Dept. for Speech, Music and Hearing
Quarterly Progress and Status Report

Formant amplitude measurements

Fant, G. and Mártony, J.

journal: STL-QPSR
volume: 4
number: 1
year: 1963
pages: 001-005

KTH Computer Science and Communication

http://www.speech.kth.se/qpsr

I. SPEECH ANALYSIS

A. FORMANT AMPLITUDE MEASUREMENTS

A survey of various measures of formant amplitudes and their interrelations was made in an earlier quarterly report [1,2] and in a paper to the Stockholm Speech Communication Seminar [3]. It is the purpose of the following report to summarize some of the results and present a revised system of symbols for the various measures together with additional material illustrating the superposition effects at a gliding pitch.

The various concepts of formant amplitude we have studied are illustrated by Fig. I-1. These are

- A_s Spectrum envelope amplitude
- A_e Root mean square amplitude
- A_a Average amplitude
- A_i Initial voice period peak amplitude
- A_p Peak amplitude

In common for all of these measures is that they are taken at the frequency of the formant and not necessarily at the exact frequency of a spectral maximum. The formant frequency F_n and the formant bandwidth B_n of formant number n are simply defined by the imaginary and real part of the corresponding pole.

$$p_n = -\pi B_n \pm j2\pi F_n \qquad (1)$$

The amplitude of formant number n is denoted by A_n if no reference is made to how it is measured and otherwise with an additional subscript, s, e, a, p or i. Thus A_{2s} is the spectrum envelope amplitude of the second formant. If amplitudes are to be expressed in dB it is recommended to adopt the symbol L for level.

$$L_n = 20\log_{10}(A_n/A_o) \text{ dB} \qquad (2)$$

denotes the level of formant No. n where A_o is a reference amplitude.

The A_s measure is confined to the envelope of a line spectrum. The A_e measure can be calculated from an r.m.s. summation of a suitable number of harmonics within the range of the

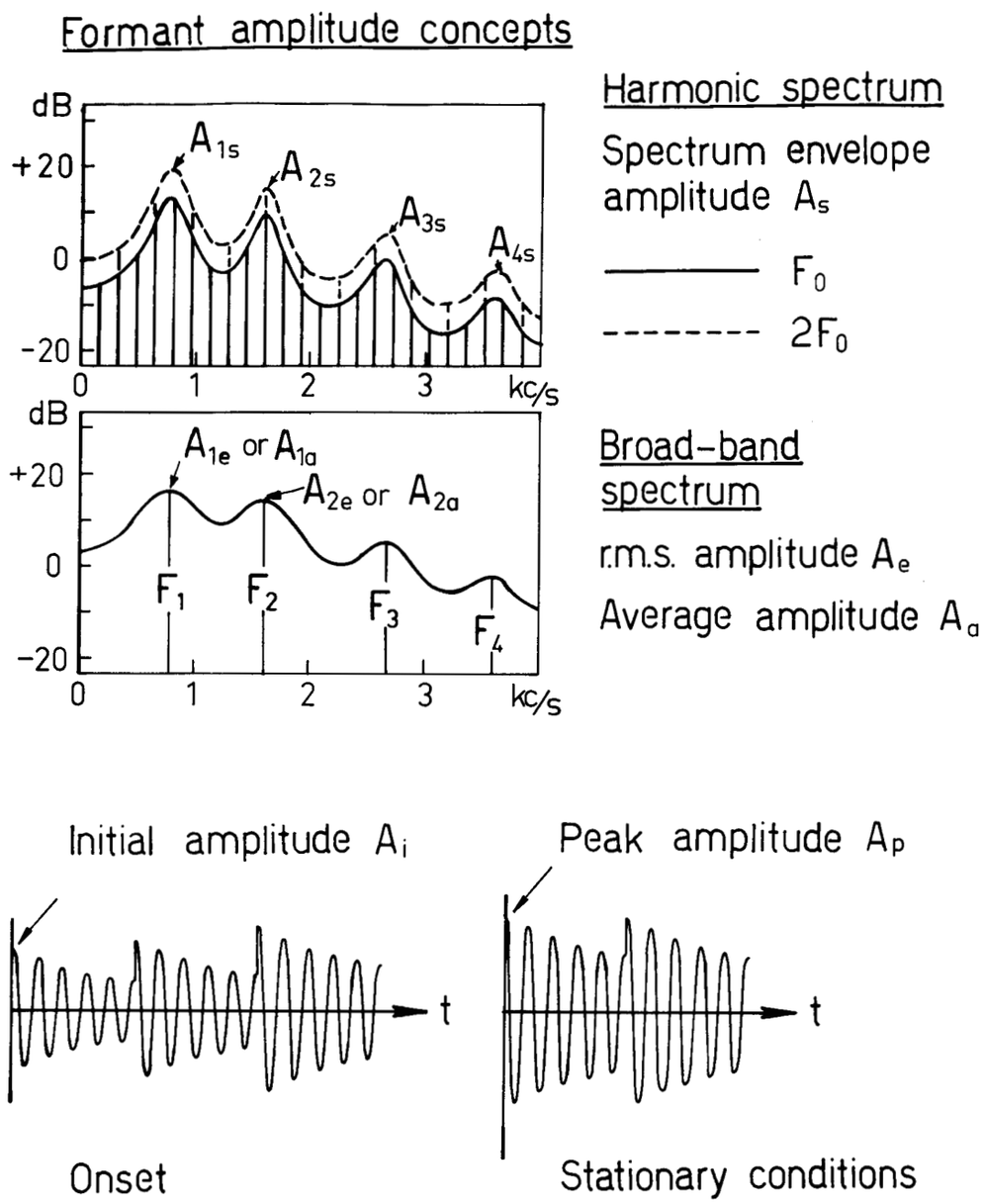

Fig. I-1. Illustrations of the various concepts of formant amplitude. Note that spectral amplitudes are taken at the formant frequency (pole frequency) which not necessarily coincides with a point of maximum amplitude.

formant peak. Especially at high F_0 this involves rather arbitrary decisions of which harmonics to include, especially if two formants lie fairly close. Both the A_e and the A_a measures are more conveniently derived from a prefiltering rectification and smoothing of the time function in which case the rectifier shall have square law characteristics for A_e and linear characteristics for A_a. The band-pass filter used for the prefiltering shall have a recommended width of 500 c/s and be centered at the formant frequency.

The peak value of the time function envelope within a voice period is denoted by A_p. Ideally assuming a single point of excitation it is also the initial value of an exponential. In general, however, because of the possibility of a distributed excitation pattern the envelope may take an arbitrary shape.

If the effect of superposition from previous voice periods is subtracted the peak value is by definition A_i. Assuming a build-up with constant periodicity and waveshape the hypothetical value A_i of the initial or reference voice period is either smaller or greater than A_p depending on whether the superposition is in phase or out of phase.

A study of the simple model adopted for standard synthesis procedure reveals the following interrelations

$$A_i = A_p\left(1 + e^{-2Y_n} - 2e^{-Y_n}\cos 2\pi\frac{F_n}{F_0}\right)^{1/2} \tag{3}$$

$$\frac{A_{p_{min}}}{A_{p_{max}}} = \operatorname{tgh}(Y_n/2) \tag{4}$$

$$A_e = A_p\sqrt{\frac{1-e^{-2Y_n}}{2Y_n}} \tag{5}$$

$$A_a = A_p\,\frac{(1-e^{-Y_n})}{Y_n} \tag{6}$$

$$A_s = A_i \cdot \frac{1}{Y_n} \tag{7}$$

where $Y_n = \pi B_n/F_o$ (8)

These measures are normalized in terms of r.m.s. amplitudes so that A_p, A_e and A_a become numerically equal if Y tends to zero, i.e. under conditions when the oscillation decays very little during a voice period.

The peak factor A_p/A_e and the form factor A_e/A_a may be calculated from the relations above.

It is of interest to study how the various formant amplitude measures vary as a function of an increasing F_0, i.e., when vocal pulses of a constant shape and size are omitted at an increasing rate. The peak amplitude of the first vocal period A_i is by definition a constant. When stationary periodic conditions have been reached the measures A_p, A_e and A_a vary in an oscillatory function with increasing F_0. Superimposed is a 6 dB/oct rise of A_a and a 3 dB/oct rise of A_e. The spectrum envelope amplitude A_s does not oscillate and increases at a rate of 6 dB/oct.

The simple behavior of A_s may thus be described as follows. An increase of F_0 by an octave is followed by a rise in A_s by 6 dB. However, the number of harmonics within any limited frequency range is halved and the net gain in A_e is thus 3 dB only. A_e/F_0 or A_i would be ideal parameters for practical work - if they could be automatically measured! Apart from the superimposed oscillations the measure A_p might be useful for the extraction of formant amplitude parameters in vocoders. It has the benefit of a relative small average increase with increasing F_0. The A_a parameter which is the most commonly used measure of formant amplitude in automatic speech analyzing systems is just as sensitive to the superposition effect as A_e and A_p and has the additional disadvantage of F_0 proportionality.

It is always of interest to see how a theory developed from a mathematical model compares with the true system. The earlier progress report[2] gave data derived from measurements of a male speaker uttering the vowel [ɛ] at a large variety of voice fundamentals. An illustration was also given of the time-frequency-intensity spectrum of a sample phonated at a gliding pitch. This particular speaker W.J. executed a high degree of control over his voice and the superposition effects were rather similar comparing this sample with that of a synthetic imitation.

A similar experiment has recently been performed with a second speaker, B.L., see Fig. I-2. His phonation with gliding voice fundamental frequency from 100 to 400 c/s was less stable than that of the previous speaker and several interesting phenomena show up comparing his utterance with that of a synthetic imitation. The first formant amplitude behaves approximately as in the synthetic utterance. There are pronounced peaks in L_1 whenever a harmonic passes through the formant peak. The extent of the positive and negative oscillation of L_1 compares well with what can be predicted from Eq.(4).

The level of the second formant L_2 does not confirm with that of the synthetic model. A reinforcement of L_2 is to be expected whenever a harmonic passes through F2 as displayed by the synthetic sample. A small tendency in this direction is observable. However, the main periodicity of the L_2 curve is the same as that for L_1. In other words, the particular relation between F_1 and F_0 governs the intensity of both L_1 and L_2 and a great part of the entire spectrum as judged by the Sonagram.

The effect can be interpreted as a register break affecting the entire source spectrum each time one of the lower harmonics sweeps through the region of F1. Under these circumstances it could be expected that the mechanical vibration of the vocal cords is subjected to a finite load from the vibrating air column above them (5) which in turn changes the waveform and spectrum of the vocal excitation function. It remains to verify the origin of the L_2 fluctuations.

The investigation is continuing with calculations of the equivalent source spectra at successive intervals within the phonation illustrated in Fig. I-2.

G. Fant and J. Mártony

References:

(1) Fant, G. and Liljencrants, J.: "How to define formant level. A study of the mathematical model of voiced sounds", STL-QPSR No. 2/1962, pp. 1-9.

(2) Fintoft, K., Lindblom, B., and Mártony, J.: "Measurements of formant level in human speech", STL-QPSR No. 2/1962, pp. 9-17.

cont.

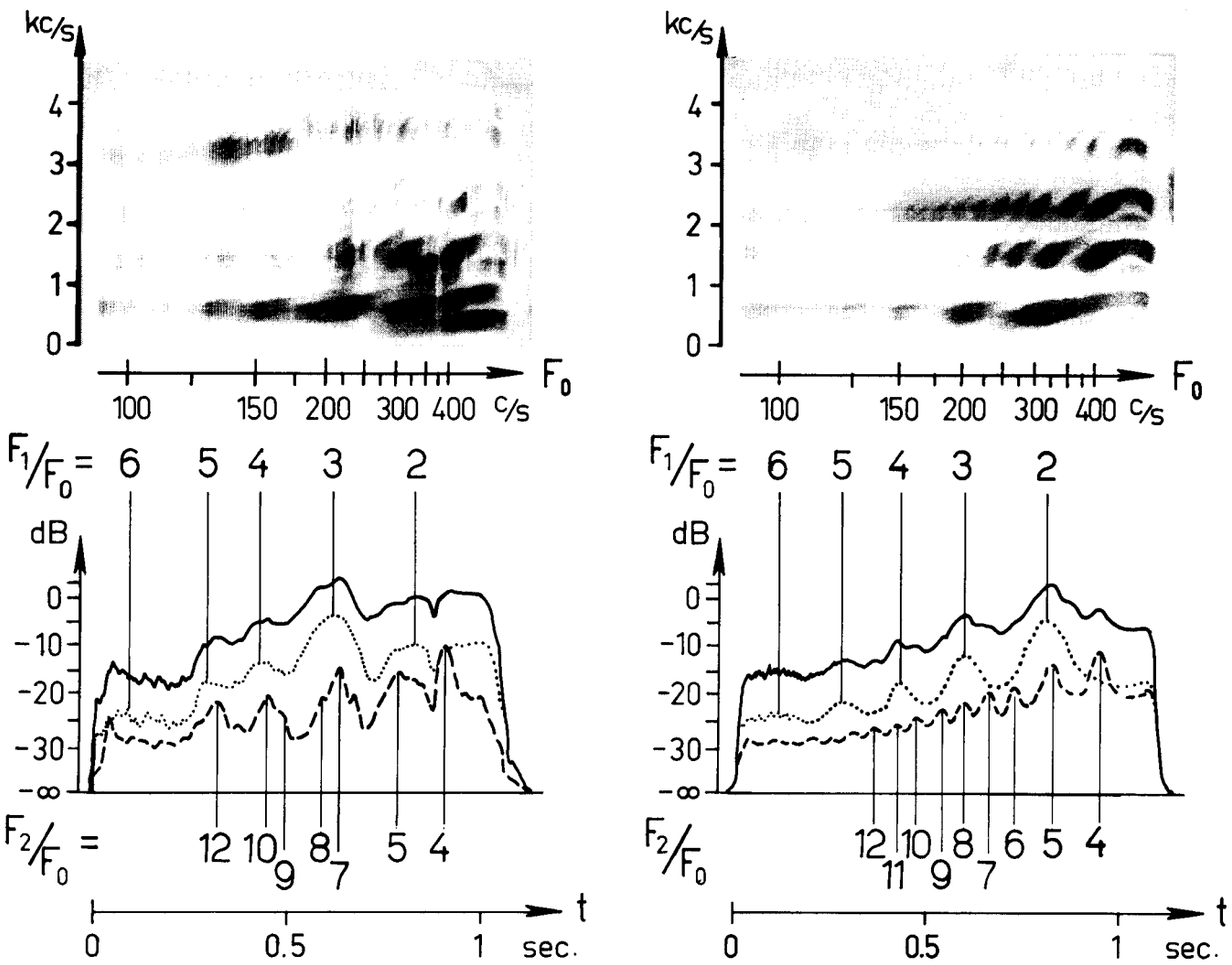

Fig. I-2. Broad-band Sonagrams and curves of overall amplitude (intensity) and selective measures of first and second formant amplitude of a sustained vowel [æ] at a gliding pitch produced by the speech synthesizer OVE II and by a human subject (B.L.). Note the behavior of second formant amplitude of the human speaker.

(3) Fant, G., Fintoft, K., Liljencrants, J., Lindblom B., and Mártony, J.: "Formant amplitude measurements", Paper C2 presented at the Speech Communication Seminar, Stockholm 1962.

(4) Fant, G.: "Acoustic Analysis and Synthesis of Speech with Applications to Swedish", Ericsson Technics 15 (1959).

(5) van den Berg, Jw.: "Myoelastic-aerodynamic theory of voice production", J. of Speech and Hearing Research, 1 (1958) pp. 227-244.

(6) Peterson, G.E. and McKinney, N.P.: "The measurement of speech power", Phonetica, 7 (1961) pp. 65-84.

B. SPECTROGRAPHIC STUDY OF THE DYNAMICS OF VOWEL ARTICULATION

A spectrographic study of vowel reduction has recently been concluded [1]. It will be briefly summarized below.

Vowel reduction is said to be a characteristic feature of languages with heavy stress but has to certain extent also been associated with rate of utterances and contextual influence. There is some evidence in the literature that, articulatorily as well as acoustically, the process of reduction amounts to centralization. Thus in the acoustic domain a reduced vowel is located somewhere along a continuum whose extreme ends are the formant pattern of the unaffected vowel and that of the neutral vowel or schwa.

An experiment was designed to test this hypothesis and to provide some insight into the dynamic properties of vowel articulation. It involved the examination of vowels pronounced under varying timing conditions and in systematically varied consonantal environment.

24 nonsense words were formed by commuting /I, ɐ, Y, æ, a, ɵ, ɔ, ʊ/ in three consonantal environments: /b-b/, /d-d/ and /g-g/. Preliminary experimentation preceded the selection of sentence frames that generated durations of these vowels within 80 to 300 msec. There were four carrier phrases. Each made up one list in which each of the 24 CVC syllables occurred 5 times.