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

Spectrum envelopes for synthetic vowels

Tappert, C. C. and Mártony, J. and Fant, G.

journal: STL-QPSR
volume: 4
number: 3
year: 1963
pages: 002-006

KTH Computer Science and Communication

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

B. SPECTRUM ENVELOPES FOR SYNTHETIC VOWELS

1. Introduction

A computer program* was written which calculates the harmonic specification of a formant pattern directly from the separate formants and in addition compares these calculated harmonics with read in harmonics which may be obtained either from a natural vowel or a machine synthesized vowel. This work was undertaken in order to study source and filter characteristics in speech production, for example during a change in the fundamental frequency.

2. Description of program with theoretical considerations

Adopted with slight modifications from Fant's work (1) on the source-filter theory and analytical constraints for idealized voiced sounds, were the following equations: (all in dB)

$$L_k = -10 \log[1 + (f/50)^2],$$

$$L_{r4}(f) = 0.54\left(\frac{7f}{F_4}\right)^2 + 0.00143\left(\frac{7f}{F_4}\right)^4,$$

$$L_n(f) = -10 \log\left[\left[1 - \left(\frac{f}{F_n}\right)^2\right]^2 + \left(\frac{B_n}{F_n}\right)^2\left(\frac{f}{F_n}\right)^2\right]$$

---

* The program was written in ALGOL and run on the BESK computer at the Swedish Board for Computing Machinery in Stockholm. A somewhat similar program was written by G. Fant in 1959. This earlier program was written for the sole purpose of computing ideal theoretical spectrum envelopes for vowels and was used as a partial check on the initial results obtained from the program discussed herein.

The function, L_k, whose cut off value was changed slightly*, represents the combined source and radiation characteristics. The vocal tract transfer function then consists of a pair of conjugate poles for each of the four considered formant frequencies, equation L_n, plus the function L_{r4} representing contributions from the 5th and higher poles up to infinity. This corrective function, $L_{r4}(f)$, which is dependent on the total length of the vocal tract, was made a function of F_4 under the assumption that the variation of F_4 is approximately proportional to the variation in the total vocal tract length during the articulation of the various vowel sounds. The formant bandwidth, B_n, was approximated as being completely dependent on the formant frequency; a parabolic function interpolated from ref. (2).

$$B_n = 50\left(1 + \frac{F_n^2}{6 \cdot 10^6}\right)$$

The accuracy of this approximation is fairly good for low frequencies but decreases roughly linearly with an increase in frequency; thus the values for B_1 and B_2 were usually satisfactory whereas for B_3 and B_4 they were often not very accurate.

Combining the above equations, one obtains the complete spectrum envelope equation in dB

$$L(f) = L_k(f) + L_{r4}(f) + \sum_{n=1}^{4} L_n(f) \quad \text{dB}$$

---

* Compare with Eq.(2.5-1)in Ref. (1).

which was used to calculate the theoretically ideal vowel harmonics directly from the formant frequencies and the fundamental frequency, F_0. For the comparison of natural vowels with the ideal ones, the formant frequencies were obtained from Sonagrams, while the fundamental frequency and the harmonics for the natural vowels were obtained from data processed by RASSLAN (3).

3. Results and conclusions

The first sets of comparisons made were between the 7 vowels /iː, eː, ɛː, aː, uː, ʉː, øː/, articulated by J.M. and their ideal counterparts - their spectrum envelopes - are shown in Figs. I-16 to I-19. On the whole these results appear to be quite satisfactory. Some of the deviations which resulted can be accounted for. There occur what appear to be errors in some of the bandwidths, for instance in B_3 for [iː]; in B_2, B_3, and B_4 for [ɛː]; in B_3 and B_4 for [aː]; and in B_3 for [uː]. This can to some extent be explained by deviations from the average bandwidth dependence on formant frequency (2). However, other sources of discrepancy are presumably present and closer investigation in this respect is needed.

The poor match obtained in the middle range of the vowel [uː] was due to background noise in the natural vowel. The deviation between natural and theoretical spectrum for [aː] could depend on some degree of nasalization. A low F_5 close to F_4 is noticeable in some of the vowels, especially [ʉː] and [øː]. Also most vowels show a vocal source zero which occurs at about 800 c/s.

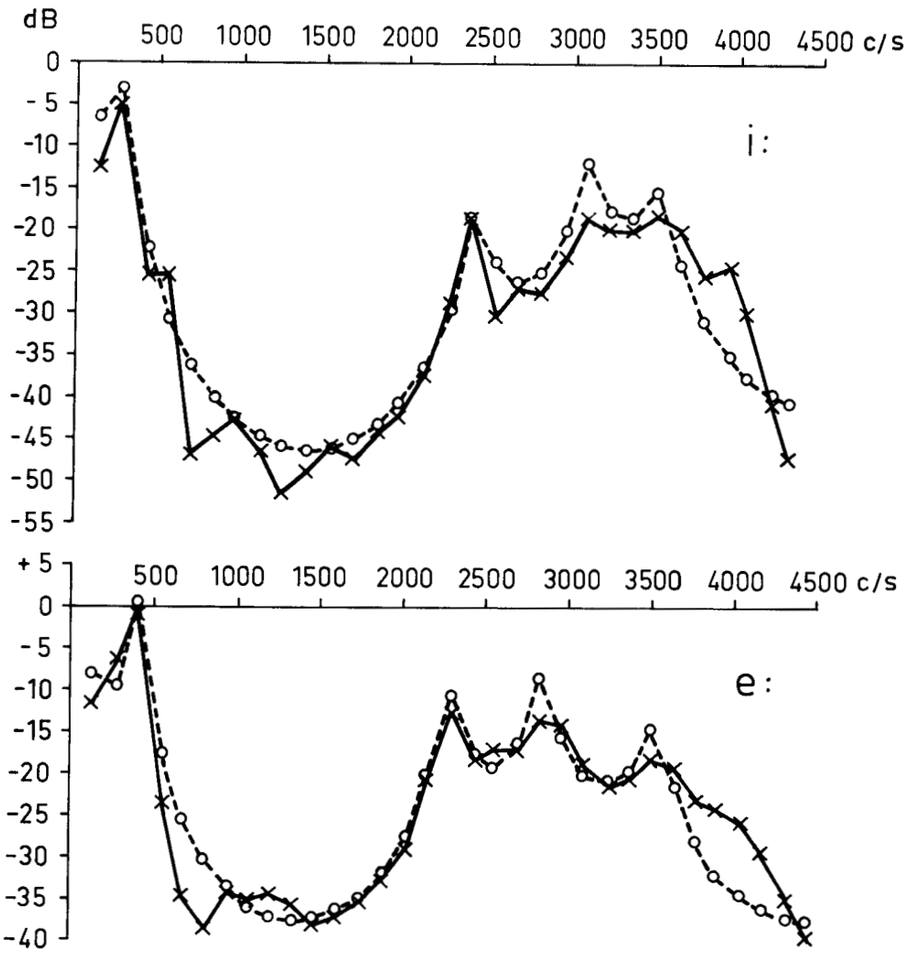

Fig. I-16. Results of spectrum matching of the vowels [i:] and [e:] speaker (J.M.) on the basis of a harmonic representation. Solid lines combine measured harmonic amplitudes and broken lines harmonics computed from the standard synthesis model with -12 dB/oct source function.

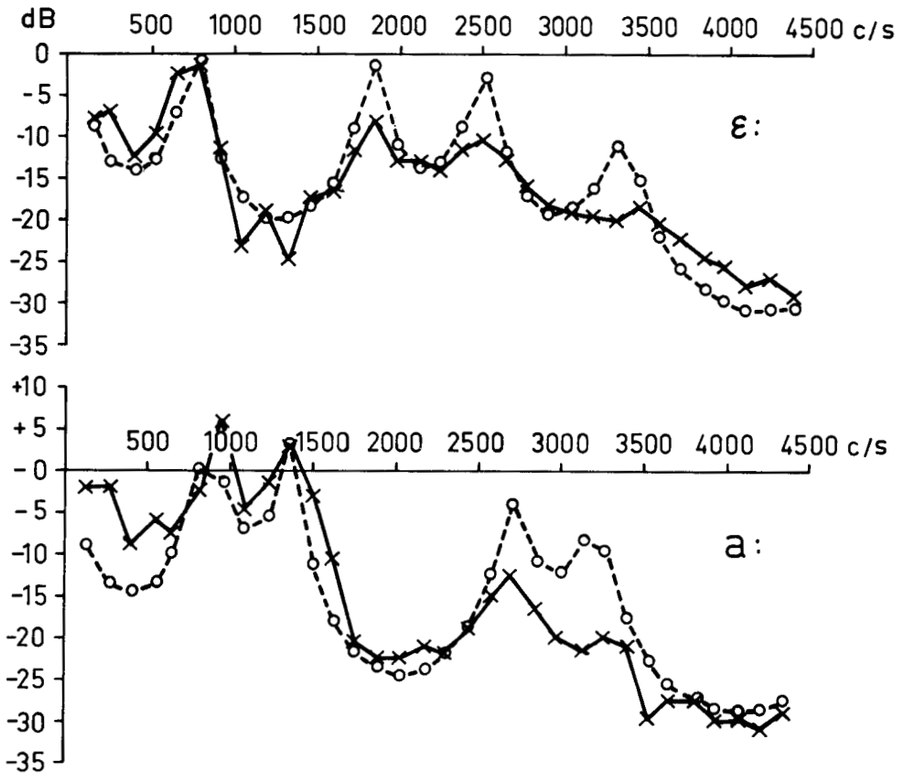

Fig. I-17. Spectrum matching of the vowels [ɛ:] and [a:]. Solid lines pertain to measured harmonic amplitudes (speaker J.M.) and broken lines to harmonics computed from the standard synthesis model.

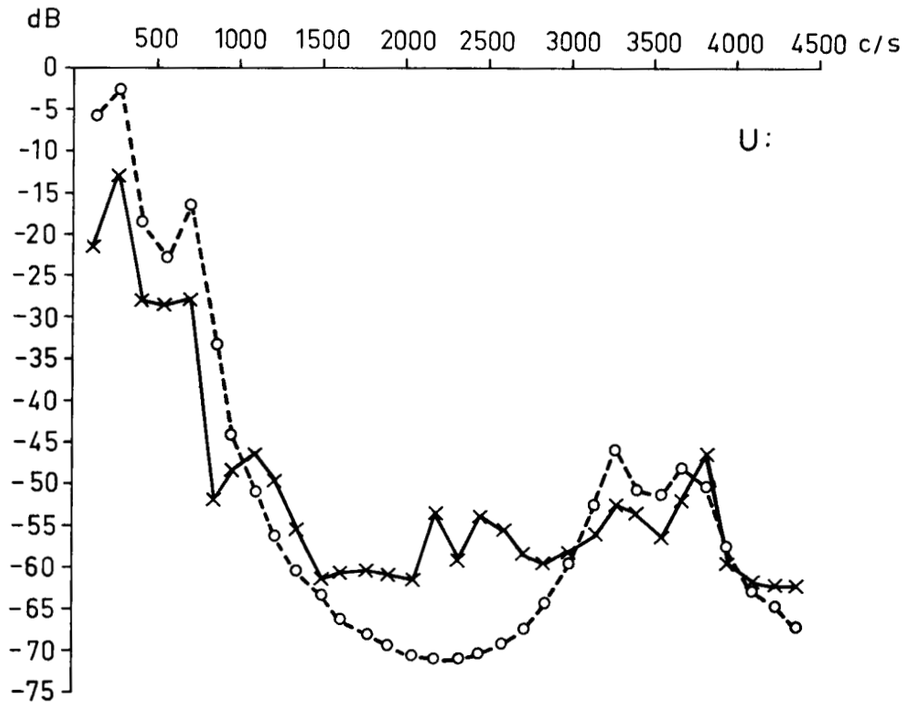

Fig. I-18. Spectrum matching of the vowel [u:]. Solid lines pertain to measured harmonic amplitudes (speaker J.M.) and broken lines to harmonics computed from the standard synthesis model. The discrepancy is due to the effect of noise limiting the spectrum level of the human sample.

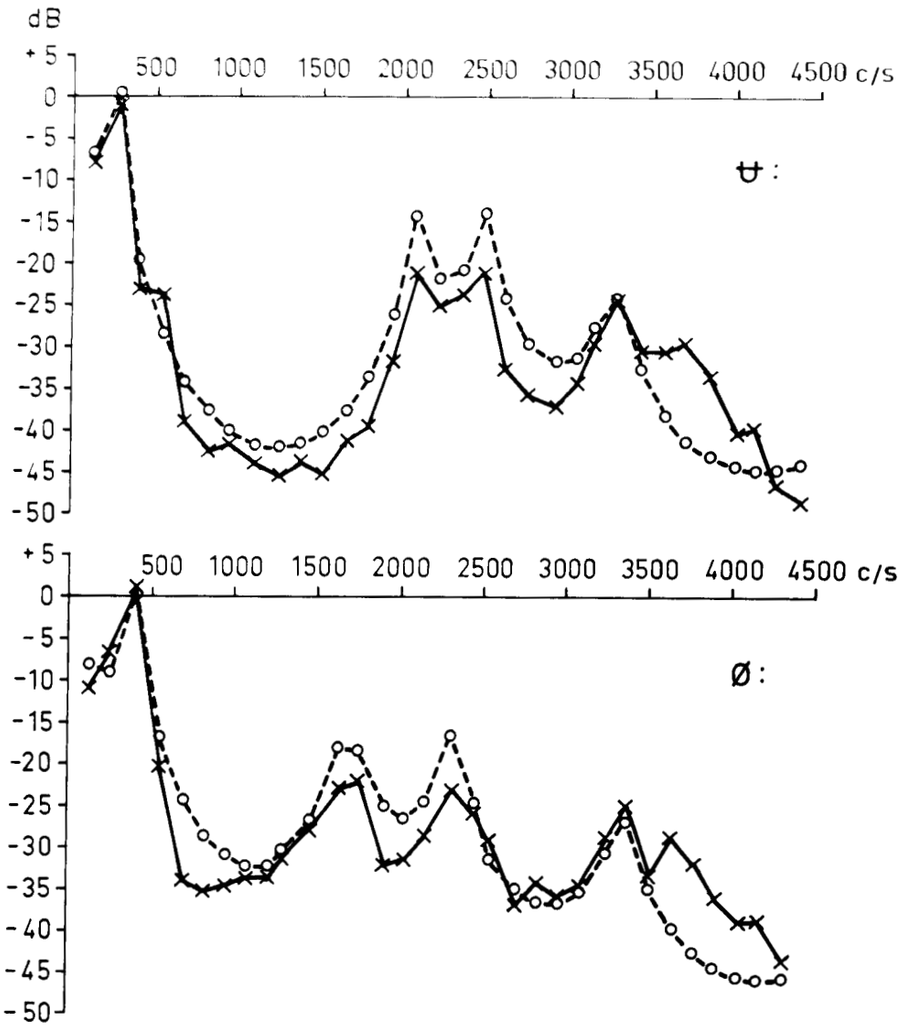

Fig. I-19. Spectrum matching of the vowels [ʉ:] and [ø:] (speaker J.M.) on the basis of a harmonic representation. Solid lines combine measured harmonic amplitudes and broken lines harmonics computed from the standard synthesis model with -12 dB/oct source function.

To further investigate the difference between the mathematical and natural source functions, five of the harmonic difference functions were averaged ([u:]and [a:] were excluded). Before averaging, the discrepancies in the bandwidths were corrected for since the objective here was to study the source function; the resulting averaged difference function is shown in Fig. I-20.a. This function approximated the difference between the natural source function and the L_k model of -12 dB/oct. It can be seen from the figure that for the lower frequencies up to about 500 c/s the natural source level is about 2 dB above the mean difference in the range 1000 - 3000 c/s and that a characteristic dip occurs in the neighborhood of 600 - 900 c/s representing a zero in the natural source function. The dip at 1800 - 1900 c/s is believed to be due to a second more damped zero in the source function. The hump above 3300 c/s is due to the particular F_5 close to F_4. Fig. I-20.b shows an averaged source spectrum deviation from L_k for a different male speaker (O.K.). This source spectrum also has a zero at 700 - 800 c/s and it falls off faster for higher frequencies.

Secondly, the natural Swedish vowel [ɛ] and its ideal counterpart were compared under a variation of the fundamental frequency. During the articulation of the vowel, the speaker B.L. started at a low F_0 and slowly increased this frequency while trying to hold the articulation constant so as to cause as little change as possible in the formant frequencies, see Fig. I-21. A series of cross-sections of this vowel were obtained with F_0 ranging

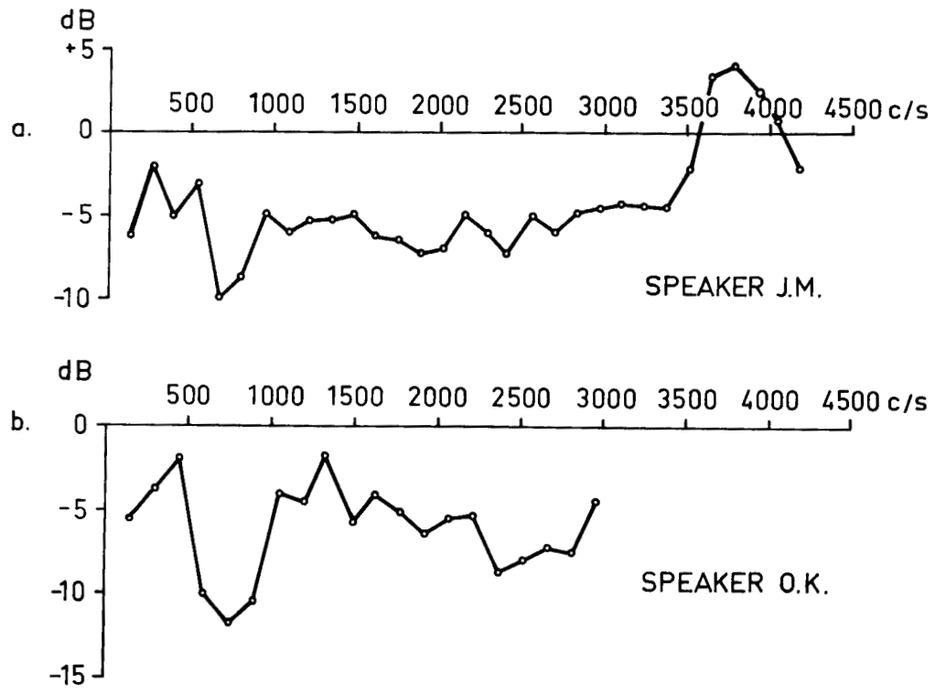

Fig. I-20. Average source spectrum normalized with respect to the standard -12 dB/oct voice source.

a. Speaker J.M.

b. Speaker O.K.

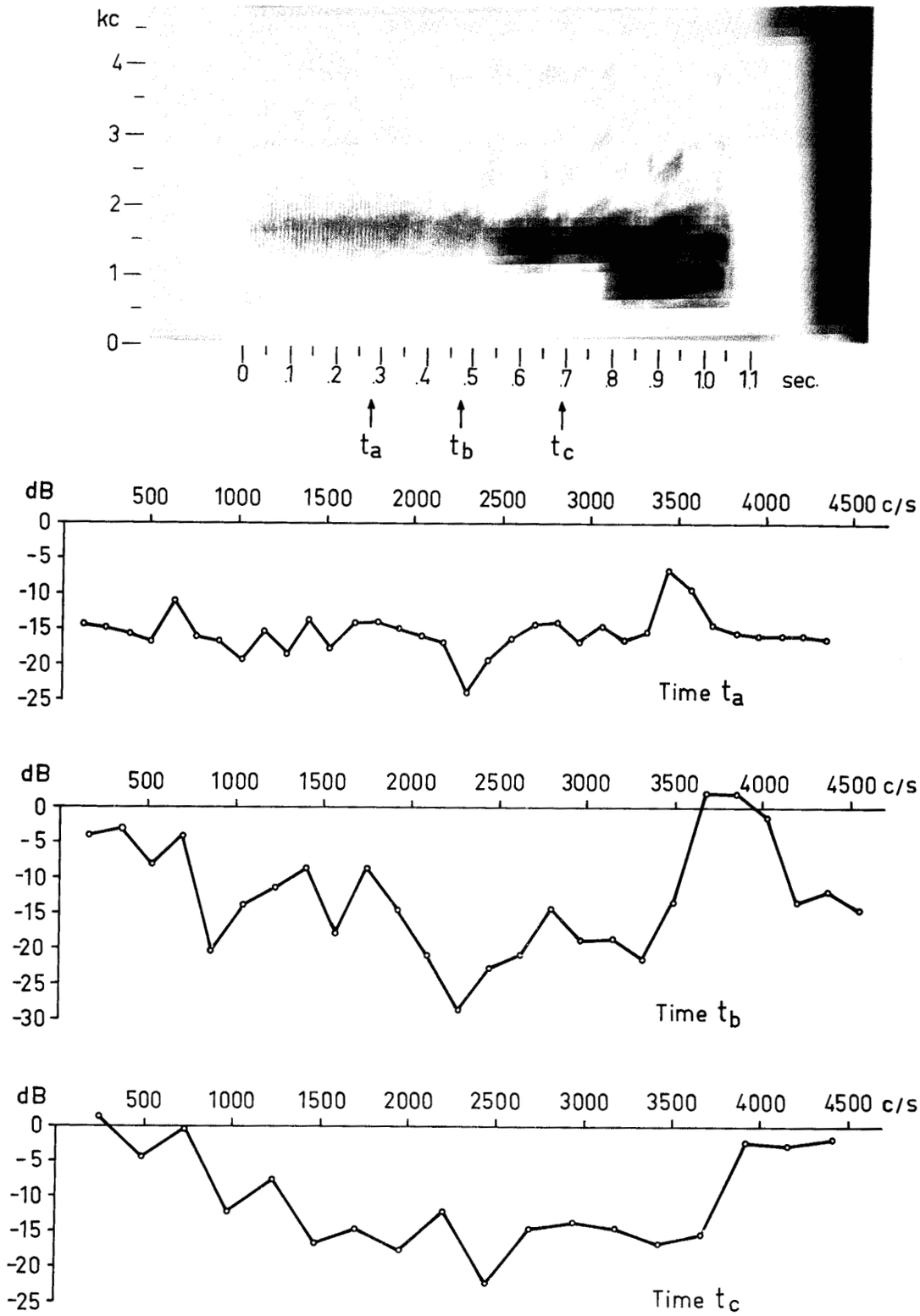

Fig. I-21. Sonagram of the vowel [ɛ] phonated at a gliding pitch and three sampling points for spectrum sections. The deviation from normalized voice spectra of these samples are shown.

from 111 to 470 c/s, and compared with a corresponding ideal series. Three of the resulting difference functions pertaining to F_0 = 128, 174, and 245 c/s respectively, are shown in Fig. I-21.

All three samples show the previously discussed effect of the spectrum emphasis in the 3500 - 4000 c/s range which may be ascribed to F5. As pitch increases there is also a relative boost of the low frequency level. The F3 level is long in all samples, possibly due to an underestimation of B_3. However, the low F2 level of the F_0 = 174 c/s sample is according to the spectrogram associated with the periodic F_0-dependent fluctuations of formant amplitudes discussed by Fant et al (4).

A control study on sampled spectrum sections of synthetic speech did not show these particular fluctuations and the estimated sampling errors were found to be of the order of $\pm$1.5 dB.

C.C.Tappert, J.Mártony, and G.Fant

References:

(1) Fant, G.: "Acoustic Analysis and Synthesis of Speech with Applications to Swedish", Ericsson Technics, 15, No. 1 (1959) pp. 3-108.

(2) Fant, G.: "Formant Bandwidth Data", STL-QPSR 1/1962, pp. 1-2.

(3) Liljencrants, J. et al: "Spectrum Sampling Instrumentation", STL-QPSR 2/1960, pp. 1-4.

(4) Fant, G., Fintoft, K., Liljencrants, J., Lindblom, B., and Mártony, J.: "Formant Amplitude Measurements", to be publ. in the J. of the Acoustical Society of America, November issue 1963.