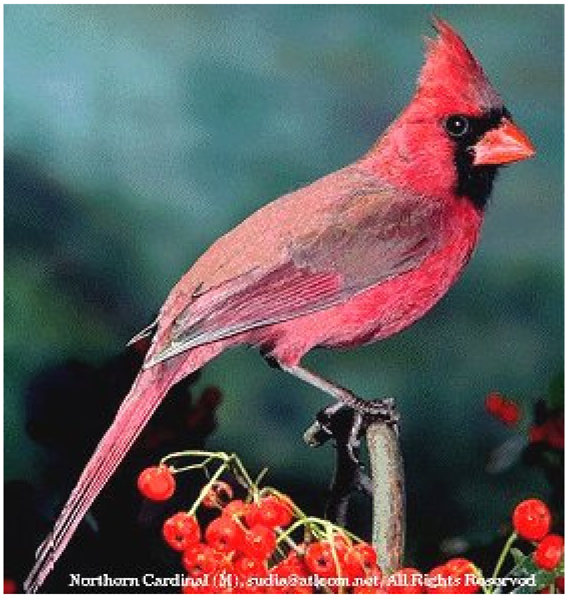
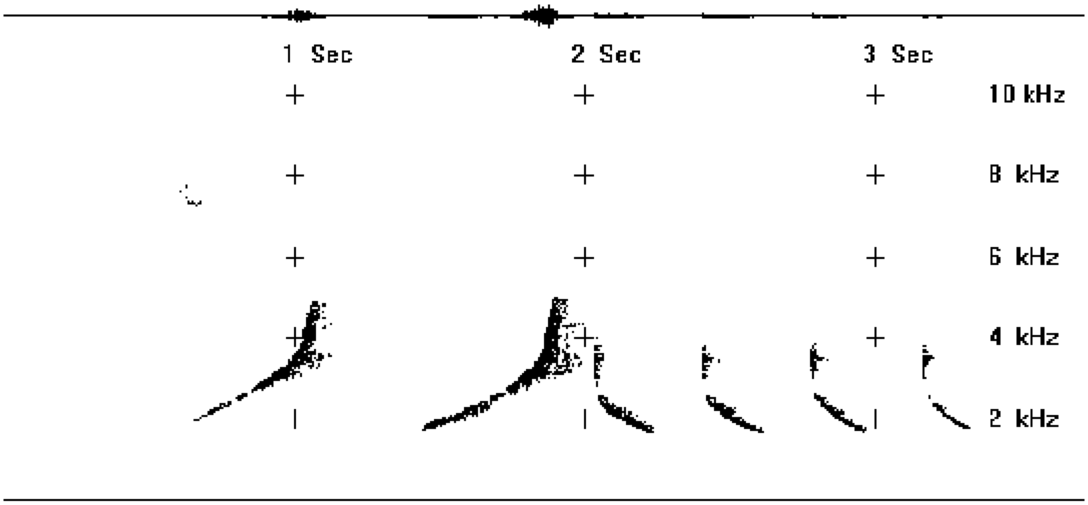
U. Spectrograms

Northern Cardinal



Courtesy Dr. Sudia



Courtesy Gregory J. Kunkel

The *spectrogram* is a special plot of sound over a short time frame. We have seen that audiograms use two axes to incorporate on a graph two basic properties of sound. The properties tested in an audiogram are amplitude (sound level) and frequency. The third basic feature of sound, the timbre, is understood by amplitudes and frequencies of the spectrum. The spectrum analyzer illustrates this. A periodic wave has a Fourier spectrum of harmonics. The frequency of the periodic complex waveform sets the base frequency for its Fourier spectrum.

However, the spectrum analyzer responds to all sounds, both periodic and aperiodic. An aperiodic sound does not have a defined frequency. So we can't express such a sound in terms of a harmonic series. Noise is a good example. It contains all frequencies. Inharmonic tones such as those produced by balanced modulation present another example. These contain non-harmonic spectral components. Nevertheless, a frequency spectrum can still describe these sounds. It just means that we need to put more in the spectrum than we have to for periodic waveforms.

The problem is this. Aperiodic tones do not remain the same as time goes on. They can be crashes, speech, music in progress, etc. For a sketch or graph to reflect these changing waves, we need to consider a time axis. This pushes us to three variables: amplitude, frequency, and time. This implies a graph in three dimensions, the axes being length, width, and height. We will see that the spectrogram is an ingenious way to reduce such a graph to two dimensions! To understand how this is accomplished, we once again return to our square wave.

Fig. U-1 illustrates the square wave we encountered earlier in our discussion of envelope shaping. Focus your attention on the sustain phase. The sustain is constant in time for the duration of the sustain phase. This is the easiest part of the wave in Fig. U-1 to consider. We will translate this section of the wave, step by step, into a spectrogram. Before considering this full-blown spectrogram which incorporates amplitude, frequency, and time information, let's look at combinations of two parameters.

First, take amplitude and time. The amplitude of a wave is a measure from equilibrium to the maximum height. The maximum heights of the square wave during the sustain phase are constant. The periodic square wave changes in displacement from crest to tough, but the boundary height (amplitude) for the crests is fixed. Fig. U-2a is a graph of amplitude (vertical axis) as a function of time (horizontal axis). For any time along the horizontal, the amplitude is seen to be the same. The value of the constant amplitude could be read on the vertical axis if there were a scale of numbers there. This amplitude is the net result of the harmonics that combine to form the overall square wave.

In Fig. U-2b we choose to plot amplitude (harmonics) and frequency. This plot is our familiar decomposition of the wave in terms of its Fourier spectrum, similar to what the spectrum analyzer on an equalizer does. The equalizer spectrum analyzer is not as detailed because there, groups of frequencies are thrown together into a series of bands. However, the equalizer shows the changing spectrum in time by providing us with spectral bars that flicker at different heights. Fig. U-2c is obtained by folding Fig. U-2a down and replacing the total amplitude with the harmonic amplitudes.

Fig. U-1. Square Wave Shaped by Envelope Generator.

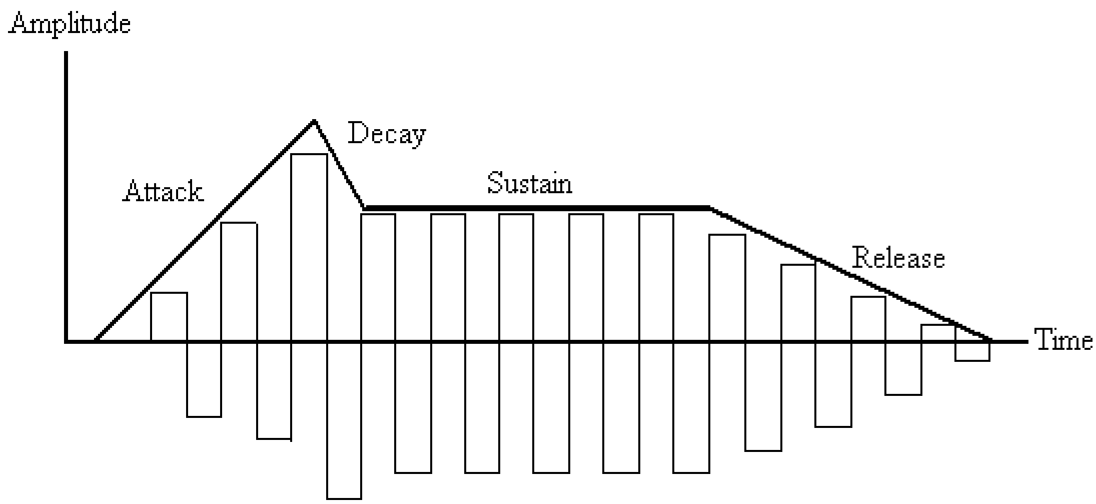


Fig. U-2d is obtained from Fig. U-2c by looking at the edges of the amplitude tops. Taller amplitudes are indicated by darker lines in Fig. U-2d. In this way, three dimensions are collapsed into two.

Fig. U-2a. Total Amplitude and Time.

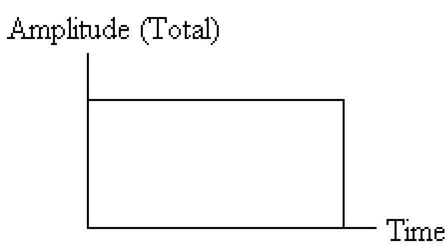


Fig. U2-b. Harmonic Amplitudes and Frequency.

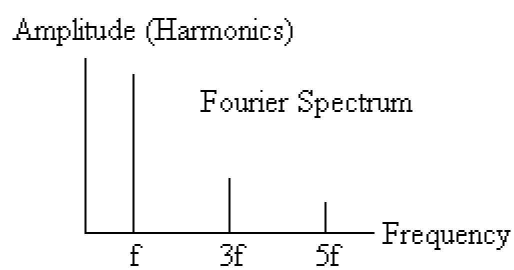


Fig. U-2c. Harmonics and Time.

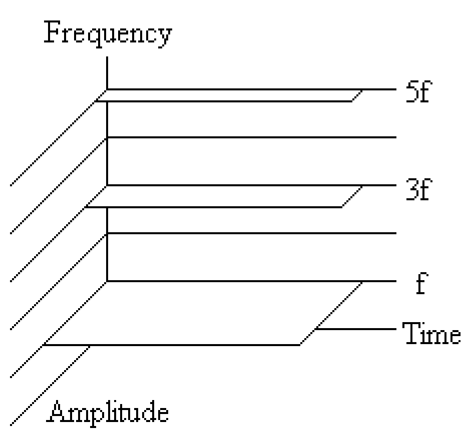


Fig. U-2d. Spectrogram.

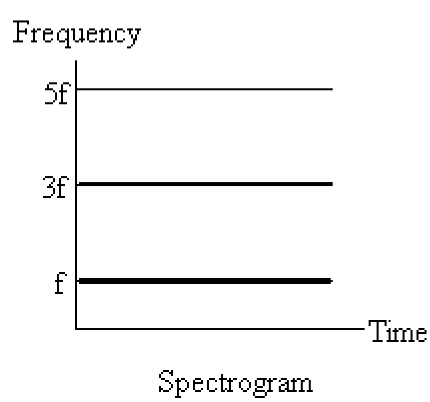


Fig. U-3 illustrates noise spectrograms. The spectrogram for white noise contains all frequencies. Frequencies are indicated along the vertical axis in a spectrogram. Therefore, all frequencies get shaded in for white noise. Shading proceeds to the right as long as the white noise is present. Moving along the horizontal axis of our graph represents movement in time.

A spectrogram of high-frequency noise describes a hissing sound. Tape hiss, which begins at around 5 kHz, is a good example. Low-frequency noise is heard more as a sigh. Let out a deep breath. Different shades of darkness represent different sound levels. The shading in Fig. U-3 is uniform for each case. White noise actually has more emphasis in the higher frequencies. The spirit of the spectrograms that follow is approximate.

Fig. U-4 is a set of spectrogram sketches. Many of these have been adapted from spectrograms made at Bell Laboratories, appearing in F. Alton Everest, *Acoustic Techniques for Home and Studio*, 2nd ed. (TAB Books, Blue Ridge Summit, PA, 1984). An excellent source of bird spectrograms can be found in Chandler S. Robbins, Bertel Bruun, and Herbert S. Zim, *A Guide to Field Identification: Birds of North America* (Golden Press, New York, 1983).

Fig. U-3. Noise Spectrograms.

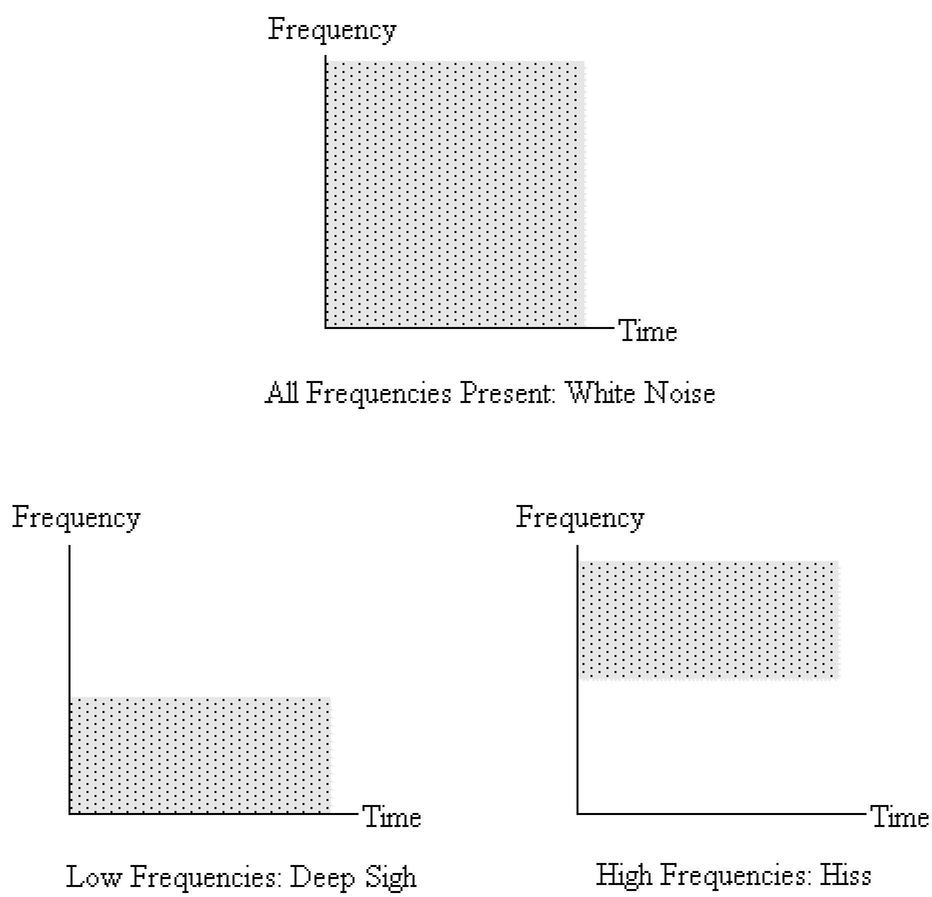
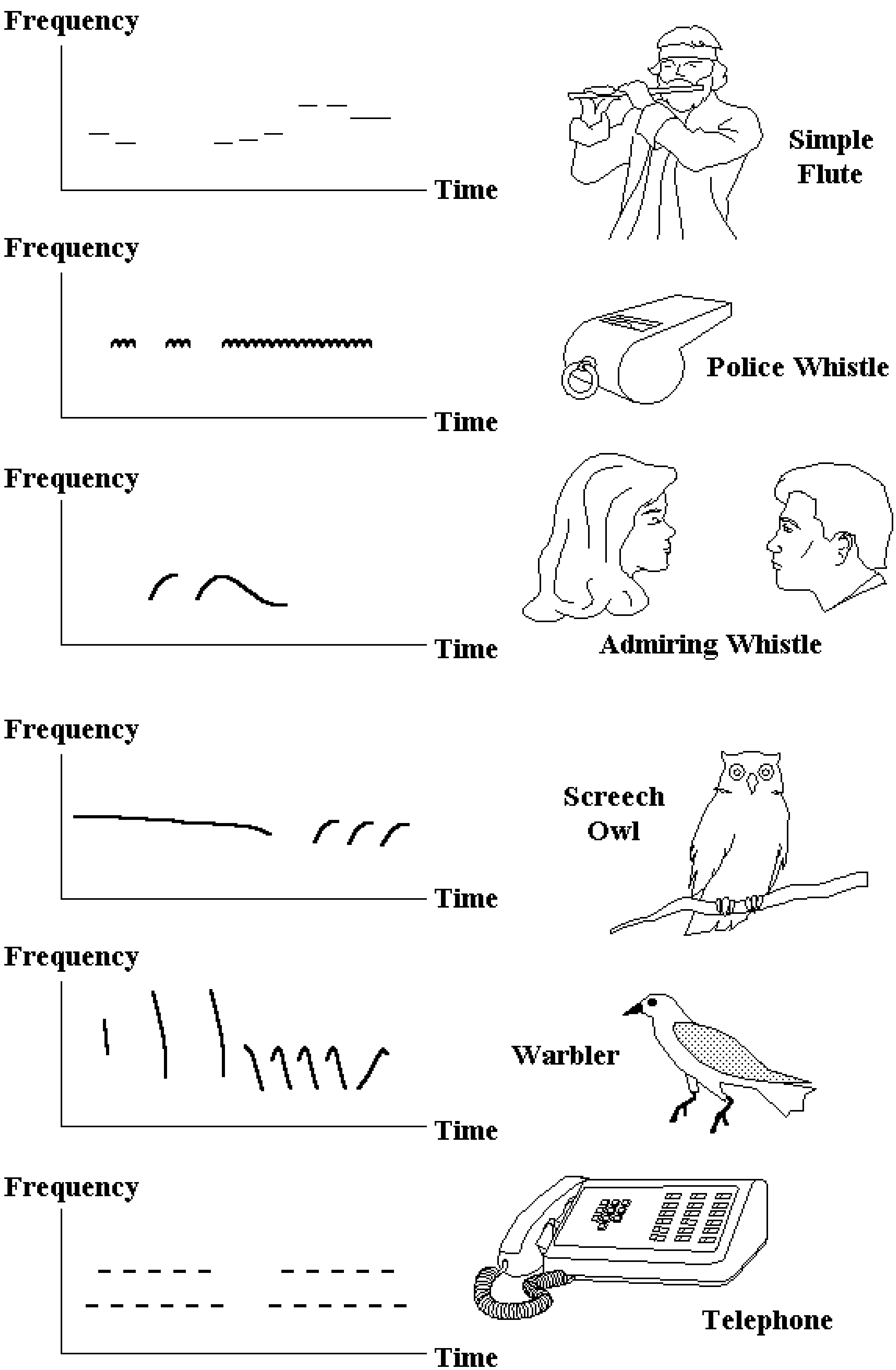
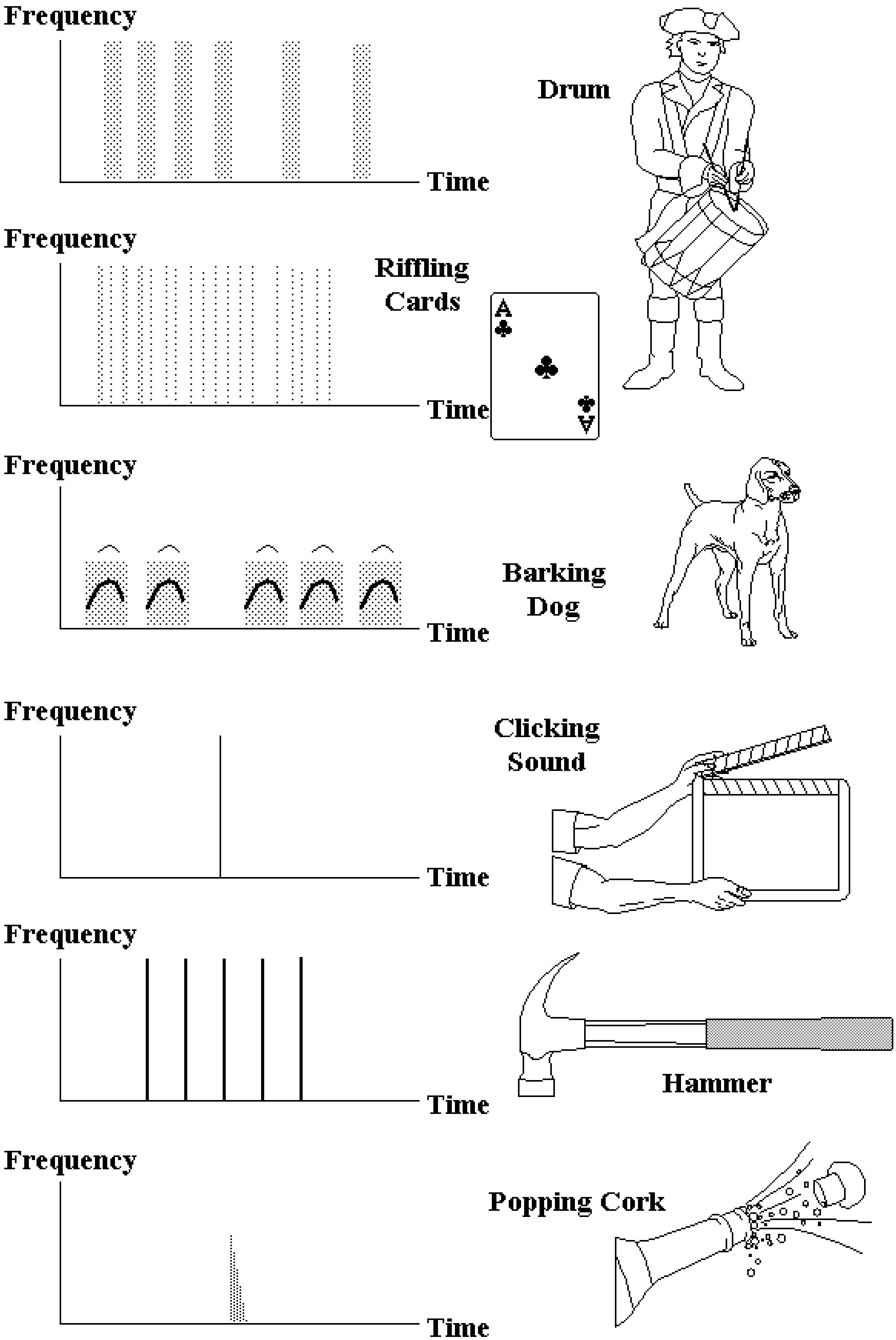
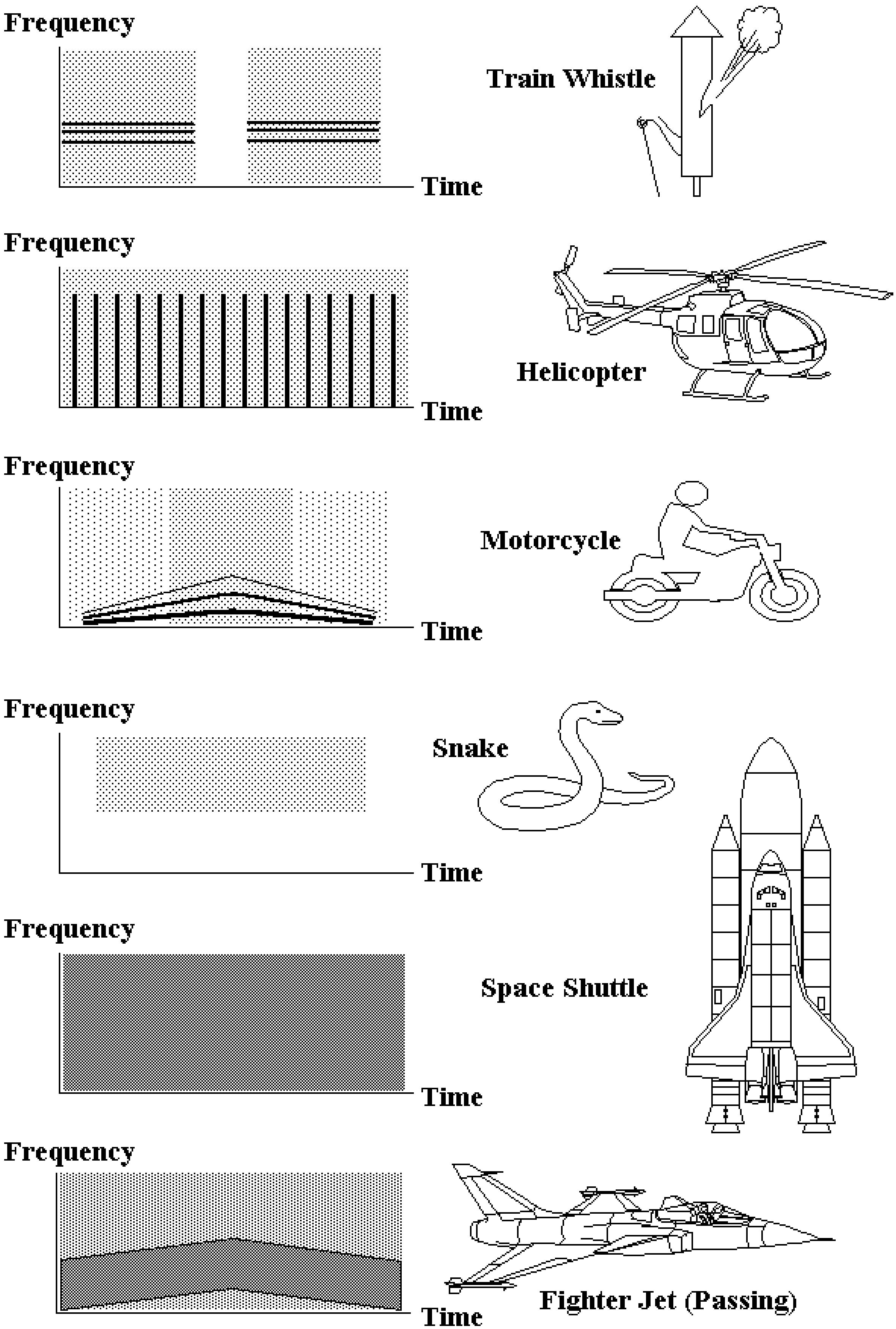
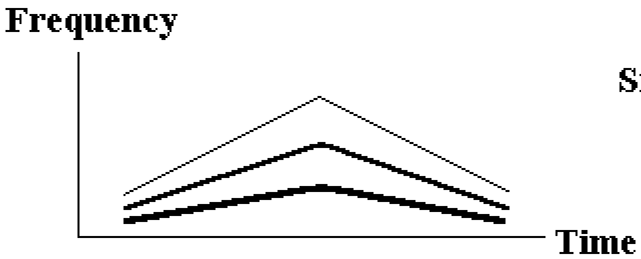


Fig. U-4. Spectrogram Sketches.

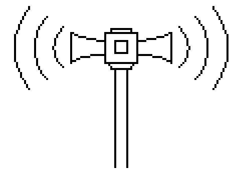


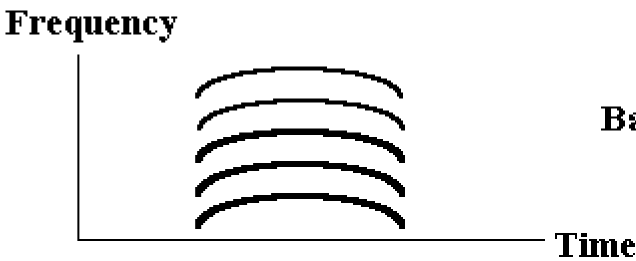

Frequency
Time
Drum
Frequency
Riffling
Cards
Time
Frequency
Barking
Dog
Time
Frequency
Clicking
Sound
Time
Frequency
Hammer
Time
Frequency
Popping Cork
Time


Frequency
Time
Train Whistle
Frequency
Time
Helicopter
Frequency
Time
Motorcycle
Frequency
Time
Snake
Frequency
Time
Space Shuttle
Frequency
Time
Fighter Jet (Passing)

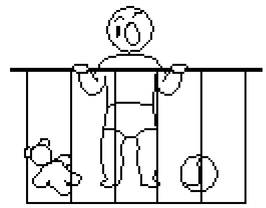


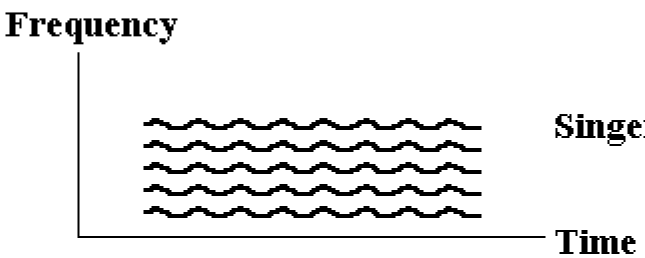
Siren





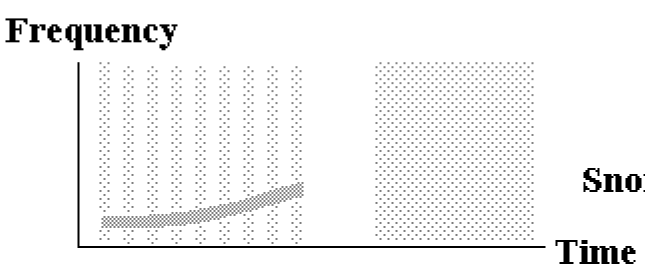
Baby Crying



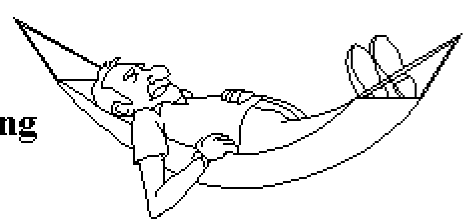


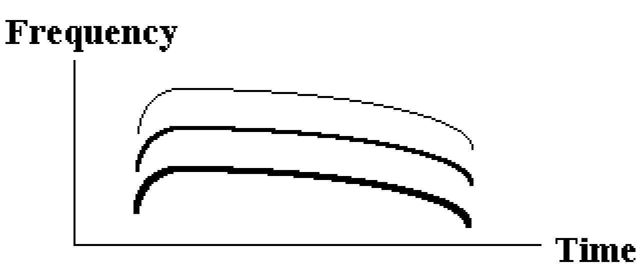
Singer (Vibrato)

Laughing

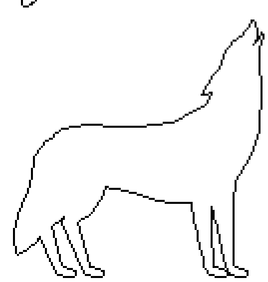


Snoring





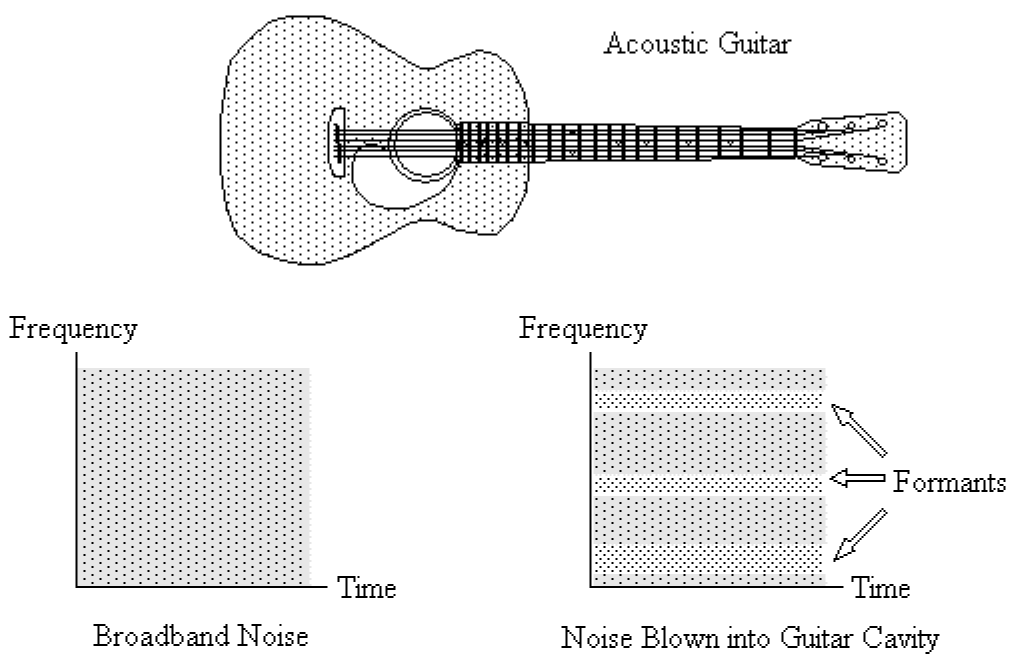
Wolf



Specific bands of frequency are enhanced in many systems such as our voice and musical instruments. A vibrating mechanism produces frequencies. Then some of these frequencies get enhanced by resonances in the system. We consider an acoustic guitar in order to illustrate this. We perform an experiment with our guitar. We do not play it at all since we want to investigate what the cavity and wood do to the sound. We blow white noise into the guitar cavity and analyze the resulting sound. The guitar modifies the white noise. It enhances certain frequency bands due to resonance. The inner cavity of the guitar has more complex resonances of air vibration than a long narrow pipe.

The wood of the guitar also vibrates in response to the noise. The wood resonance vibrations depend on the characteristics of the wood. Certain bands of frequencies of the white noise are amplified. The spectrogram of the final sound is the white-noise spectrum with the enhanced frequency bands. See Fig. U-5. These enhanced regions are called *formants*. The formant regions are determined by the shape and size of the guitar cavity, the type of wood and other materials used in the construction.

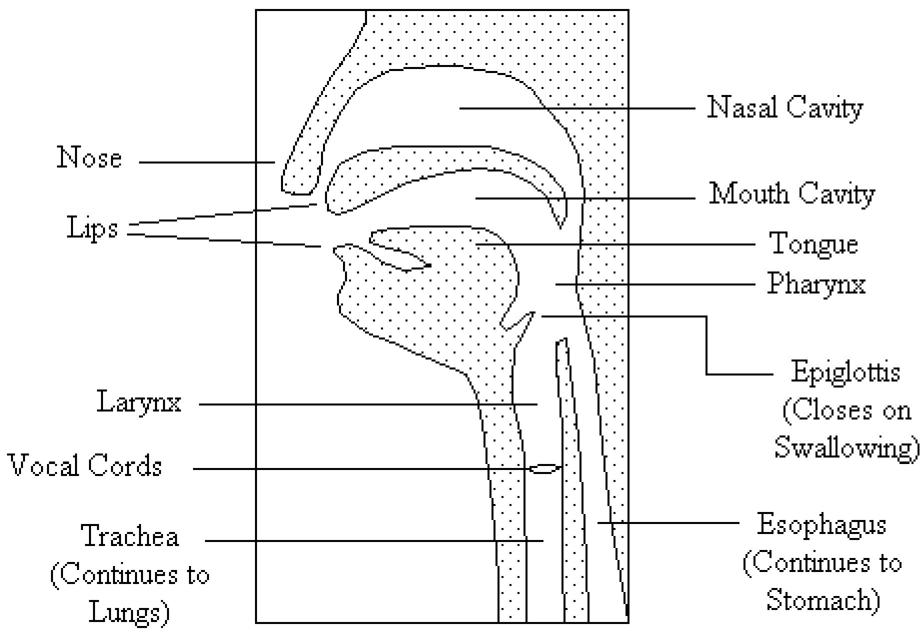
Fig. U-5. Guitar Formants.



During the normal playing of the guitar, the strings are used to produce the initial vibrations of sound. The strings produce periodic tones. Each pitch has a fundamental and associated overtones. Those overtones falling in formant regions of the guitar get enhanced. The tone becomes richer. The final sound is unique to the make of the guitar. A similar analysis can be made with the violin. The final rich spectrum, dependent on the formants, is once again, a unique timbre.

The human voice has formants that vary, depending on the shape of the vocal system. The cavities involved are the larynx, pharynx, mouth, and nasal cavity (see Fig. U-6). These depend on the size of the person. Also, the individual can control a considerable portion of the vocal system in order to change these dramatically. Colds and congestion influence the nasal cavity. The vibrations originate at the vocal cords. The sound is enhanced by the neighboring cavities in a similar way the acoustic-guitar cavities bring out the sound of a vibrating guitar string. Formants are relevant.

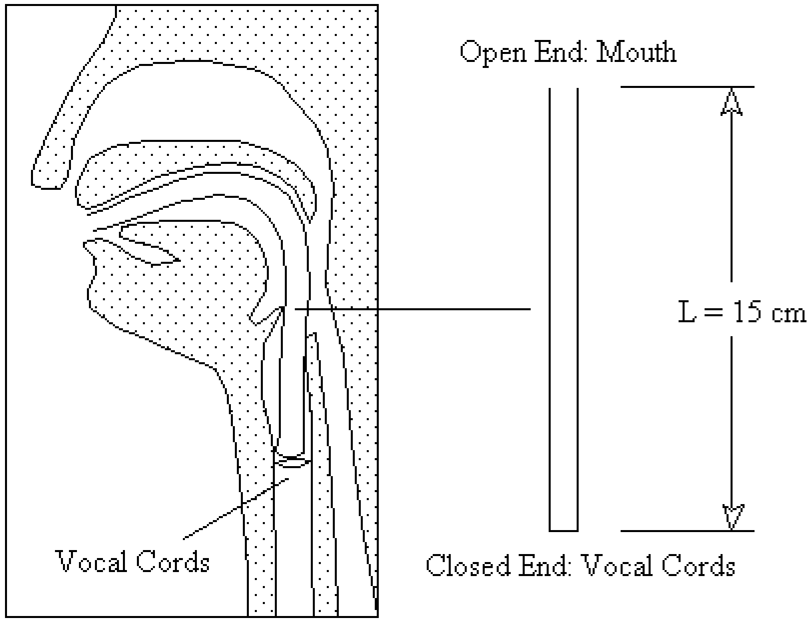
Fig. U-6. Human Vocal System.



We once again turn to biology. It is obvious from Fig. U-6 that the human voice system is quite complex. We would like to have a simple model. Here is where physics comes in. We can take the vocal cords to be the closed end of a pipe, the pipe extending all the way to the mouth opening. See Fig. U-7. Look upward, place your thumb on your Adam's apple and touch your lips with your forefinger. Take a ruler and measure this distance between your thumb and extended forefinger. The result varies from person to person, roughly correlating with height. A typical value is around 15 cm.

A singer can produce a range of pitches with considerable control over the voice system. We can determine very approximately where the formant regions are in the frequency spectrum. The formant frequencies are related to the natural frequencies of vibration in the cavity of the vocal tract. In our model we approximate this tract as a closed pipe. Therefore, the vocal formants will be set by the resonances of a closed pipe. These resonance frequencies are odd harmonics.

Fig. U-7. Modeling the Vocal Track as a Closed Pipe.



The steps for determining the resonance frequencies of the closed-pipe vocal tract are given below. The key issue is finding the fundamental. We know that the resonances are the odd harmonics. Once we know the fundamental frequency "f" as a number, the odd harmonics 3f, 5f, 7f, and so on are then known explicitly. These are the vocal formant regions of our model.

1. The length of the closed pipe. $L = 15$ cm

2. The fundamental wavelength. One quarter wave is fitted to the length. $\frac{1}{4}\lambda = L$

The wavelength is four times the length. $\lambda = 4\ L = 4 \times 15\ \text{cm} = 60\ \text{cm}$

3. The frequency. Wavelength x frequency = speed. $\lambda f = v = 340$ m/s

Express sound speed in centimeters per second since wavelength is in cm (must be consistent). $(60\ \text{cm})\ f = 34{,}000$ cm/s

Divide 34,000 by 60 to find frequency. $f = (34{,}000 / 60) = 3400 / 6 = 1700 / 3$

$f = 1700 / 3 = 567$ Hz (but we will use 500 Hz since our model is very approximate).

The next two modes are then 1500 Hz (3f for H3) and 2500 Hz (5f for H5).

4. Vocal Formants. The first three vocal formants are 500 Hz, 1500 Hz, and 2500 Hz.

Vocal formants depend very much on the shape of the cavities in the vocal system. For the ideal closed narrow pipe, we find approximately 500 Hz for the first vocal formant. An adult's first vocal formant can vary a couple hundred hertz either way from 500 Hz, depending on the sound produced. The second formant can vary over an even wider range. Since vocal formants are dependent so much on the sounds we pronounce, the best way to determine vocal formants in specific cases is from spectrograms. Some professional singers have been found to have a strong formant region between 2500 and 3000 Hz. This formant is called the *singer's formant* or *singing formant*. It assists in projecting the voice. Is it a coincidence that trained singers can achieve this formant, which happens to be right where our ears are most sensitive (consulting the Fletcher-Munson Curves)?

Simplified spectrograms are given in Fig. U-8 for six vowels spoken with a steady voice at a fundamental of 150 Hz. The steady voice producing a vowel sound is virtually a periodic tone. We can then analyze the spectrum in terms of harmonics. The equally-spaced horizontal lines are the harmonics H1, H2, H3, etc. for the 150-Hz wave. They have frequencies 150 Hz (1f), 300 Hz (2f), 450 Hz (3f), 600 Hz (4f), etc. The spacing between harmonics is simply 150 Hz since each time you rise to the next harmonic, you add another 150 (i.e., f). Note the richness of the overtones. Some of the sounds have over 20 harmonics.

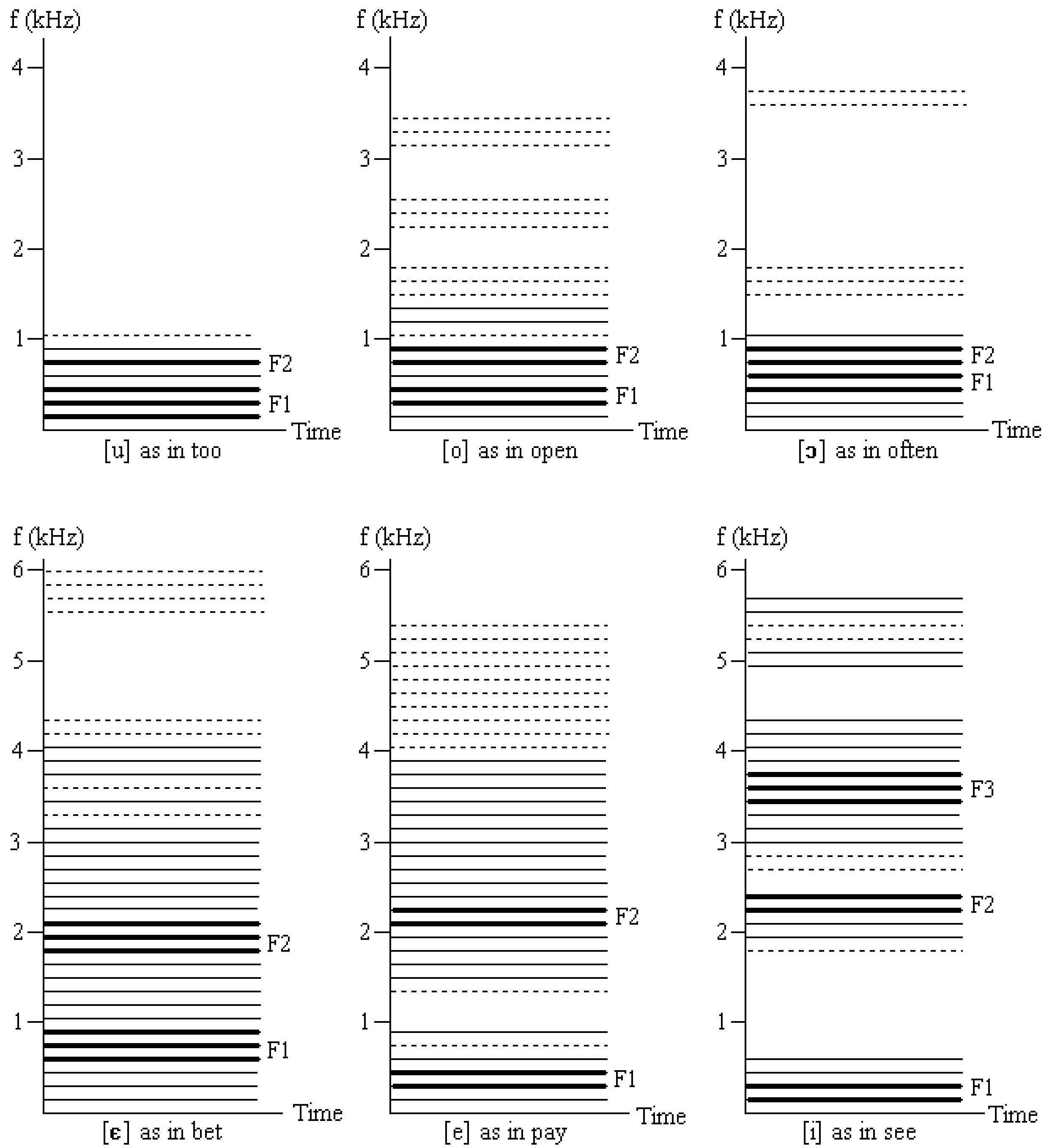
The more intense harmonics are darker. Sometimes a harmonic is not intense enough to see. The formants are indicated as F1, F2, etc. These are enhanced harmonics due to the shape of the internal cavities of the vocal system. The first enhanced region is called the first formant region and so on. There is no correlation in many cases to the formants of our ideal pipe since the voice can change formants by inner shaping of resonance cavities.

However, note that the [i] sound has formant regions near 500 Hz, 2500 Hz, and 3500 Hz. These correspond to formant regions of our model (the 1st, 5th, and 7th harmonics of the simple closed-pipe model of the voice system). But the formant corresponding to the 3rd harmonic (1500 Hz) has been suppressed. The 3rd harmonic has been filtered out by the voice system. The shape of the system for this vowel makes it difficult for the 3rd harmonic to resonate. The 2nd formant region (F2) for [i] is near the 5th harmonic of the "closed-pipe" (2500 Hz), which is close to H16 of the 150-Hz wave being produced (making H16 and also H15 more intense). We know 2500 Hz is close to H16 because the frequency of the 16th harmonic is 16 x 150 Hz = 2400 Hz.

The symbols employed as pronunciation guides for the sounds in Fig. U-8 are from the International Phonetic Alphabet (IPA), devised in the late 1800s. The IPA consists of basic units of speech called *phonemes* (FOE-neams) that can be put together to indicate how any word in almost any language is pronounced. The IPA is used by teachers of the hard-of-hearing, speech pathologists, actors, singers, radio and TV announcers, and others. A singer armed with the IPA can know how to sing a phrase in a foreign language. The IPA enables you to pronounce words in other languages you don't understand, if you have mastered the basic sound units (about 40) of the IPA.

The data in Fig. U-8 comes from doctoral research in an area where psychology, physics, and medicine converge. You should consider it an achievement that you can understand and appreciate such advanced research data.

Fig. U-8. Spectrograph Sketches for Six Vowel Sounds Spoken with a Steady Voice.

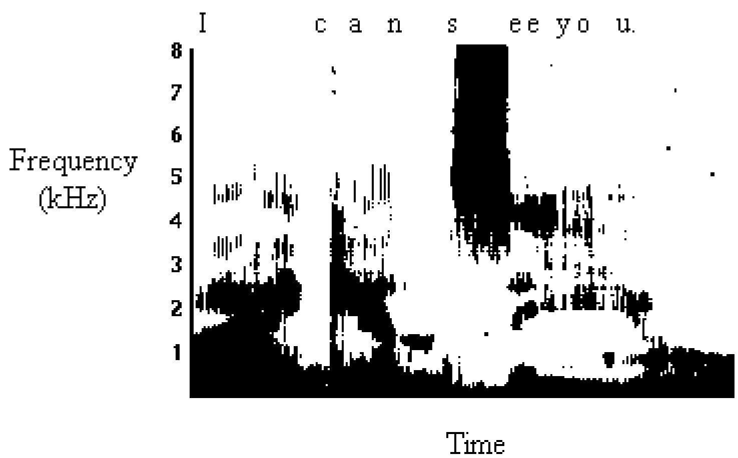


Sketches are adapted from spectrograms found in Colin Painter, *An Introduction to Instrumental Phonetics* (University Park Press, Baltimore, 1979).

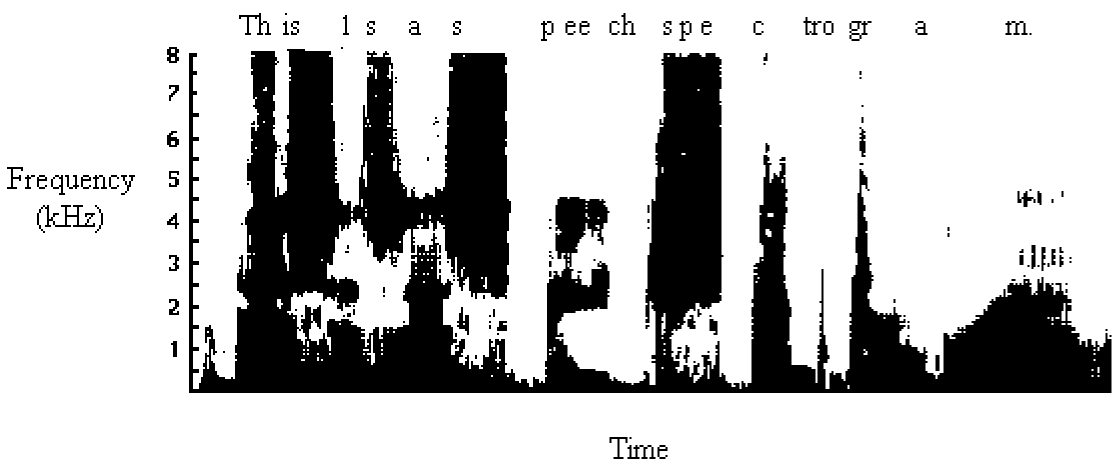
Two speech spectrograms can be seen in Fig. U-9. The pictures are examples of those produced by a sound *spectrograph*. The spectrograph analyzes the sound by playing it over and over again in order to scan the frequencies up to 8 kHz. A drum spins and an ink pen records the presence of frequencies on a paper rolled on the drum. When finished, the unwrapped paper becomes the spectrogram.

Fig. U-9. Speech Spectrograms.

Spectrogram of the spoken sentence: I can see you.



Spectrogram of the spoken sentence: This is a speech spectrogram.



Reference: F. Alton Everest, *Acoustic Techniques for Home and Studio*, 2nd ed. (TAB Books, Blue Ridge Summit, PA, 1984).

--- End of Chapter U ---