8. ANALYSIS OF COMPLEX SOUNDS AND SPEECH ANALYSIS

Amplitude, loudness, and decibels

A complex sound with particular frequency can be analyzed and quantified by the amplitude spectrum: the relative amplitudes of the harmonics. Our goal today is to understand how the pitch, loudness, and timbre of a sound are represented in a spectrum. However, rather than use an analog Band Filter, we will use the ***PRAAT*** program, which has powerful analysis software built in to calculate and graph the spectrum of a recorded sound.

A. "Warm up": Analysis of a measured sine wave

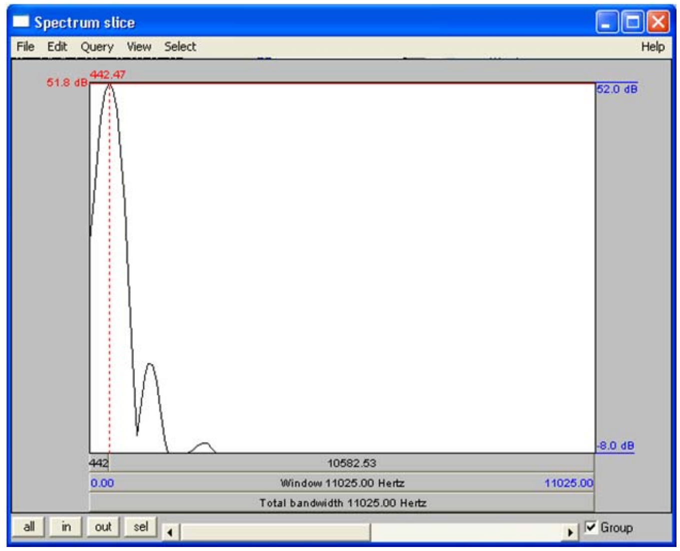
Before using ***PRAAT*** to analyze a more complex wave, first analyze a simple sine wave to better understand frequency spectra. Take the output from the frequency synthesizer ground (black socket) to the headphone box ground input (also black socket). Use another banana plug cable from the top red 8 Ω output socket of the synthesizer to the other input on the headphone box.

Use the synthesizer to generate a sine wave (use oscillator 1-Left, change the function using the switches at the top left corner of the synthesizer box; make sure to switch all other oscillators out).

1. Send the signal from the headphone box to the microphone input of the computer using the cable supplied. When you start the program, two windows should have appeared: a **Praat objects window** and a **Praat picture window**. Recall how you record a sound with ***PRAAT***: choose **Record mono Sound...** from the **New menu** in the Praat objects window. A **SoundRecorder** window will appear on your screen.
2. Use the **Record** and **Stop** buttons to record a few seconds of your sine wave signal. You can repeat this several times until you are satisfied with your levels. Adjust the gain of the synthesizer output so that the level meter of the SoundRecorder window registers and is "green", or occasionally "yellow". Type in a name for your sound file (e.g., "sine") in the text box below the **Save to list:** button. hit the **Save to list**: button and the text string "sound sine" should appear in the **Praat objects window** that indicates the file where your sound is recorded.
3. With the name of the file highlighted in the Praat objects window, hit the **Edit** button to see the waveform that you just recorded. To simplify the display (if it is not already done), go across the menu of the Edit window (that should be titled "Sound sine"),
   - **Spectrum** menu, pull down, **Show spectrogram** deselect (i.e., if a check mark is shown, select it. If no check mark, exit without doing anything).
   - **Pitch** menu, pull down, deselect **Show pitch**
   - **Intensity** menu, pull down, deselect **Show intensity**
   - **Formants** menu, pull down, deselect **Show formants**
   - **Pulses** menu, pull down, deselect **Show pulses.**
4. Set up by going to the **Spectrum** menu, going to **Advanced Spectrum settings**..., and in **Pre-Emphasis** (dB/out) type "0.0" if it is "6.0". This latter option will be used when we analyze speech in the next lab.

5. Check that you have a clean sine wave by using the **View menu** and combinations of **Zoom in** (<CTRL>-i), **Zoom out** (<CTRL>-o), or selecting and **Zoom to selection**.

6. Now generate a frequency spectrum of this input waveform. From the **Spectrum menu**, choose **View spectral slice**. A **Spectrum Slice window** should appear as shown in the graph below.



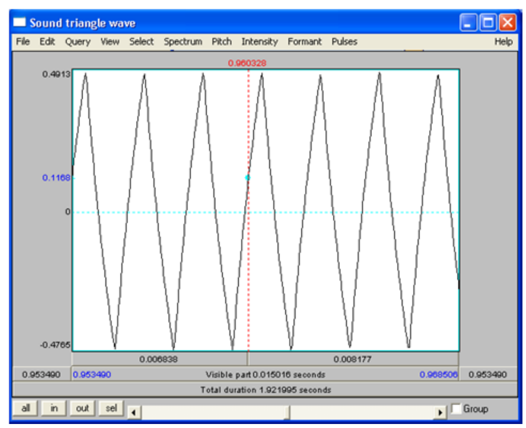
Frequency spectrum of a sine wave.

This frequency spectrum shows frequency in Hz along the horizontal axis and power of that frequency component on the vertical axis.

Use the cursor to click at the tip of the first frequency peak (the fundamental). The frequency of the peak is given as a red number at the top of the display (in this case, 442 Hz, very close to the 440 Hz sine wave expected from the synthesizer). Its power level is given as a number at the left hand side. Note that the frequency spectrum of a pure sine wave is dominated by the power of the frequency of the sine wave. The other harmonics simply show that this is not a perfect sine wave; an ideal one would have *only* the first peak.

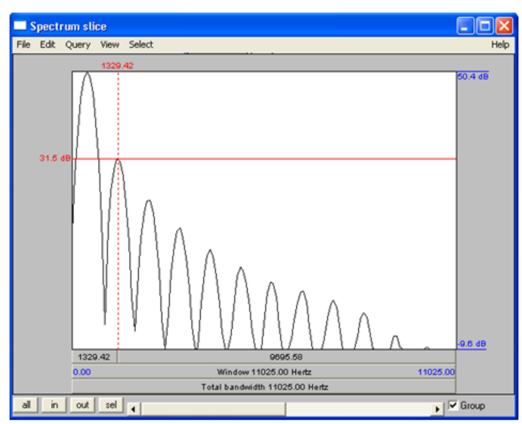
B. Analysis of a measured triangle wave

Now use the synthesizer to generate a triangle wave (use oscillator 1-Left, change the function using the switches at the top left corner of the synthesizer box; switch other oscillators out).



a) Follow the same procedure as with the sine wave to record the wave with ***PRAAT.*** In the window obtained from using Edit, use the **Spectrogram menu, View spectral slice**, to generate a power spectrum of the triangle wave. It should appear similar to the example below.

Frequency Spectrum of a triangle wave.



b) Notice that the frequency power spectrum is made up of a series of peaks. Each peak represents one of the harmonics. You can use the mouse to click at the tip of each peak. The number above the display shows the frequency and the number to the left the power (in decibels) for that frequency. In the example above, the frequency peak of 1329 Hz has a power of 31.5 dB. **Record** the frequency and power for each peak in the table below. To obtain the last line of the table, you need to subtract the power level of the fundamental (**P**$_1$) from each harmonic power level (**P**$_n$).

Table 3: Measured power spectrum for a triangle wave

	Harmonic								
	1	2	3	4	5	6	7	8	9
frequency (Hz) = fn									
Harmonic number=fn/f1									
Power (dB) = Pn									
Relative Power (dB) = Pn - P1									

c) How do your **relative** power numbers compare to what you calculated in the Table 2 of this PRELAB?

d) Return to the initial **Sound object window** obtained from the **Edit** button. Zoom in to a few cycles of the triangle wave. Use the mouse click on a peak and an adjacent valley in order to measure the peak-to-peak amplitude of the triangle wave.

_______________________ Pascals.

e) Now reduce the amplitude of the oscillator; record the sound again using ***PRAAT***. Measure the peak-to-peak amplitude of the sound signal:

_______________________ Pascals.

Record the power levels of all of the harmonics in the table below.

Table 4: Measured power spectrum for a triangle wave, new amplitude

	Harmonic								
	1	2	3	4	5	6	7	8	9
frequency (Hz) = fn									
Harmonic number=fn/f1									
Power (dB) = Pn									
Relative Power (dB) = Pn - P1									

f) Compare your results to those from Table 3:

Do all of the power levels change?

By how much do they change?

Do they change by different amounts or the same amount?

Do the relative power levels change?

g) The change in the power levels for ***each harmonics*** present should be the same as the change in the overall power level, which you can calculate from your measurements of the **new** and **old** amplitudes.

Change in Power (dB) $= 10 \log_{10}[(\mathbf{A_{new}}/\mathbf{A_{old}})^2]$

Are your measurements consistent with this observation?

C. Analysis of a superposition of a triangle wave and another harmonic, not present in the triangle wave

Now switch in oscillator **4** (summing it to the triangle wave). Turn its amplitude halfway up of the oscillator **1** level (make this adjustment by recording signal for oscillator 4 first, before summing the two signals. You will probably have to go through several cycles of recording, adjusting amplitude, recording, adjusting amplitude, etc.). Note that this harmonic was not present in the triangle wave signal

a) Record the sound signal. How do the sound signal and the power spectrum change?

b) Change **the phase** of oscillator **4** by **180°** and record the signal again. Does the sound change? Qualitatively, how do the sound signal and power spectrum change?

D. Analysis of a superposition of a triangle wave and another harmonic, already present in the triangle wave

Now switch out oscillator **4** and switch in oscillator **3**. Adjust its amplitude (by recording the signal separately first) to be approximately half of that from oscillator 1-Left. Note that this harmonic **was** present in the original triangle wave.

a) Record the sound signal. Qualitatively, how do the sound signal and the power spectrum change?

b) Change the **phase** of oscillator **3**. Does the sound change? Record the signal again. Qualitatively, how do the sound signal and power spectrum change?

E. Discussion of expected effects.

If the Power Spectrum of the sound changes, you can probably hear the difference. If the Power Spectrum does not change, you can probably not hear the difference. For example, you should have found that changing the phase of the fourth harmonic did not change the power spectrum, but changing the phase of the third harmonic did matter. Why? The third harmonic was present in the original sound. Phase *may* matter when you add sounds that are close together (or the same) in frequency.

Switch Oscillator 1-Left back to a sine wave.

a. Now switch in or out additional oscillators, one at a time, listen to the differences and record the signal after each switch. The sound should change every time, and new peaks should appear or peaks should disappear in the Power Spectrum. If you see no change, the amplitude for that oscillator is probably close to zero.

b. Changing the phases of any of the oscillators should not change the sound or the power spectrum. The lesson is that the shape of the power spectrum determines the timbre.

We can now give very precise ways to define how to measure sounds made up of a harmonic series using the Power Spectrum:

Pitch: the frequency of the fundamental of the harmonic series **(f_1)**.

Timbre: the relative heights of peaks in the power spectrum (compared to the fundamental)

Loudness: the overall heights of the peaks in the power spectrum

SPEECH ANALYSIS

In the previous section of this lab we found that we could analyze a complex sound with particular frequency using the power spectrum. We gave very precise ways to define how to measure sounds made up of a harmonic series using the power spectrum:

In this part of the lab we will try to better understand the acoustical features of speech. Speech sounds, called **"phonemes"** are classified either as **vowels** or **consonants**. Frequency analysis of a vowel sound always reveals a clear harmonic spectrum, with a pitch (fundamental frequency) varying from individual to individual, with the pitch of female speakers being on the average twice as high as that of male speakers. The pitch is determined by the vibration of vocal cords.

The features that distinguish specific **vowels** are called **formants**. Formants are vocal tract resonances, the frequency of which depends on such effects as the tongue height or tongue advancement. The first two formants are particularly important in speech recognition. The frequency of the first formant increases as we open our mouth wider and lower the tongue.

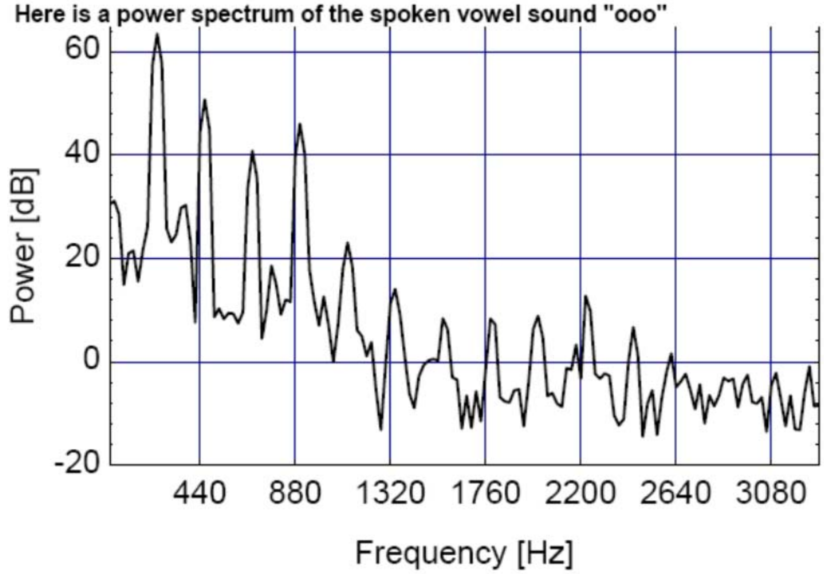
The frequency of the second formant increases as we advance our tongue (see positions of formants for selected vowels in the attached figures pp19-21.). Frequencies of formants change only within 15% between female and male speakers.

Most **consonants** do not have harmonic frequency spectra. The features that distinguish consonants are periods of silence, voice bars, noise, and the consonant's effects on the frequency spectra of adjacent vowels. Consonants are classified by: **(i)** manner of articulation, **(ii)** place of articulation, and **(iii)** as voiced or unvoiced. The consonant types, classified by manner of articulation, include:

1. **plosive or stop** (p, b, t, d, k, g) – produced by blocking the flow of air somewhere in the vocal tract
2. **fricative** (f, s, sh, h, v, th, z) – produced by constricting the air flow to produce a turbulence
3. **nasal** (m, n, ng) – produced by lowering the soft palate
4. **liquid** (l, r) – generated by raising the tip of the tongue
5. **semi-vowel** (w, y), always followed by a vowel.

A. Looking at a power spectrum of a spoken sound

Figure 1:
Here is a power spectrum of the spoken vowel sound "ooo"



The power spectrum is quite complicated; nevertheless, we can spot some of the same features as those found in the spectrum of a simpler sound (such as a triangle wave):

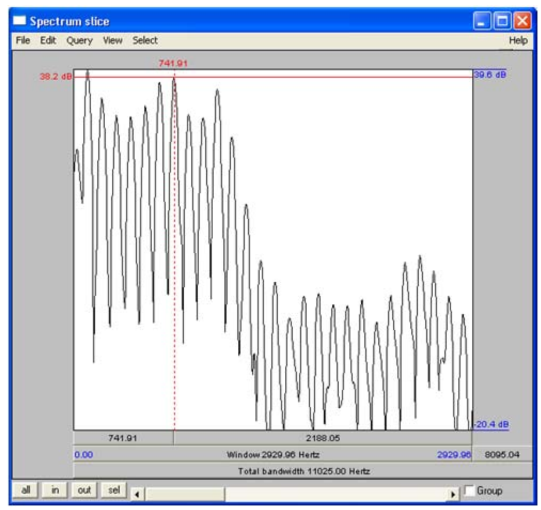
a. **Identify** on the graph the position of the peak corresponding to the fundamental (hint: it is the biggest peak on the graph); label it as "1." Note that the horizontal position of the peak shows the frequency corresponding to the peak. The fundamental peak is about halfway between 0 and 440 Hz; therefore, a good estimate for the frequency of the fundamental would be 220 Hz.

b. Now label the other peaks (**important**: label only the peaks that correspond to **harmonics** of the fundamental frequency - ignore the small peaks before the first big peak or between the other big ones). You should be able to find 11 distinct harmonic peaks. What are the **prominent harmonic numbers** for this sound? (with peak heights above the 0 dB line).

c. Generally, the lowest harmonics have higher power. Do you notice that, **some** higher harmonics have greater power than some lower harmonics? Give some examples:

d. We will now use ***PRAAT*** to record and analyze your own vowel sound:

- Use the microphone as input to the computer. Following the same instructions as in the previous lab, record a few seconds of someone in your group singing the continuous vowel "**aw**" (the sound of "a" in "call").
- Use the Zoom controls as before in zoom in to see a number of consistent cycles with repeating shape. Drag and select an approximate 0.1 seconds of the sound. In the **Spectrum** menu, choose **View spectral slice**.
- In the **Spectrum slice window**, click and start dragging at the left and select the approximate region of 0 to approximately 3000 Hz. From the **View** menu, use **Zoom to selection** to magnify the interesting frequency region. Below is shown an example frequency spectrum of someone saying the vowel "**aw**":



Frequency spectrum of the vowel "aw".

What is similar about your **"aw"** power spectrum and mine? What is different?

B. Quantifying the power spectrum of a spoken sound Formants.

We want to understand what makes each sound identifiable. We know that different periodic sound signals have different harmonic power spectra. Let's look at one way to quantify the power spectrum.

The sound power **measured** by the microphone is the **output of the vocal cords** as **filtered through the rest of the vocal tract and as transmitted out of the mouth.** The net output of the vocal cords (called **"Source"** in the table) represents the combined effects of the vocal cords and the transmitted efficiency of the mouth:

This Source spectrum resembles triangular wave form, which diminish in amplitude $1/n^2$, i.e. −12dB/octave. The mouth radiates more efficiently at high frequencies. This increases the transmission power by 6dB/octave. The resulting change is

6 dB per octave {net} = **–12 dB per octave** {vocal cords} + **6 dB per octave** {mouth}.

Therefore (other than telling us the pitch), the Source spectrum does not tell us much about what distinguishes the different sounds of speech. What changes when we talk is the **filter function**. This is a function, representing relative power changes as a function of frequency. It describes how we modify the sound produced by the vocal cords. The filter function changes as we shape our mouth, adjust tongue position etc to pronounce a given sound. In particular, a sound corresponding to a specific vowel is characterized by a specific filter function that changes little from one individual to another. A filter function for a given vowel, plotted as a function of frequency, exhibits characteristic maxima, called **formants**.

The filter function for your sound can be approximated by subtracting the "Source" from "Power" for each harmonic:

Filter function (dB) = Observed Power (dB) – Source Power (dB)

The ***PRAAT*** program can be set up to do this automatically for you. **Set it up**: in the **Spectrum menu**, go to **Advanced spectrogram settings...**, and under **Pre-emphasis (dB/oct.)**, type in "6.0" if it is not already at this value. Re-analyze your "**aw**" vowel with this setting.

Table 1a: Here is a sample table for the example spoken "aw" sound:

Harmonic No.→	1	2	3	4	5	6	7	8	9	10	11	12
Frequency [Hz] = f_n	105	210	315	420	525	630	735	840	945	1050	1155	1260
Power [dB] = P_n	39.6	34.5	31.0	31.6	33.0	37.2	38.4	32.0	31.6	36.1	28.1	17.1
Relative Filter Magnitude [dB] = $P_n - P_1$	0.0	-5.1	-8.6	-8.0	-6.6	-2.4	-1.2	-7.6	-8.0	-3.5	-11.5	-22.5
Approx. Formant Position (peak)							F_1			F_2		

Now Use the mouse and cursor to measure the power of each harmonic in your "**aw**" sound.

a) First determine the fundamental frequency for your sound by moving the cursor by clicking the mouse at the tip of the first large peak in the spectrum. The red numbers that appear show the frequency and power (in dB) for that peak. Record these numbers in the first column of the table following.

b) Now move the cursor so that it is aligned with the next big peak in the spectrum. Make sure that the peak really corresponds to a harmonic of the fundamental (the frequency should be very close to a whole-number multiple of the fundamental frequency that you recorded from part a.) Record the frequency and power of each peak in the table following:

Table 1b: Measured power spectrum for your "aw" sound

Harmonic No.→	1	2	3	4	5	6	7	8	9	10	11	12
Frequency [Hz] = f_n												
Power [dB] = P_n												
Relative Filter Magnitude [dB] = $P_n - P_1$												
Approx. Formant Position (peak)												

To get the relative filter magnitude, subtract the filter magnitude of the first harmonic from the filter magnitude of the other harmonics. Scan across the row of "Relative Filter Magnitude", and "peaks" or areas of enhancement should be evident. The points of greatest enhancements can be identified as the approximate positions of formants (and in the example of Table 1a, are observed roughly at harmonic no. 7 and 10).

This method of obtaining the filter function is only an approximation because we are only obtaining the function at a small set of frequency values (the harmonics). The true filter function is a continuous function of frequency; we need to draw a smooth line between the points in our approximation.

As mentioned earlier, the peaks in the filter function are called **formants**. Two are apparent in Table 1a (at harmonic numbers 7 and 10). We can compare our numbers to those in Rossing's *The Science of Sound* [shown in brackets].

Table 1c: Formants obtained from the Power Spectrum of spoken "aw" sound:

Formant number	Frequency	Relative Magnitude (dB)
F_1	735 [730]	-1.2 [-3]
F_2	1050 [1090]	-3.5 [-5]

The ***PRAAT*** program will also calculate the frequency locations of these formants by performing a fit to the filter function across all frequencies. Click anywhere in your waveform window (instead of selecting), and under the **Formant menu**, choose **Formant listing**. The first five formants should be given. Table 2a gives the formants calculated by ***PRAAT*** for the "**aw**" sound analyzed in Table 1a. These differ slightly from the numbers in Table 1a, but they were much easier to obtain! For bandwidth (the "width" of each formant peak; some are wider than others), use the **Formant menu**, **Get bandwidth**..., and fill in for which formant you want it calculated.

Do the same for your own recording of your "**aw**" sound in Table 2b.

Table 2a: Formants for spoken "aw" sound, corresponding to the power spectrum recorded in Table 1a

	Formant number		
	F_1	F_2	F_3
Frequency [Hz]	625	1049	2560
Bandwidth [Hz]	112	91	71
Relative Magnitude [dB]	-1.0	-5.0	-28.0

Table 2b: measured formants of your "aw" sound, corresponding to your power spectrum recorded in Table 1b

	Formant number		
	F_1	F_2	F_3
Frequency [Hz]			
Bandwidth [Hz]			

How do these peaks compare with your numbers from the power spectrum (Relative Filter Magnitude line) recorded in Table 1b?

3) **Try a few more vowel sounds** (recommend **"ee"** as in "heat", **"u",** as in "blue", "**i**" as in "bit", etc) . Record your data in the tables below. Record formant frequencies only – let the program do the calculations for you. **Compare** your results to the "averages" found in the attached Table 15.3 from Rossing's textbook. **Write** your comments (does it agree/disagree with the "averages") next to the relevant Table.

Table 3: measured formants of your /i/ ("ee"), as in "heat", sound

	Formant number		
	F_1	F_2	F_3
Frequency [Hz]			
Bandwidth [Hz]			

Table 4: measured formants of your /u/ , as in "blue", sound

	Formant number		
	F_1	F_2	F_3
Frequency [Hz]			
Bandwidth [Hz]			

Table 5: measured formants of your "…….." sound

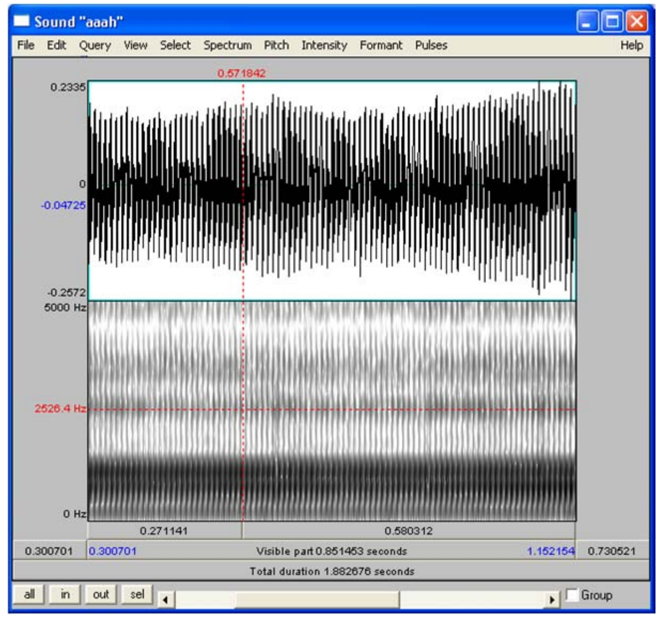
	Formant number		
	F_1	F_2	F_3
Frequency [Hz]			
Bandwidth [Hz]			

Table 6: measured formants of your "…….." sound

	Formant number		
	F_1	F_2	F_3
Frequency [Hz]			
Bandwidth [Hz			

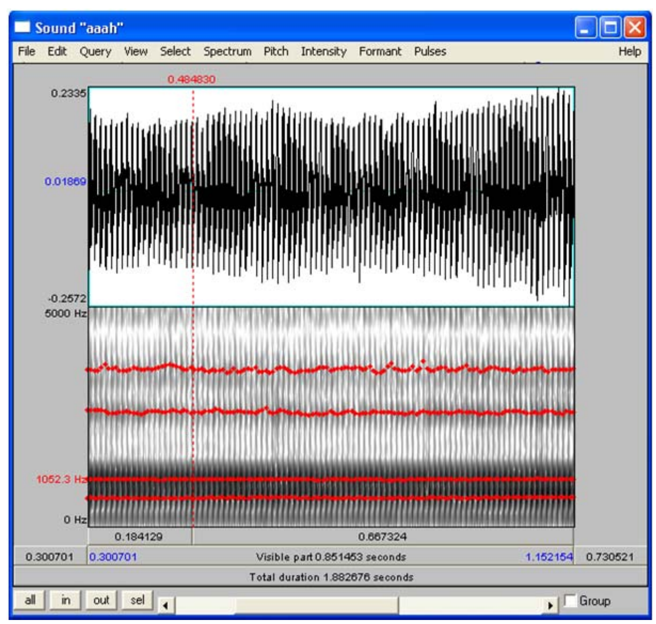
C. Vowel/Consonant Recognition using Spectrograms (optional)

We can also the ***PRAAT*** program to plot spectrograms of our waveforms, that is, frequency plotted versus time with the intensity or level of grey scale representing the relative power of contributing frequencies. Start with recording several vowels, e.g. "ooo","ee","aw","u". In the window obtained with the **Edit** button, under the **Spectrum** menu, select **Show Spectrogram**. A frequency spectrogram should show up below the waveform as shown below. You can zoom in and out as before.



Spectrogram of the vowel "aw".

Regular dark, horizontal bands, representing contributing harmonics should be clearly visible. In the example above, the cursor in centered at approximately the third formant at 2560 Hz. The power spectrum (spectrogram structure) should look different for different vowels. To identify different parts of the spectrogram click and drag mouse on the waveform display to select a portion of the recording, and then play back the selected part. In addition, for the recorded vowels, you can also display their formants. To do this, under the **Formant menu**, select **Show formants** and they will appear as red dots overlaid on the frequency spectrum. In some cases, they will change slightly with time, but generally should be relatively constant for vowels. The figure at right shows frequency spectrogram of spoken "aw" vowel, with formants overlaid.



Record the spectrograms and formants of several vowels and then print them out. Unfortunately for these spectrogram menus, printing is not directly supported. The way around it:

- Open Microsoft Word.
- Go back to the window showing the spectrogram you would like to capture and make sure it is "active" by clicking on it.
- Do a "screen shot" of that window by <ALT>-Print Screen.
- Go back to Microsoft Word, and under the **Edit** menu, select **Paste,** and the spectrogram should now be inserted in to the Microsoft Word document.

Collect the spectrograms of a number of vowels in the same Word document, print it out, and identify the recorded vowels on each plot.

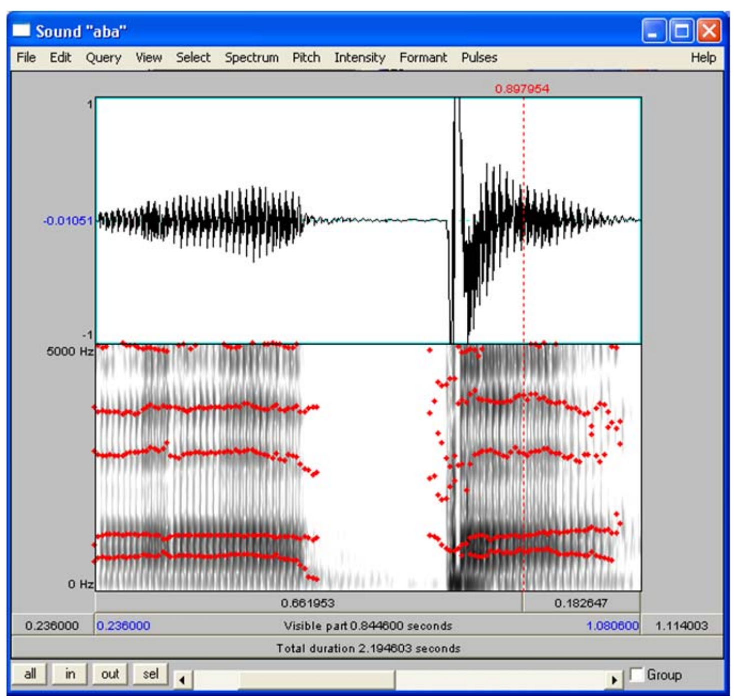
Consonants also have distinctive features that define them on a spectrogram. These do not include formants like for the vowels, unless a consonant is combined with a vowel and one observes formant transition. Formant transition is a formant frequency change associated with phonetic transcription moving from a vowel to a consonant. **Consonants** are classified by manner of their articulation. This manifests itself by such features of a spectrogram as: ***periods of silence, burst, noise, and presence of voice bars***.

Consider ***stop type consonants*** as an example. The spectrogram should reveal a period of silence, followed, right after, by a powerful burst. Finally, a detection of voice bars would indicate whether the stop sound is a voiced consonant or not.

Fricative consonants are characterized by noise, recognized on a spectrogram by its aperiodic and irregular pattern (i.e., "ssssh" or "sss" should fill all frequencies). ***Affricates*** (a combination of stops and fricatives) are described as having a brief period of silence, followed by a weak burst, and some noise.

Record and **make spectrograms** of different consonants pronounced between two “a” vowels. Zoom into the relevant part of the waveform and spectra in each case. Suggested combinations are listed in the Table below. **Print** two or three of the relevant graphs (using the procedure described earlier) above as an illustration. **Identify** and **mark** on the spectrograms features like periods of silence, bursts, voice bars and **list** them in the Table below (see an example for “/aba/”, with an example spectrogram shown below).

Consonant type	Voiced	Voiceless
Stops	/aba/ = "boy" silence, voice bar, burst	/apa/ = "pot"
Fricatives	/aza/ ="measure"	/asa/ = "sheet"
Affricates	/aja/ = "judge"	/aca/ = "church"



Frequency spectrogram with formants overlaid of the vowel-consonant combination "/aba/".The first "a" vowel is clearly seen, then a period of silence with a faint voice bar, a sharp burst, and then the second "a" vowel.