

# Spectrum and spectrogram

## What is sound? A recap...

Sound is audible mechanical wave motion. A medium is needed in order for sound to be transmitted. In the usual case, when the medium is air, sound propagates as very small changes in air pressure.

Upon reaching the human ear, small variations in air pressure make the tympanic membrane oscillate. It transmits the movement to the chain of small bones in the middle ear that amplifies the sound and transmits it to the inner ear. The tiny pressure waves propagating in the jelly-like substance within the inner ear move the so-called basilar membrane. At different regions of the basilar membrane, there are auditory cells that are sensitive to sounds of different frequencies. If the cells are excited by a wave, they send a signal to the auditory nerve and to the brain, in order to be perceived and further interpreted.

The speed of sound in air is about 330 meters per second. However, not all substances conduct sound at the same rate. The speed of sound also depends on the temperature of the substance. When a sound wave encounters an obstacle, some of the energy of the sound wave is usually reflected back. The shape of an item affects its vibration properties.

An oscillating body that initiates waves perceived as sound is called a **sound source**. The sound source can be a solid oscillating body, but for example the column of air inside a pipe or cavity can also form an oscillating mass. In nature, friction and other forces affect both objects and air particles and dampen their movement. Thus, the oscillation and the resulting sound wave will gradually get weaker and they will eventually stop, unless a force keeps driving the oscillating movement.

The simplest possible sound can be described as a **sine wave**. The name comes from the sine function, which is a mathematical model of the shape of a sine wave or a **pure tone**. A pure tone has a certain **frequency**; a certain **amplitude**; and a certain **phase**, ie. the relative time at which a single oscillation begins and ends. In nature, however, pure tones are practically never present. Speech and all other natural sounds are always **complex**. Thus, you can think of any sound as consisting of a number of pure tones of different frequencies and amplitudes.

Using a method known as the **Fourier transform**, the complex sound can be decomposed into its **frequency components** or **spectral components**: a number of pure tones, each of which represents one frequency only. These components are also referred to as the partial frequencies, or simply **partials**. The partials are not perceived individually but together as one sound.

The perceptual impression of a specific sound, the **timbre**, depends on what partial frequencies it includes, what the relative intensities of the partials are, and how they change in the course of time. You should also bear in mind that in computers, digital sound is often analyzed as brief, temporally overlapping clips, whereas humans perceive sound as a continuous signal. What we hear and perceive will also affect what we expect to hear – and what we expect will affect what we end up hearing.

## Spectral analysis

We have learned that the partial frequencies and their intensities can be determined from an audio clip by **spectrum analysis**. However, virtually all sounds that can be recorded from nature - including speech - are constantly changing. The shape of the spectrum depends on the type and duration of the sound sample that was selected. To some extent, we are facing a dilemma: If we analyze an audio clip that is too long, the possibly interesting changes that take place **inside** the audio clip will not be detected in the resulting spectrum. If, on the other hand, we analyze an audio clip that is too short, we will not be able to detect those events that last longer than the duration of the analyzed sample. We will see that this problem is also reflected in spectrograms, where we need to make a choice between good time resolution and good frequency resolution.

### How to inspect the sound spectrum in Praat

In the Sound or TextGrid editors, you can calculate a spectrum either from a narrow, automatically selected window around the cursor or from the selected portion of the sound by selecting the command **Spectrum: View spectral slice**. The horizontal axis of a spectrum denotes the **frequency** of the partials and the vertical axis shows the relative **intensity** (in decibels, *dB*). Thus, time is completely excluded from the spectral slice, and the spectrum only displays the overall information about the partials detected within the selected time window in the original sound. **Spectrum** settings can be adjusted in the editor from the same point as the **Spectrogram** settings. More information: [https://www.fon.hum.uva.nl/praat/manual/Intro\\_3\\_7\\_Configuring\\_the\\_spectral\\_slice.html](https://www.fon.hum.uva.nl/praat/manual/Intro_3_7_Configuring_the_spectral_slice.html)

## Spectrogram: consecutive and overlapping spectra over time

A spectrogram is a visualization method that can be used to illustrate the changes that occur in the frequency structure of sound over time. From the spectrogram, an overall understanding of the frequency and energy distribution of the audio signal can be quickly obtained. The spectrogram is based on a series of spectra that are calculated from time windows of a given width that are moved in small steps from the beginning to the end of the sound sample.

Spectrograms can be drawn in many ways. The traditional spectrogram, available in Praat by default, shows the energy density (average intensity) of sound in grayscale: the darker the area in the spectrogram, the higher the average intensity found in that frequency range at that time in the sound signal ([https://www.fon.hum.uva.nl/praat/manual/Intro\\_3\\_1\\_Viewing\\_a\\_spectrogram.html](https://www.fon.hum.uva.nl/praat/manual/Intro_3_1_Viewing_a_spectrogram.html)). Sometimes intensity is described in different colors instead of grayscale. You might also come across spectrograms where the time axis is plotted along the "depth" axis in the image, i.e., from front to back. Depending on the purpose, spectrograms can even be animated or calculated in real time, as shown in this video by Mark Newman:

<https://youtu.be/ZRZlZ81nXo4>

## Broadband and narrowband spectrograms

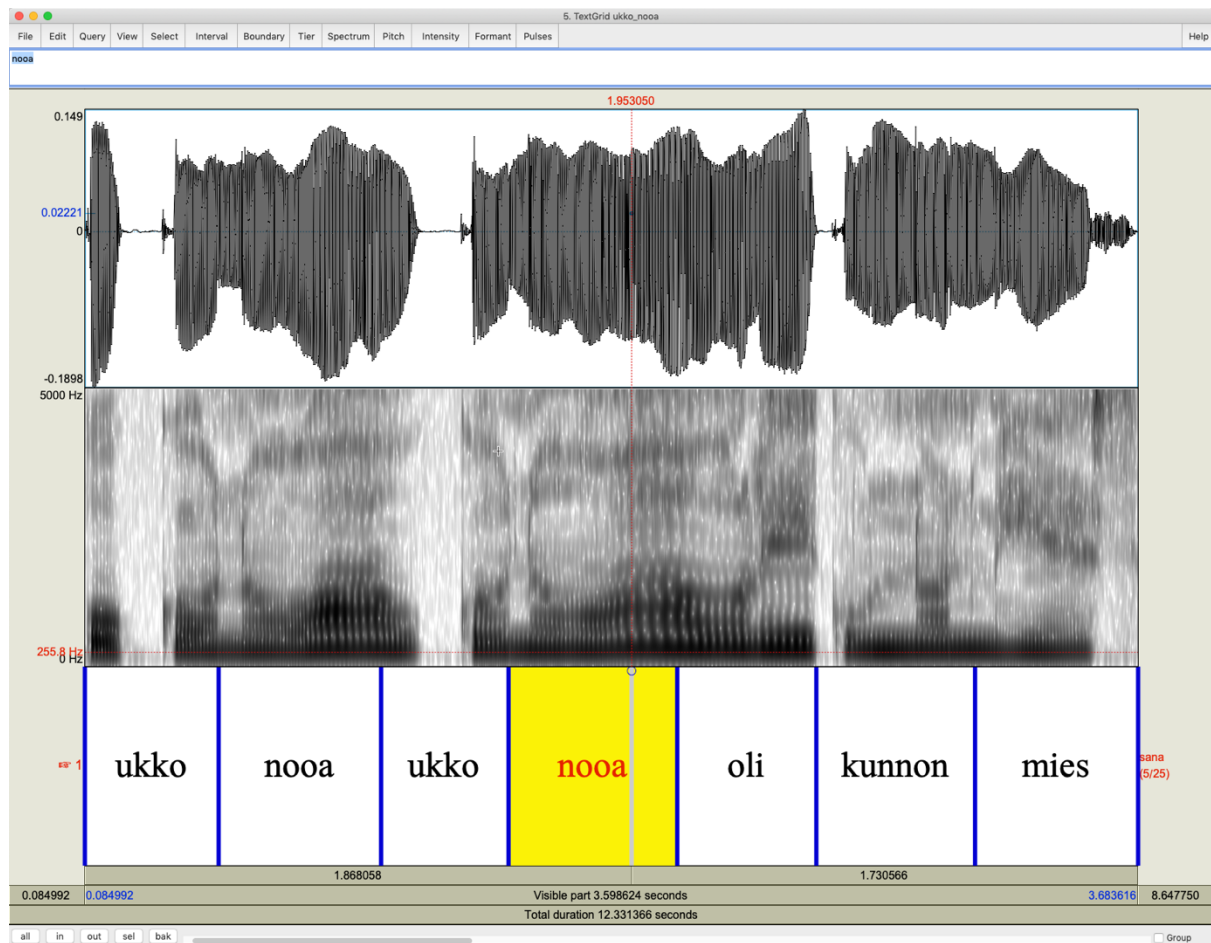
Two types of spectrograms, wideband and narrowband, have traditionally been used for different purposes. By applying different settings, the spectral properties of speech can be viewed from different perspectives.

A **wideband spectrogram** (analysis window length approx. 5 ms; default in Praat) has good time resolution, i.e., it can reflect changes over time more accurately. For example, in a wideband spectrogram, the energy pulses originating from individual glottal periods are usually reflected as visible vertical stripes during the voiced portions of the speech signal.

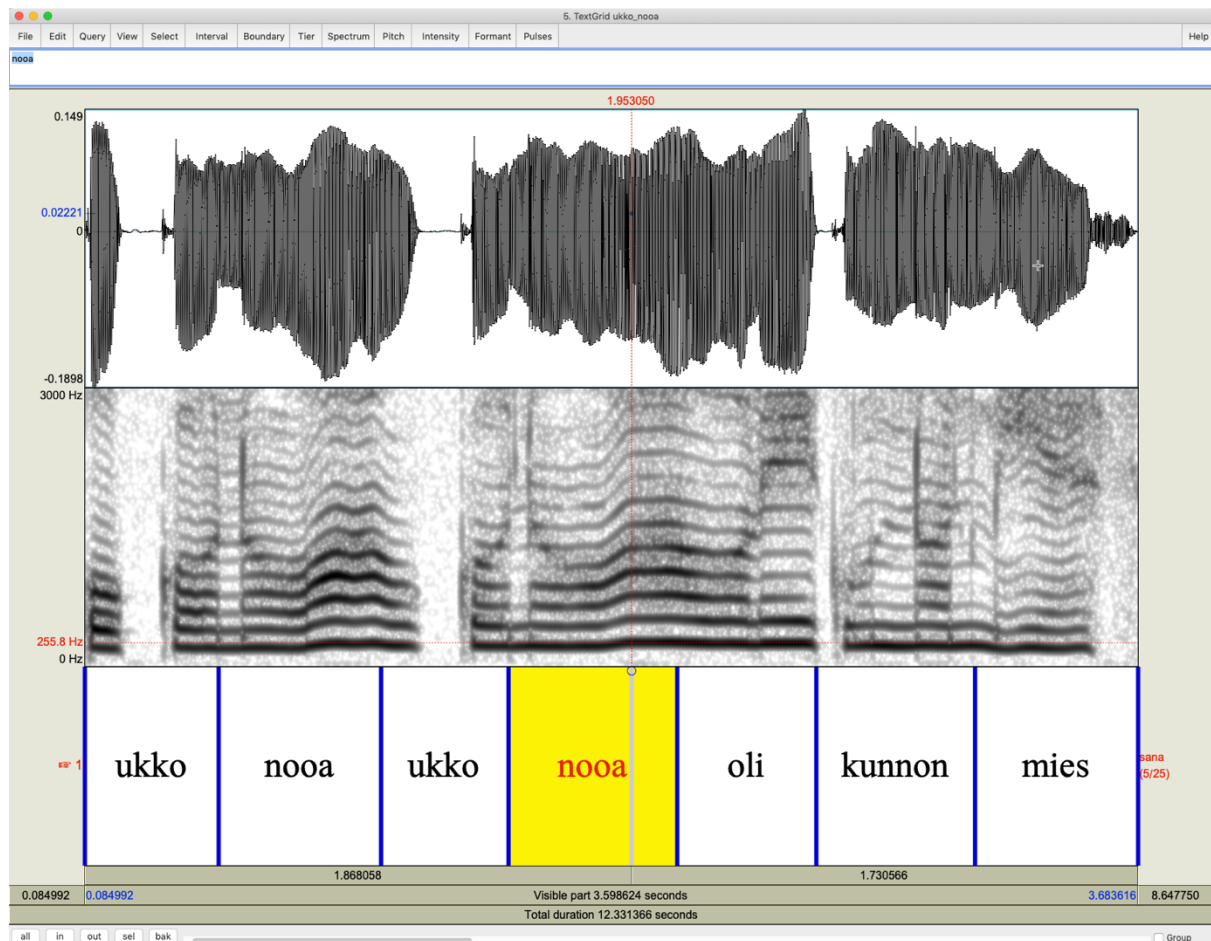
The wideband spectrogram is useful when you are segmenting and labeling intervals in a speech sample, for example, for locating the boundaries of individual speech sounds. Since spectral changes over time are accurately reflected, you are able to find potential spectral “turning points” more quickly, and these are often good candidates for segment boundaries. (Of course, the boundary locations should still be confirmed by listening.)

In a **narrowband spectrogram** (analysis window length approx. 30 ms), the frequency resolution is better than time resolution, i.e., the spectral components will be located more accurately with respect to the frequency scale. Especially in vowels and other voiced sounds, you can see the individual partials of the sound (in voiced sounds, the partials can also be referred to as the **harmonics**). It may be easier to locate the center frequencies of **formants** (displayed as darker, curved horizontal bands) that can be important if you are studying vowel quality and articulatory phonetics. However, a narrowband spectrogram will not display fast temporal changes.

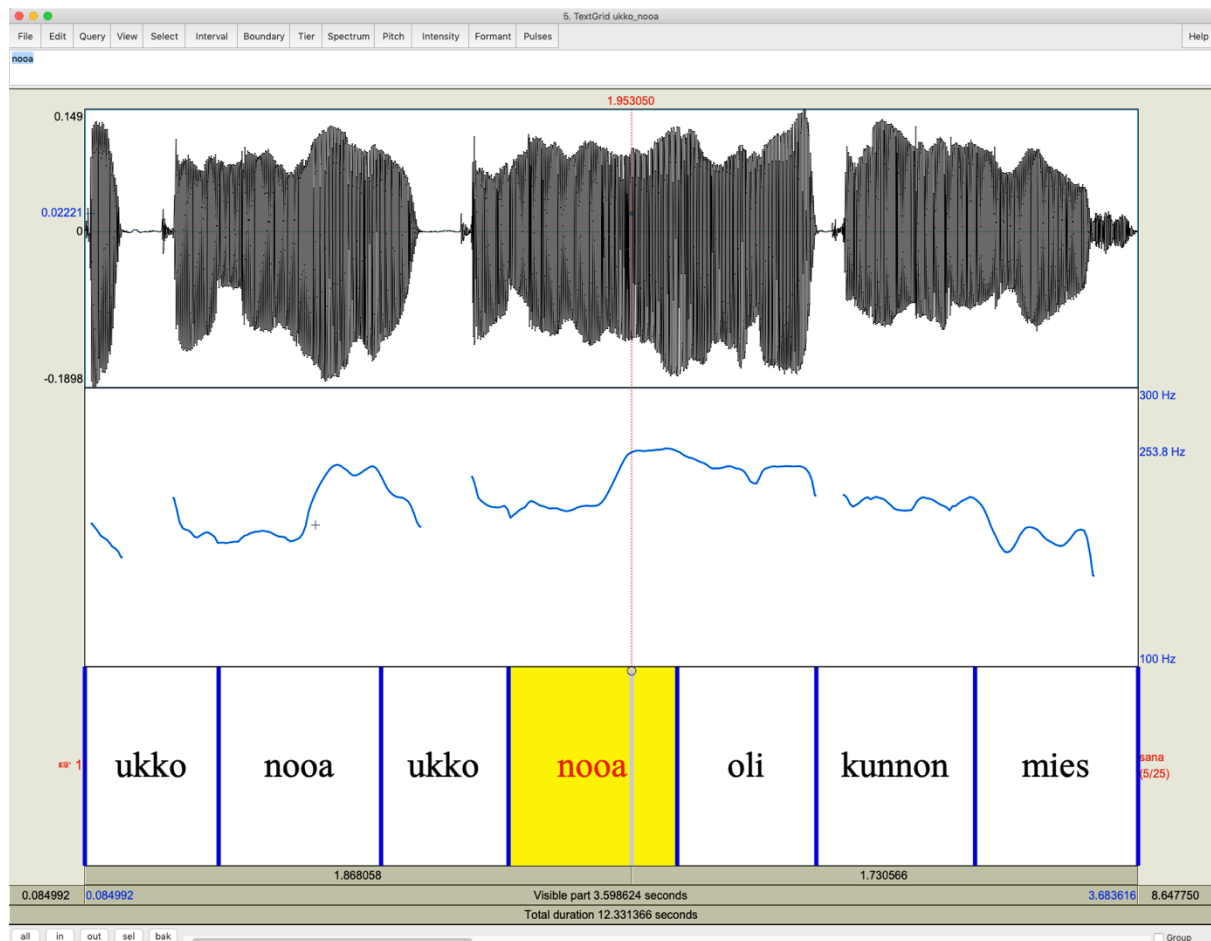
The narrowband spectrogram can be useful as an additional confirmation step for your pitch analysis. You can set only the lowest frequencies of the spectrogram to be displayed in the editor window, so that the fundamental frequency and a few of the lowest harmonics will clearly stand out. You can then display the pitch curve simultaneously with the spectrogram and set an identical display range for both the spectrogram and the pitch plot. In principle, the pitch curve should follow the lowest harmonic, i.e., it should be detecting the fundamental frequency.



A wideband spectrogram of a verse from the song Ukko Nooa sung by a female speaker. The width of the analysis window is set to 0.005 s (= Praat default).



A narrow-band spectrogram of a verse from the song Ukko Nooa sung by a female speaker. The length of the analysis window is set to 0.03 s. The visible frequency range is set to 0-3000 Hz in order to better distinguish the lowest harmonics. The red vertical cursor is set within the last syllable of the word *nooa* at the time point of 1.953050 s. By visual approximation, the horizontal line has been placed in the center of the lowest harmonic in the spectrogram, which should correspond to the fundamental frequency of the sound. The frequency of the horizontal red line corresponds to the frequency 255.8 Hz in the spectrogram image, which is displayed in red on the left side of the window. Note that this is not an exact measurement but an estimate based on the spectrogram. This information can be compared with the result of the fundamental frequency analysis shown in the following figure, as the cursor is positioned at the same time point in the sound sample in both figures:



Waveform and fundamental frequency curve (Pitch) from a verse of the song Ukko Nooa sung by a female speaker. The minimum frequency for pitch analysis was set to 75 Hz in the **Pitch settings** of the Praat Sound editor window, and the maximum frequency was set to 500 Hz. The visible frequency range is set to 100-300 Hz in **Advanced pitch settings** to better distinguish between fundamental frequency movements. The cursor (red vertical line) remains at the same time point as in the previous two images. The fundamental frequency value 253.8 Hz measured from the pitch curve at the cursor is displayed in blue text on the right side of the window.

## Viewing spectrograms in Praat

### In the editor window

To display the spectrogram in the Sound editor or TextGrid editor, select **Show spectrogram** from the **Spectrum** menu. The most important settings of the spectrogram can be changed in **Spectrogram settings...** under the **Spectrum** menu.

- If you wish to calculate a **wideband spectrogram** (the default in Praat), the length of the analysis window (window length) should be about 0.005 s (ie., 5 milliseconds, which results in a frequency bandwidth of 260 Hz in the spectrogram).

- If you want to calculate a **narrowband spectrogram**, the length of the analysis window should be about 0.03 s (ie., 30 ms; in this case the so-called bandwidth is 43 Hz).

### In the Object window

In the Object window, a new *Spectrogram* object can be created by first selecting the desired *Sound* object and pressing the **Analyze Spectrum: To Spectrogram...** button in the dynamic menu.

## How can spectrograms be used?

Spectrograms are just one possibility for visualizing the frequency structure of sound as a function of time. A spectrogram provides a nice overview of a given audio signal, but it does not constitute an actual research result or statistical data. Nevertheless, spectrograms of individual examples extracted from a speech corpus can be used as illustrations in research articles, as long as you are able to interpret the figures in a meaningful way.

In my experience, the spectrogram is best used as an exploratory instrument in research. It allows you to browse through your speech material and find more elaborate ideas for your research. As you gain more experience, you may learn to recognize different speech sounds and other specific acoustic phenomena just by looking at the spectrogram. Phoneticians even like to organize "Mystery Spectrogram" contests where you need to guess what a spectrogram is "saying"