A Voice And Compare screen. To exit the vi, the user clicks the Quit button. The acquisition must be paused before the vi can be quit. A cursor and cursor display box are provided so the user can observe the values of the harmonics.

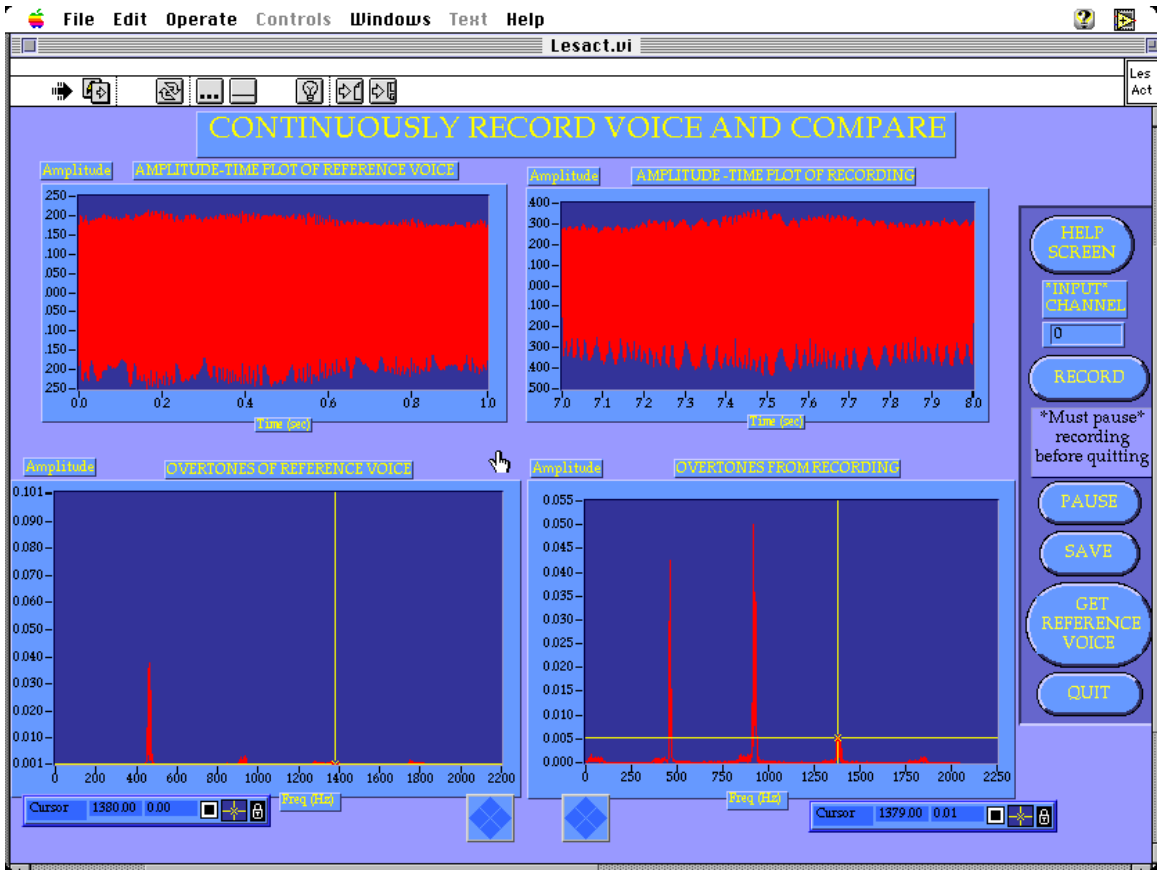


Fig 6.7 Continuously Record A Voice And Compare vi front panel. (BMP, 509KB)

6.3 Continuously Record A Voice And Compare

This screen works exactly like the Continuously Record A Voice And Compare vi, except it allows the user to retrieve a voice file for comparison by clicking the Get Reference Voice button. Its front panel is shown in Fig 6.7. The comparison voice must have been acquired with this vi, or with the Continuously Record A Voice vi. With this vi, the user can not only compare his harmonics with someone else's, but also watch his voice change while trying to emulate the reference voice. The screen if Fig. 6.7 compares an "oo" on the left to an "ah" on the right, sung by the same singer at the same fundamental frequency.

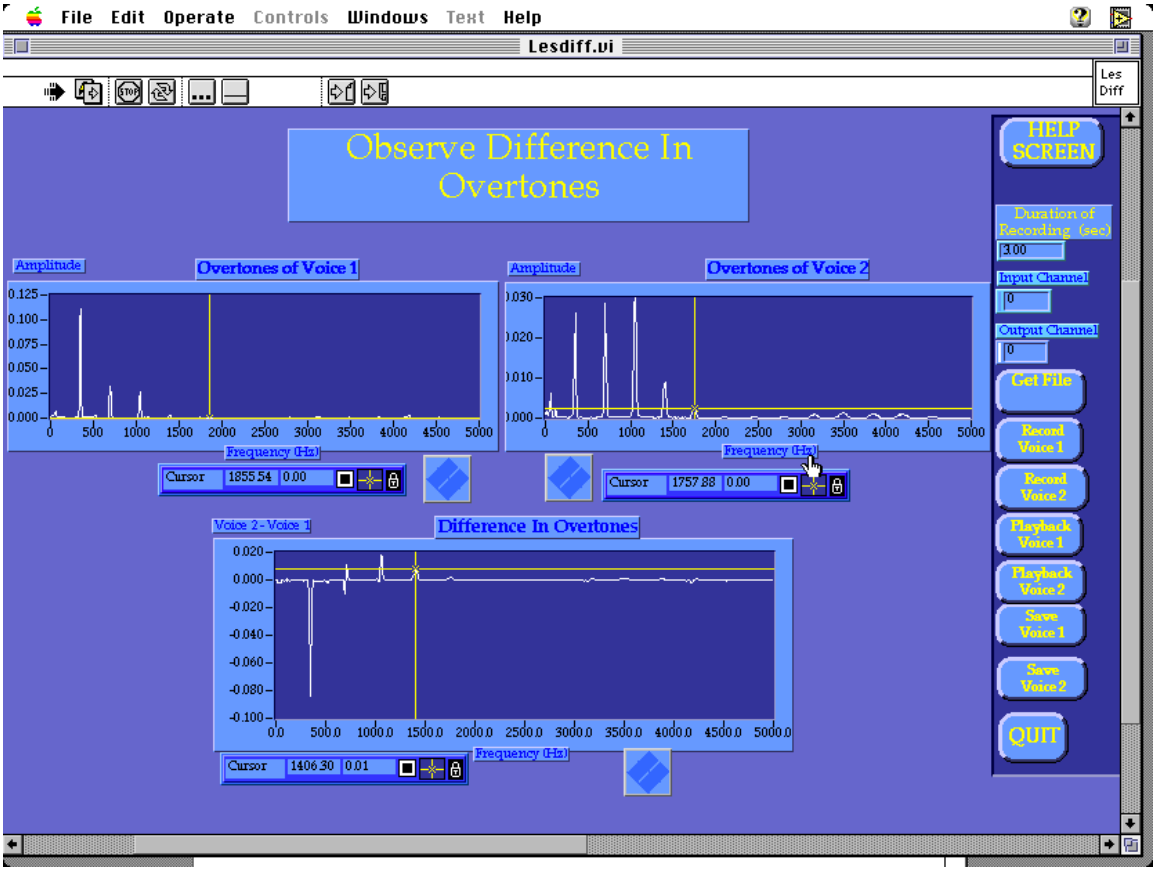


Fig 6.8 Observe Difference In Overtones vi Front Panel (BMP, 509KB)

6.4 Observe Difference In Overtones

With this vi, the user can compare the differences between the overtones of two voices by calculation. The front panel is shown in Fig 6.8. The screen calculates the amplitude spectra of two voices, and then calculates the difference between the two as Voice 2 - Voice 1. These differences are then plotted on the graph at the bottom of the screen.

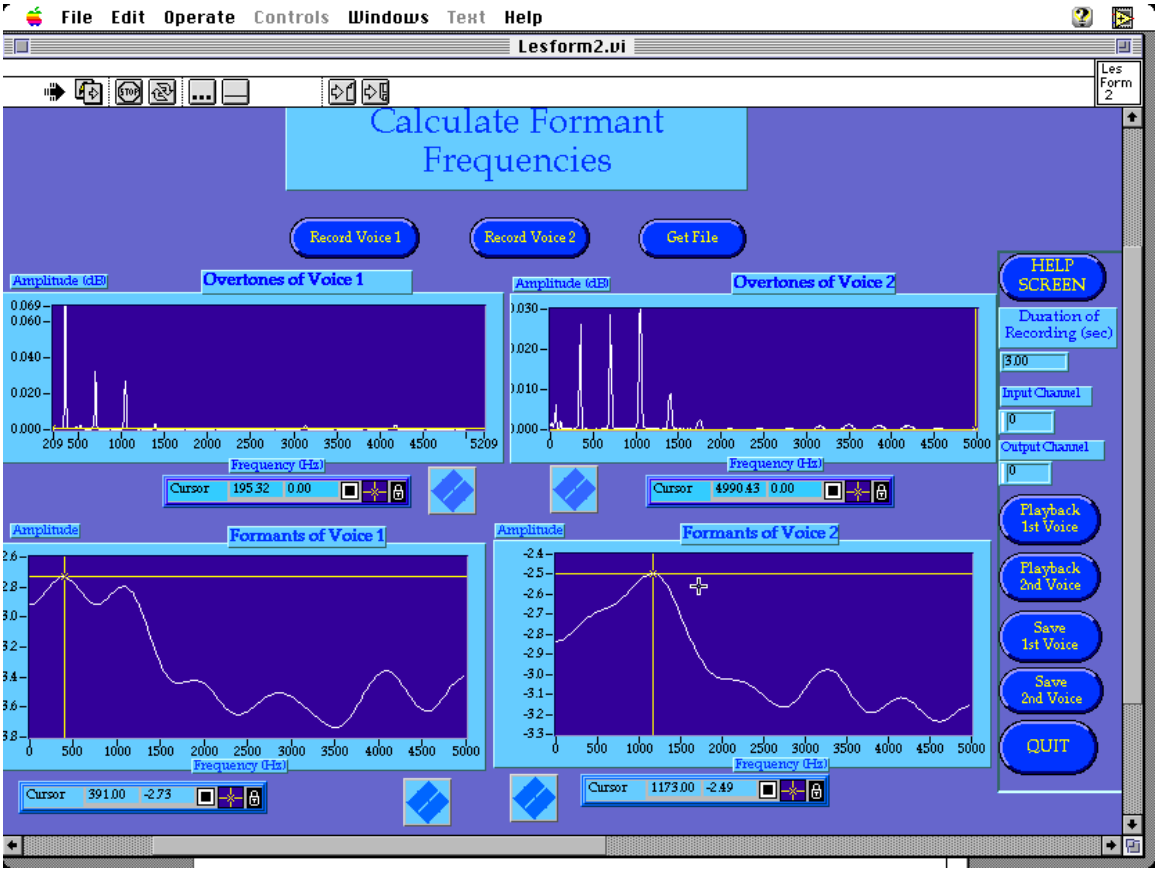


Fig 6.9 Calculate Formant Frequencies vi front panel. (PCX, 77KB)

With this screen, the user can see immediately the differences in the two spectra. The user can also re-acquire data and watch the differences increase or decrease. The voice data that is analyzed on this screen can be recorded on any screen except for either of the Continuously Record screen. To record a voice with this screen, the user begins singing and then clicks on the appropriate Record button. To retrieve a voice that has already been saved as a file, the user must click the Get File button and then follow the directions on the resulting dialogue box. The user can also save a voice as a file by clicking on Save and following the directions.

6.5 Calculate Formants

The formant vi calculates the FFT of a signal and displays the amplitude spectrum of the data, and then calculates the formant

frequencies and plots them under the corresponding amplitude spectrum. Its front panel is shown in Fig 6.9. The formant frequencies are the source of voice timbre because they dictate which harmonics have the highest amplitudes and which have the lowest. The formants frequencies are the peaks in the formant plots, and are located at the frequencies where the amplitudes are the highest in the corresponding overtone plots. This vi analyzes data recorded using every Music Muse vi except for the two Continuously Record vi's. For the formants calculations, the data is divided into blocks of 256 data points, corresponding to 25.6 ms intervals. The formants for each block of data are then calculated by cepstral analysis, as was explained in Chapter 3. The formants from each data block are then averaged together and plotted.

To prove that the formants calculated by the Music Muse are correct, a signal with known formant frequencies was generated and entered into the Calculate Formants vi. The calculation results were then plotted on a different screen, which is shown in Fig 6.10, where the following signal was generated and entered into the program:

$$y = \sin(Fo) + \sin(2Fo) + \sin(3Fo) + 8\sin(4Fo) + \sin(5Fo) + \sin(6Fo) + \sin(7Fo) + 8\sin(8Fo) + \sin(9Fo) + \sin(10Fo)$$

where F_o is was the fundamental frequency, equal to 256 Hz. The harmonics at 1024 Hz and 2048 Hz (4Fo and 8Fo) are known to be amplified, so theoretically the formant plots should show peaks at the two amplified frequencies. The Music Muse calculated results do show obvious peaks at 1024 Hz and 2048 Hz as was predicted, thus the calculations correctly suggest the existence of formants at these frequencies.

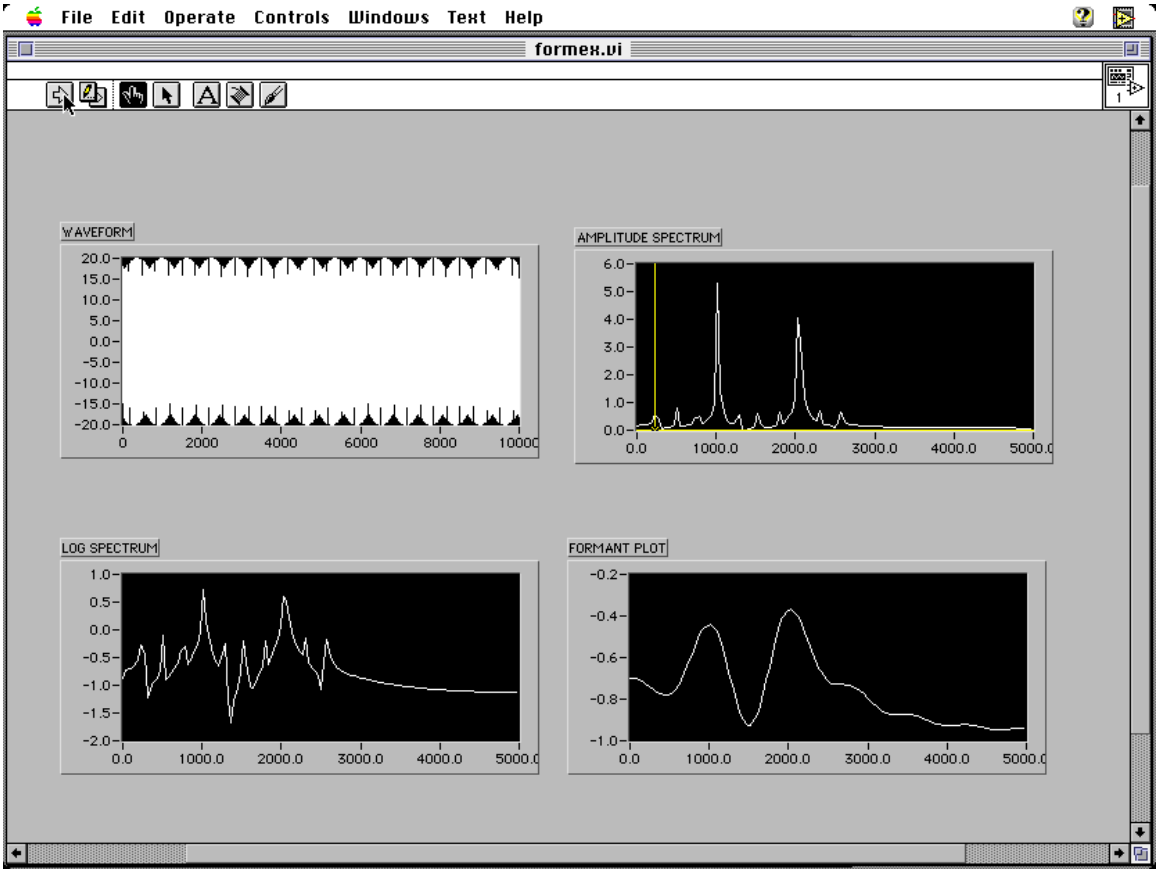


Fig 6.10 Music Muse calculations of formants of a generated signal with known formant values at 1024 Hz and 4096 Hz. (BMP, 509KB)

6.6 Areas For Future Improvement

The Music Muse does have some shortcomings that must be overcome before it can be marketed. For starters, the program is slow to process the recorded information. The Music Muse requires the execution of a plethora of calculations and commands, which take more than a few

seconds to complete. The problem can be rectified by the use of a more recent version of Labview®. Another solution to this problem might be to recreate the Music Muse using a different source of code.

The Continuously Record vi's have several problems that are caused by Labview® limitations. The vi's collect data at a rate of 4096 samples/sec, and plots 4096 data points at a time. However, the program takes longer than 1 second to render all 4096 points onto the graph, especially the data that must first be converted to frequency information via the Fourier transform. As a result, the data is not plotted in real time. What the user sees on the graph was actually acquired a few seconds earlier. The data that has been acquired but not plotted collects in a buffer that can hold up to 80,000 data points. The problem is, after a while the buffer fills up and an error message appears and stops the acquisition process.

The problem with the backlog of data in the buffer leads to another problem. Once the data leaves the buffer, it goes directly to the amplitude-time graph to be plotted, and then it also goes through an FFT block to be processed and then plotted on the overtone plot. The FFT process takes a little extra time. As a result, the data in the time graph is not plotted at the exact instant as the FFT data, so the graphs do not completely correspond. In other words, what the user sees on the overtone plot does not always correspond to what he sees on the time plot. These problem must be rectified either by Labview® with the production of updated versions, or with the implementation of another source of code.

As a measure of improvement, the Music Muse should divide all frequency information into octave bands that correspond to the notes on the piano. This would make the program a little simpler for the non-technical musician to understand. The axes could then be labeled with letters instead of frequencies.

CHAPTER 7: RELATED SOFTWARE

Although the Music Muse is an innovative tool, it is not the only one. There are many software packages available that are similar to the Music Muse. One company in particular, Kay Elemetrics Inc, is a manufacturing company that specializes in software for speech pathology, and manufactures a few similar state-of-the-art versions of the Music Muse. Also, similar software is now being used in the automobile industry. This chapter will present these and other related software packages currently on the market.

7.1 Pro-Audio Analyzer

Intelligent Devices is a manufacturing company that produces software packages and plug-in computer accessories for computer music and studio-type editing. One such accessory, the Pro-Audio Analyzer, is an acoustical analysis tool that displays the amplitude-time waveform of a music sample, as well as the FFT, all in real time. This software is similar to the Music Muse Continuously Record vi. The Pro-Audio-Analyzer displays the FFT of a signal on a bar graph with 1/3 octave bands. Floating above the bars are peak indicators, that show time averaged peaks for the individual center frequencies. The peaks are averaged over a manually set interval of time.[120] The Waveform Monitor and Spectrum Analyzer screens from the Pro-Audio Analyzer are shown in Appendix C.1.

7.2 Digital Performer 1.7 with PureDSP

Mark of the Unicorn, Inc, is another corporation that manufactures computer products for musicians. One such product is the Digital Performer 1.7 with PureDSP. It can be used to analyze the voice, like the Music Muse, but its intended purpose is editing musical samples. The Digital Performer is a Macintosh compatible, MIDI Sequencer and Audio Recorder. However, additional hardware is required for recording. Compatible hardware is manufactured by a company called Digidesign. The Digital Performer is essentially a home studio for editing musical samples. It uses spectral analysis not only to display spectral effects,

[120] Intelligent Devices

like the Music Muse, but to change spectral effects. With the Digital Performer, musicians can pitch shift, time compress and stretch, and gender bend.[121]

The pitch shifting feature allows musicians to transpose voices or modulate whole samples into higher or lower keys, without losing sound quality. In other words, voices can be raised an octave without sounding like chipmunks, and can be lowered an octave without sounding like Darth Vader in the end. For example, the melody of a sample can be transpose up a third and down an octave to create harmonies of the same voice. Also, if a particular voice is flat, it can be transposed into the correct key. Also, an entire sample can be modulated into another key all at the same time.[122]

The time stretch and compress feature uses spectral analysis to change the tempos of samples without changing pitch. For example, a 5 second sample can be stretched out to 7 seconds without the fundamental frequencies or formants being lowered, or it can be compressed into 3 seconds without the fundamental frequencies or formants being raised.[123]

Finally, the gender bending feature uses spectral analysis to change timbral characteristics. Male voices can be transformed spectrally into their female counterparts, and vice versa, by manually changing the formants frequencies. For example, a tenor could be transformed into a bass, an alto and a soprano, and a whole chorus of one voice can be created. Also, since the timbral changes are manual, voices can be played with to create an infinite number of new timbres.[124]

---

[121] Mark of the Unicorn
[122] ibid.
[123] ibid.
[124] ibid.

7.3 Kay Elemetrics Corp.

As was mentioned before, Kay Elemetrics is a manufacturing company for speech pathology software. They specialize in acoustical analysis and imaging instrumentation. One such product is the Computerized Speech Lab, CSL™. CSL™ is a completely integrated system that includes both hardware and software, and was designed to work in conjunction with a host PC personal computer.[125]

One of its Music Muse related options is the Sona-Match, Model 4327, shown in Appendix C.2. The Sona-Match records the voice, and displays a duel-screen of vowel formant frequencies in real time. The two screens show the same data, but with different frequency ranges that can be manually altered. The feedback is superimposed on top of a graph showing a targeted vowel formant graph, for comparison. Unlike the Music Muse, which calculates formants via cepstral analysis, the Sona-Match uses linear predictive coding to calculate formants.

Another CSL™ option is the Real-Time Spectrogram, Model 4329. This option graphs the FFT of a signal in real-time in the form of a spectrogram. A sample screen is shown in Appendix C.3. The spectrogram plots frequency vs time, as opposed to the way the Music Muse plots frequency vs amplitude. In addition, the plot has a duel-screen, so that a target spectrogram can be plotted for comparison. Also, unlike the Music Muse, the Real-Time Spectrogram data can be scrolled back to view previous data.[126]

An additional CSL™ feature is the Multi-Dimensional Voice Program, Model 4305. This option calculates over 22 voice quality parameters, including average fundamental frequency, fundamental frequency variation, and noise-to-harmonic ration. These parameters are displayed graphically, as shown in Appendix C.4, against a set of normal parameter values. These vocal quality parameters indicate far more than tonal

[125] Kay Elemetrics
[126] ibid.

beauty. This program, to a much greater degree than the Music Muse, is intended to help diagnose vocal abnormalities.[127]

7.4 NVH Applications

Software similar to the Music Muse has now found its way into the automobile industry by way of a new genre of acoustics called sound quality, or NVH, for noise, vibration, and harshness. For a long time, car makers have been designing cars with minimal noise. Until recently, however, they only concerned themselves with noise levels. Now, cars are being designed to have good noises, as opposed to bad noises. For instance, the low rumble of twin exhausts is a good noise for sports cars, but valve clatter is definitely bad.[128]

Laboratories such as Structural Dynamics Research Corporation (SDRC) have been investigating which noises are considered bad, and which are good, and are developing software to identify such noises. Walter Esser, manager of the NVH lab at Chrysler, says that "we have engineered cars to be so quiet that customers now hear noises that were previously masked. It's a hunt for the source of undesirable sound. The trick is not so much to eliminate sound from the vehicle, but to tune and enhance them to give the car pleasing acoustics."[129]

Software similar to the Music Muse Continuously Record vi's is used to identify the spectral components of the unpleasant noises of automobiles. A sample screen from such software is shown in Appendix C.5. takes recordings from the sounds inside cars, and shows the FFT of the signals in real-time. Suspected annoying frequencies can then be edited out, and the signals can be played back to see if the sounds improve. The most widely used method of identifying good vs bad frequencies is to organized panels of people, called juries, to listen to recordings of car sounds with this software, and give their opinions.[130]

---

[127] ibid.
[128] Machine Design
[129] ibid.
[130] ibid.

CHAPTER 8: CONCLUSIONS

The Music Muse program is a tool that helps singers train their voices by showing them the components in their voices that contribute to timbre: the number of harmonics in the voice, and their amplitudes relative to one another. The Music Muse accomplishes this by applying Fourier analysis and cepstral analysis to voice signals, and displaying the resulting harmonic spectra and formant frequency plots. The harmonic spectra plots reveal the harmonic content in the voice, and the relative harmonic amplitudes. The formant plots display the frequency response of the throat, from which the formants can be determined. By using the Music Muse, singers can learn how to manipulate their formants and change their spectra vocally, to produce desired sounds.

The Music Muse program contributes to traditional voice training because it supplies the user with objective feedback. With this program, singers do not have to rely solely on their own biased perceptions of their sound, nor on the perceptions and subjective feedback of others. This program provides a visual image of the voice in its raw form, before it has been altered by the ear. The beauty of a singing voice has traditionally been rated on a scale of quality. The Music Muse program brings just enough science to the art of singing to calibrate the quality scale, and quantify vocal beauty.

APPENDIX A: REFERENCES

1. 1991. *American heritage dictionary.* 2nd College Ed. Boston: Houghton Mifflin Company.

2. Bartholomew, Wilmer T. 1942. *Acoustics of Music.* Englewood Cliffs, NJ: Prentice Hall, Inc.

3. Beranek, Leo L. and Istvan L. Ver. 1992. *Noise and vibration control engineering.* New York: John Wiley & Sons, Inc.

4. Coppens, Alan B., Austin R. Frey, Lawrence E. Kinsler, and James V. Sanders. 1982. *Fundamentals of acoustics.* 3rd Ed. New York: John Wiley & Sons, Inc.

5. Dvorak, Paul. 1996. *Good vibrations.* Machine Design Vol 68, no 4 (22 February). pp 68-72.

6. Fowler, Leslie and Leslie Willson. 1995. *Investigations of frequency spectra and time responses of the voice.*

7. Giancoli, Douglas C. 1989. *Physics for scientists and engineers with modern physics.* 2nd Ed. Englewood Cliffs, NJ: Prentice Hall, Inc.

8. Intelligent Devices, 7 Hickory Ridge, Baltimore, MD 21228.

9. Keily, Bill, Product Specialist: Kay Elemetrics Corp, 2 Bridgewater Lane, Lincoln Park, NJ 07035-1488

10. *Labview® for Macintosh user manual.* 1994. Austin Texas: National Instruments.

11. *Product newsletter of Mark of the Unicorn Inc.* No 13 (Winter 1996).

12. Newland, D. E. 1993. *An introduction to random vibrations, spectral and wavelet analysis.* 3rd Ed. New York: John Wiley & Sons, Inc.

13. Oppenheim, Alan V. and Ronald W. Schafer. 1989. *Discrete-time signal processing.* Englewood Cliffs, NJ: Prentice Hall, Inc.

14. O'Shaughnessy, Douglas. 1987. *Speech communication: human and machine.* New York: Addison-Wesley Publishing Company.

15. Parsons, Thomas W. 1987. *Voice and speech processing.* New York: McGraw-Hill Book Company.

16. Plomp, Reinier. 1976. *Aspects of tone sensation: a psychological study.* New York: Academic Press.

17. Rao, Singiresu S. 1990. *Mechanical vibrations.* 2nd Ed. New York: Addison-Wesley Publishing Company.

18. Roederer, Juan G. 1975. *Introduction to the physics and psychophysics of music.* 2nd Ed. New York: Spring-Verlag.

19. Sundberg, Johan. 1987. *Science of the singing voice, The.* Dekalb, Illinois: Northern Illinois University Press.

20. Vennard, William. 1968. *Singing: the mechanism and the technique.* 5th Ed. New York: Carl Fischer, Inc.

APPENDIX B: SCREENS FROM MUSIC MUSE

(IN CHRONOLOGICAL ORDER)

B.1 WHAT DOES IT ALL MEAN?

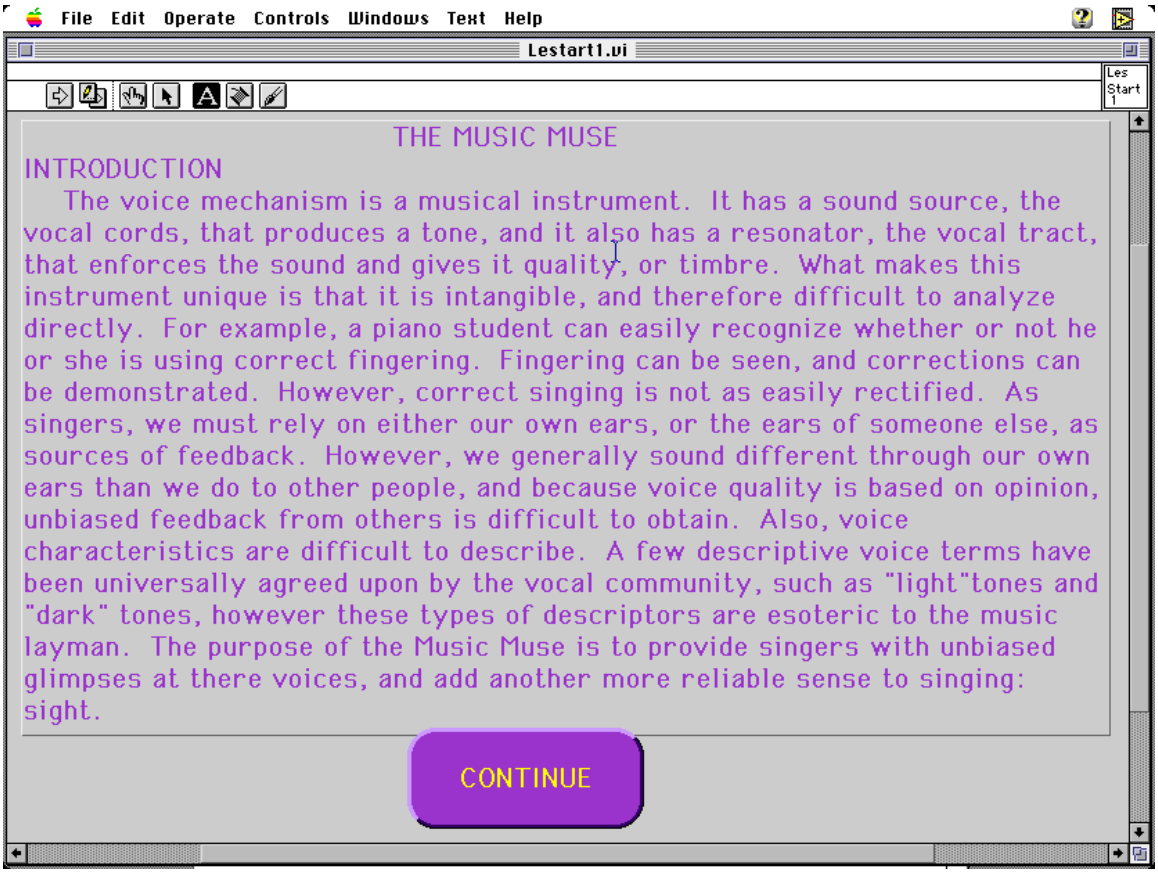


Fig B.1.1 First WHAT DOES IT ALL MEAN? help screen (BMP, 509KB)

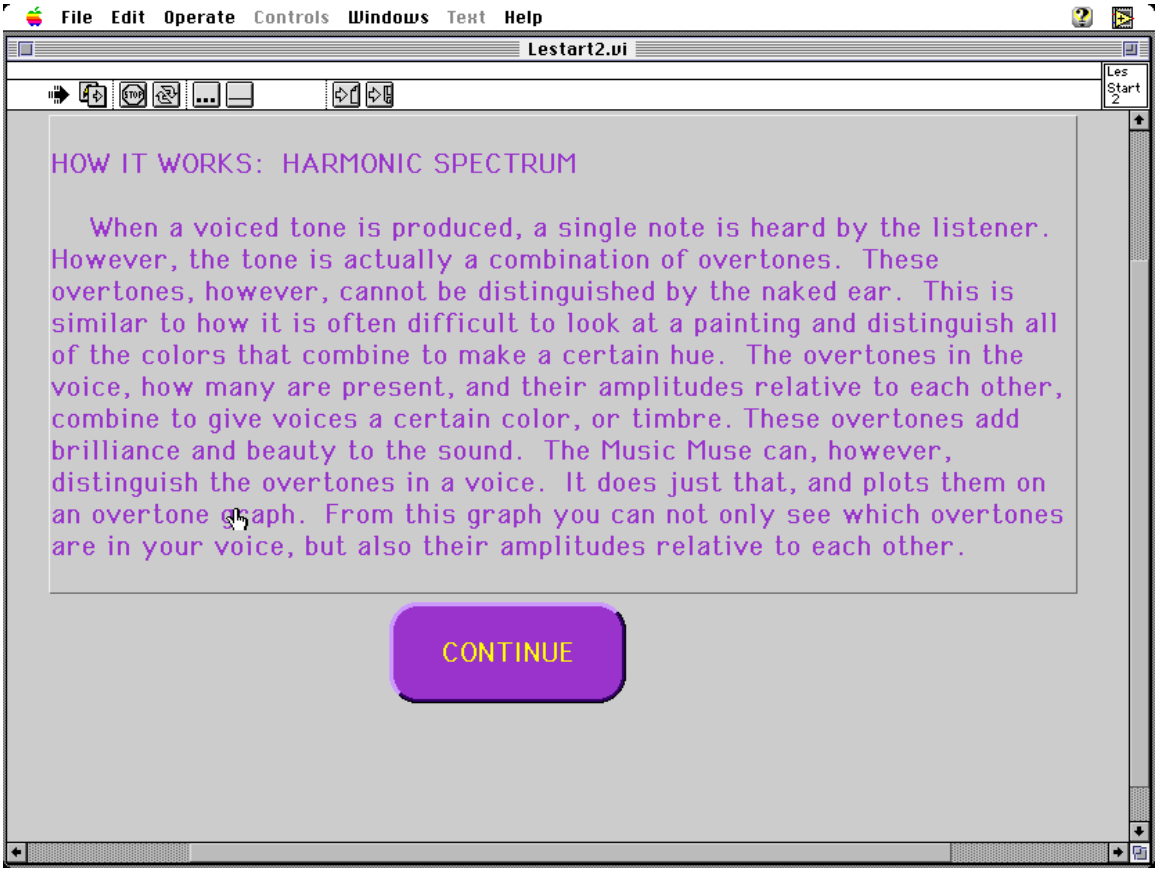


Fig B.1.2 Second WHAT DOES IT ALL MEAN? help screen (BMP, 509KB)

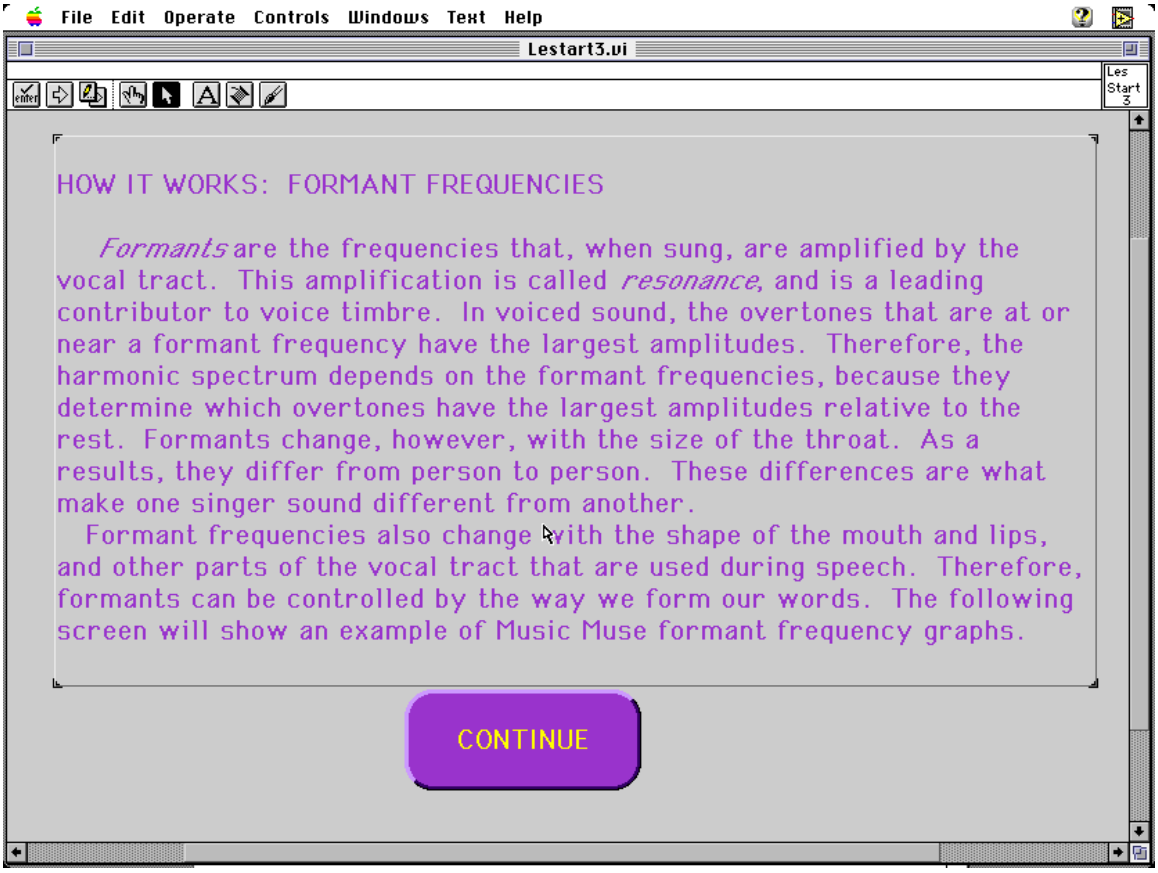


Fig B.1.3 Third WHAT DOES IT ALL MEAN? help screen (BMP, 509KB)

B.2 RECORD A VOICE HELP SCREENS

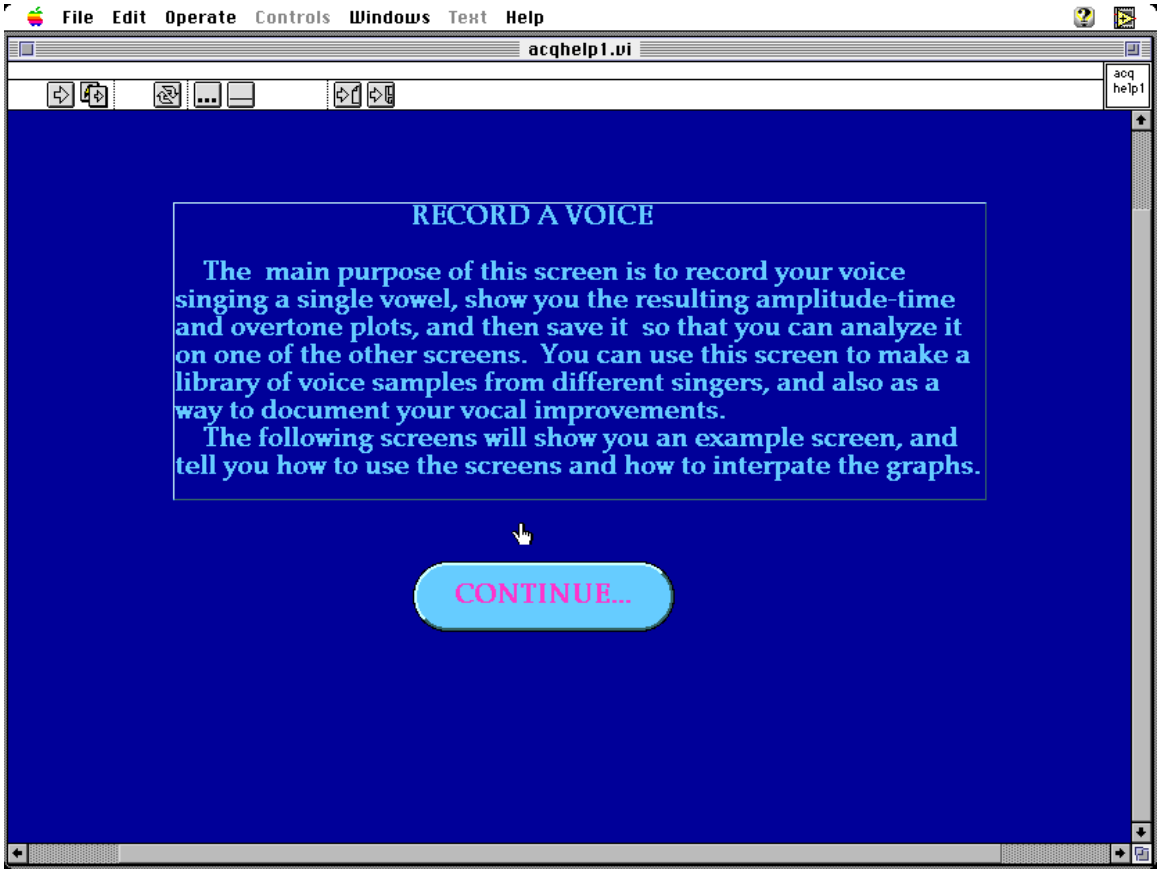


Fig B.2.1 First RECORD A VOICE help screen (TIFF, 509KB)

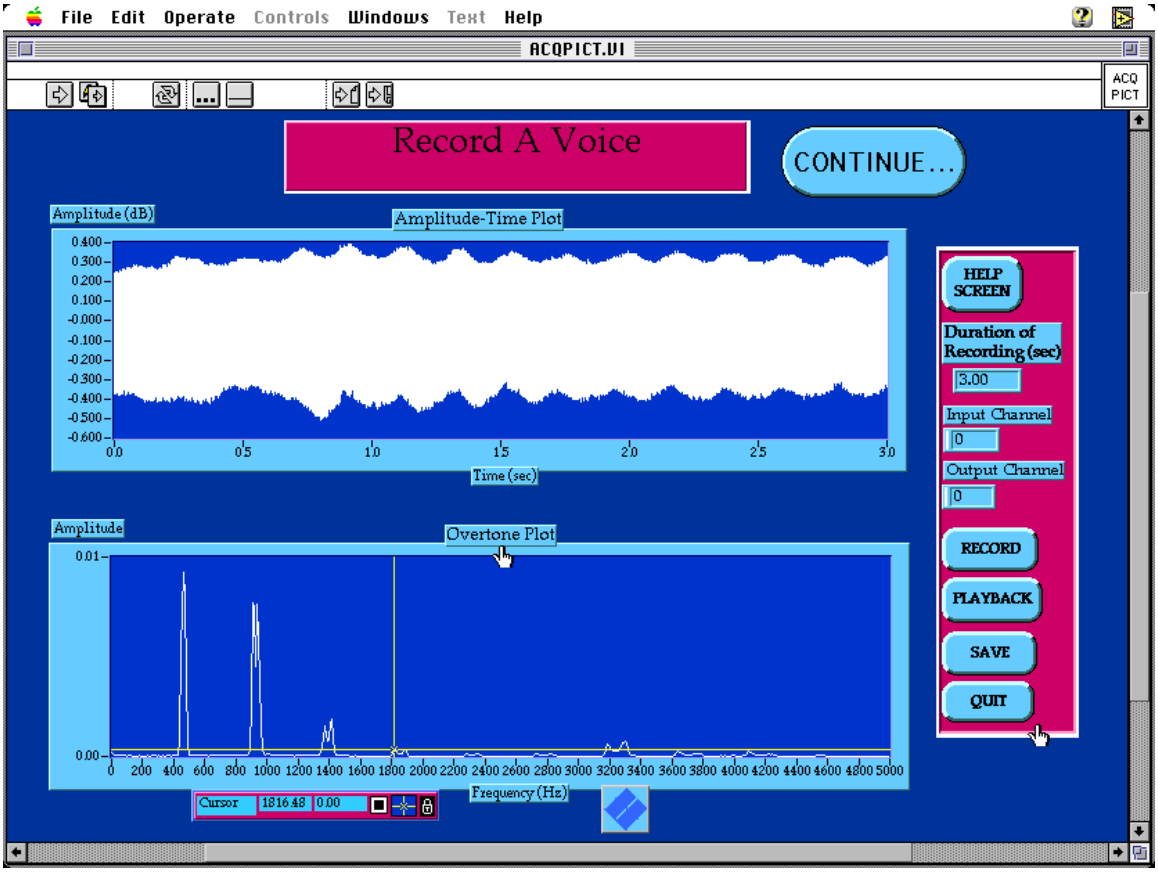


Fig B.2.2 Second RECORD A VOICE help screen (TIFF, 509KB)

Fig B.2.3 Third RECORD A VOICE help screen (TIFF, 546KB)

B.3 CONTINUOUSLY RECORD A VOICE HELP SCREENS

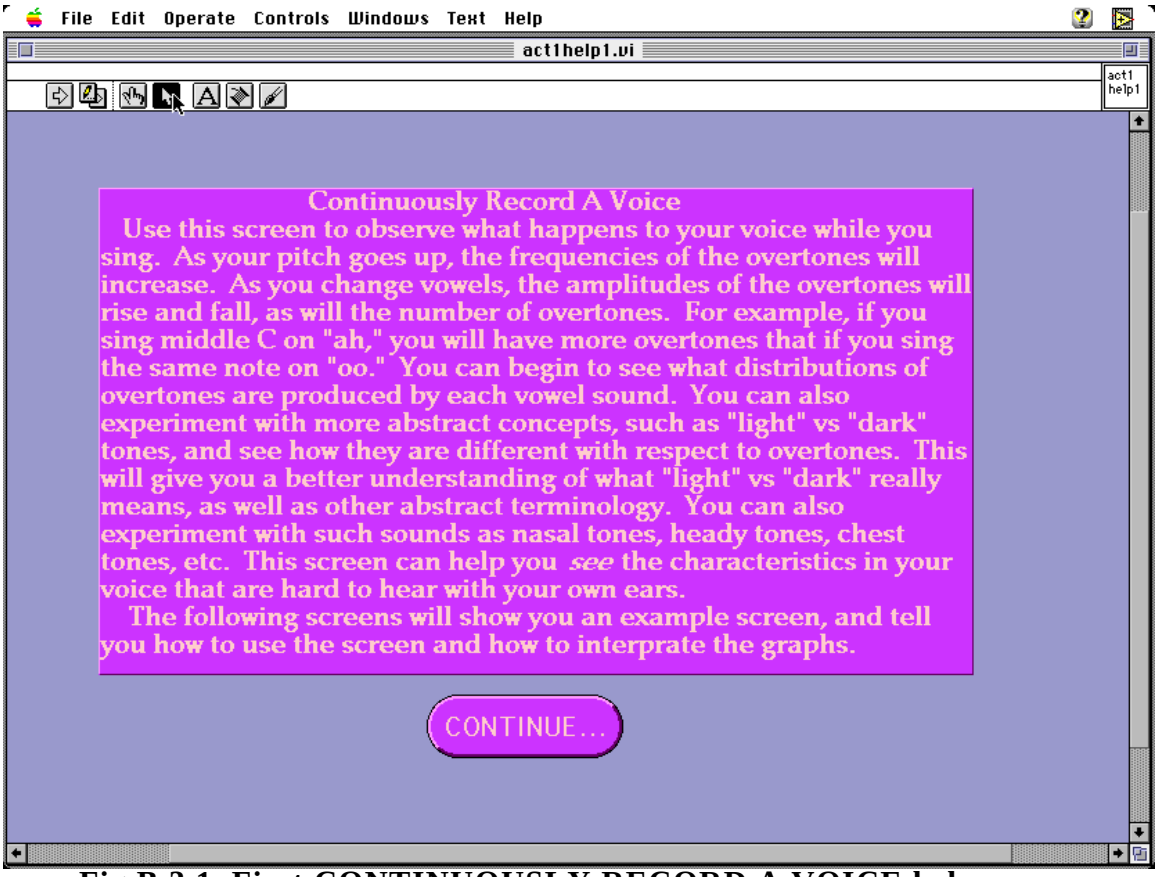


Fig B.3.1 First CONTINUOUSLY RECORD A VOICE help screen (TIFF, 546KB)

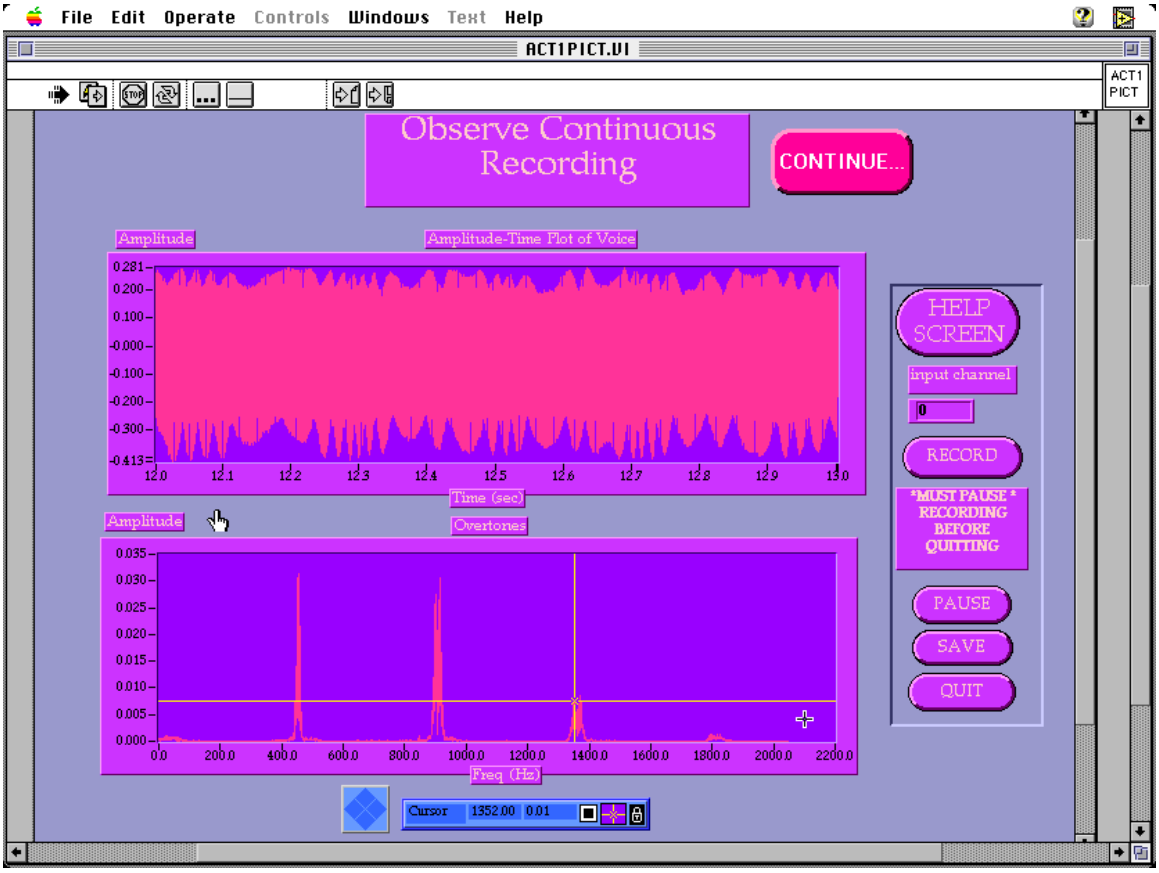


Fig B.3.2 Second CONTINUOUSLY RECORD A VOICE help screen (TIFF, 546KB)

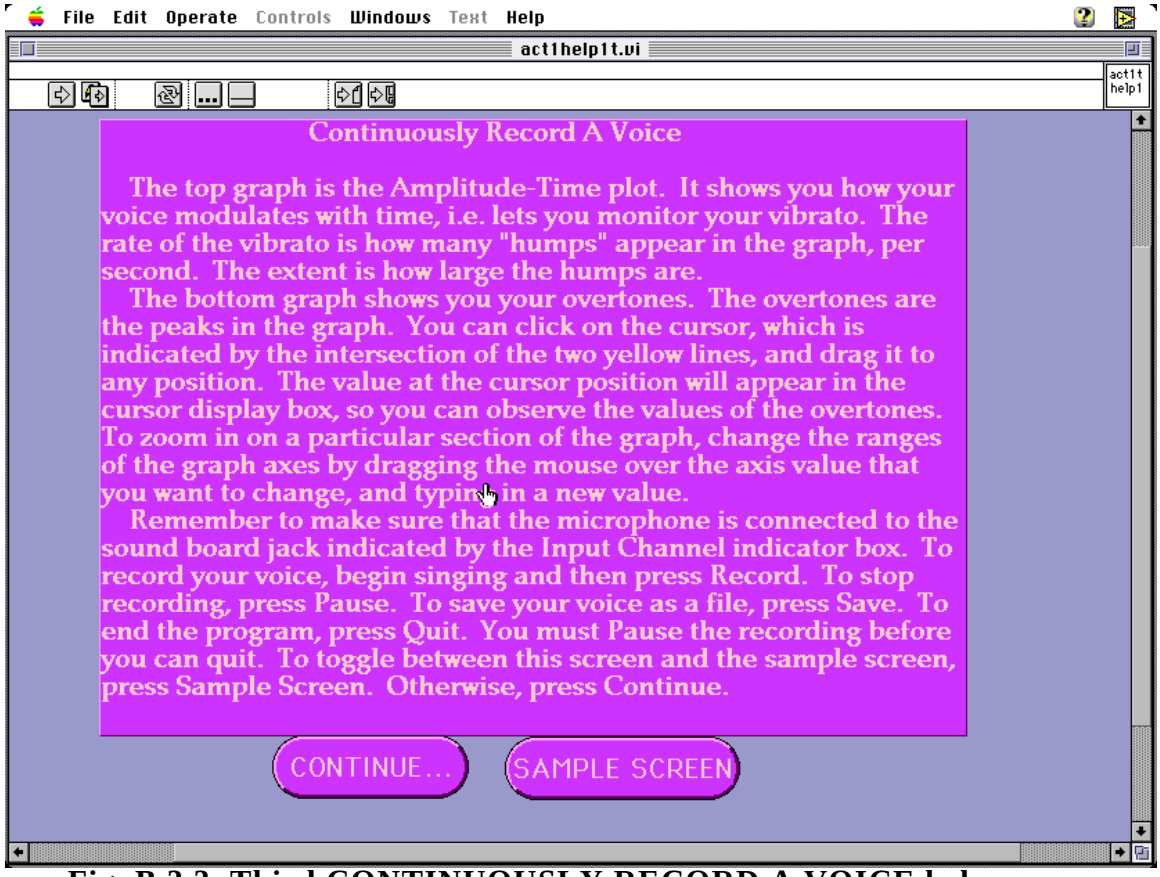


Fig. B.3.3 Third CONTINUOUSLY RECORD A VOICE help screen (TIFF, 546KB)

B.4 CONTINUOUSLY RECORD A VOICE AND COMPARE HELP SCREENS

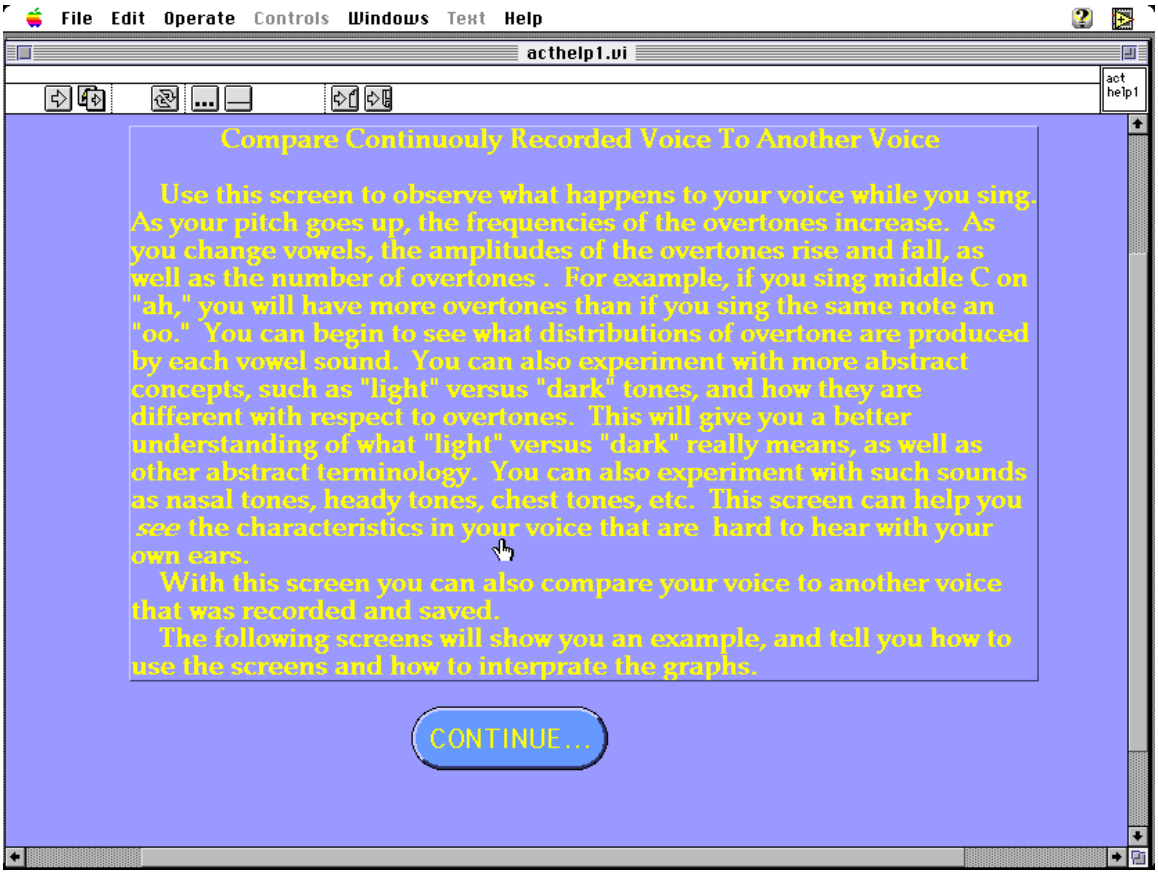
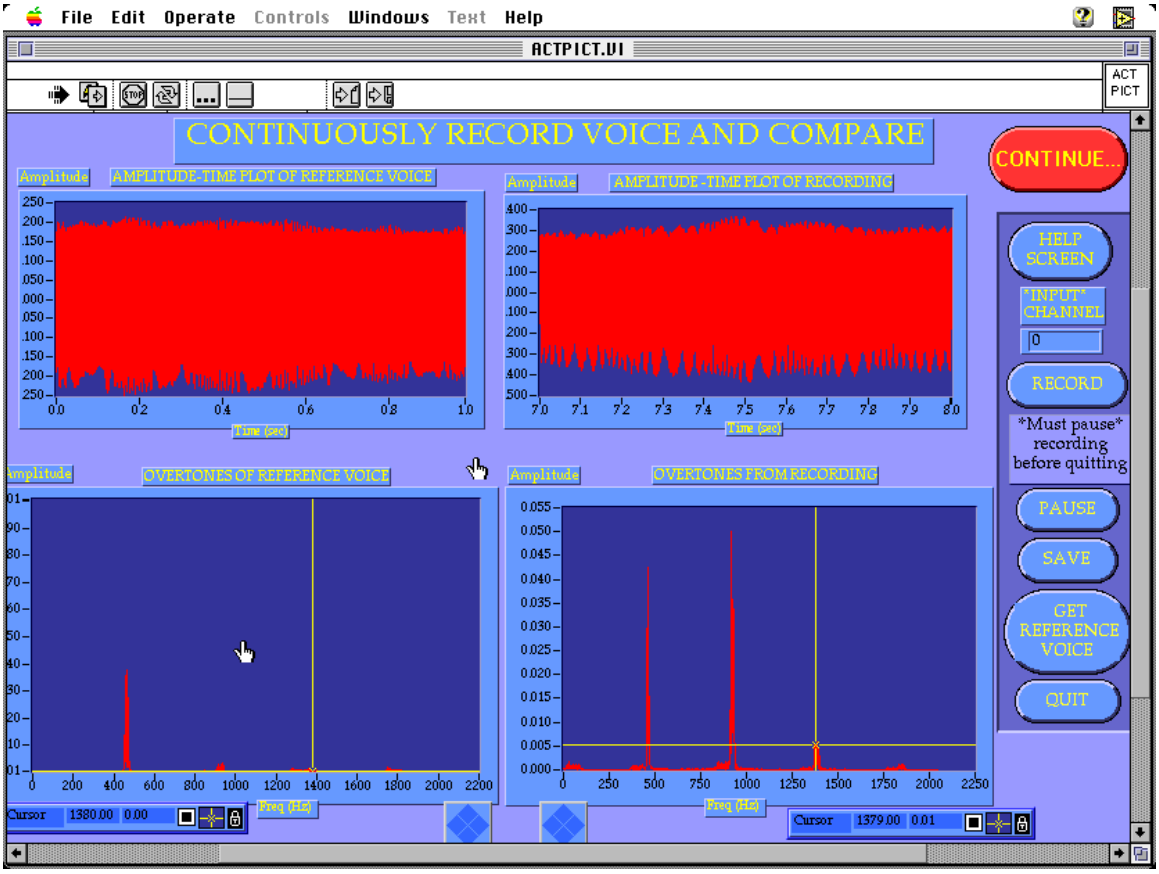


Fig B.4.1 First CONTINUOUSLY RECORD A VOICE AND COMPARE help screen (TIFF, 546KB)



B.4.2 Second CONTINUOUSLY RECORD A VOICE AND COMPARE help screen (TIFF, 546KB)

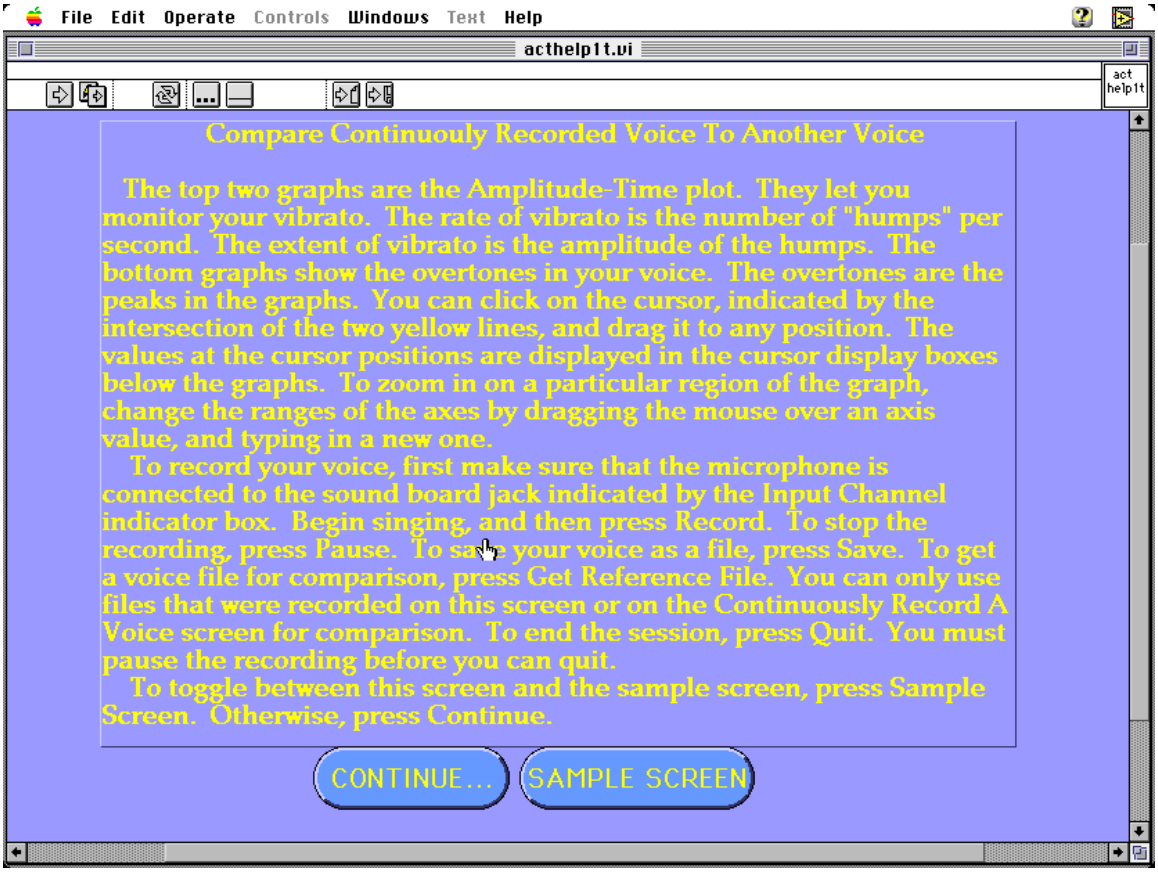


Fig B.4.3 Third CONTINUOUSLY RECORD A VOICE AND COMPARE help screen (BMP, 509KB)

B.5 CALCULATE DIFFERENCE IN OVERTONES HELP SCREENS

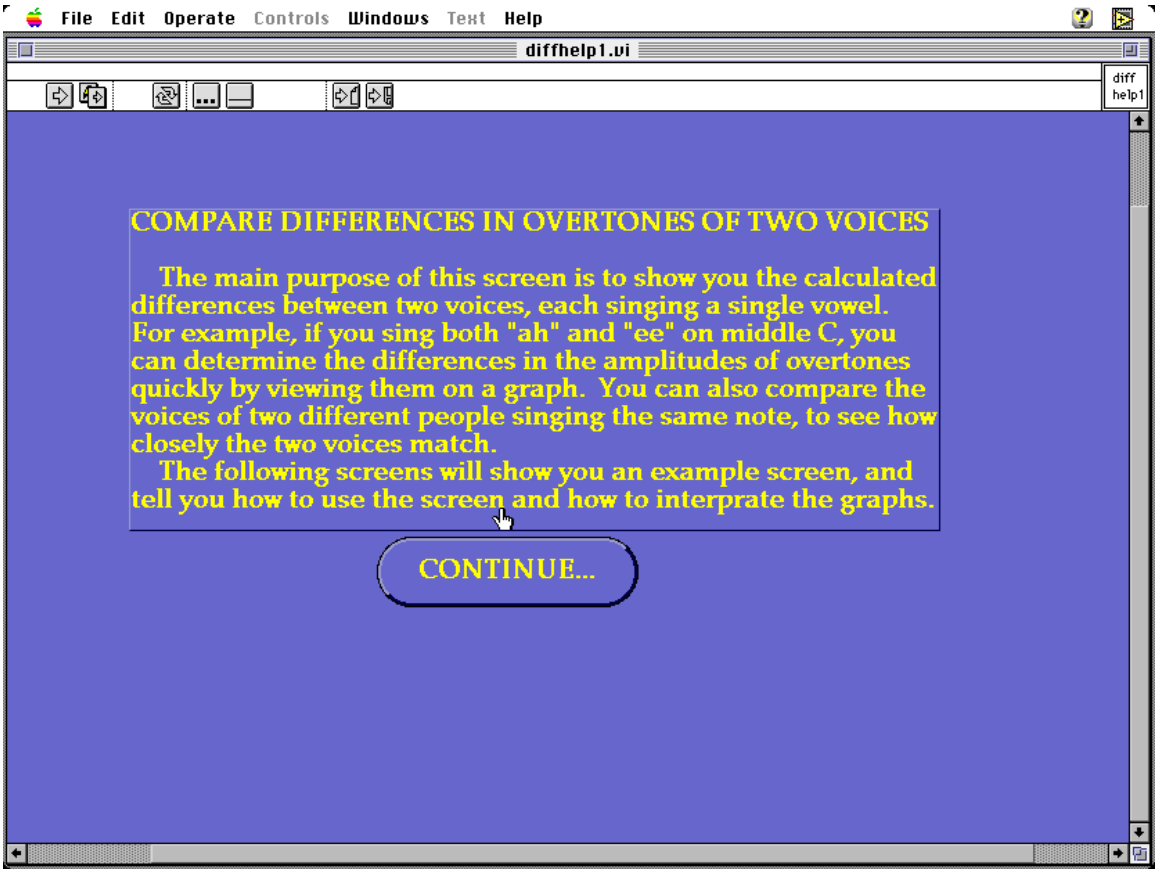


Fig B.5.1 First CALCULATE DIFFERENCE IN OVERTONES help screen (TIFF, 546KB)

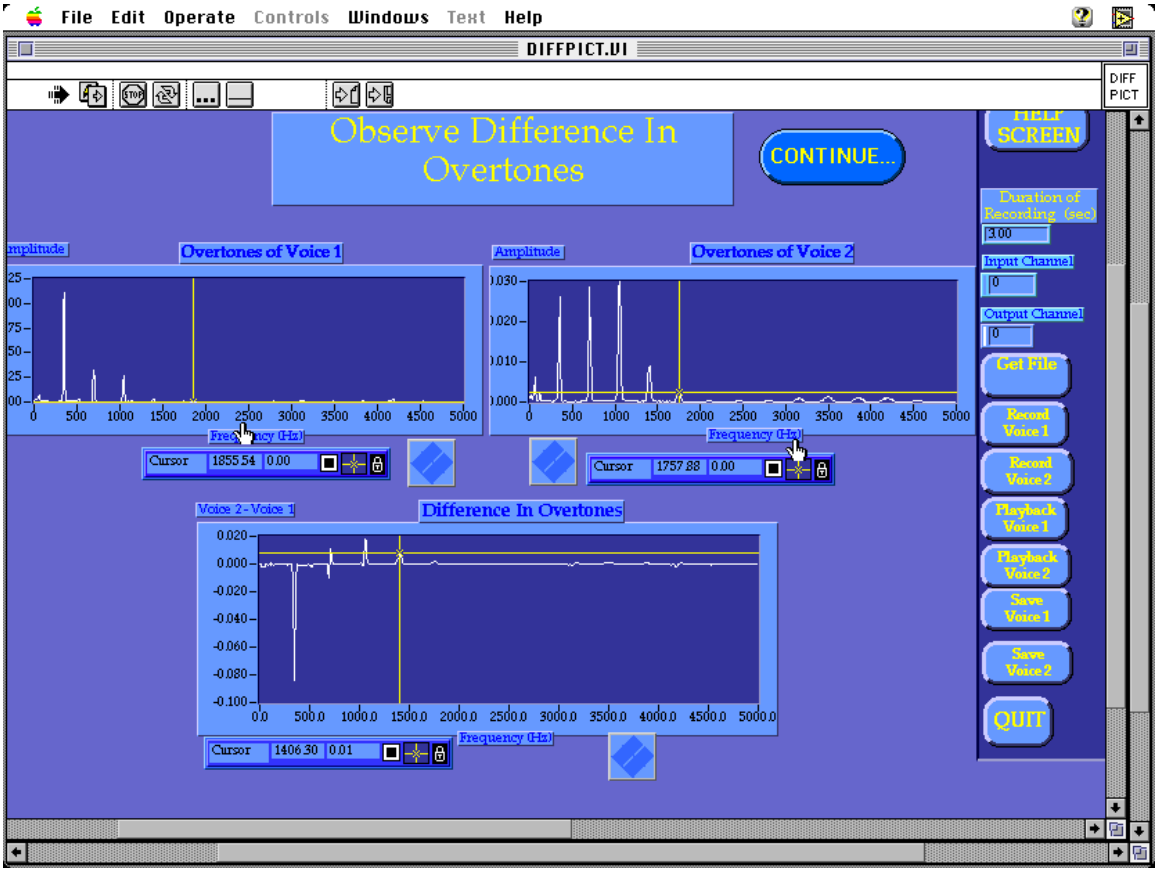


Fig B.5.2 Second CALCULATE DIFFERENCE IN OVERTONES help screen (TIFF, 546KB)

Fig B.5.3 Third CALCULATE DIFFERENCE IN OVERTONES help screen (TIFF, 546KB)

B.6 CALCULATE FORMANTS HELP SCREENS

Fig B.6.1 First CALCULATE FORMANTS help screen (TIFF, 546KB)

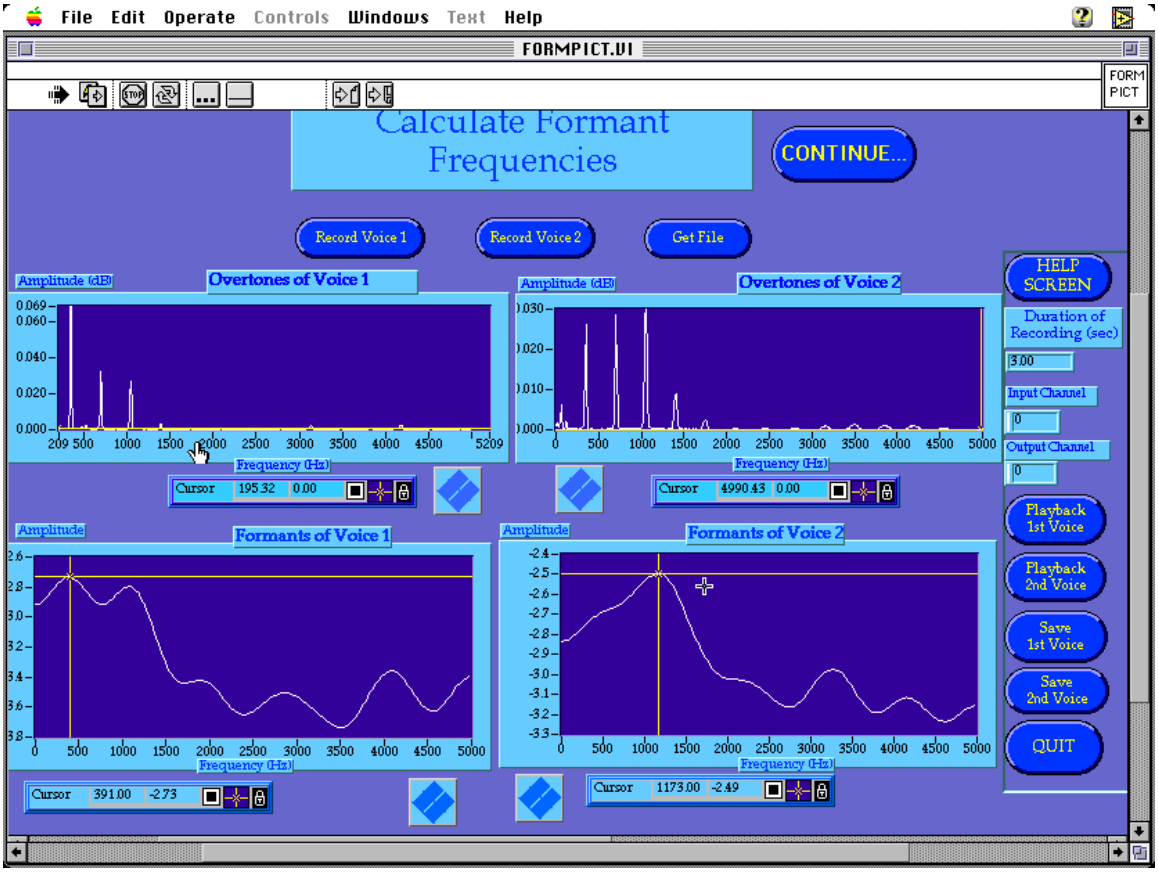


Fig B.6.2 Second CALCULATE FORMANTS help screen (TIFF, 546KB)

Fig B.6.3 Third CALCULATE FORMANTS help screen (TIFF, 546KB)

APPENDIX C: MUSIC MUSE BANDWIDTH COMPARISONS

Fig B.1 These overtone plots of a female singing a fundamental around 900 Hz, were calculated using the Music Muse. The data for the top plot was sampled at 10 kHz, while the data for the bottom plot was sampled at 20 kHz. It can be seen from the plot that even for a fundamental as high into the upper register of the voice as 800 Hz, there is insignificant harmonic information above 5000 Hz. Therefore, the 10 kHz sampling rate suffices.

APPENDIX D: ILLUSTRATIONS OF SCREENS FROM RELATED SOFTWARE

D.1 PRO-AUDIO ANALYZER

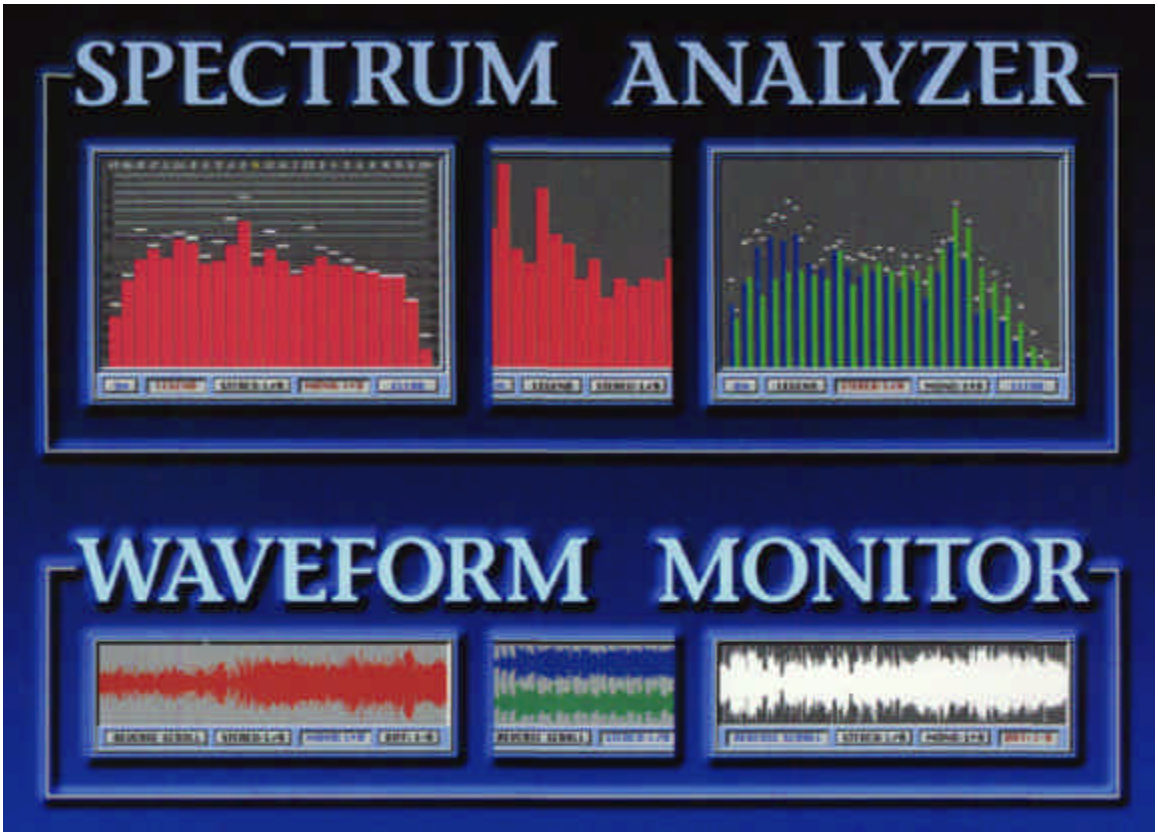
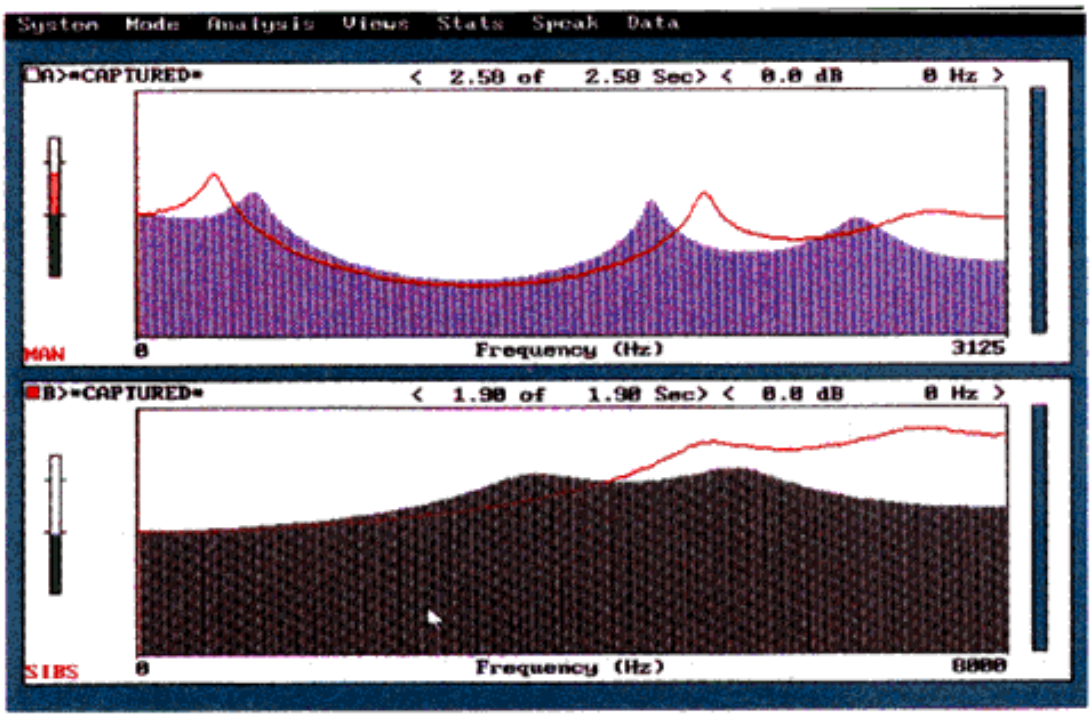


Fig D.1 The Pro-Audio Analyzer from Intelligent devices has a Spectrum Analyzer screen and a Waveform Monitor screen, that are similar to the graphical displays of the Music Muse Continuously Record vi's. The waveform monitor screens show the time plots of the signals in real time. The spectrum analyzer shows the real-time FFT of the signals, broken into 1/3-octave bands (Intelligent Devices).

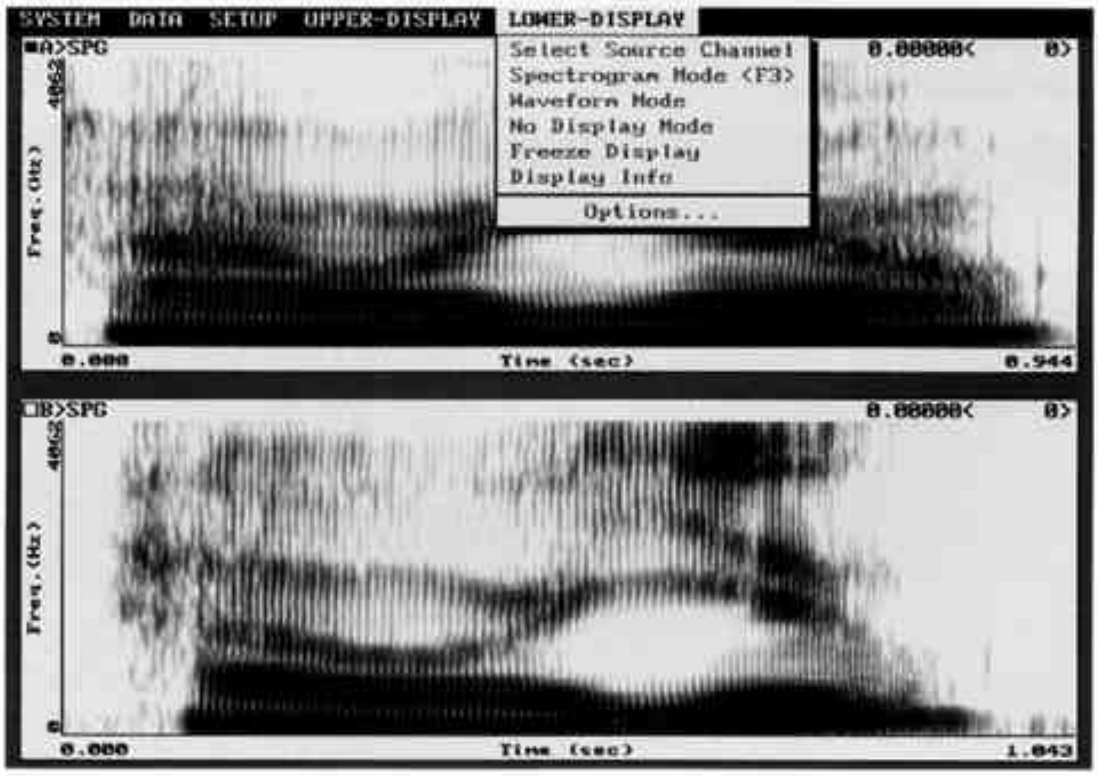
D.2 KAY ELEMETRICS SONA-MATCH, MODEL 4327



A split screen display of Sona-Match with the upper screen displaying attempt (in filled display) against target (in outline) for vowel production "e". The bottom screen is set to wide frequency range (0-8,000 Hz) and shows "s" in outline, and "sh" in filled display. Sona-Match displays real-time LPC-derived frequency response used for speech training applications.

Fig D.2 The Sona-Match from Kay Elemetrics is similar to the Music Muse Calculate Formants vi. It calculates the formants of a signal, indicated by the filled display, and plots them against a target signal, indicated by the outlined display, in real-time. Unlike the Calculate Formants vi which uses cepstral analysis to extract the formants, the Sona-Match uses linear predictive coding.

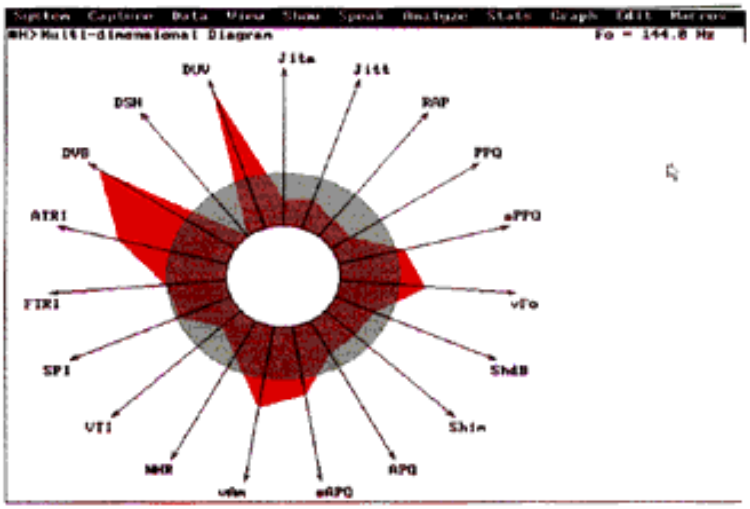
D.3 KAY ELEMETRICS REAL-TIME SPECTROGRAM, MODEL 4329



Split-screen display with spectrogram of target on the top screen and attempt on the bottom screen.

Fig D.3 The Real-Time Spectrograph is similar to the Continuously Record vi's, except it displays the graph in a time-frequency format. Just like the Music Muse, it shows a real-time signal and a target signal for comparison.

D.4 KAY ELEMETRICS MULTI-DIMENSIONAL VOICE PROGRAM, MODEL 4305



MDVP can graphically display up to 22 extracted parameters against a normative database. Parameters plotted inside the green circle (on the screen itself) indicate normal values, while the red area indicates parameters with higher than normal values. This display provides a convenient multi-parameter, graphic "snapshot" of the patient's performance in relation to normal subjects. The subject above has significant voice tremor.

Parameters Extracted

- Average Fundamental Freq. (Hz)
- Average Glottal Period (msec)
- Duration (sec)
- Absolute Jitter (microseconds)
- Relative Jitter - RAP (%)
- Relative Jitter - PPQ (%)
- Fundamental Freq. Variation (%)
- Shimmer (dB)
- Relative Shimmer 1 (%)
- Relative Shimmer 2 (%)
- Peak Ampl. Variation (%)
- Noise-to-Harmonic Ratio
- Turbulence Index
- Soft Phonation Index
- Degree of Voice Breaks (%)
- Degree of Subharmonics (%)
- Degree of Voiceless (%)
- Number of Voice Breaks
- Degree of Subharmonic Segments
- Number of Unvoiced Segments
- Number of Segments Computed
- Tremor Rate (Hz)
- Tremor Intensity (dB)

Fig D.4 The MDVP from Kay Elemetrics displays 22 different voice quality parameters, each of which is listed above. These parameters are better indicators of speech quality, however, rather than quality in singing.

D.5 NVH SOFTWARE FROM SDRC

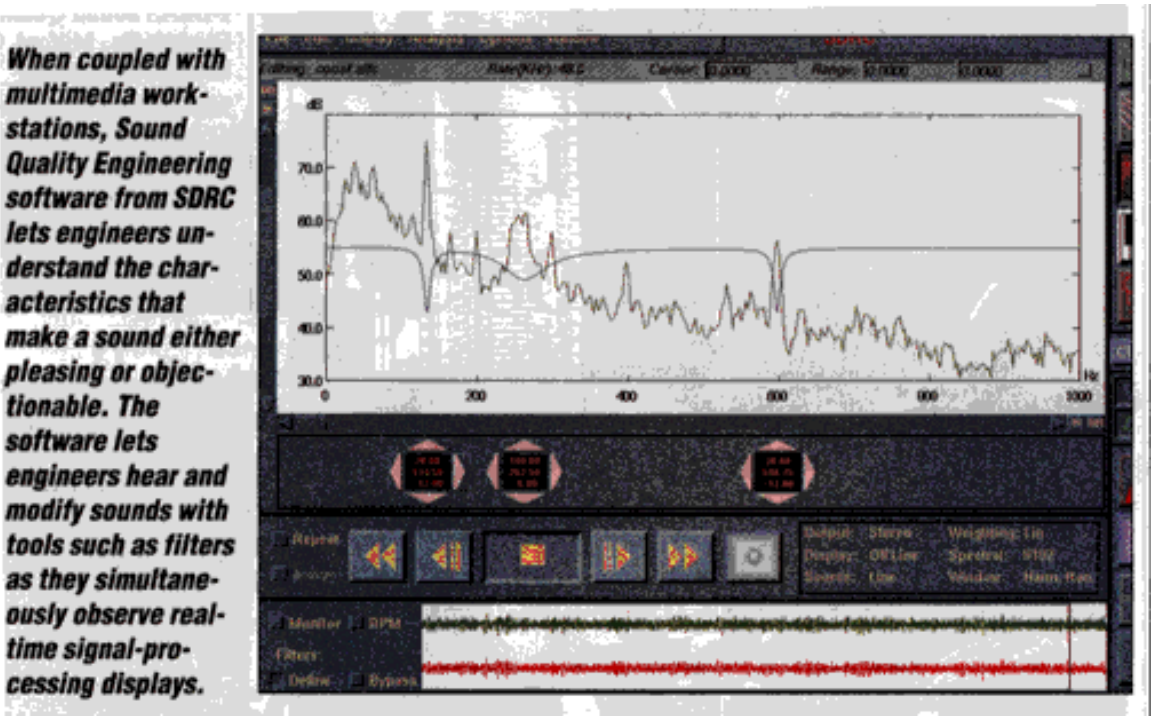


Fig D.5 This software uses the principles of the Music Muse to quantify the sound quality of cars. Juries of people listen to samples of noises from cars, and try to identify the less pleasing sounds by their spectral characteristics. This software is the NVH equivalent to the Music Muse Continuously Record vi's.

VITA

Leslie Willson

Leslie Willson is the only child of Etta and Wayne Willson. She was born in 1973 in Fairfax, VA, where she remained until she entered college. Leslie enrolled in Virginia Tech In 1991, as a freshman in engineering. As an undergraduate student, Leslie was a member of UTAP (Undergraduate Training Assistance Program), sponsored by the Defense Intelligence Agency. This program paid for her entire undergraduate school education, and provided her with summer employment as an engineer in training.

Leslie received her bachelor's degree in mechanical engineering in 1995, and entered the master's program that same year. She worked under Dr. A. L. Wicks, and completed her master's degree requirements in November of 1996. Upon completion, she returned to the Defense Intelligence Agency as a mechanical engineer.