Formants

Formant analysis of speech has a long tradition, especially in the study of vowels. The formants are related to the shape of the speaker's vocal tract (the pathways between the larynx to the lips and the nostrils) during speech production, and so the relative changes in formant frequencies can, to some extent and under certain conditions, be used for detecting some articulatory properties of individual sound segments. However, caution is required in interpreting the results.

In the following, I will try to explain the general principle of formant analysis, without going into too many engineering details or mathematical formulae…

What is a formant?

Speech is always a complex sound in which a large number of spectral components or partial frequencies can be detected by means of spectrum analysis. The sounds generated by the speaker may originate in the larynx and/or other parts of the vocal tract, depending on the type of sound. Depending on the shape of the vocal tract where the sound travels and is reflected back and forth, different partials of the sound may be either amplified or dampened, i.e., some frequencies will increase whereas some frequencies will decrease in their amplitude. Some proportion of the sound energy will eventually radiate to the outside world through the mouth and/or nose.

If we try to simplify things a bit, the vocal tract can be compared to a chain of tubes of different lengths and thicknesses. Within certain limits, a person can change the shape of his vocal tract, for example by moving his tongue, lips, etc., thus changing the number, lengths and thicknesses of the individual "tubes" or sections within the "pipeline" .

Any tube amplifies all sound waves whose wavelength (inversely proportional to the frequency of the wave) has a suitable proportion to the length of the tube. When sound travels in a tube, each end of the tube - even an open end! - will cause a certain proportion of sound energy to be reflected to the opposite direction. Thus, in the vocal tract, sound waves are constantly moving in different directions, and waves keep meeting other waves.

> Imagine you are sitting on a swing and your friend is giving you a push. When your friend is standing in front of you or behind you and periodically pushing the swing at just the right moment, your swing will gradually swing higher and higher. It would be pointless for him to try and push the swing when it is not in the right place! If your friend moves to the middle of the path of the swing and tries to push you back when the swing approaches him, he will hardly succeed, but the swing will most likely stop (and your friend might fall and hurt himself). So, timing is crucial.

If two sound waves pass through each other at a point where both happen to be in the same phase (e.g., the air pressure peaks of both waves occur simultaneously at the same location), a wave with temporarily increased amplitude occurs at that point. So, the colliding waves add up. If, on the other hand, the waves happen to be in opposite phases when they meet, the total amplitude of the waves at that point is reduced, i.e. they will cancel each other. Under given conditions, the waves inside the tube are amplified. This property of the tube is called **resonance**. The length of the tube determines the sound frequencies that can resonate in it.

> Imagine your friend pushing the swing only at every second or third swing. That is a kind of resonance, too!

A formant

corresponds to one or more resonances of the vocal tract, i.e., a frequency at which sound waves are amplified in some part of the vocal tract. The formants can be distinguished as "ridges" or "peaks" in the spectrum of the speech signal. In the spectrogram, formants appear as horizontal dark bands that are most clearly visible during voiced sounds, especially vowels.

The number of formants and their center frequencies change almost constantly during speech - sometimes the changes are faster, sometimes slower. The frequencies and movements of the formants have been shown to be linked with the identification of speech sounds. This is only natural: formants reflect the ways in which the speaker has shaped her vocal tract.

The resonances of the vocal tract will of course affect any sound that passes through the vocal tract. Therefore, formants can sometimes also be distinguished during unvoiced sounds, such as fricatives. In vowels, however, it is much easier to inspect formants, since the speaker's vocal tract remains fairly open and the number and frequencies of the formants are relatively predictable. There are also many previous studies of vowel formants to which new results can be compared.

In principle, the vocal tract has several resonances at any given time, but only a few of them can clearly affect, for example, the perceived quality of vowels. Usually, no attempt should be made to identify more than five lowest formants in the spectrum of an individual vowel, since the human ear can hardly distinguish changes in formants in higher frequencies.

In formant analysis, a simplified model of the spectrum of the speech signal is first calculated by using the so-called LPC method, which can be used to estimate the center frequencies of formants (the frequencies that the vocal tract has amplified the most) and the bandwidths of the formants (how wide a frequency range is affected by the amplifying effect of each individual formant).

In practice, speech involves a nearly constant motion of the vocal tract. LPC is a mathematical model of the sound spectrum, so formant analysis can only provides a rough approximation of the likely shape of the vocal tract at that point in time. Automatic formant analysis with any piece of software is a computational estimate of the most dominant peaks in the spectrum. The operating principle and sources of error of formant analysis should be understood if the method is to be applied to a particular research question.

Performing formant analysis in Praat

As a result of the formant analysis performed with the Praat program, a Formant object is created, which represents the spectral structure of the sound object as a function of time. It consists of a series of evenly spaced samples, each of which may contain frequency and bandwidth information from several formants and the maximum intensity of the sound within that time window. The Formant object is thus a kind of simplified version of the spectrogram.

The default formant analysis of the Praat program is the **Burg** algorithm used with the **Analyze Spectrum: To Formant (Burg)...** command.

How formant analysis works

Initially, the audio signal is resampled to a sampling frequency that is twice the upper frequency limit of the formants given in **Maximum formant**. The signal is then pre-emphasized so that the peaks in the higher frequencies that are inherently attenuated in speech will also end up having the same weight as the lower formants. From the resulting signal, short-term spectra are calculated at certain intervals (the distance between the spectra and the width of the spectrum analysis window are defined in **Time step** and **Window length**).

The spectrum obtained in each analysis window or frame is then approximated by the *Linear Predictive Coding (LP)* method, which uses the Burg algorithm here. In linear prediction, the aim is to describe the shape of the spectrum with a small number of peaks, for each of which the center frequency and bandwidth are estimated. These peaks can be considered to represent vocal tract resonances, i.e., the formants. The result of LPC is, in fact, a set of coefficients that would not in themselves be at all illustrative for a human. Therefore, in the final stage of analysis, the information included in the coefficients is refined into the frequency and bandwidth values of the formants.

> Note. A Formant-type object produced as described above contains only those formants that were detected in the signal in each window, and the frequency values may fluctuate a great deal between successive analysis windows. This basic analysis does not attempt to search for "continuity" between successive formant values. If you want to study, for example, formant movements within a vowel segment, you can first perform the basic formant analysis and proceed with the **Tracking** feature, see below.
>
> The formant algorithm initially finds formants at very low and high frequencies, so the normal Burg algorithm also removes formants below 50 Hz and formants that are higher than the parameter *Maximum formant* - 50 Hz. If for some reason you absolutely want to include these frequency bands (in which case you will hardly get traditional-looking F1 and F2 values), try A**nalyze Spectrum: To Formant (hack): To Formant (keep all)…** If you want to always get the same number of formants evenly distributed over the entire frequency range , you can try the otherwise unreliable Split-Levinson algorithm with the command **Analyze Spectrum: To Formant (hack): To Formant (sl)…** However, the previous commands under the **Analyze Spectrum: To Formant (hack)** menu are not generally recommended.

Performing a Burg formant analysis

Method 1:

If you just want to view the formant values computed by Praat, for example, along with the sound waveform and spectrogram, do the formant analysis inside the sound editor window.

1. From the object list, select the sound object (of type **Sound**) for which you want to perform formant analysis.

2. Click the **Edit** button on the right side of the object list to display the audio editor window.
3. In the sound editor window, select Show formants from the Formant menu. The formant analysis appears below the sound waveform as red dots. The formant points are always displayed in the same frequency scale as the spectrogram. Read the frequency values on the left side of the window!
4. Check the analysis settings in the Formant settings section of the Formant menu… For the time being, only the standard Burg algorithm can be used to perform formant analysis in the sound editor. The settings are otherwise the same as in analysis method 2.
5. If a long portion of the audio signal is visible in the window, the formant analysis might not be displayed for the entire window. If you want the analysis to be calculated over a longer period of time, select the command Show analyses… in the View menu and change the number of seconds shown in Longest analysis (s). Note, however, that the formants are recalculated each time you scroll or zoom in on the editor window, so a long formant analysis can slow down your work. It's a good idea to turn off formants in the editor whenever you don't need them.
6. If you want, you can also take measurements in the audio editor as follows:
    - approximate measurements by clicking with the mouse on a red formant point (the frequency at the mouse location is shown in red on the left side of the window), or
    - for more precise measurements, click on the waveform or spectrogram and then select, for example, Get first formant from the Formant menu, in which case the Info window will display the frequency value of the 1st formant that is closest to the time at cursor. (These formant commands only work if formant analysis is shown.)

Method 2:

When you want to analyze in more detail, take accurate measurements, draw images, or use formant analysis within a script, create a formant object separately in the object list.

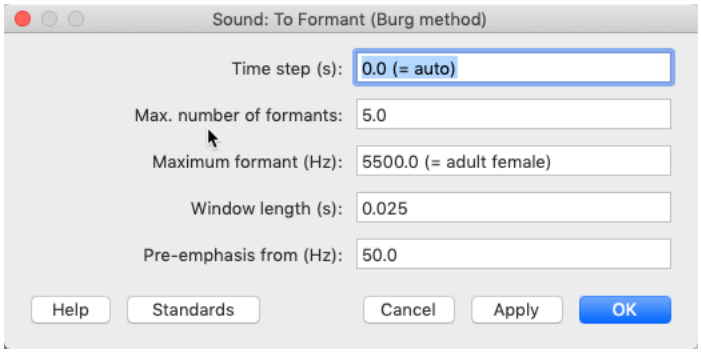
1. From the object list, select the sound object (type Sound) for which you want to perform formant analysis.
2. Click the Analyze spectrum button on the right side of the object list and select To Formant (Burg) from the drop-down menu.
3. Make sure the formant analysis settings are appropriate:
    - **Time step (seconds):** Time between the centers of successive analysis frames or windows. If the audio object to be analyzed is 2 seconds long and the time step is 0.01 seconds, a total of about 200 frames are analyzed. However, the actual number is slightly smaller because measurement is more difficult at the edges of the sound sample.
    - **Max. number of formants:** The maximum number of formants to search for. For human speech analyzes, you should usually use a value of 5. If the Maximum formant parameter is also set correctly, this is the only way to get sensible results.
    - **Maximum formant (Hz):** The upper limit of the frequency of the detected formants. This value must be set according to the speaker being analyzed. A value of 5000 Hz can be used for a low-pitched speaker. The default value of 5500 Hz is suitable for a slightly higher-pitched speaker, for example, an

average adult woman. For the speech of young children, on the other hand, it may be justified to use clearly higher values, e.g. 6000-8000 Hz.

- The optimal upper frequency limit can be found by experimenting with for instance isolated vowels and looking for a setting where the formant strips formed by the red dots appear to follow the smoothest possible paths during the vowels under study. If the upper limit setting is not optimal, the upper formants in particular will tend to "wander around" a bit.
- Too high an upper limit, on the other hand, can cause two formants in lower frequencies to be interpreted as a single formant, as the algorithm tries to find the 5 formants set in the previous section so that they are separated as well as possible. For example, for a [u] vowel pronounced by a low-pitched man, two adjacent formants should in principle be found below 1000 Hz, but setting the maximum formant too high may result in a combination of two formants for F1 and the remaining formants will be pushed too high.

- **Window length:** The effective duration of the analysis window or frame. (The actual calculation window is twice as long because Praat uses a Gaussian-shaped window with edges close to zero.)
- **Pre-emphasis from (Hz):** Lower limit of the spectrum pre-emphasis (+ 3dB limit for an inverted low-pass filter with an angle of + 6dB / octave). Usually, the vowel spectrum attenuates towards the high frequencies by about 6 dB per octave. However, the formant analysis also tries to find local peaks at upper frequencies, even if they are weaker than the formants at the low end of the spectrum. Therefore, the spectrum is filtered before formant analysis so that the upper frequencies increase and the tilt of the spectrum decreases.

4. Finally, press OK. A new formant object appears in the object list.



Formant analysis settings.

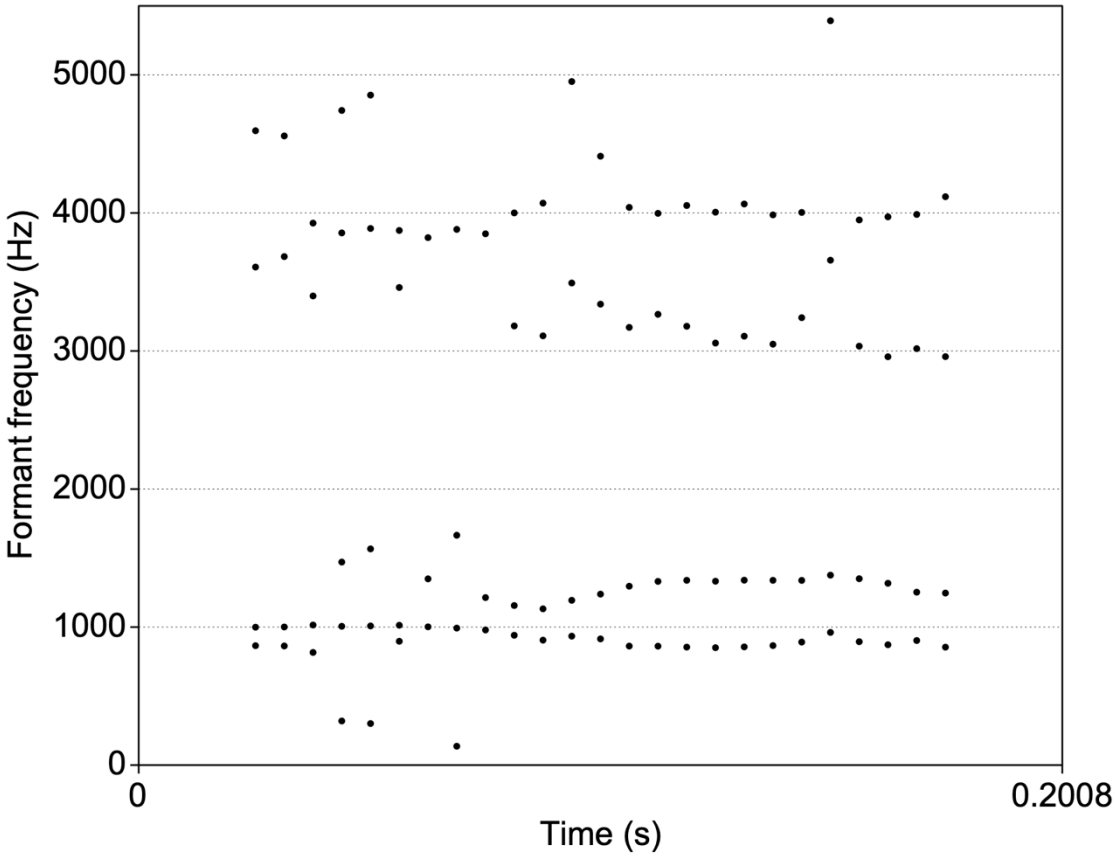
Tracking

If you want more continuous and consistent-looking formant curves, calculate the Formant object first as described above in method 2, and select the object from the object list. Then press the **Track…** button on the right side of the object list. This command seeks to detect the same number of formants within every analysis window (frame) and tries to find the most direct "path" between the formant points in adjacent windows. In order to display, for

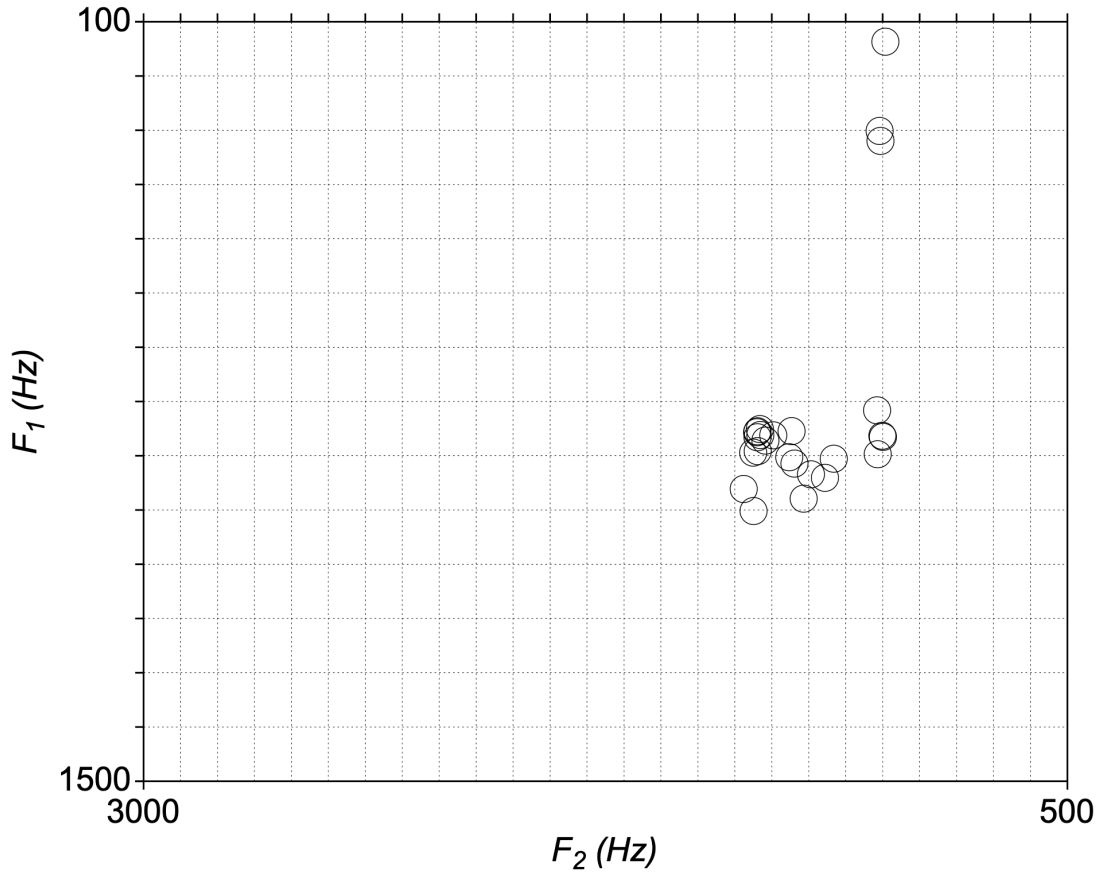
example, 3 formant tracks, each analysis window of a Formant object must contain at least three formant candidates (ie., you should calculate the original formant object with e.g. 5 formants and then use the Track command).

Drawing formant pictures

Here are a few example plots created in Praat:



Select the Formant object and press Draw: Speckle…



Select the Formant object and press Draw: Scatter plot…

Tip: With Praat scripts, you could also plot vowel characters inside circles, for instance!

Drawing an F1 / F2 formant chart

It is possible to draw a traditional F1/F2 vowel chart, which often appears in the literature, if you first calculate a Formant object from a vowel or some other uniform sound segment, for example by using the Burg analysis. Then do the following:

1. Select the Formant object and select the **Draw - Scatter plot...** command.

2. In the form that appears, specify the lower and upper limits for F1 and F2 in Hertz. The boundaries should be chosen so that they just cover the range of all vowel qualities for that speaker, so that the different vowels will be as clearly separated as possible and in places that are easy to interpret. For a low-pitched adult, such as the average male speaker, a range of 200 to 1000 Hz for F1 and a range of 500 to 3000 Hz for F2 are usually OK. Normally, the boundaries should be given in the default order, i.e., “backwards”: the second formant on the horizontal axis and the first on the vertical axis, and the maximum frequency first and the minimum frequency second. This way, the image is drawn so that the measured formant points will roughly correspond to the order of the so-called vowel quadrilateral: front vowels to the left,

back vowels to the right; close vowels up and open ones down.

3. The shape and size of the character to be drawn can be selected at the bottom of the form (**Mark size** and **Mark string**).

4. You can change the drawing color from the Pen menu in the Picture window before accepting the form. Also, be sure to select an area of appropriate size and shape in the Picture window, so the formant chart will be scaled in the way you want.

5. Click OK.

The result should be a vowel chart with the minimum of each formant in the upper right corner i.e. the image should roughly correspond to the charts found in the literature. If a Formant object is calculated from a sound sample that is long enough to hold multiple calculation windows, each of these formant points is drawn in the image. This is useful if you want to draw a formant chart from a diphthong, for example, so that the movements of the formants during the vowel are visible.

You can overlay other Formant objects in the same image. Since the “vowel spaces” of different speakers are slightly different in size due to physiological differences, it usually makes sense to draw the formants of only one speaker in the same picture.

The formant chart can be saved to an image file.

Sources of error in formant analysis

Formant analysis is based on a theoretical model in which an acoustic speech signal consists of a source sound (e.g., the “buzzing” noise produced by the larynx) and the filter properties of the vocal tract (e.g., formants). If the measured sound has been affected by some external factors (such as the properties of the room, other speakers, or a technical disturbance), you may end up getting distorted results.

Formant analysis always requires some interpretation: the researcher assumes that the formants are found "where they should be found." The reference values used by researchers are based on numerous studies, in which the formants of clearly pronounced vowels have generally been measured from data controlled by certain parameters, or on a schematic model of the structure of the "average" vocal tract. The results of the formant analysis must not be considered as purely objective figures, but must be taken as relative to the speaker in question, considering the phonetic environment, the analysis parameters and the quality of the recording. Many contextual factors may alter the frequencies and bandwidths of formants, and some features of speech (e.g., nasality) may further complicate the interpretation of formants. Speakers are always individuals, and the results may not always be what you expect.

Formants should only be measured at points where there is no overlapping speech. This ensures that the voices of other speakers do not interfere with the analysis, as formant analysis cannot distinguish between different audio sources. Other background noise or strong reverberation in the recording room can also cause erroneous results.

Conventional formant analysis settings are best suited for voiced sounds, especially vowels. Of course, there are formants in all sounds, but with unvoiced sounds the analysis parameters may not make sense and interpreting the results is challenging. It is also not advisable to use the Track… command at time points where a transition between the consonant and the vowel is close by, because the number of formants may rapidly change, making it impossible to find formant tracks that make sense.

When formants occur in close proximity, such as F2 and F3 for the vowel [y] or F1 and F2 for the vowels [u] and [o], there is a risk that formant analysis will interpret adjacent formants as one single peak. This often happens if the usual five formants are searched in a wide frequency band, for example if you set the Maximum formant too high for a low-pitched male speaker.

The computationally common reason for formants "merging" or "fluctuating" is that a sufficient number of spectral coefficients has not been used for the LPC analysis (this number is proportional to the number of formant peaks). A sufficient amount is at least the sampling frequency of the signal in Hertz divided by one thousand (e.g. 16 for a 16 kHz signal). Praat's Burg formant analysis automatically samples the signal first to a sample frequency twice the Maximum formant, and if (2 * Max. number of formants) is proportional to this frequency, a sufficient number of coefficients are automatically calculated and the analysis should be reasonable. However, you need to consider even more carefully the right number of coefficients if you are not using Praat's formant analysis directly but are doing a separate LPC analysis of the Sound object. In this case, you must either re-sample the signal yourself to the appropriate sample frequency before the LPC analysis, or supply a sufficient Prediction order for LPC, e.g. for a 16 kHz audio sample you need the order of 16.

Further information about formant analysis

In Praat's internal manual you may want to read the tutorial page Source–filter synthesis (search for the keyword "source-filter" to find the page), which describes the source-filter theory of speech. The manual also provides instructions on how Praat can be used to calculate the sound source and/or the vocal tract filter functions from a speech sample. The manual page is especially useful if you are interested in the speech synthesis features available in Praat.