Speech Analysis using PRAAT

(A brief guide prepared by Pranav Jawale)

1. Praat installation

Windows users can install the latest version of Praat from http://www.fon.hum.uva.nl/praat/download_win.html.
Double-click the setup file to install Praat. This will create a praat.exe file.

Linux users can install Praat from http://www.fon.hum.uva.nl/praat/
It is better if you run Windows version of Praat by running it under 'wine' in Linux.

Additional installations –
Doulos SIL Font - Has phonetic symbols, can be downloaded from
http://www.fon.hum.uva.nl/praat/DoulosSIL4.106.zip or
http://scripts.sil.org/cms/scripts/page.php?site_id=nrsi&item_id=DoulosSIL_download#FontsDownload

When you open Praat (double-click praat.exe) two windows, *Praat Objects* and *Praat Picture*, would appear. The Praat objects window contains all the analysis options. Using the Praat picture window, pictures of various aspects of the *objects* can be exported. For now, the picture window is not needed, you can close it.

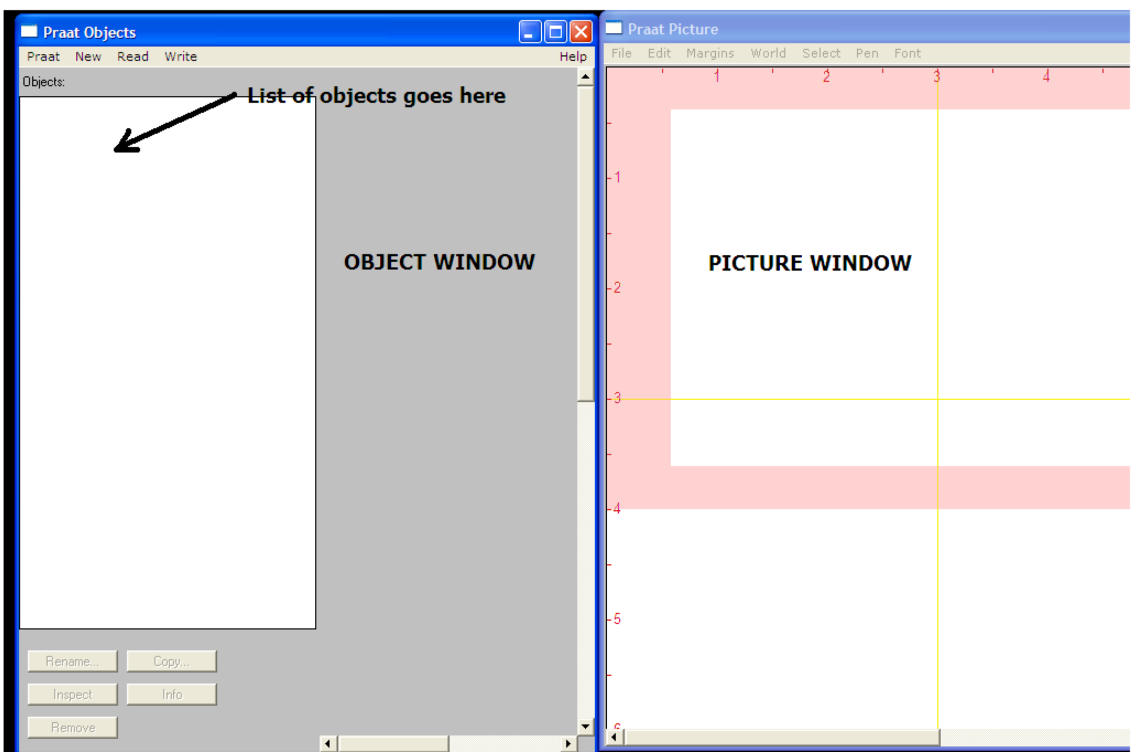


Figure 1 Praat Object and Praat Picture windows

In the Praat Object window click **Read → Read from file**
Then a file selection window will appear. Browse, locate and select '**machali.wav**'.
Click **Open.** Highlight its name in the list of objects and click **Edit.**

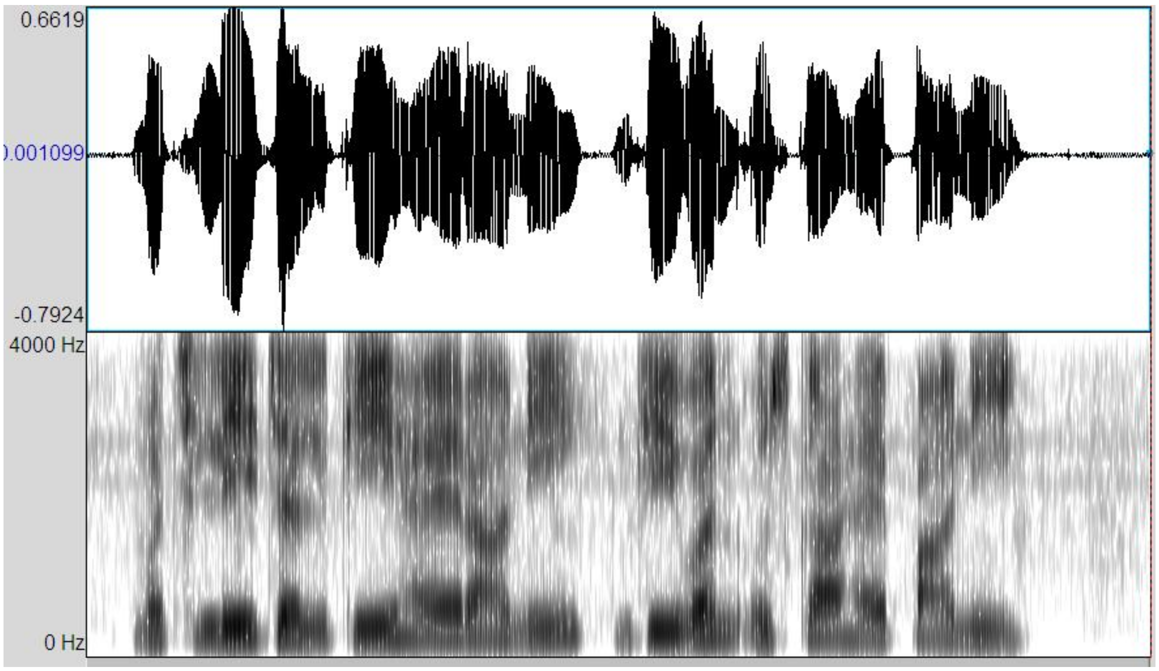


Figure 2 Waveform and spectrogram of machali.wav

Play the file. What is it playing?

Zoom into shorter segments of the waveform and observe the acoustic properties of the various phones: periodicity, energy, average zero-crossing rate, formants.

2. Various spectrogram settings in Praat

Now we will look into the spectrogram settings.
Click **Spectrum → Spectrogram settings**

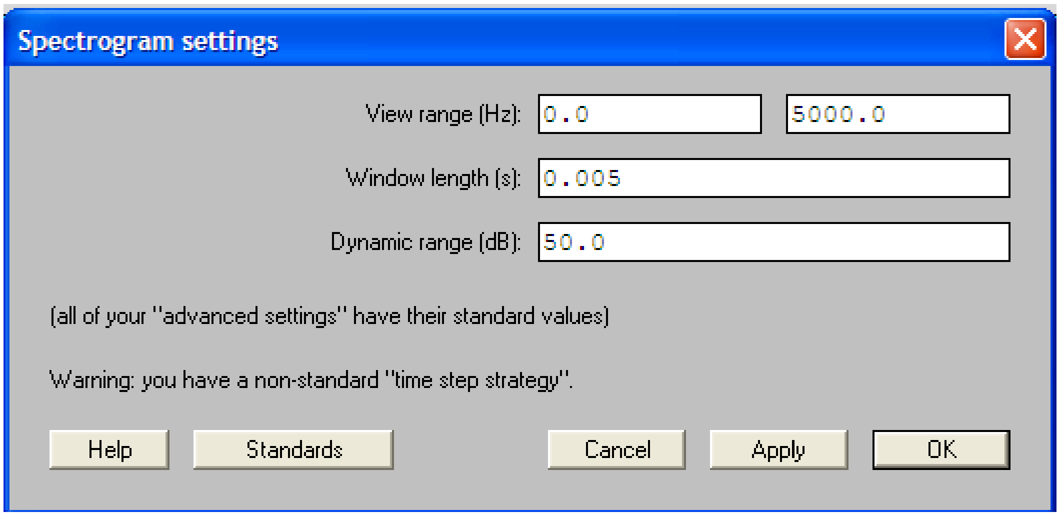


Figure 3 Default spectrogram settings

<u>Explanation of parameters</u> **[1]**:

- **View range** (Hz) **-** Range over which the spectrogram is shown. To view the spectrogram over whole frequency range, max frequency should be half the sampling rate (8000/2 = 4000 Hz). Spectrogram will be blank for frequencies higher than the Nyquist frequency if they are displayed.
- **Window length** (s) **-** To compute short time Fourier transform Praat uses analysis windows (frames) each of length as specified by the *window length* parameter. For a window length of 0.005 s, Praat uses for each frame the part of the sound that lies between 0.0025 seconds before and 0.0025 seconds after the centre of that frame.
- **Dynamic range** (dB) **-** All values that are more than *Dynamic range dB* below the maximum will appear white. Values in-between have appropriate shades of grey. Thus, if the highest peak in the spectrogram has a height of 30 dB/Hz, and the dynamic range is 50 dB (which is the standard value), then values below -20 dB/Hz will be drawn in white, and values between -20 dB/Hz and 30 dB/Hz will be drawn in various shades of grey.

Now click **Spectrum -> Advanced spectrogram settings**

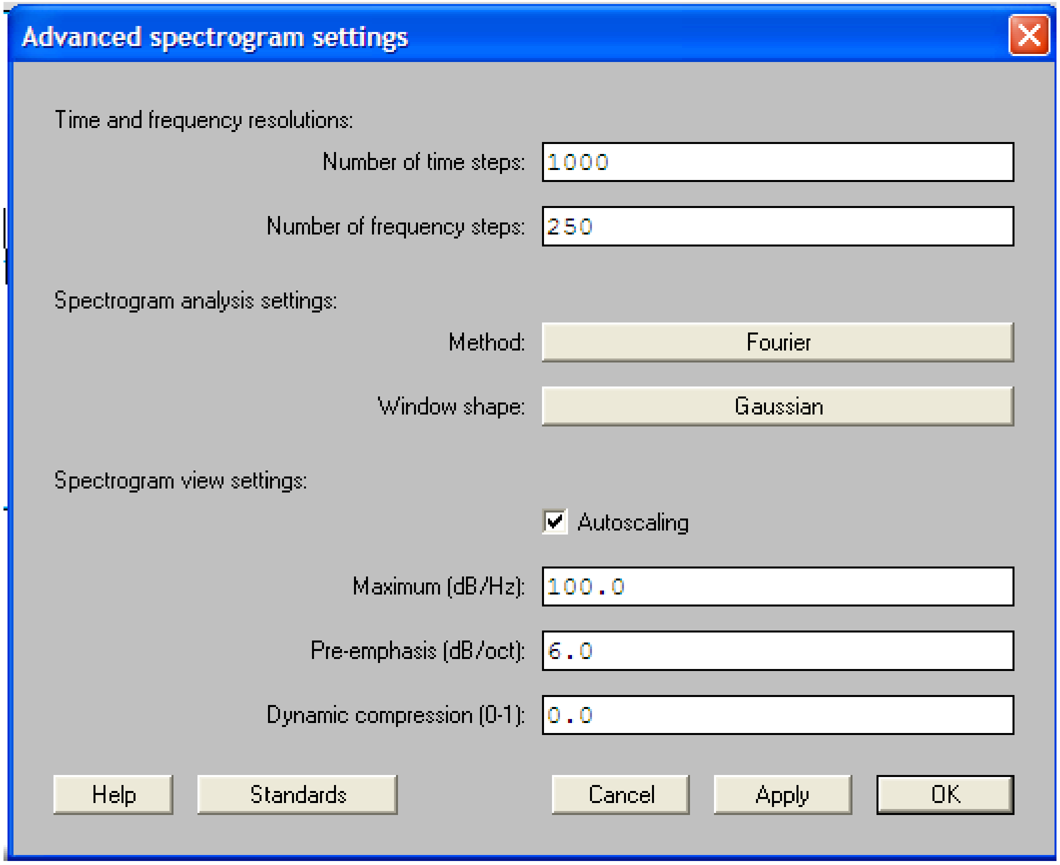


Figure 4 Default advanced spectrogram settings

<u>Explanation of some parameters:</u>

- **Number of time steps -** The maximum number of points along the time window for which Praat has to compute the spectrum. If your screen is not wider than 1200 pixels, then the standard of 1000 is appropriate, since there is no point in computing more than one spectrum per one-pixel-wide vertical line.
- **Number of frequency steps -** The maximum number of points along the frequency axis for which Praat has to compute the spectrum. If your screen is not taller than 768 pixels, then the standard of 250 is appropriate, since there is no point in computing more than one spectrum per one-pixel-height horizontal line.
- **Window shape -** Default is Gaussian window. You can change it to Rectangular, Hamming, Hanning etc. (See **[1]** under Further Reading)

3. Wideband and narrowband spectrograms

Speech spectrograms are called as wideband or narrowband based on what window length is used. Shorter the window, larger its bandwidth and the spectrogram is a wideband one. Similarly if the window is larger, the spectrogram is called as narrowband. Now, how much window length corresponds to either of these? There is no clear cut boundary but for a speech sample, if the window length is around 3 - 5 ms (bandwidth ~ 200 - 300 Hz), the resulting spectrogram is called as wideband. For the window length around 20 – 30 ms (bandwidth ~ 30 - 50 Hz), the spectrogram is called as narrowband.

What is the significance of these two types of spectrograms?

Wideband spectrogram is used to observe the formant structure while narrowband spectrograms reveal the harmonic structure (pitch information). Let's say you use a window length of 20 ms. With this window length, voice harmonics are resolved much clearly as frequency resolution between them is generally > 50 Hz (1/0.02). If the window is shortened to say 3 ms, then the spectra of two or three voice harmonics will merge and instead of individual harmonics you will see structure of formants in the spectrogram.

As an example, zoom in to a section of machali.wav depicting phoneme /m/. Play it.

Keep advanced spectrogram settings at their default values (by clicking **Standards** button) and observe the waveform and spectrogram for these two settings. Note: changing dynamic range only changes the spectrogram intensity settings. To view the formants clearly, we have reduce it to 30 dB.

Table 1 Window lengths for narrowband and wideband spectrograms

Parameter	**Setting 1**	**Setting 2**
Window length (s)	0.02	0.005
Dynamic range (dB)	50	50

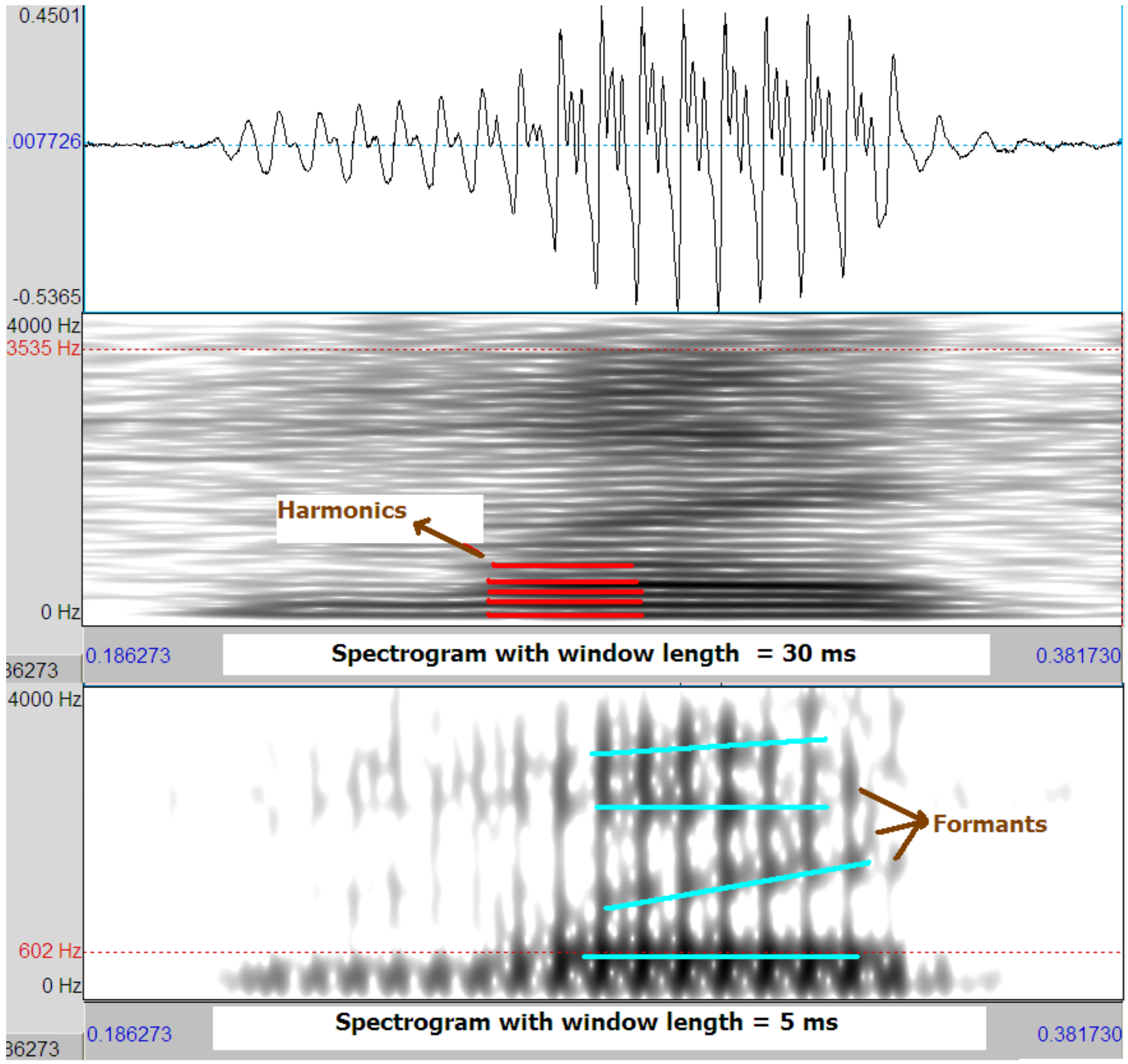


Figure 5 Spectrograms of /m/ in "machali" for two different window lengths (30 ms and 5 ms)

In Fig 5, when the window length is 30 ms, we can see many horizontal black lines in the spectrogram which correspond to harmonics of the pitch frequency. As the window length is made 5 ms, we can no longer distinguish between various harmonics, but we can see the formants. The vertical dark lines in the spectrogram indicate glottal pulses.

Now zoom out (click **all**) and observe the whole spectrogram for the two settings in Table 1.

4. Observing Phonemes and their formants

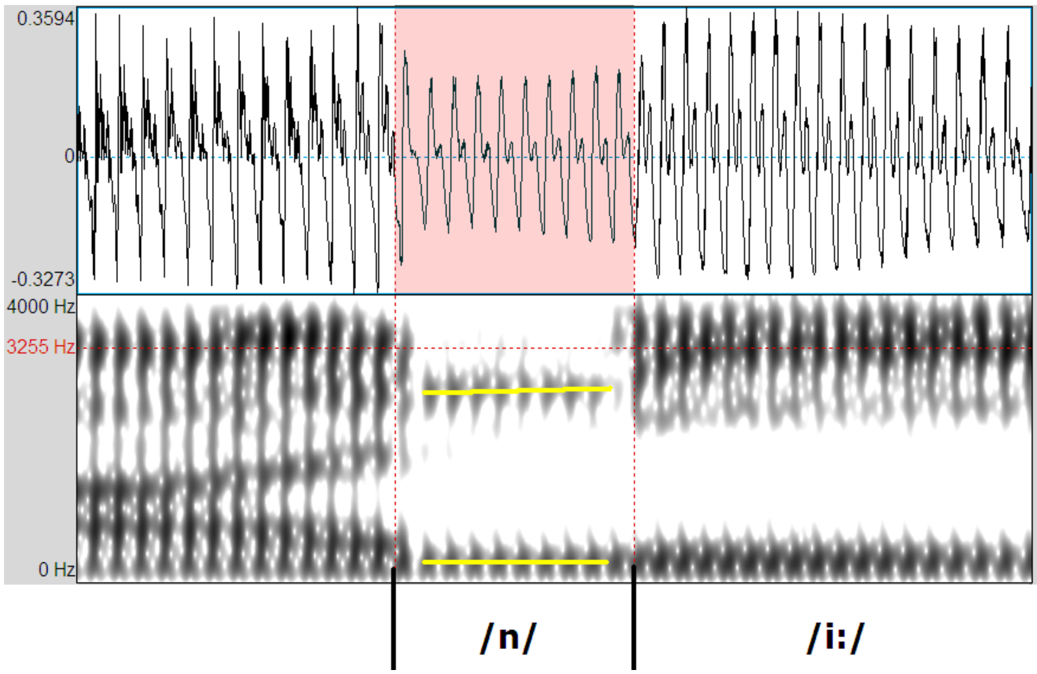


Figure 62 Waveform and wideband spectrogram for /n/ in paani. Window length = 5 ms.

Zoom in to the part in waveform where /n/ in *paani* is located. It is characterized by abrupt discontinuities in the spectrum at the beginning and end. Note the loss of intensity in the spectrogram due to the anti-resonances. Nasals are waveform-wise differentiated by their lower amplitudes (as compared to vowels).

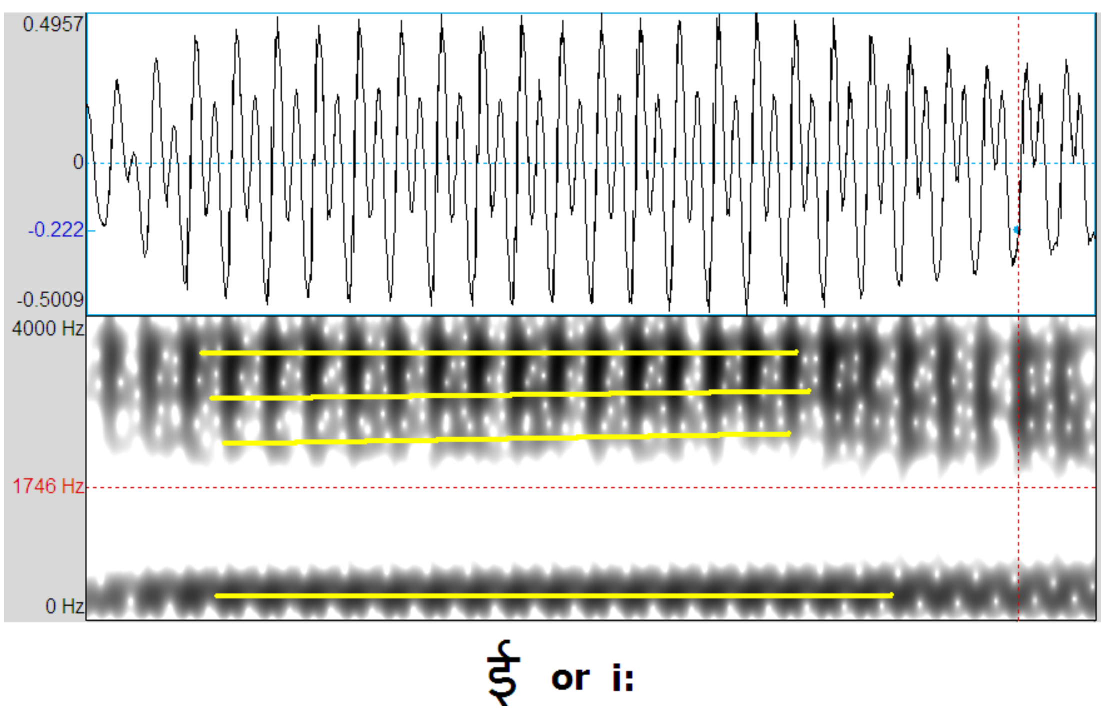


Figure 7 Waveform and wideband spectrogram of /i:/ in "paani"

Now observe /i:/ in “jal ki”. Observe the different formants. What are their frequencies? The phoneme, /i:/, has the maximum distance between F1 and F2 of all the vowels.

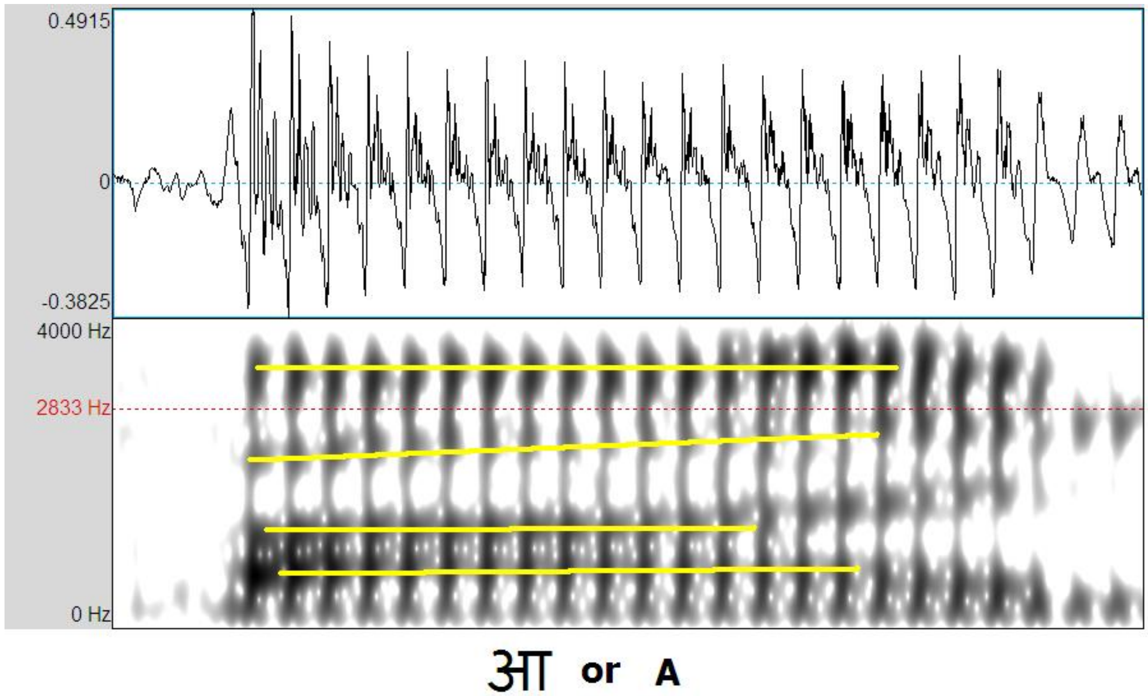


Figure 8 Waveform and wideband spectrogram of A in “paani”

Now look at the spectrogram of /A/ in “paani”. Where do its first three formants lie?

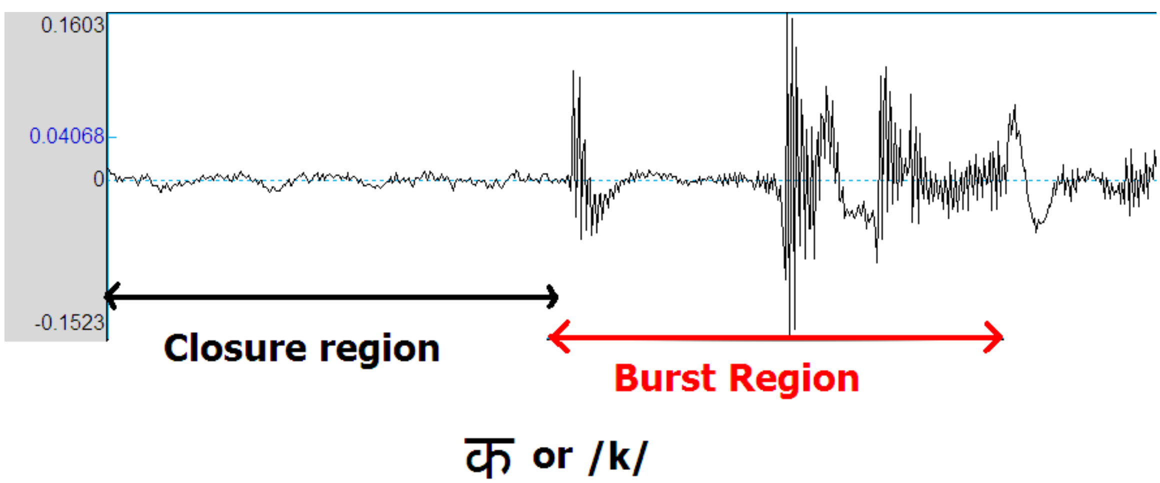


Figure 3 Closure and burst regions of /k/ in “jal ki”

Let’s look at the waveform of /k/ in “jal ki”. We can divide into two regions. First, the closure and second the burst region. In this example we can see two bursts along with some frication in the burst region.

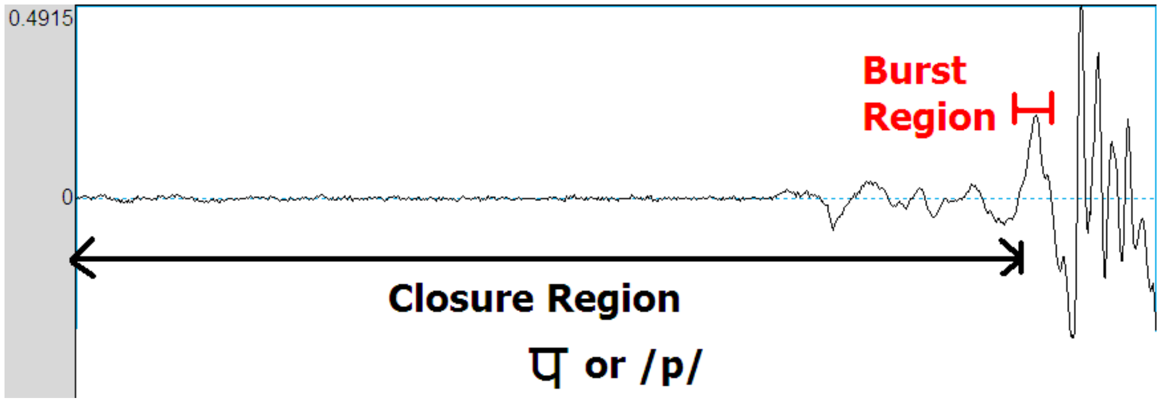


Figure 40 Closure and burst regions in /p/ in "paani"

Zoom in to waveform of /p/ in "paani". As can be seen, /p/ has a very short burst region.

5. Pitch

Now we will look at ways to find out the pitch.
Select Pitch → Show Pitch to show the pitch contour. You can remove the spectrogram from the view by selecting Spectrum → Show Spectrogram to get a clear view of the cpitch contour. The pitch contour is shown in blue.

You will note that this contour is not continuous. The points at which the pitch is absent are usually where the speech is not voiced or not present at all. For example, in Figure 11, the word "uska" has been zoomed into. The pitch contour is shown. The phonemes /s/ does not have any pitch because it is unvoiced. Similarly the pitch contour of closure region of phoneme /k/ is not shown.

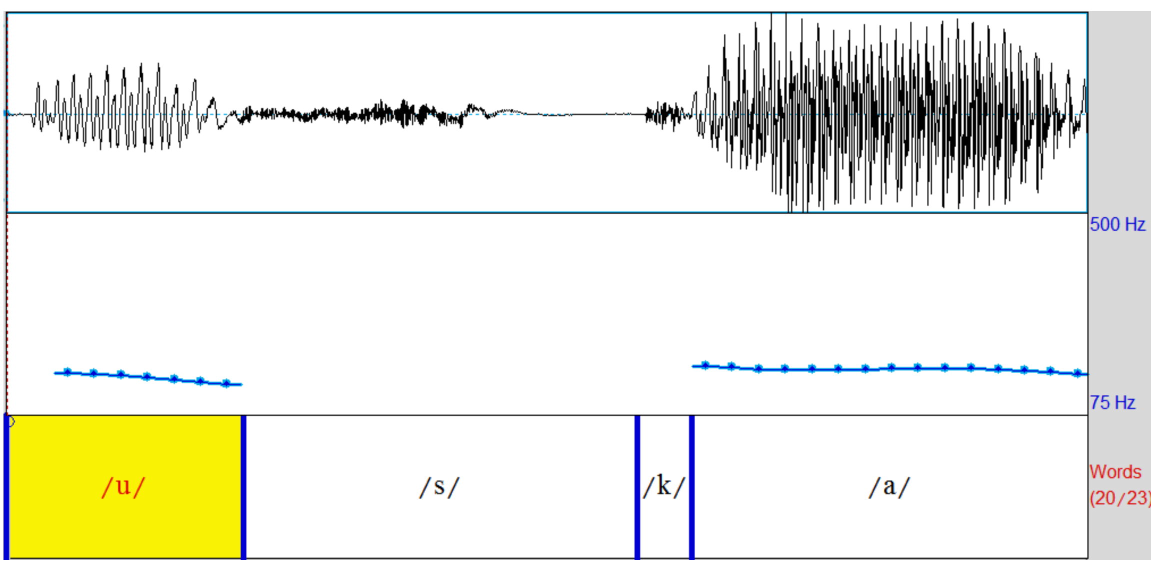


Figure 11 Pitch contour of 'uska'

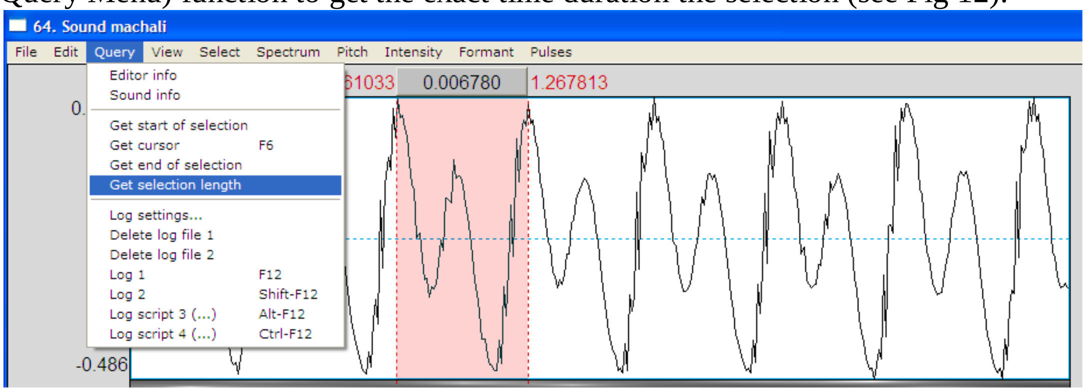
Finding the pitch of a section

Suppose we want to find out the pitch of /i/ in ‘paani’. Zoom into the phoneme and select a few waveforms of the phoneme. Go to Pitch → Pitch listing to get the pitch of the selected segment.

Sometimes, because of the algorithm used by Praat calculate the pitch [4], the pitch calculated may be half or double of the actual pitch. This can be avoided by varying the Pitch Settings. Go to Pitch → Pitch Settings. You can change the Minimum and Maximum Pitch Settings. This defines the range in which Praat will look for pitch values. Since this sound is a male voice, we will keep the Minimum Pitch to be 75 Hz and Maximum Pitch to be 300 Hz.

Method 1:

Zoom in to the phoneme /i/. It appears to be periodic. So the inverse of the time period should be the pitch. Select one wave form of the phoneme. Use **Get selection length** (in Query Menu) function to get the exact time duration the selection (see Fig 12).



ई or i:

Figure 52 Finding selection length of section of a waveform of /i:/ in “paani”

What value of pitch did you find out?

Now go to Pitch → Pitch Listing to get the value of pitch calculated by Praat and compare it.

Method 2:

Let’s change the spectrogram settings so that it shows the harmonics. Change the window length to 0.03 s. The difference in Hz between adjacent harmonics is the pitch as shown in Fig 13. You can find an approximate value of pitch using this method. Is it the same as you found out from waveform analysis?

However, note that pitch is equal to the difference in Hz between the harmonics *only when all the harmonics are present in the audio sample*. If some harmonics are missing, you will compute a wrong value of the pitch. Waveform analysis can almost always tell you the correct value of pitch.

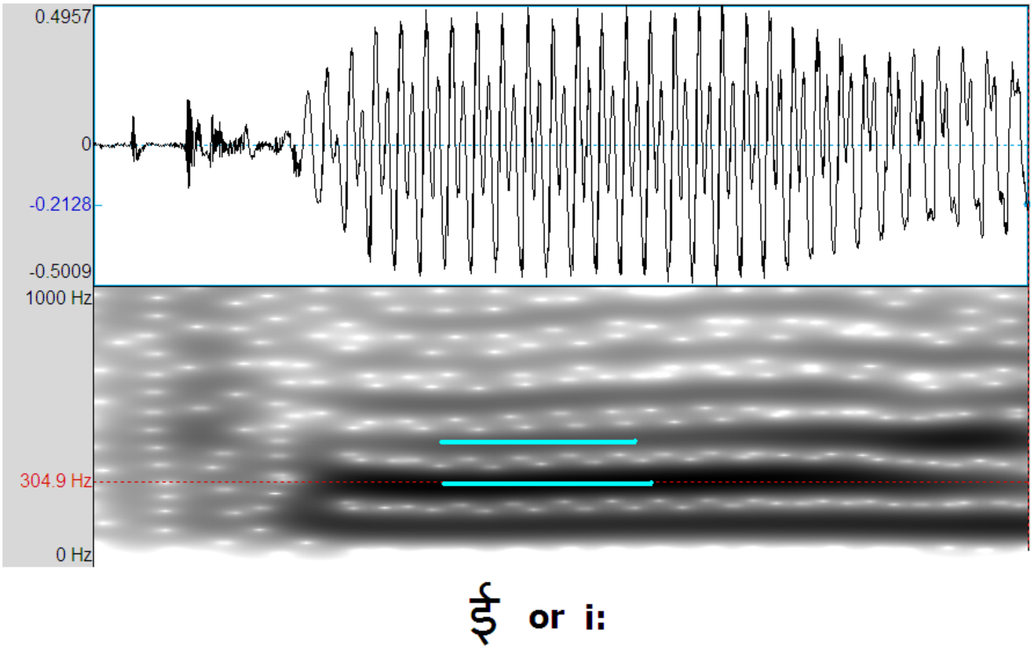


Figure 6 Waveform and narrowband spectrogram of /i:/ in "paani"

Method 3:
A more accurate way to detect the pitch from the spectrum: Select two/three periods of waveform and click **Ctrl+L** (Or go to **Spectrum ->View spectral slice**). This shows a Magnitude Vs Frequency plot drawn after computing FFT. The difference in Hz between any of the high energy adjacent peaks will give you the pitch value.

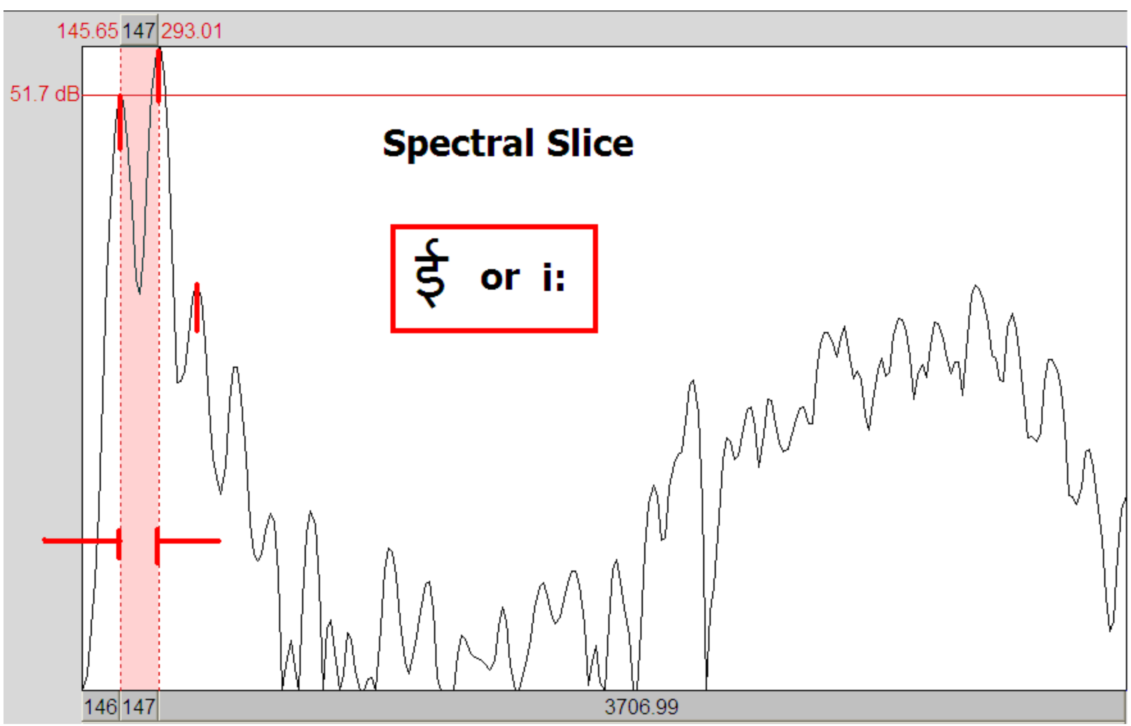


Figure 147 ***Spectral slice of /i:/ in "paani"***

Now use the waveform and spectral slice methods to find out pitch of /m/ in "machali" (Fig 15) and compare these values with value calculated by Praat.

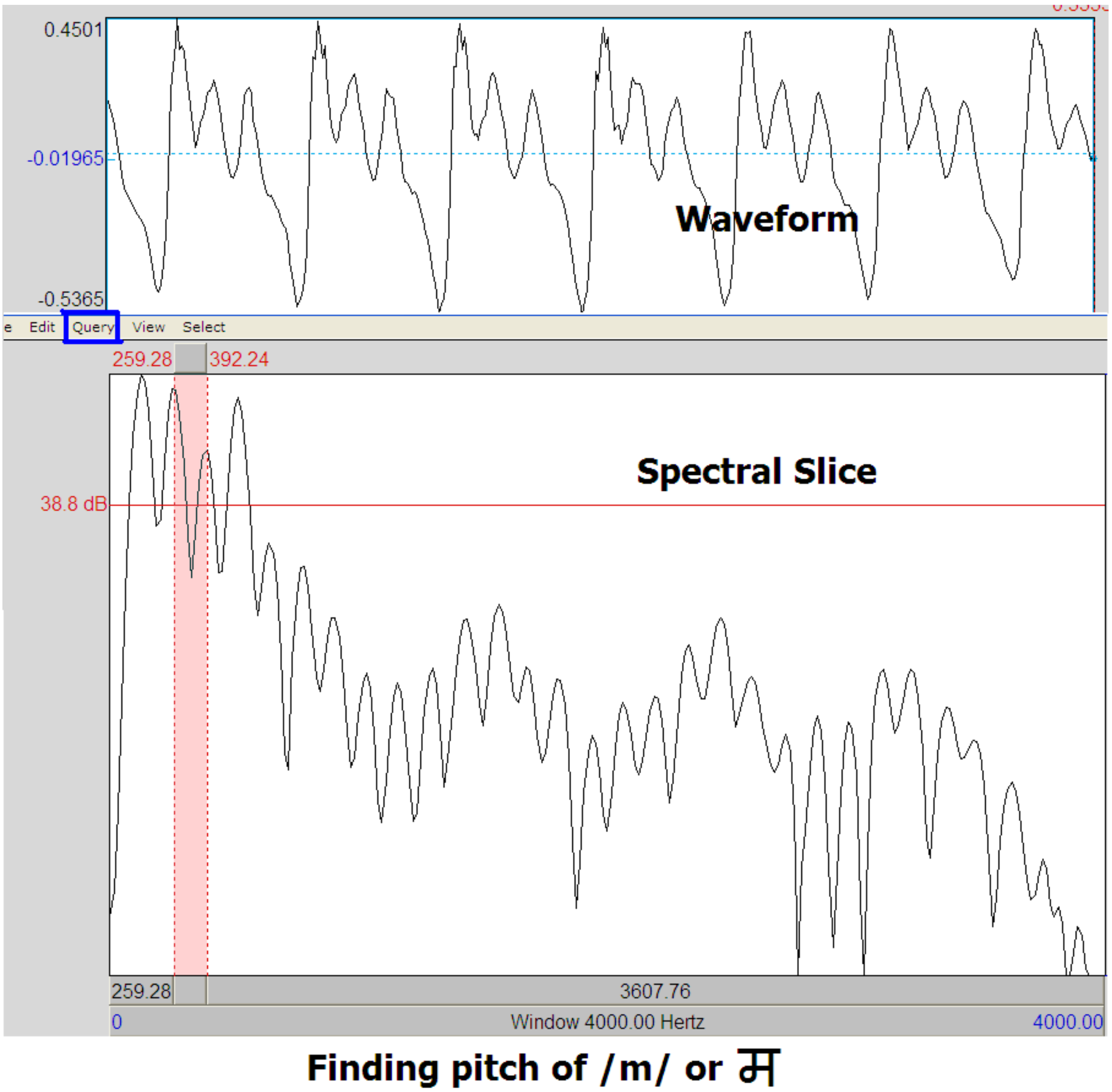


Figure 85 Waveform and spectral slice for /m/ in "machali"

Measure the pitch at several points in the voiced segments of the utterance. See how it varies with time across the utterance.

6. Prosody

Not all information conveyed by speech is encoded in the vocabulary and grammar. Aspects like age, gender, emotional state of the speaker and other elements like sarcasm, emphasis on an idea may also be present in the speech. These features make the prosody of the speech. These can be identified, to an extent using features such as phoneme duration, pitch and energy. Following few examples show some prosodic features in different occurrences of speech:

6.1 Segmenting

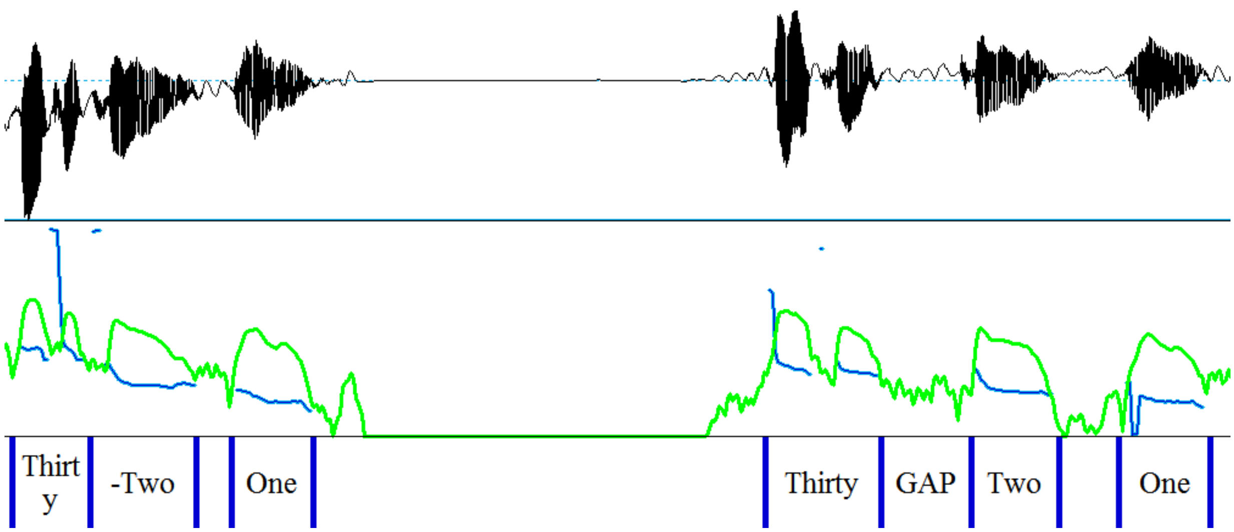


Figure 16 Waveform, Energy (Green) and Pitch (Blue) for the utterance 'Thirty-Two, One, Thirty, Two, One.'

In the above example, there are two utterances of 'thirty' and 'two'. However, the first instance corresponds to numerical 32 and the second instance is for numbers 30 and 2. The two instances can be differentiated by the gap between the 'thirty' and 'two' in both cases. The duration of the first instance of 'thirty' is less than then that of second instance. Also, the pitch in the trailing sequence of first instance of the word 'thirty' is different from the pitch in second instance.

6.2 Emphasis

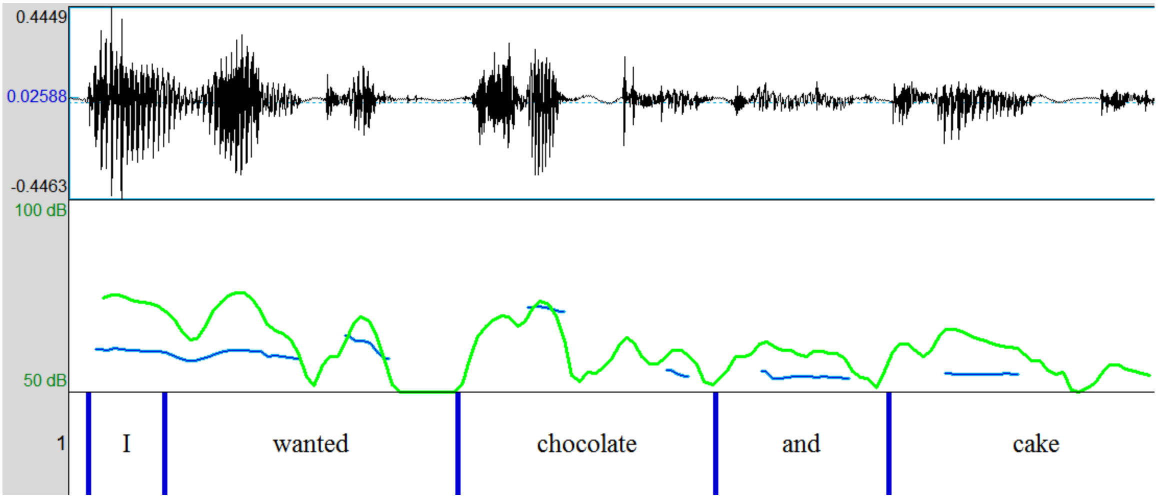


Figure 17 Waveform, energy, pitch of 'I wanted ***chocolate*** *and cake.'*

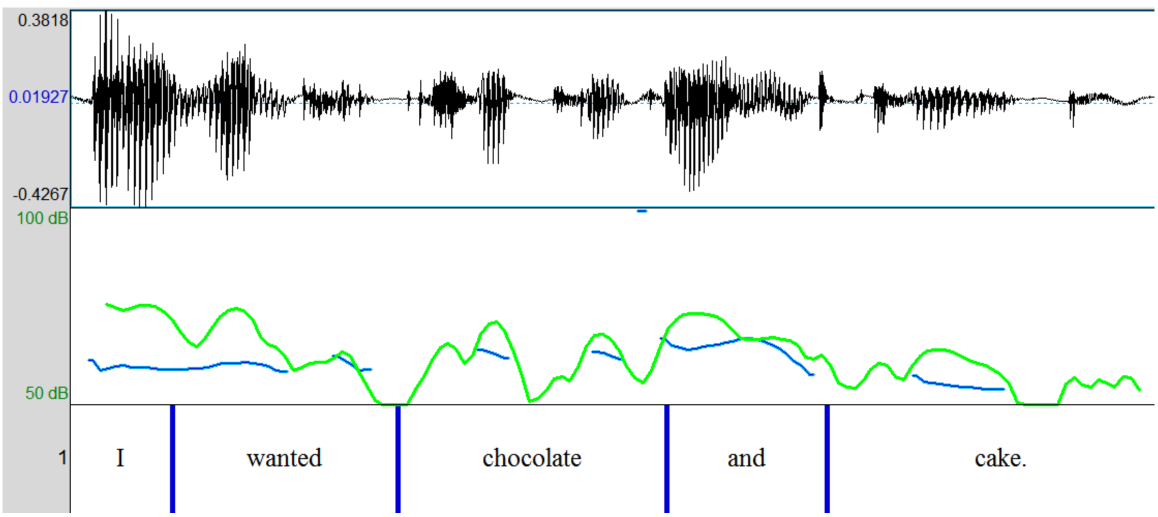


Figure 17 Waveform, energy, pitch of 'I wanted chocolate ***and*** *cake.'*

In the first sentence, the speaker emphasizes on the word 'chocolate' and in second word the emphasis is on 'and'. The pitch of the word 'and' is higher in the second case. For the word 'chocolate', pitch is largely absent because of the presence of two stops (/ch/ and /c/).

The intensity of the word which is emphasized is higher in both cases at the beginning of the word.

6.3 Tone

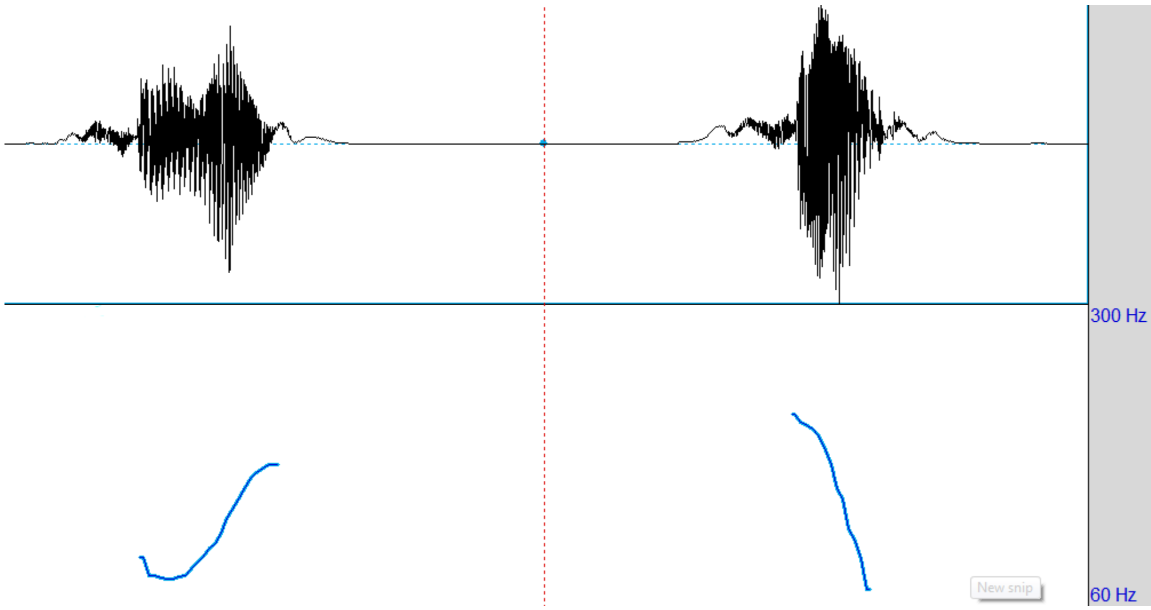


Figure 18 Pitch of two instances of 'so'

The variation of the pitch also conveys the tone of the speaker. In the example above, the speaker utters the word 'so' twice. The pitch rises in the first case and falls in the second case.

These are just a few examples of the features being used to describe the prosody. Some other examples may to find the end of a sentence using such features or to identify if a given utterance is a question or not.

References:

1. http://www.fon.hum.uva.nl/praat/manual/Intro_3_2__Configuring_the_spectrogram.html from **Introductory tutorial to Praat** available at http://www.fon.hum.uva.nl/praat/manual/Intro.html

2. http://www.fon.hum.uva.nl/praat/manual/Advanced_spectrogram_settings___.html

3. **Praat usergroup** - http://uk.dir.groups.yahoo.com/group/praat-users/message/2431

4. Accurate short-term analysis of the fundamental frequency and the harmonics-to-noise ratio of a sampled sound. IFA Proceedings 17: 97-110.

Further Reading:

1. Harris, F.J. "*On the Use of Windows for Harmonic Analysis with the Discrete Fourier Transform*." Proceedings of the IEEE. Vol. 66, No. 1 (January 1978)
2. Rob Hagiwara, *How to read spectrograms?* http://home.cc.umanitoba.ca/~robh/howto.html