

IMPROVED FORMANT FREQUENCY MEASUREMENTS IN PRAAT

David Weenink

University of Amsterdam

David.Weenink@uva.nl

ABSTRACT

We present a way for calculating and visualising different formant frequency measurements of the same sound. This will enable a researcher to easily pick, for any interval in the sound, a desired analysis from the available alternatives. To achieve this, we first calculate the alternatives by performing multiple formant frequency analyses of the same sound with different formant ceilings. By varying the ceiling we can adapt the frequency interval of the spectrum to the fixed number of formant peaks that are used to model it. The results of these multiple formant analyses are collected in a FormantPath data structure. A FormantPathEditor makes it possible to view and edit the available formant options. Optionally a corresponding TextGrid can also be shown in the FormantPathEditor. Both the FormantPath as well as the FormantPathEditor were recently added to the sound processing program Praat[1].

Keywords: Formant frequency measurement, variable ceiling, visualisation, trajectory smoothness.

1. INTRODUCTION

Formant frequency measurements have a long tradition in phonetic research and are considered to be important for determining vowel quality. However, formant frequency values are notoriously difficult to measure automatically. Most of the time they are measured by performing some type of linear predictive analysis (lpc), sometimes accompanied by manual intervention.

The *source-filter* model of speech production on which lpc is based assumes that the speech signal is the result of filtering a source signal by the filter characteristics of the vocal tract. The number of filters to model the vocal tract is fixed during the lpc analysis and has to be chosen beforehand. Reasonably good results are obtained when the model and the data fit well. If model and data do not fit well, poor formant frequency estimates may result. For example, if the number of filters is chosen larger than the actual number of formants present

in the sound, spurious formant tracks will appear. On the other hand, if the number of filters was chosen too low, then some formants might not be measured at all, or, two or more formants might be averaged which would result in formant frequency estimates at the wrong frequency. Many of the deficiencies of lpc analysis have been known for a long time, see for example [2] or more recently [3]. Often after the measurements a lot of post-processing has to be done. For example, [4] mention their tedious work to measure the acoustic characteristics of 12 American English vowels in h-V-d syllables spoken by 45 men, 48 women and 46 children. They used various linear predictive analyses with different number of coefficients to estimate the formants and very often based their final decision on the steadiness of the formant contour and sometimes also on the experimenter's acoustical phonetic knowledge about the identity of a particular vowel. In another example [5], the authors mention that in their estimation of vowel formant frequencies of Northern and Southern Dutch they had to manually adjust the results in 20 to 25% of the cases. Many more examples could be cited. They all show that a fully automatic perfect formant frequency measurement does not seem to exist and therefore different ad hoc procedures had to be used to improve upon the results obtained by automatic measurements.

In this article we build on a promising lpc-based method that was presented in [6] to overcome some of the deficiencies of standard linear predictive formant frequency analysis, notably the possible mismatch between the number of analysis filters, i.e. the number of lpc coefficients, and the number of formants present in the sound. Essentially our method, instead of performing one analysis with fixed parameters on a sound, performs a multiple of analyses each with a different analysis parameter setting. For each of these analyses a kind of stress value is calculated which can function as an overall smoothness measure of the formant tracks. The analysis parameter that we vary between the different analyses is the so called *formant ceiling* which is the maximum frequency of the formant search range. Because a formant can be

associated with a peak in the spectrum, lpc analysis is equivalent to the matching of a fixed number of peaks to the frequency spectrum of a sound. Because we cannot know beforehand the optimal width of the frequency interval on which a fixed number of formant peaks fits best, we just have to use multiple interval widths. This is equivalent to the use of multiple ceilings in the formant analysis. We have named the data structure that contains all the different formant analyses a *FormantPath*. The visualisation and manipulation of a *FormantPath* is taken care of in a *FormantPathEditor*.

In the following sections we describe the *FormantPath* and the *FormantPathEditor* in somewhat more detail and we show an example of its use.

2. THE FORMANTPATH

A *FormantPath* object in Praat maintains a path through a collection of *Formant* objects, where each *Formant* is the result of a formant frequency analysis of the same sound but with a different setting of the formant ceiling parameter. These *Formant* objects all have the same time domain, the same time at the first frame, the same time step and the same number of frames.

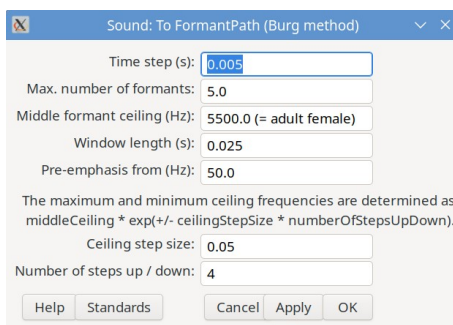


Figure 1: The form to obtain a *FormantPath* object from a *Sound* object.

To obtain a *FormantPath* object from a selected sound we use the **To FormantPath (burg)...** command. Fig. 1 shows the associated form. The first five entries in the form are almost equivalent to the entries in the default **To Formant (burg)** analysis. Two parameters, *Ceiling step size* and *Number of steps up / down* are extra. The last parameter determines how many analysis will be done. For the default number in the form above, 4, there will be 4 analyses with a ceiling higher than the *Middle formant ceiling*, 4 analyses with a ceiling lower than the *Middle formant ceiling* and one analysis with exactly this ceiling, making nine

analyses in total. The *Ceiling step size* defines how much the increase or decrease in formant ceiling frequency between two successive analyses will be. The ceiling frequency for the k^{th} step up will be calculated as $middleFormantCeiling \cdot \exp(k \cdot ceilingStepSize)$.

For the form given above which shows the default middle frequency of 5500 Hz for a female speaker, a ceiling step size of 0.05 and 4 steps up and down, this would result in the following rounded nine values for the formant ceiling in hertz: 4503, 4734, 4977, 5232, 5500, 5782, 6078, 6390 and 6718. The top ceiling for a female voice was calculated as $5500 \exp(4 \cdot 0.05)$ and this resulted in 6718 Hz; for a male speaker with a middle ceiling of 5000 Hz its top ceiling would be $5000 \exp(4 \cdot 0.05)$ which rounds to 6107 Hz. If you would want fifteen analyses instead of the previous nine but with approximately the same top and bottom ceilings you calculate the desired step size by solving $\exp(7 \cdot stepSize) = \exp(4 \cdot 0.05)$ for *stepSize* which gives $stepSize = \frac{4}{7} \cdot 0.05$.

What effect do these different ceilings have on the analyses? Suppose we have a sound with a sampling frequency of 44100 Hz. In the formant analysis, starting with the lowest ceiling, 4503.0 Hz, the sound will first be downsampled to the new sampling frequency of 9006 Hz (two times the ceiling). Pre-emphasis takes place, after which the Burg method of lpc analysis with ten coefficients will be performed and formant frequencies and bandwidths calculated and stored in a *Formant* object. This *Formant* object will be collected in the *FormantPath* object. Next this whole procedure is repeated but now with a 4734 Hz ceiling. This means first downsampling of the original sound to a 9467.8 Hz sampling frequency and then applying the other steps we just described. This will go on until we have collected nine *Formant* object in the *FormantPath*.

Since we have guaranteed that all *Formants* have the same sampling this implies that for each frame number we have nine different analysis frames to our disposal. The *FormantPath* data structure offers the possibility to assign to each frame number one of the nine available analysis frames. A simple *path interval tier* handles this association. In the default case there is only one interval that has the index of the middle ceiling in the *Formant* collection as its label (which happens to be "5" here). Following the label indices in this tier would constitute a *path* in time along possibly different *Formant* objects.

3. THE FORMANTPATHEDITOR

The FormantPathEditor enables the user to view and edit the path taken along the different Formant objects. The editor appears when we select the Sound, the FormantPath and an optional TextGrid together and choose **View&Edit** from Praat's dynamic menu. We use the publicly available Hillenbrand et al.[4] corpus with carefully measured formant frequency values, the original sound files as well as listener evaluations of the vowels, to demonstrate this editor's use.

Fig. 2 shows the FormantPath of the /had/ sound, spoken by female speaker 25 (file w25ah). The FormantPath was calculated with the default values of Fig. 1. The left part of the editor shows the sound oscillogram with a selection on top, the sound's broad-band spectrogram that is overlaid with formant tracks and a TextGrid. The view range of the spectrogram was set to 6800 Hz which is just a little bit higher than the top ceiling used in the analysis. The TextGrid was produced by a simple Praat script that converted all corpus data into TextGrids with one interval tier and one point tier. The interval tier has an interval labeled with the vowel /a/. The two time points in the point tier correspond to the most stable position in the vowel according to two different judges that Hillenbrand et al. used. The figure also shows that the vowel interval has been selected because it is masked by a rectangular shaded area.

The right hand side of the editor is what makes it very different from other editors. The nine panels in this *selection viewer* show the formant tracks from each of the nine different Formant objects in the FormantPath object. In each panel the tracks with even numbers, F_2 and F_4 , are drawn in a somewhat different colour than the uneven ones, F_1 , F_3 and F_5 , to be better distinguishable. In all the panels time runs horizontally from left to right starting at time 0.151 s and ending at 0.581 s; these times correspond to the start and ending time of the selected part of the sound on the left. The vertical frequency scale is linear and tops at 6800 Hz, just like the scale of the spectrogram. Above each panel its ceiling frequency is displayed. The top-left panel shows the lowest ceiling, 4503 Hz, while the bottom-right panel shows the highest ceiling, 6718 Hz. Each frame's formant frequency is displayed as a dot while its bandwidth is drawn as a vertical line centered at its frequency. This might help us in getting a better impression of the analysis results since well defined formants have small bandwidths and, therefore, show shorter

vertical lines.

Fig. 2 hints that the panel in the middle has been selected because it shows a different background shading. The panel selection can be changed by clicking with the mouse in any of the other eight panels. This panel selection has some consequences. In the first place a new interval is created in the FormantPath's path tier with the corresponding start and end times and with the selected index as a label. The formant dots displayed on the spectrogram always correspond to the formants of the Formant object at the corresponding index in the FormantPath. This means that selecting one of the panels will also automatically update the dots displayed on the spectrogram. As a second consequence, the selected ceiling of an interval will be shown in the spectrogram area as a fat horizontal line with the ceiling frequency above it. The fact that in the figure there is just one straight horizontal line at 5500 Hz means that we didn't make any selection yet that differed from the default.

Fig. 2 also shows that the default analysis for this particular sound might not be the optimal choice: although the first formant track displayed in the spectrogram area seems to be a rather smooth, some values deviate clearly. Closer inspection reveals that most deviates have a frequency lower than the value of the point in the track. From the middle panel we also note that these deviates have extreme large bandwidths as is shown by the long vertical lines centered at their frequency values. Actually it also shows that the dot in the "first formant track" corresponds to the second formant in this deviate frame. It would be better to completely ignore these deviates. The more because they influence the formant frequency statistics of the interval. If only one value were necessary to represent the interval then taking the median of the values on the interval instead of the mean would be necessary to cure for these outliers. However, if we are interested in the more dynamic aspects of formants, problems might occur.

Inspection of the other panels shows that better alternatives exist. For example, the panel with a ceiling of 6078 Hz at the bottom left. The tracks are much smoother than the ones in the panel at the middle. The start of the interval shows some irregularities because of the influence of the preceding /h/ sound. At reference point 1 the formant values with the 6078 Hz ceiling would be $F_1 = 1000$ Hz and $F_2 = 1580$ Hz. Our values are not too far off from the $F_1 = 1053$ Hz and $F_2 = 1677$ Hz as reported by Hillenbrand et al. Approximately 50 Hz off for the first formant and 100 Hz off for the second

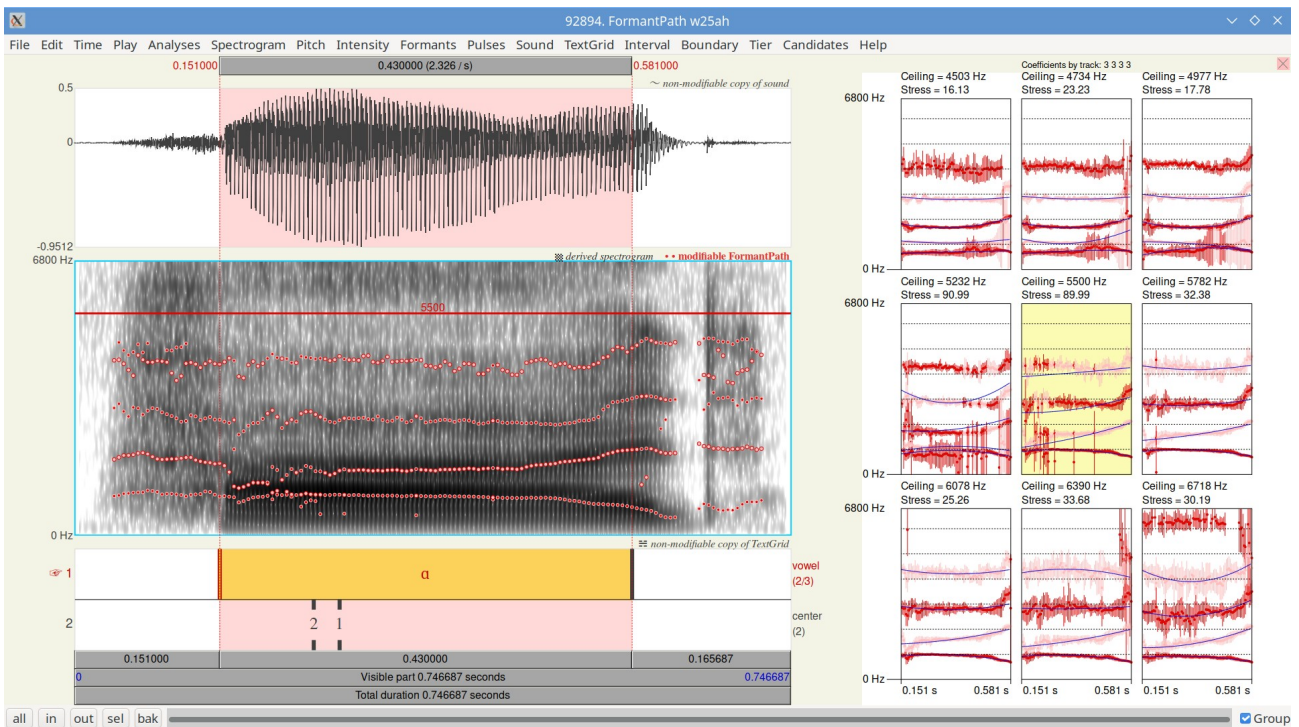


Figure 2: The FormantPathEditor showing the analyses for the /ɑ/ sound of file w25ah from the Hillenbrand corpus.

is not too bad with an lpc-based analysis.

The panel at the top-left shows an interesting example of a formant split: too many peaks have to be fitted on too small a frequency interval. Although its stress value calculated from [6] is the lowest of the nine panels its first three formant frequency values are not realistic for a female /ɑ/ sound.

4. AFTER EDITING

After leaving the editor the FormantPath object contains the changes in the path that have been made. The formant frequency values that would result from following the path can be accessed by using the **Extract Formant** method on the selected FormantPath. A new Formant object will be created and it can be queried just like any other Formant object.

5. SUMMARY

We have described only some of the possibilities of a new method to substantially improve the formant frequency analysis of a sound. However, the new method cannot cope with one of the fundamental limits of every analysis based on linear prediction: the attraction to strong harmonics in

the speech spectrum. Despite this deficiency, linear predictive analysis still is a very good practice to get reasonably accurate formant frequency measurements in a fast and reproducible way as this article tried to demonstrate.

6. REFERENCES

- [1] P. Boersma and D. Weenink, "Praat: doing phonetics by computer [computer program]," Version 6.3.03, 2023, available from <http://www.praat.org/>.
- [2] J. D. Markel and A. H. Gray, Jr., *Linear prediction of speech*. Springer Verlag, Berlin, 1976.
- [3] D. Whalen, W.-R. Chen, C. Shadle, and S. A. Fulop, "Formants are easy to measure; resonances, not so much: Lessons from Klatt (1986)," *Journal of the Acoustical Society of America*, vol. 152, pp. 933–941, 2022.
- [4] J. Hillenbrand, L. A. Getty, M. J. Clark, and K. Wheeler, "Acoustic characteristics of American English vowels," *Journal of the Acoustical Society of America*, vol. 97, no. 5, pp. 3099–3111, 1995.
- [5] P. Adank, R. Van Hout, and R. Smits, "An acoustic description of the vowels of Northern and Southern Standard Dutch," *Journal of the Acoustical Society of America*, vol. 116, pp. 1729–1738, 2004.
- [6] D. Weenink, "Improved formant frequency measurements of short segments," *Proceedings of the ICPhS XVII 2015*, 2015.