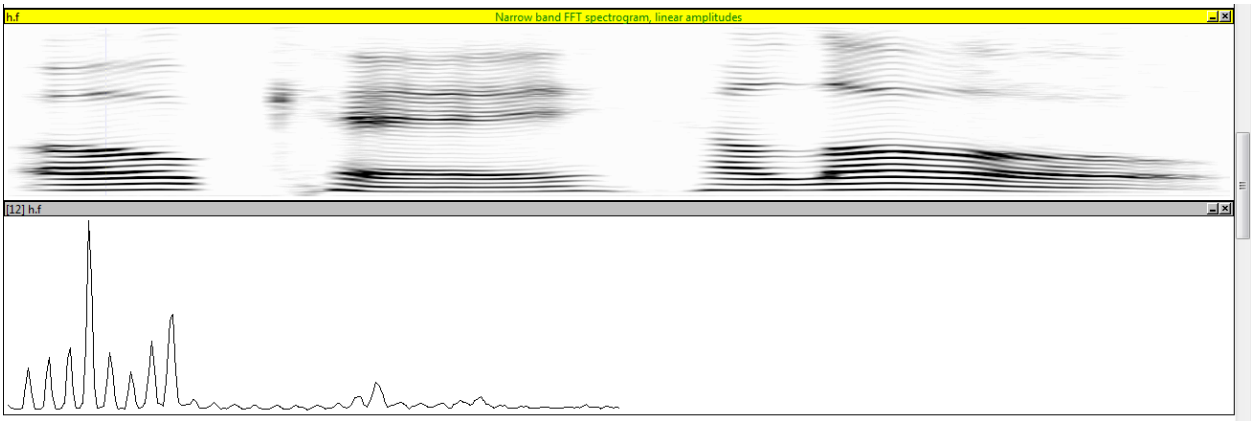
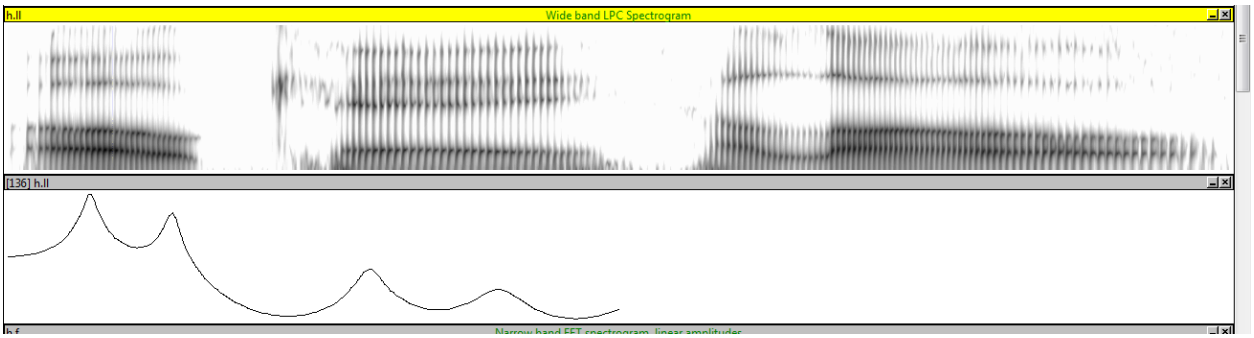
Narrowband versus Broadband Spectrograms

The big picture: Narrowband spectrograms are created with fine, high-resolution frequency analysis. At typical speaking fundamental frequencies (~80-300 Hz or so) the analysis is fine-grained enough to resolve individual harmonics of the voice source. The display below shows a narrowband spectrogram of the utterance, "Ok, hello.". Below that is a spectral slice (amplitude spectrum) taken from about the middle of the first vowel. Notice the harmonics that are visible in both the spectrogram and the spectral slice during voiced segments of the signal.



Broadband spectrograms are created with a more coarse frequency analysis. The idea is to smear over a large enough band of frequencies to display the ***collection*** *of frequency components* that comprise the broad spectrum envelope peaks that correspond to vocal tract formants. The most common and straightforward analysis method that is used to calculate broadband spectrograms computes the spectrum over a short segment of the time signal. Because of this short time segment (called a "window"), the analysis is able to capture rapid changes in amplitude. For this reason, *during voiced speech segments*, the broadband spectrogram shows vertical lines corresponding to the rapid increase in amplitude that occurs when the vocal folds slap together. Notice that these vertical lines are not visible in the narrowband spectrogram above. The narrowband frequency analysis is calculated over a much longer time window – too long to capture the rapid increase in amplitude that occurs at the time of vocal fold closure. The display below shows a broadband spectrogram of the same utterance ("Ok, hello."). Below that is a spectral slice taken from the middle of the first vowel. Notice the broad horizontal bars corresponding to the formants in the spectrogram, and the envelope peaks in the spectral slice. These peaks also correspond to vocal tract formants, Also notice the vertical lines – seen only during voiced segments – that occur when the vocal folds slap together.



The narrowband/broadband spectrogram exercise

There are several test signals in the ztool directory:

h.wav
m01ae.wav
avacado.wav
aba.wav
apa.wav
si1002nm.wav

For each of these test signals:

1. Use File/Open to open the sound file (e.g., c:\ztool\m01ae.wav).
2. Bug the “BBGram” button.
3. Bug the "Slice" button.
4. Click on the .wav file again. The title bar will turn yellow.
5. Bug the “NBGram” button.
6. Bug the "Slice" button.

(You might want to pull down on the bottom of the narrowband (NB) gram pane to make it a little bit taller.)

Use the "double wave" button at the top, right to set SpeechTool in continuous-update mode -- i.e., set it so that the slices are updated automatically as the mouse is moved left to right over the sound wave or either of the spectrograms.

The spectral slices are just individual amplitude spectra – frequency on the x, amplitude on the y. The spectrograms are created from these amplitude spectra: the spectra are rotated 90 degrees so that frequency is now on the y, and amplitude is converted to a gray scale (black = high amp, white = low amp, medium amps are the appropriate shade of gray in between). NB grams are created from very fine, high resolution spectra. BB grams are created from spectra that are purposely lower in resolution; i.e., these spectra smear across a wider band of frequencies. This explains why you no longer see harmonics. Although it's not always quite that simple, NB grams are based on a frequency analysis that is fine (high resolution) enough to resolve individual voice harmonics, while BB grams are based on a frequency analysis that is coarse enough to collect energy across the group of frequency components that correspond to the envelope peaks that (usually but not always) correspond to formants. (Why “usually”? The short answer, which will have to do for now, is that nothing about speech is simple.)

a. Which of the spectrograms – BB or NB – shows well defined, *regularly spaced* harmonics? You'll see the harmonics in one of the slices but not the other one. Which one? Why?

b. Which of the spectrograms – BB or NB – show (more clearly) thicker, broader bands of energy corresponding to formants? You'll see these broader, irregularly spaced spectral peaks (again, more clearly and more simply) on one of the slices but not the other one. Which one?

Why?

c. Which of the spectrograms – BB or NB – show vertical lines corresponding to individual glottal pulses?

d. When you finish with the first signal, go to SpeechTool's “Window” menu and choose "Close all". Open the next signal and go through the same steps. Continue until you're out of signals, or until you get what's going on, but ***do at least three***.

Bonus question: On the ***broadband*** spectrogram above, the first speech sound you see is the [o] of “Ok”. This is followed by a silent interval corresponding to the closure for the [k^h]. During this closure interval air continues to flow through the glottis and, since the oral cavity is entirely occluded by the contact between the tongue dorsum and the velum, pressure will build behind this obstruction (this is what the term “stop consonant” means). The tongue then drops, immediately releasing this pressure, and the tongue begins to move to a position appropriate for the [e] of “Ok”. At the moment of articulatory release for the [k^h], the BB spectrogram shows a very ***sharp*** pulse of energy – a transient – with energy spread across a wide range of frequencies. The sharp pulse (sharp = very narrow ***in time***) is followed by a patch of noise called aspiration (a brief [h]), but we’re not concerned with that for this discussion. Now, finally to the main point: take a look at that same feature (the transient following articulatory release) on the NB spectrogram. Is it a sharp, narrow (again, narrow *in time*) pulse, or is it sort of smeary? It’s more smeary, right? Why is this feature more smeary on the NB gram than the BB gram? Hint: take a look at the comments in the second paragraph on page 1 having to do with short time windows vs. longer time windows. Is this NB/BB difference related in any way to the fact that vertical lines corresponding to glottal closure can be seen on BB grams but not NB grams? Hint: Yes, it is. But how so?