# FIELDWORK TECHNIQUES FOR RELATING FORMANT FREQUENCY, AMPLITUDE AND BANDWIDTH

*Peter Ladefoged*
Phonetics Laboratory, Linguistis Department, UCLA, Los Angeles, CA 90095-1543, USA
(e-mail: oldfogey@ucla.edu)
and
*Gunnar Fant*
Department of Speech, Music and Hearing, KTH, Box 70014, S–10044, Stockholm, Sweden
(e-mail: Gunnar.Fant@speech.kth.se)

## ABSTRACT

An analysis-by-synthesis method for finding formant bandwidths from vowel spectra has been implemented on a solar-powered computer used in fieldwork, thus enabling linguists to test hypotheses about differences between sets of vowels while working with speakers of the language. The procedure has been tested on the two sets of vowels that occur in Degema, a language spoken in Nigeria.

## INTRODUCTION

In most of the world's 7,000 languages, differences in vowel quality depend almost entirely on the frequencies of the first two or three formants. In these languages, from a physiological point of view, vowel quality depends simply on the shape of the vocal tract. In some languages, however, vowels are distinguished by differences in the mode of vibration of the vocal folds; vowels may have a breathy or creaky phonation in contrast with a modal phonation type. These physiological differences in phonation type affect the overall spectral slope and the bandwidths of the formants. Some languages may also contrast vowels in yet another way, namely, by differences that affect just the formant bandwidths. Finally, a point which we will not consider further in this paper, in many languages there are differences among vowels due to nasalization.

## COMPUTER MODELING OF VOWEL SPECTRA

Formant frequencies and amplitudes are comparatively easy to measure using standard acoustic analysis techniques. But formant bandwidths can seldom be measured directly from representations of spectra. To understand why this is so, it is useful to consider how the spectral curve representing a non-nasalized vowel is constructed. As shown in (1), the spectral curve of a vowel in dB vs. frequency, L(f), can be taken to be the sum of curves representing the four varying formants, $\sum H_n(f)$, together with a curve representing the contributions of higher, invariant formants, $K_{r4}(f)$, generally referred to as a higher-pole correction, and a curve, S(f), corresponding to the spectrum of a glottal pulse modified by the lip radiation transfer.

$$L(f) = \sum H_n(f) + K_{r4}(f) + S(f) \quad (1)$$

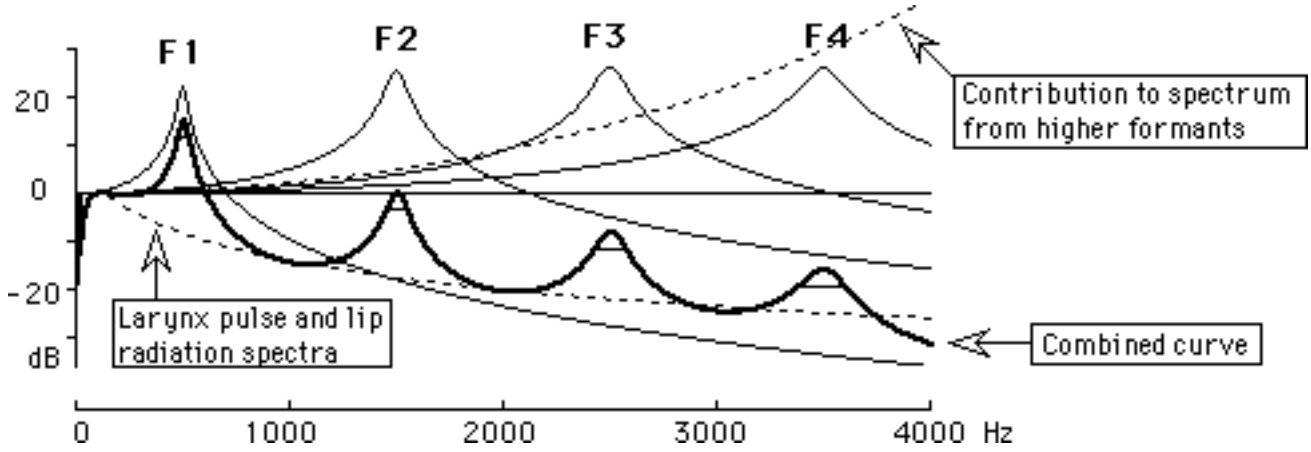


Figure 1. The combination of four formants, and curves representing the contributions of higher formants, the larynx pulse spectrum and lip radiation to form the observed spectrum in a non-nasalized vowel.

The four formant curves, $H_n(f)$, correspond to the four lowest resonances of the vocal tract. The contribution in dB of each formant to the spectrum envelop is as in (2).

$$H_n(f)= -10\log_{10} ((1 - f^2 /F_n^2)^2+(B_n^2 /F_n^2)(f^2 /F_n^2)) \quad (2)$$

For each formant, $F_n$ is a resonant frequency of the vocal tract, and $B_n$, which is the bandwidth of this resonance. is considered to be not larger than the corresponding formant frequency [1,2]. The amplitude, $A_n$, which can be observed in the spectrum, is not an explicit part of the definition of a formant. It is dependent on the particular pattern of the formant frequencies, the higher pole correction, and the modified source spectrum.

These relations are shown graphically in Figure 1. When we add the curves representing the first four formants together, the peak of each formant curve is raised slightly by the contribution of each of the higher formants, and, for all formants except formant one, is lowered quite considerably by the contribution of each of the formants below it. To get a true summation of the first four formants we must include the raising contributions of higher formants — the lowering by lower formants is automatically included by the summation of the curves for these formants. The raising contributions are represented by the upper dashed curve in Figure 1. As shown by Fant (1960:50), when four formants are specified, the equation for the additional contributions is as in (3), assuming a total vocal tract length of 17.65 cm and thus a first resonance of 500 Hz for a neutral vowel. When analyzing fieldwork data we often do not have reliable information on the fourth formant. Given data on only three formants, the contributions of the fourth and higher formants are as in (4).

$$K_{r4} =0.54 (f/500 )^2 + 0.00143 (f/500 )^4 \quad (3)$$
$$K_{r3}=0.72 (f/500 )^2 + 0.0033 (f/500 )^4 \quad (4)$$

The shape of the larynx pulse depends on the phonation type. It is often taken as having a spectrum that falls at a rate of –12 dB per octave, but in our analysis it must be considered to be a variable. It can, however, be combined with a fixed curve representing the acoustic effects that occur when the vibrations of the air in the vocal tract are converted into a sound wave propagated from the lips. These effects contribute a positive slope of +6 dB per octave, so that the combined curve typically has a slope of –6 dB per octave. The equation we will use is shown in (5), where G(f) is a modification to the basic – 6 dB/octave spectrum slope, simulating differences in phonation type. This is the equation for the lower dashed curve in Figure 1, with $g = 1.0$.

$$S(f) = 20 \log_{10} G(f)(2(f/100)/(f + 100)^2) \quad (5)$$

As we have noted, the phonetically interesting features of the spectrum of a vowel are the formant frequencies, which reflect the vocal tract shape, and the bandwidths, which reflect the phonation type and the losses within the vocal tract (and nasalization, which is not our concern in this paper). Figure 1 makes it clear that the formant frequencies are usually readily derivable from the spectral curve. The locations of the peaks in the combined curve are the same as those for the individual formants. The only problem that arises in determining formant frequencies is when two formants come close together so that only one peak is identifiable in the combined curve.

The situation is not the same for formant bandwidths, in that the bandwidths in the combined curve are not the same as those of the individual formants. The matter is further complicated in analyses of actual utterances. The shape of the spectral curve may be distorted by artifacts arising from a glottal source that does not have a spectrum similar to the idealized smoothly falling curve in Figure 1. There may also be extraneous noises that affect the shape of the combined curve, but do not affect the location of the peaks corresponding to the formants. The formant frequencies and amplitudes can be determined from the peaks in FFT or LPC spectra. But the shape of the spectral curves is seldom such that we can measure the bandwidths.

Our solution to this problem is to determine the bandwidths of the formants by an analysis by synthesis procedure. The frequencies and amplitudes of the formants in a vowel are found by Fourier analysis. The formant frequencies are then used to calculate a vowel spectrum, assuming default values for the formant bandwidths and glottal source spectrum. The calculated formant amplitudes are then compared with the observed amplitudes, and adjustments are made to the glottal source spectrum to find the best overall spectral slope. Next, the bandwidths of the formants are adjusted until there is no further improvement in the match between the calculated and observed formant amplitudes. The glottal source spectrum is then adjusted again to see if the match can be further improved. Finally the bandwidths are readjusted in an effort to find an even better match.

## ANALYZING THE VOWELS OF DEGEMA

The procedure has been tested by analysis of the vowels of Degema, a Niger-Congo language spoken in Nigeria. Degema has 10 vowels, arranged in two sets of five, with the restriction that, in general, words contain vowels from one set or the other, but not both. This constraint, known as vowel harmony, is fairly common among languages spoken in West Africa. The two sets of Degema vowels are said to be distinguished by the position of the tongue root, the one set being said to have an advanced tongue root {+ ATR], and the other set a retracted tongue root [– ATR]. In this and other similar languages, however, there are sometimes said to be other differences, notably a more breathy phonation in [+ ATR] vowels as discussed by Ladefoged and Maddieson [3]. Somewhat contrasting with this view is the notion (also described in [3]) that [+ ATR] vowels have stiffer vocal tract walls, and thus a ‘brighter’ quality. Finally, there is the possibility that [+ ATR] and [– ATR] vowels differ simply in formant frequencies — differences that are ascribable to just the shape of the vocal tract, with no

consistent differences in phonation type or vocal tract losses.

When studying a language such as Degema, our normal fieldwork practice is to use a battery operated DAT recorder in conjunction with a Macintosh computer that can be recharged by solar power. We analyze vowels with a 512 point FFT, and a low sample rate so as to obtain greater accuracy in the frequency domain [4]. The recordings are re-digitized at 11,400 samples per second, so that a 512 point FFT has a window length of 45 ms (the longest we like to use in a vowel in which the quality may be changing). There will then be 22 Hz between harmonic components in the FFT spectrum.

In the case of the present investigation, the recordings were re-analyzed using the Kay CSL package in the UCLA Phonetics Lab. For this preliminary report we will consider the productions of only one speaker. A plot of the relation between the first two formants for this speaker is shown in Figure 2. Each [+ ATR] vowel is clearly distinguished from its [– ATR] counterpart, with the exception of the high front pair. These two vowels are also not distinguished by F3. Our concern in this investigation is to see whether there are any consistent relationships dependent on the bandwidths or the overall spectral slopes that apply to each set of vowels.

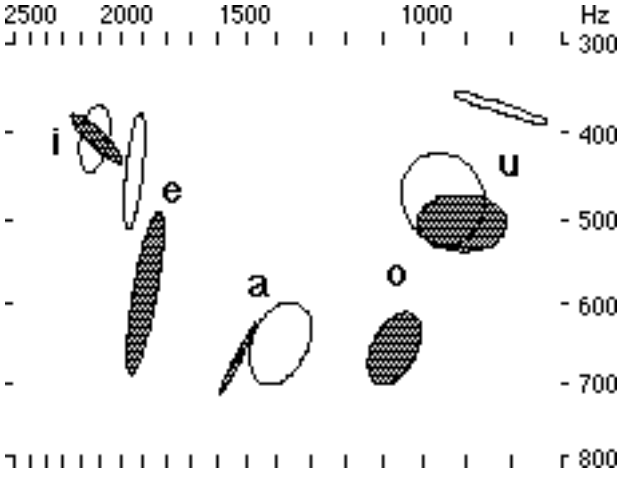


Figure 2. The relation between F1 (ordinate) and F2 (abscissa) for the vowels of a speaker of Degema. Ellipses enclose vowels that are within two standard deviations of the mean. Shaded ellipses denote [– ATR] vowels.

As there might be some complications introduced by different degrees of lip rounding which would affect the lip radiation, we restricted these preliminary observations to four tokens of each of the three unrounded vowels in each set. The recordings were made in typical fieldwork circumstances, with cocks crowing in the distance, children playing outside, and cicadas cheerfully singing away. Although the close-talking, noise canceling, microphone and DAT recorder enabled us to maintain a fairly high signal to noise ratio, we were uncertain of our observations of the fourth formant, and therefore used a three formant model in our synthesis.

Our initial assumption was that [+ ATR] vowels might have one type of phonation and [– ATR] vowels another. Accordingly the first task was to find the most appropriate glottal slope for each group separately. The formant frequencies of the two sets of 12 vowels were used as input to the computer model, which was also supplied with default initial estimates of the formant bandwidths: B1 = 30 Hz, B2 = 50 Hz, B3 = 70 Hz. The slope of the modified larynx pulse spectrum was varied from 3 dB per octave to 9 dB per octave, and the relative amplitudes of the formants calculated. Two measures of the spectral slope were calculated for each vowel, the difference in amplitude between the first formant and the second (A1 – A2), and the difference between the first formant and the third (A1 – A3). Using the (A1 – A3) measure as the indicator of the spectral slope, we found virtually no difference in the best fitting spectral slope for [+ ATR] and [– ATR] vowels, but the (A1 – A2) measure indicated that [+ ATR] vowels have a 1.5 dB per octave greater slope, and therefore may have a more breathy phonation type. This point was re-examined at a later stage of the investigation.

The next stage was to allow the model to generate the most appropriate bandwidths for each of the formants for each of the vowels. The measure used was the rms difference between the calculated and observed formant amplitudes, for each formant and for each vowel considered separately. For these calculations we chose a spectral slope of 5.75 dB per octave for both [+ ATR] and [– ATR] vowels. The results of this process are shown in Figure 3. In general, for all three formants the bandwidth increases as the frequency increases (see Figure 2 for F1 and F2 frequency data); but there are no systematic differences between [+ ATR] and [– ATR] vowels.

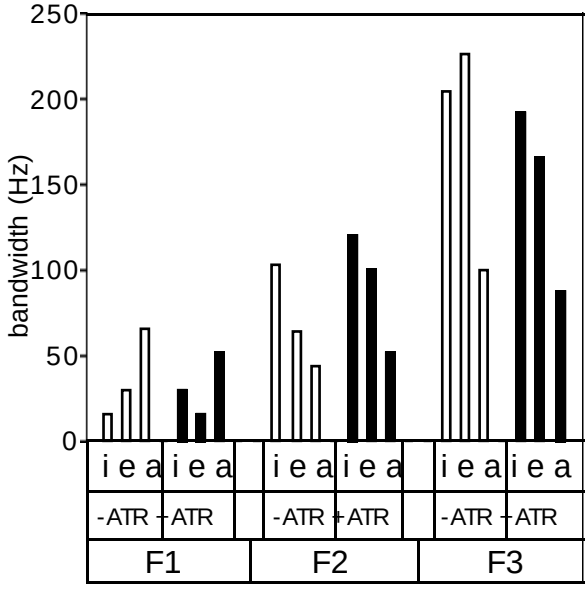


Figure 3. The bandwidths of each vowel when the model produced spectra in which the calculated formant amplitudes most closely matched the observed formant amplitudes.

The first formant bandwidths in Figure 3 do not follow the pattern reported by Fant [5], in that in our data [i] has a smaller bandwidth than [a] in both sets of vowels. In a report on Akan, another vowel harmony language spoken in Ghana, Hess [6] also notes differences from the pattern reported in [5]. Hess, however, found that all the [+ ATR] vowels had significantly smaller bandwidths than the corresponding [– ATR] vowels. In her study as in the present study, bandwidth generally increases with frequency.

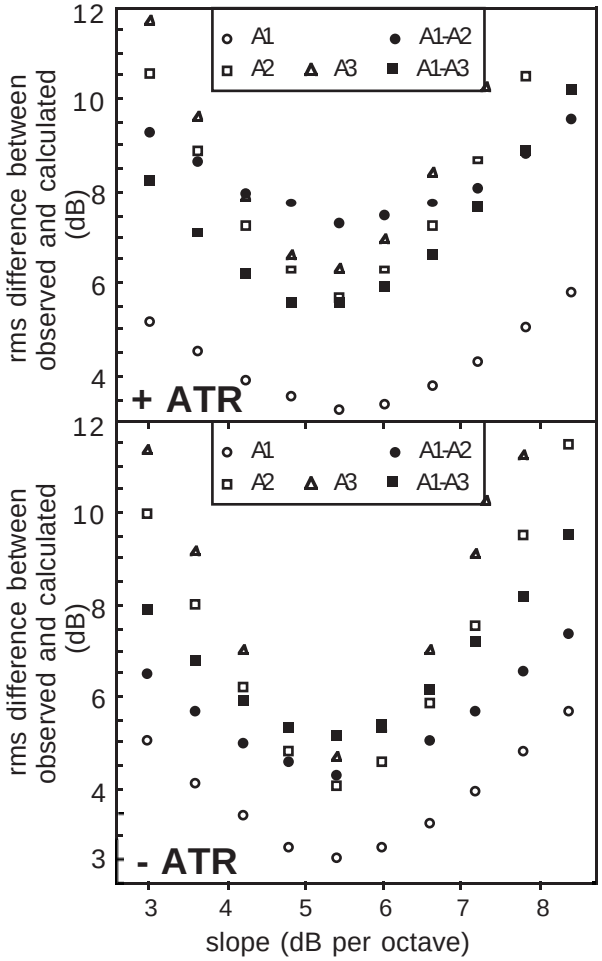


Figure 4. The effect of varying the larynx pulse slope on the difference between the observed and calculated vowels, each of which has formant bandwidths that produce amplitudes that best match the observed amplitudes.

Finally, each set of 12 vowels was synthesized again with varying glottal slopes, to see if a better match could be found now that the formant bandwidths had been determined. The results are show in Figure 4. Good matches were obtained for each of the first three formant amplitudes. For both sets of vowels, for each measure of spectral slope, (A1 – A2) and (A1 – A3), the best match was achieved with a slope of slightly less than 6 dB per octave. We can therefore be reasonably certain that there are no phonation type differences between these two sets of vowels; and, as we have seen, there are no consistent differences in bandwidth. At the moment we must conclude that for this speaker there is no evidence that the two sets of Degema vowels differ consistently in anything other than their formant frequencies; and in this set of words, there is no phonetic difference between the phonologically distinct [+ ATR] and [– ATR] high front vowels.

## CONCLUSION

We have shown that by varying the formant bandwidths in a model of vowel spectra, the observed formant amplitudes in a set of data can be matched. But a word of caution is necessary: the model we are using has too many degrees of freedom to be certain that the solution that we have obtained is unique. This speaker had no detectable difference in phonation type. But, as we will show at the meeting for another subject, if there is an apparent difference in phonation type, then different results may be obtained if the bandwidths or the glottal slopes are adjusted first. This model needs to be more fully tested before its results are accepted.

## ACKNOWLDEGEMENTS

Thanks to Sean Fulop for analyses of Degema vowels, and to E. Kari for organizing and recording Degema. This work was upported by NSF grant SBR 9319705

## REFERENCES

[1] G. Fant, “Acoustic Theory of Speech Production”, Mouton, The Hague, 1960. (Reprinted: Walter de Gruyter, Berlin)

[2] G. Fant, “The LF-model revisited. Transformations and frequency domain analysis,” *STL-QPSR* 2-3, pp. 119-156, 1995.

[3] P. Ladefoged and I. Maddieson, “The Sounds of the World’s Languages”, Blackwells, Oxford, 1996.

[4] P. Ladefoged, "Elements of Acoustic Phonetics". 2nd. ed. Chicago University Press, Chicago. 1996.

[5] G. Fant, “Vocal tract wall effects, losses, and resonance bandwidths,” *STL-QPSR* 2-3, pp. 28-52, 1972.

[6] S. Hess, “Assimilatory effects in a vowel harmony system: an acoustic analysis of Akan advanced tongue root vowels,” *J. Phonetics.* 20, pp. 475-492, 1992.