Guide for measuring stops/affricates

First of all, each token should look something like this:

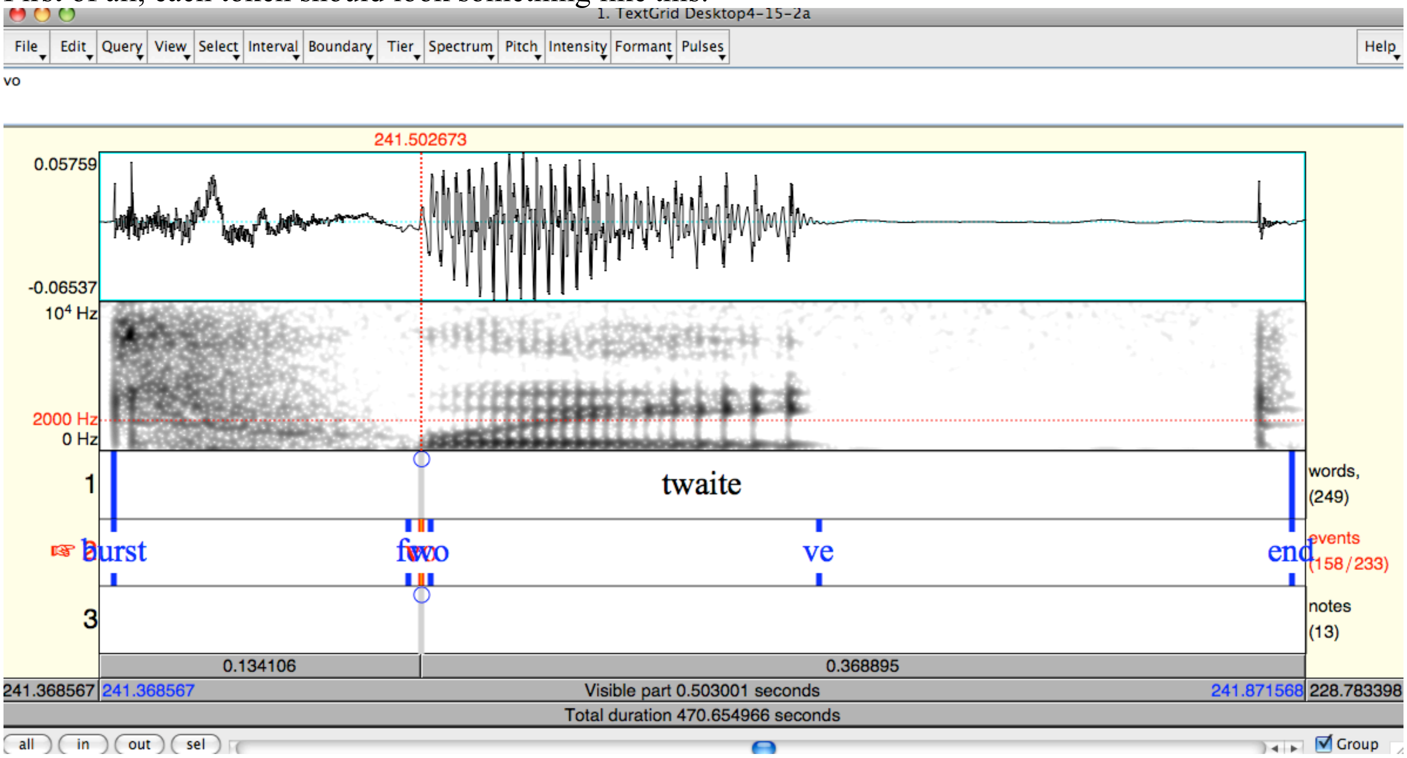

Figure 1: proper placement of points in event tier.

with a point in the events tier for each:
<burst> (at the beginning of the stop/affricate, which is usually a burst), <fe> (frication/aspiration end), <vo> (voice onset), <wo> (/w/ onset, where we can see formants), <ve> (vowel end), <end> (end of word)

I. Duration: The easiest measurements to take are duration. Simply select two points and calculate the time difference between them.

<u>Word duration</u>: Select <burst> (by clicking on the boundary), Get cursor (in the Query menu as shown in the figure below). Select <end>, Get cursor. Subtract the burst time from the end time.

<u>Stop/Affricate VOT/duration:</u> Stops are usually measured in terms of Voice Onset Time, that is how long after the burst voicing begins, while affricates may be measured by duration. Things get really different when dealing with voiced affricates/stops, but luckily, we are looking at voiceless ones so both stops and affricates can be measured by calculating the difference between the burst and the onset of voicing (<vo>). Select <burst>, Get cursor. Select <vo>, Get cursor. Subtract the burst time from the <vo> time.

<u>Vowel duration:</u> We will want a way of normalizing our affricate duration based on how long the vowel is following it. The idea here is that if somebody is talking really fast, both the vowel and the affricate will be shorter, but if they are talking more slowly, or articulating more carefully, the vowel and affricate will lengthen. Select <wo>, get cursor. Select <ve>, get cursor. Subtract the <wo> time from the <ve> time.

<u>Relative duration:</u> Divide the stop/affricate duration by the fricative duration. A larger proportion means a longer stop/affricate, and a smaller proportion means a shorter one, relative to the length of the vowel.

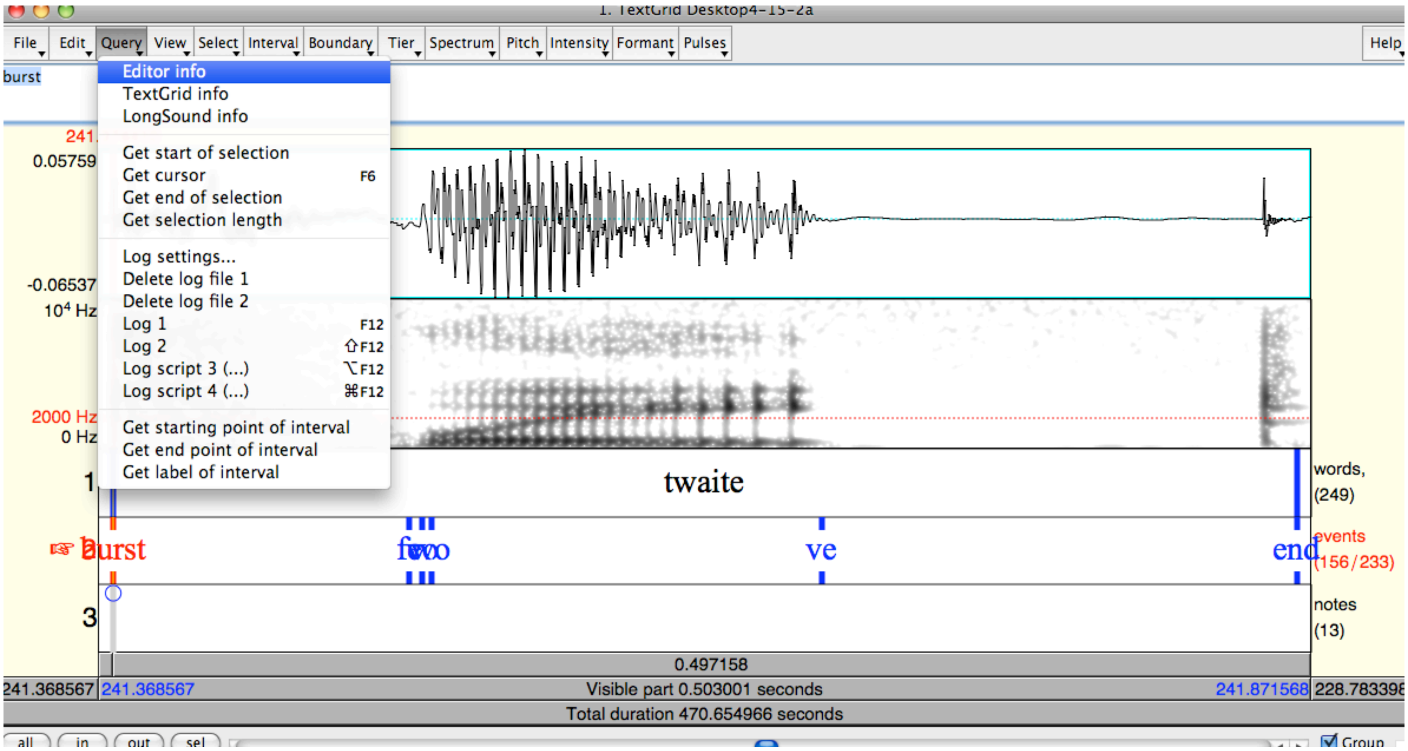

Figure 2: Choose Query > Get cursor to find the time point for each relevant event.

II. Intensity: Next we want to figure out how much energy is in the affricate/stop, especially compared to how much energy is in the vowel. The idea is that an affricate will have greater energy than a stop. The sonority hierarchy also works for intensity. That is, the more sonorous a sound is, the more energy is in that sound. Think about if you want to yell something, like OW! Then think about trying to "yell" Shhh! Which one can be louder? The one composed of vowels.
Sonority hierarchy: Vowels > Approximants > Fricatives > Affricates > Stops

<u>Affricate/stop intensity:</u> Select the interval between <burst> and <fe> (Take the time point for <burst> from your duration measures, and enter this as the first value, and find the time point for <fe> and enter it as the second value into the query box that is created from selecting the Select button and choosing Select… as in the figures below). Then choose Intensity (make sure that Show intensity has a check mark next to it) and choose Get intensity, as in Figure 5. This will get you the average intensity for the highlighted area.

<u>Vowel intensity:</u> Select the interval between <wo> and <ve>, Get intensity.

<u>Relative intensity:</u> Divide the affricate/stop intensity by the vowel intensity. Usually this will result in a number less than zero, giving a proportional difference in intensity between the vowel and affricate.

<u>/w/ intensity:</u> Select the point marked <wo>, Get intensity. This will give us an idea of how powerful the onset of the /w/ is.

<u>Relative /w/ intensity:</u> Divide the affricate/stop intensity by the /w/ intensity. I'm not sure what to expect here, but it seems reasonable to think that the less intensity a stop/affricate has, the more force will be left in the lungs for the onset of voicing, which would mean a smaller proportion than if a strong affricate makes a weaker voicing onset.

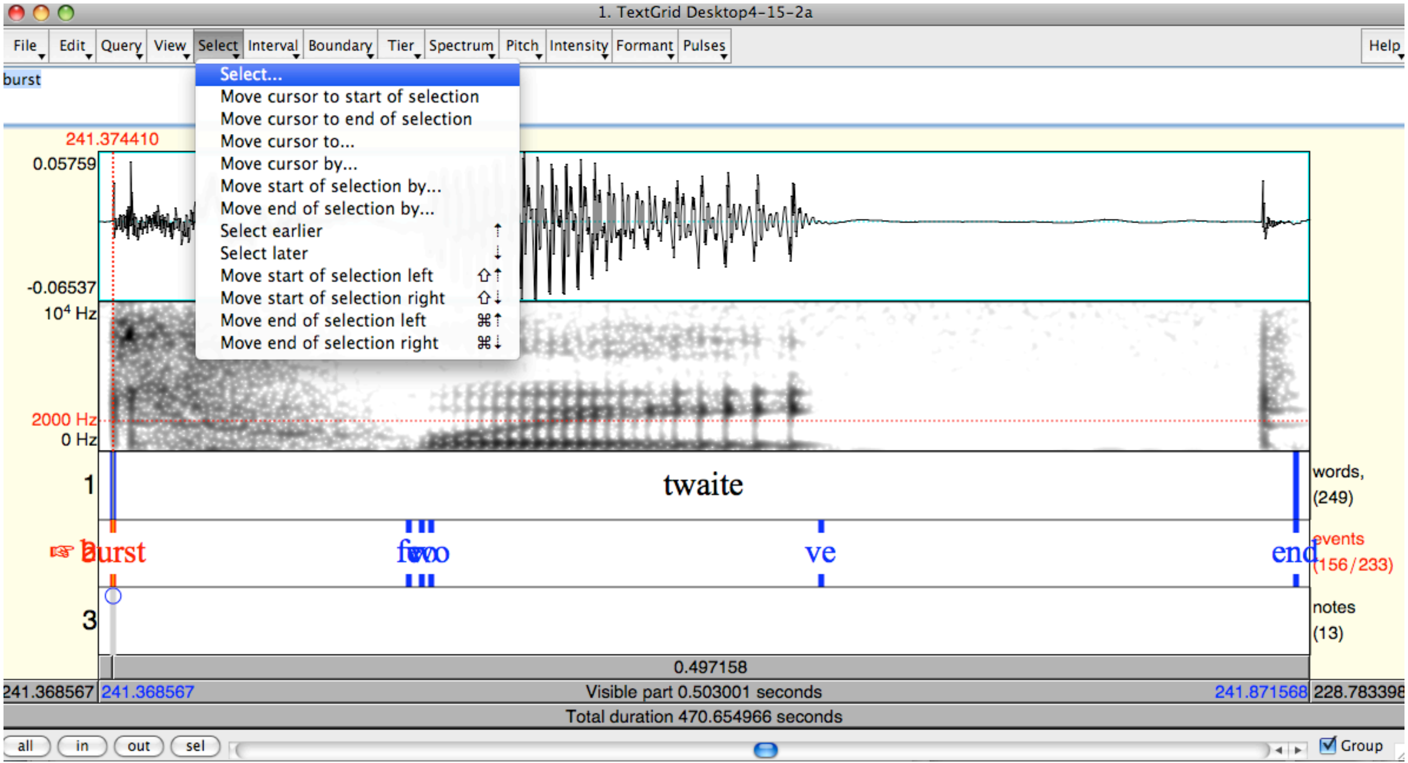

Figure 3: In order to select the area between two events, choose Select > Select…

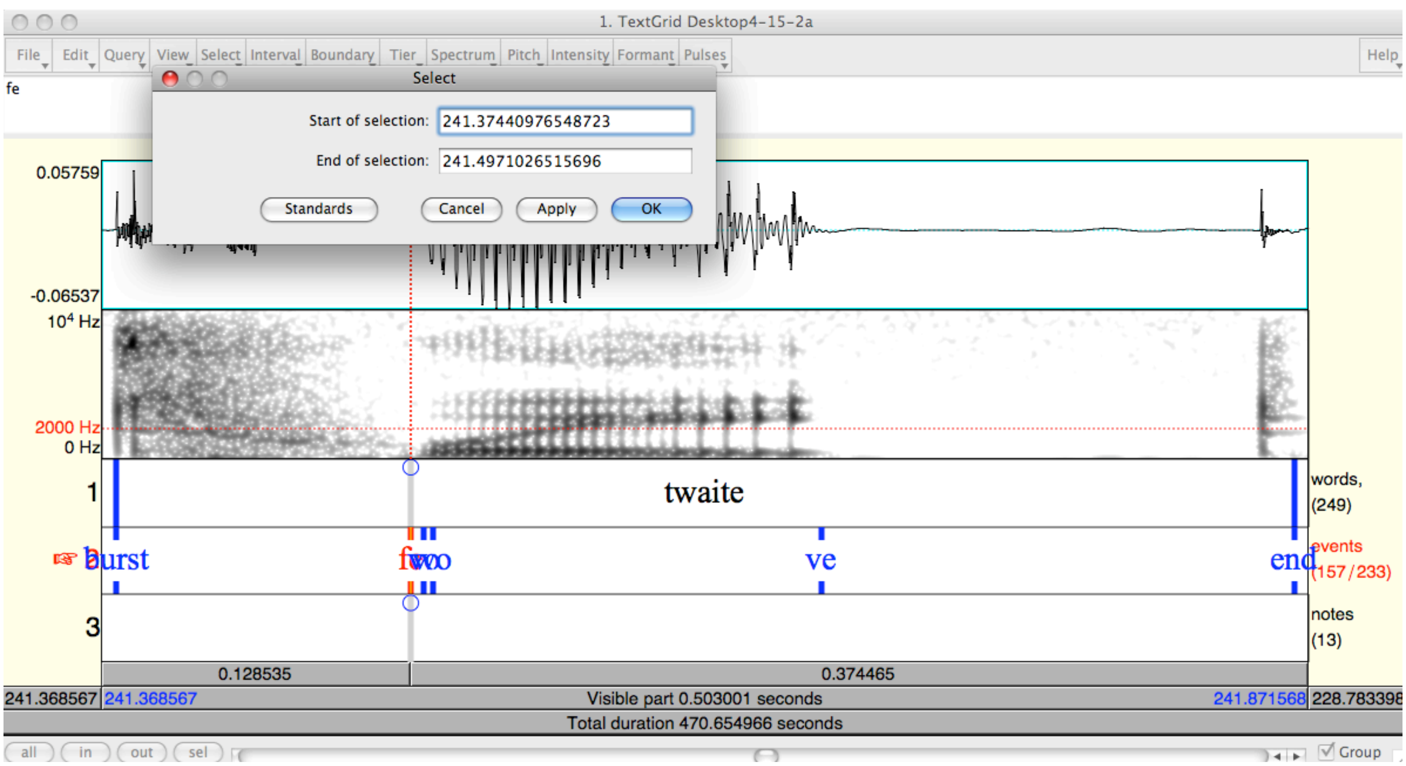

Figure 4: After choosing Select > Select… type in the time values you got from using Query > Get cursor as in Figure 2.

Burst intensity: Starting at the <burst>, select a window of 20 ms, and Get the intensity for this window.

Relative burst intensity: Divide the burst intensity by the affricate/stop intensity. This will show how much energy is at the beginning versus the later portion. In a stop, the energy should be stronger at the beginning and decrease, but an affricate will have less energy in the beginning than in the middle.

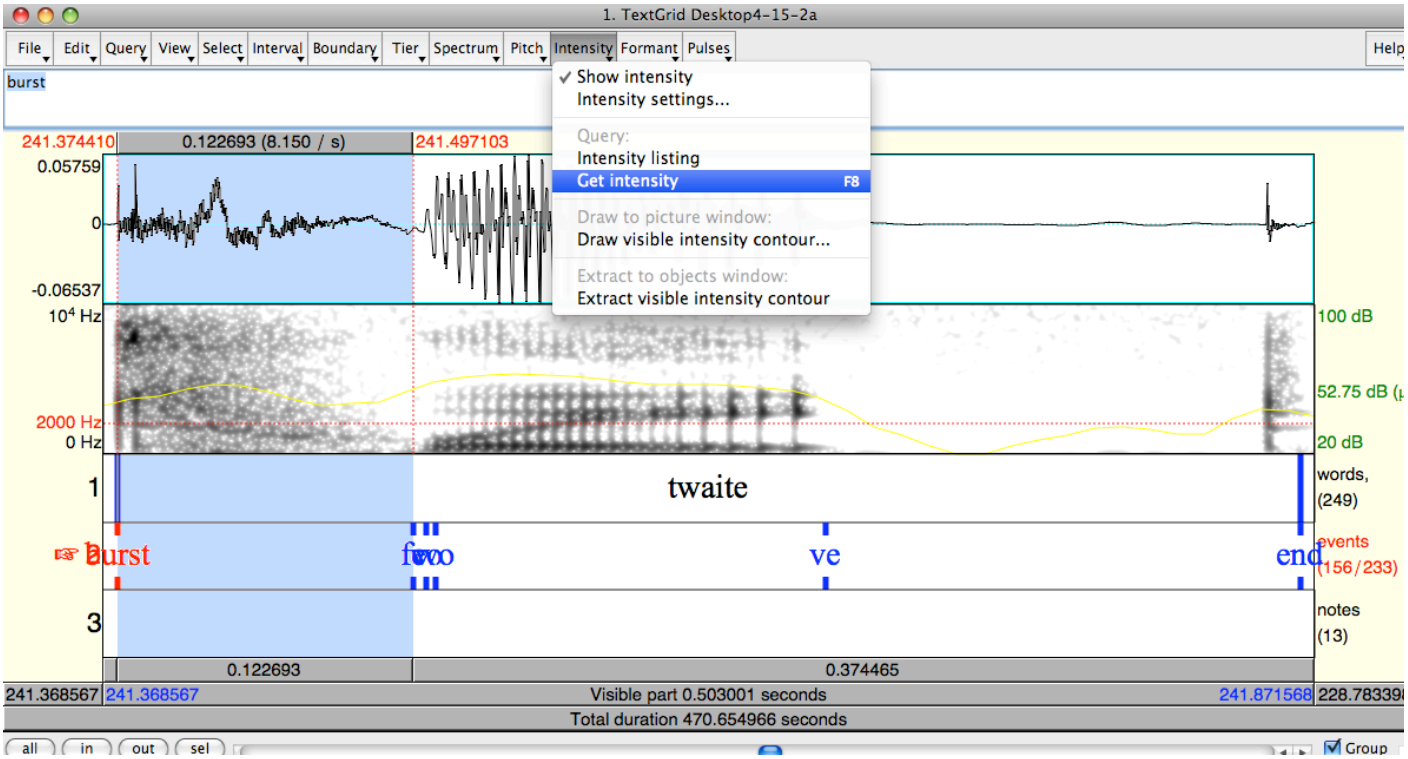

Figure 5: After you have highlighted your selection, choose Get intensity, in order to get the average intensity over the highlighted area.

Max intensity: Select the interval between <burst> and <fe>. First we need to extract this section so we can find the maximum intensity just in the stop/affricate part, so choose File > Extract selected sound (preserve times). Navigate back to the objects list and select the sound you just extracted. Then click on the button to the right labeled To intensity. The settings should look like those in Figure 6, with minimum pitch set to 500 Hz so that we can see smaller fluctuations in intensity. (The higher the min. pitch, the greater the detail in intensity because praat will do the averaging over a smaller window, which is necessary if we want to isolate only 20ms). This will create a new object in the list called Intensity untitled. Now you can query the intensity object (choose the box to the right marked Query) to get the maximum (Get maximum…). This will give you a value in dB.

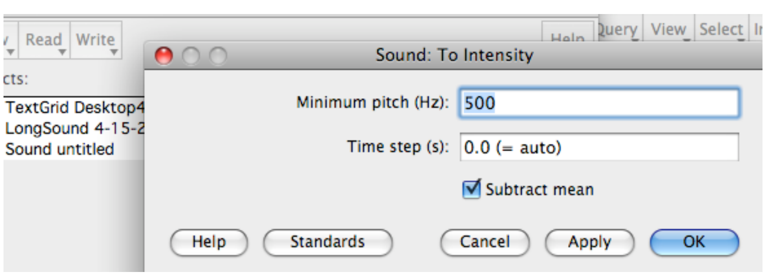

Figure 6: Creating an intensity object

Relative Max intensity: Divide the max intensity by the vowel intensity.

Rise Time: With the intensity object still selected in the objects window, click on Query again, this time select Get time of maximum. Subtract the burst time from the max time to get rise time, which is how long it takes to get to the loudest part of the affricate/stop. Stops should have short rise time, while affricates should have longer rise time.

III. Formants: Now, we are curious to see if the formant structure of the /w/ and following vowel has anything to do with how the affricate patterns. Before beginning this part of the analysis, you need to make sure that show formants is checked in the Formants dropdown menu. Then select Formant settings. Here you will need to decide if you are looking at a male or female voice, but more specifically how long the talker's vocal tract is. The reason this matters is that praat uses Linear Predictive Coding to find spectral peaks of a given time window of the sound. So, it is looking for the specified number of peaks (i.e., formants) over the specified frequency window (starting at 50Hz up to the maximum you specify). Men have longer vocal tracts, so their formants are all lower and closer together. Women have shorter vocal tracts, so their formants are higher and further apart. So, the default for men is to set the window to max out at 5000Hz for 5 formants. Women should have a window set to 5500 for 5 formants (as in Figure 7). However, you should look at the formants and see how well the red dots match up, that is how well praat's LPC algorithm is finding the appropriate spectral peaks. You may have to adjust the window higher or lower. Or you can adjust the number of formants to 5.5 or 6 (but not fewer than 5). After you have looked at several tokens to make sure the settings are correct, you may proceed:

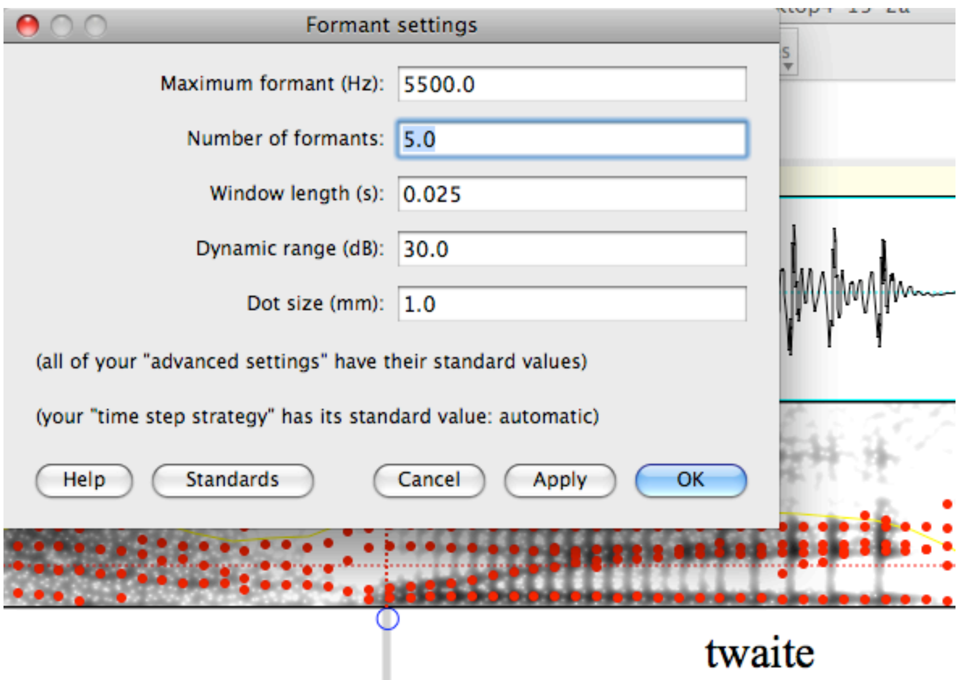

Figure 7: Formant settings for a female.

<u>/w/ F1, /w/ F2, /w/ F3:</u> select <wo>, then from the top menu, pull down the Formant menu. You can choose one at a time to Get first formant, Get second formant, Get third formant, or you may select Formant listing, which will list all the formants.

<u>vowel F1, vowel F2, vowel F3:</u> We want to move away from the /w/, in case the following vowel is having an effect on the affricate., but in many cases there are diphthongs, /r/s, triphthongs, and nasals involved. Also, the /w/ lasts longer in some cases than in others. So, we will take measurements at 25%, 50%, 75% and 90% along the interval between <wo> and <ve>. Here you need to find your vowel duration value, then calculate 25%, 50%, 75% and 90%. Use Select > Select… to place your cursor exactly x% plus the time point of <wo>. Take formant measures from all 4 locations.

IV. Spectral measurements: Now we are going to look at the spectral shape of the stop/affricate. Basically, we take a series of small slices of the sound wave and measure how peaky it is, where most of the energy is located, and the general shape of it. /t/ should have a fairly diffuse (flat) shape compared to /ts/ and /tS/. /ts/

should have more energy in higher frequencies because the front cavity is smaller due to the frontness of the tongue. /tS/ should have slightly lower frequencies because the cavity is lengthened if the tongue is retracted. But before we can do this we need to pre-emphasize: choose Filter (menu in the bottom right), drop down to Filter (pre-emphasis). In the box, you should specify from frequency (Hz) 80. This boosts the intensity of higher frequency sounds, at a rate of 6 dB per octave (doubling each frequency, so 80, 160, 320, etc…) so the high frequency components don't get hidden beneath all of the low-frequency sounds like the whispery but strong resonances of the vocal tract during aspiration. The justification for doing this is that lower frequency sounds naturally have more energy due to the way sound waves are constrained; if you think about how high a bouncing ball can go when it's going slow versus when it is bouncing fast, you will have an idea. The more up-and-down movement (or amplitude), the slower the ball seems to bounce (less frequency) and vice versa. If the ball is bouncing really, really fast, it can't get all the way to the ceiling, and if it goes to the ceiling, it bounces slower. Sound waves work similarly. Faster sound waves (that is, those with a higher frequency) have less energy or intensity just by virtue of having higher frequency. There is evidence to suggest that the hearing mechanisms take this into account when we're listening to speech sounds, so it makes sense to transform the spectrum in this way when we are talking about how speech sounds are perceived.

Now you will have a sound object called Sound untitled_preemp. This is the same as the sound you had selected, except with more energy in higher bandwidths. Use this pre-emphasized sound to get the spectral moments and spectral peak:

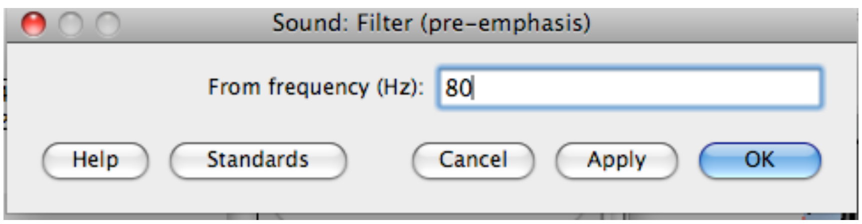

Figure 8: pre-emphasizing higher frequencies.

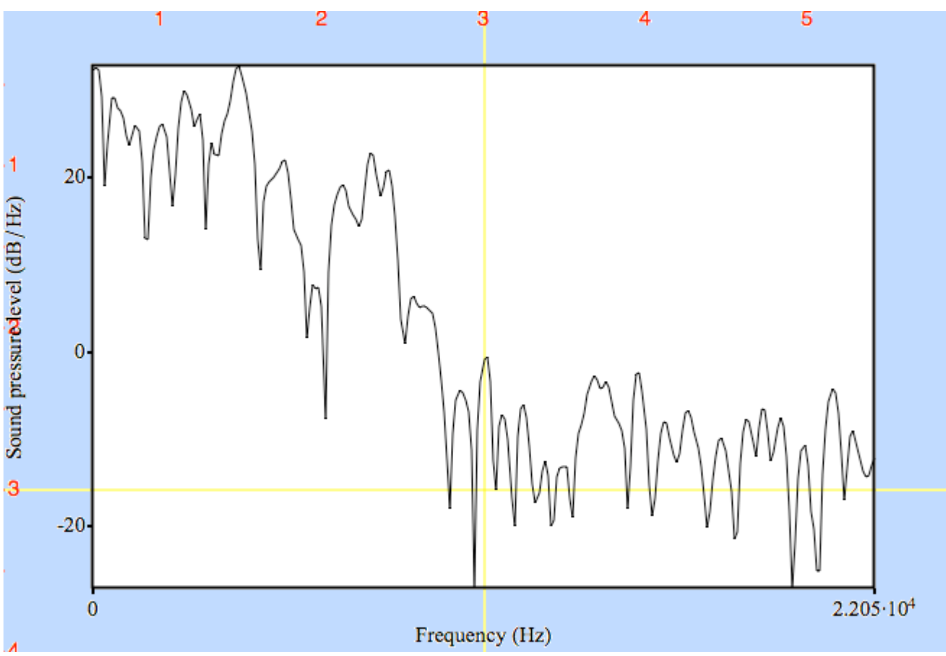

Figure 9: A single spectral slice

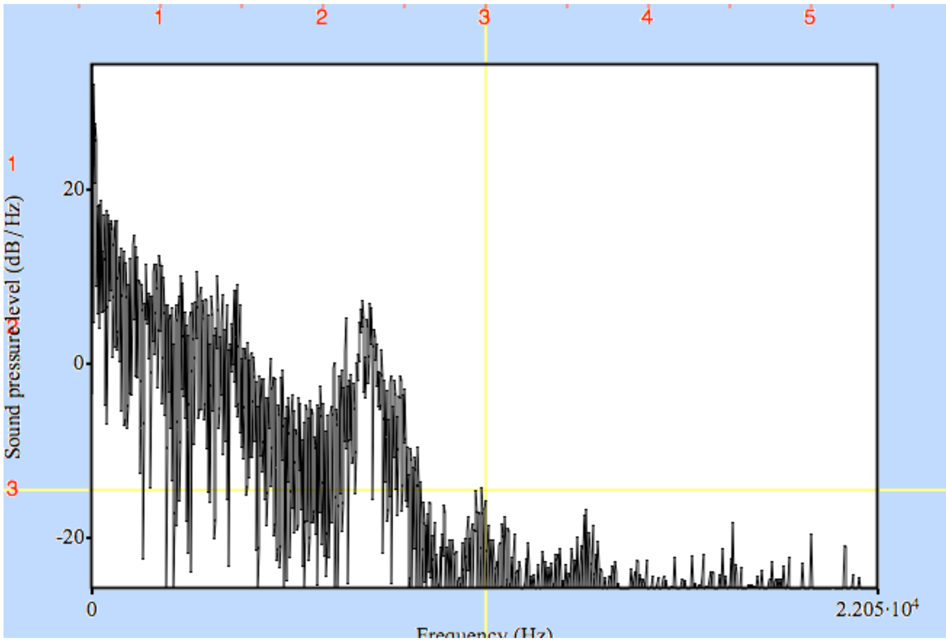

Figure 10: A spectrum taken from a 20 ms window. Notice how much rougher it looks than a single spectral slice. That's because there is a lot more information from many slices.

Center of Gravity: (1st spectral moment) This is the average frequency of the entire spectrum. If you think of the spectrum as being a histogram with really tiny bins at each Hz value (22050 bins if you've sampled at 44,100 Hz), if you add up each token (or multiply each bin by its dB/Hz value as if that were the token count), then divided by how many tokens there are, that would be your average.
We can go back to the objects menu and find the section that we extracted and pre-emphasized earlier, called Sound untitled_preemp, and take average measurements over the entire duration of the affricate/stop. Select the sound then go to the menu on the right that says Spectrum, pull down the menu and select To spectrum… This will generate a spectrum object in your objects list. Select Spectrum untitled, then choose Query from the menu on the right, pull down the menu and select Get center of gravity.

Standard deviation: As the center of gravity is the average, the standard deviation (also known as the 2nd spectral moment) tells how far each frequency value is from that average, so basically whether there is a large peak centered near the average, or if the distribution is relatively flat, with values distributed evenly all across the spectrum. Perform the same routine as for center of gravity, except in the dropdown menu, choose standard deviation.

Skewness: (3rd spectral moment) tells whether the average is close to the middle, with values falling more-or-less evenly to the far ends of the spectrum, or if the average is closer to one side or the other (has a higher or lower frequency than the center). A voiced sound is likely to have a very skewed distribution, with much greater energy in the lower frequencies. A fricative is less likely to be skewed. Same as above, but select Skewness from the drop-down menu.

Kurtosis: (4th spectral moment) tells whether the distribution around the average is like a normal distribution, and how far it actually is from that ideal distribution. Same as above, but select Kurtosis from the drop-down menu.

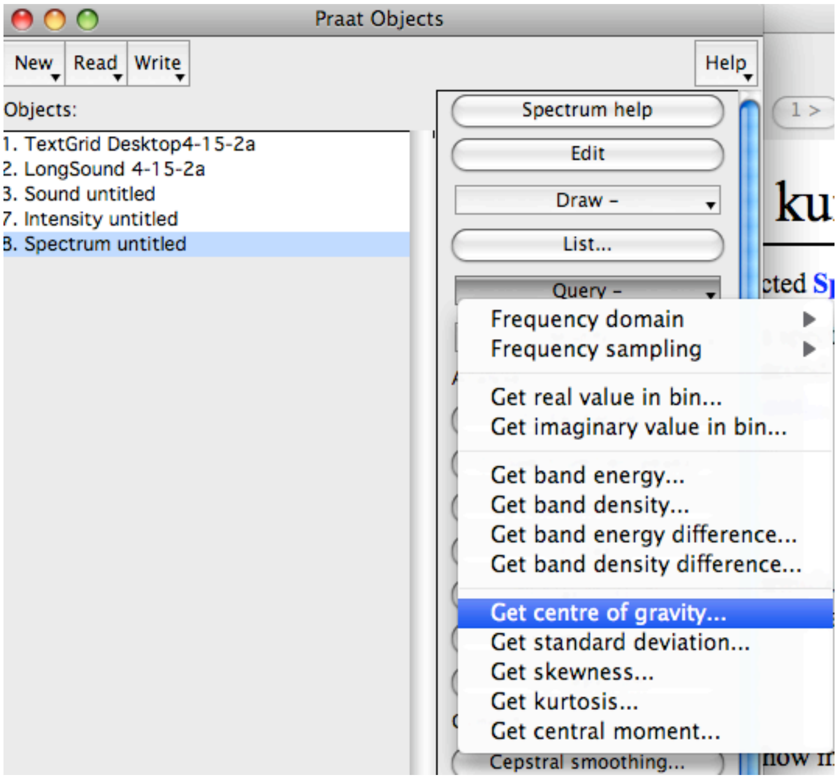

Figure 11: finding spectral moments from a Spectrum object

<u>Spectral moments from smaller windows:</u> The averaged values we get from taking spectral moments across the entire stop/affricate don't tell us very much because there is a lot of change going on during production of the sound. If you look at the spectrogram, you can often see a dark band of intensity decreasing in frequency as the lips round and increase the length of the front cavity during the transition from /t/ to /w/. So, we want to take a 20 ms window starting from the <burst,> 10 ms from both sides of the 25% point of the interval between <burst> and <fe>, 10 ms from both sides of the 50% point of the interval between <burst> and <fe>, and 20 ms backwards from <fe>. Follow the same procedure for extracting each 20 ms window, pre-emphasizing higher frequencies, and obtaining the four spectral moments as you did for the whole affricate/stop. Additionally, for each of these four smaller windows, you should calculate:

<u>Spectral Peak:</u> The frequency with the highest intensity. Again, if we can envision a spectral slice as a histogram, the spectral peak would be the mode, that is the tallest bar, or the bin with the most tokens. This has been frustrating trying to figure out since nobody spells out how they found "spectral peak" in their research articles, or even from where they started their pre-emphasis. Here is how I believe it should be done in our case: After you have extracted the 20 ms window and pre-emphasized the higher frequencies, select it in the objects list and from the menu on right, go to Spectrum, and this time choose To Ltas (which stands for long term average spectrum). This creates a smoothed spectrum by averaging the values from all of the spectra in the window. Set the bin size (see, it really is like a histogram) to 250 Hertz.

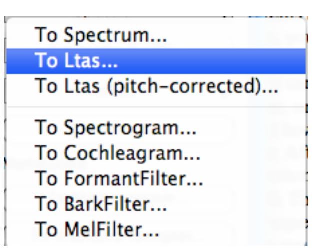

Figure 12: To Long Term Average Spectrum

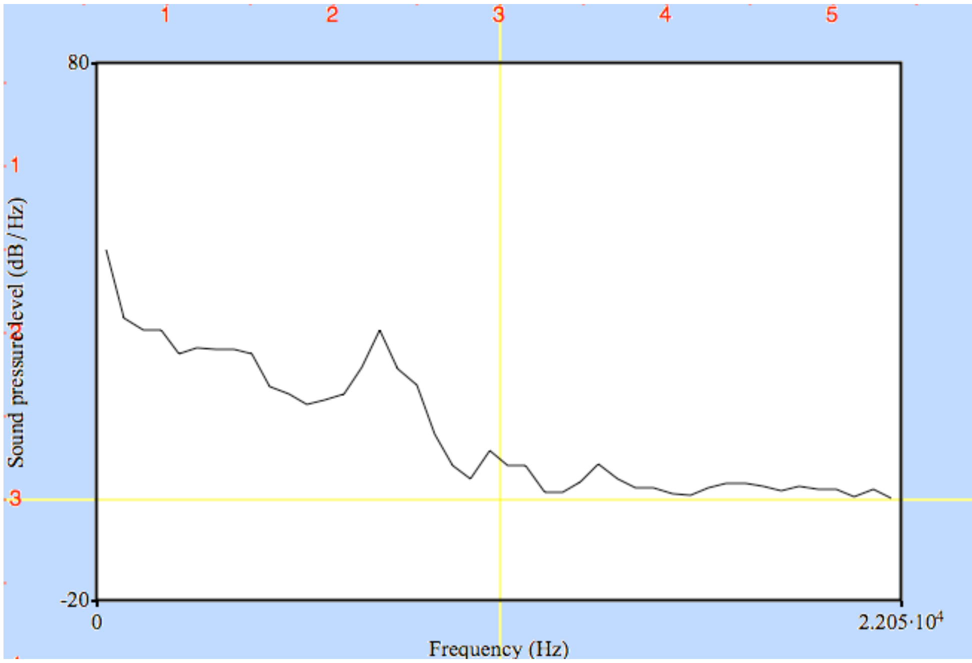

Figure 13: Long term average spectrum taken from the same window as in Figure 9. Notice how much smoother this looks than 9. The larger the bin size, the smoother it gets, and smaller bins make it bumpier.

Now you will have an Ltas object in your list. Select Ltas untitled_preemp. Go to the Query drop-down menu on the right and select Get frequency of maximum. Now, even though we preemphasized the higher frequencies, we don't want to risk getting a value in the low frequency range, so we will instruct praat to look for this peak between 2800 and 12000 Hz, as in Figure 14, with no interpolation (it's already averaged, so it doesn't matter). The value that praat returns will be the tallest bin in the Ltas, that is the frequency band of the highest intensity peak.
Also repeat this measurement over the four 20 ms windows.

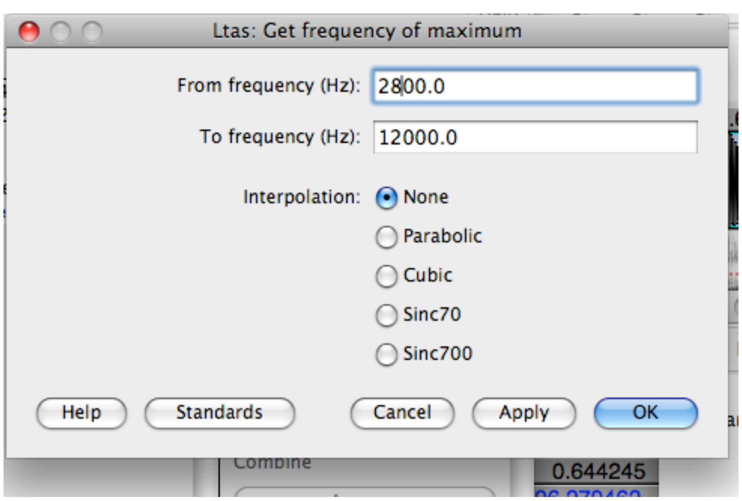

Figure 14: Searching for peak intensity between 2800 and 22000 Hz.

So, for each word, you should have 53 measurements. This is why, after you learn how to do the measurements by hand, and what they mean, why we're doing them, we will use a script to let praat collect these measurements for us.