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

Quantization of formant coded synthetic speech

Fant, G. and Mártony, J.

journal: STL-QPSR
volume: 2
number: 2
year: 1961
pages: 016-018

KTH Computer Science and Communication

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

III. SPEECH SYNTHESIS

A. QUANTIZATION OF FORMANT CODED SYNTHETIC SPEECH

Experiments devoted to maximally exact synthetic reproduction of natural speech by means of formant coded synthesizers have given promising results as stated in the previous quarterly report[1]. Although a trained listener can hear the difference between the synthetic speech and the natural speech the difference can be made rather small if a sufficient effort is put into the matching procedure.

The question then arises to what degree the control signals from the function generator may be quantized without causing a substantial loss in the intelligibility and quality of the speech. Recent experiments on analog and quantized coding of a few sentences have shown [x] that the perceptual difference is very small providing the information rate of the control signal in the quantized coding is of the order of 1000 - 1200 bits/sec.

The main set up of the synthesizer is that shown in Fig. III-1. The synthesized speech is the summed output from three separate filter units, one for vowellike sounds, one for nasal sounds, and one for fricative noise sounds. The noise generator can be connected to the fricative filter through the amplitude-modulated gate A_C and to the vowel filter by means of the gate A_H. The voice generator is connected to the vowel filter via A_O and to the nasal filter via A_N.

Each of the three filter units contains a set of series connected resonant circuits. The first three resonance frequencies of the vowel filter, F_1, F_2, and F_3, are controlled from the function generator whereas F_4 and the correction for higher formants[2] are fixed. In the nasal filter all resonances are preset before the synthesis. The fricative filter comprises two resonances K_1 and K_2 and one anti-resonance K_O.

Formant frequencies F_1, F_2, and F_3 are traced directly onto the coding sheet from a sonagram. All other spectrum defining parameters are calculated by means of spectrum matching techniques.

---

x) Demonstrated at the 61st meeting of the Acoustical Society of America, May 10, 1961.

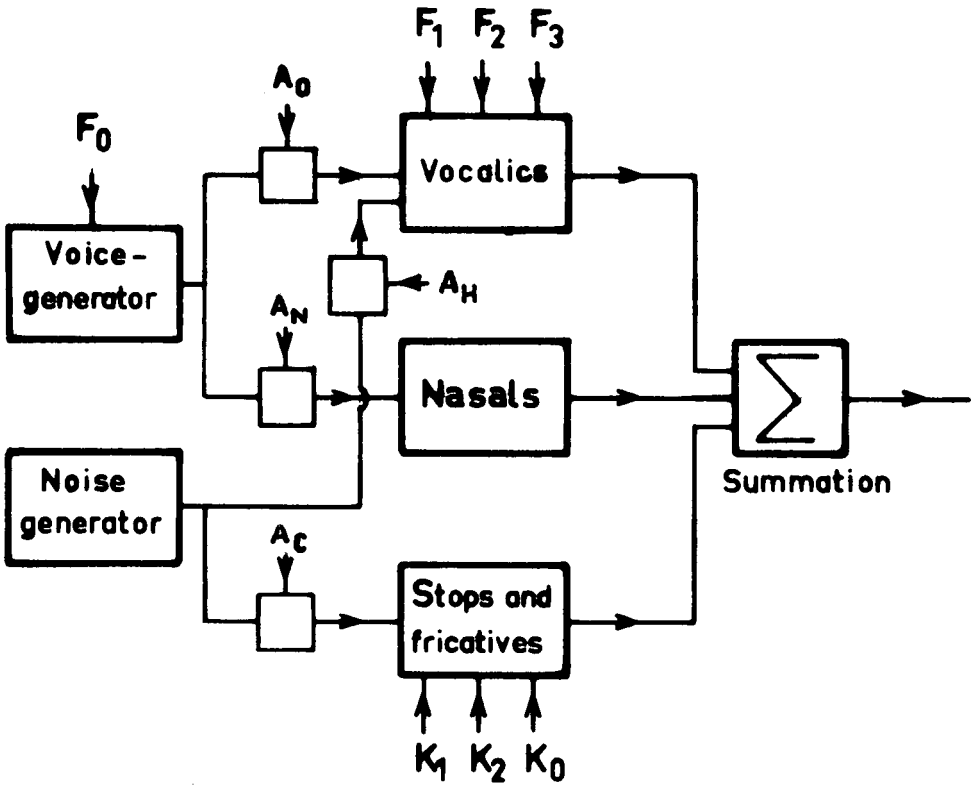

Fig. III-1 Block diagram of the speech synthesizer OVE II.

The initial analysis comprises the taking of amplitude frequency sections and oscillographic curves of F_0 and various intensity parameters for obtaining a first approximation to the amplitude parameters A_0 A_H A_C and A_N.

When the final match was achieved after a series of checks and corrections the quantization proceeded by producing a tracing of the coding sheet, see Fig. III-2, and preparing a new sheet with staircase curves, see Fig. III-3. The 11 parameters were sampled at a rate of 40 times per second and were quantized as follows:

Parameters	Number of bits	Number of levels	Comments
F_0	4	16	First voiced sample following voicelessness was coded in 1/6 octave steps covering the range 60 - 340 c/s. Next and following samples within voiced portions of speech were coded in 1/24 octave steps of change versus the pitch of the previous sample. Total possible range of F_0 46.3 - 428 c/s.
F_1	4	16	The range of F_1 was 150 - 900 c/s covered in 50 c/s quantal steps.
F_2	4	16	The range of F_2 was 550 - 2800 c/s covered in 150 c/s quantal steps.
F_3	3	8	The range of F_3 was 1550 - 4000 c/s covered in steps of 350 c/s.
K_1	2	4	The range of K_1 was 3000 - 6000 c/s covered in quantal steps of 1000 c/s.
K_2-K_1	1	2	Two alternative values of K_2-K_1 = 2000 and 3000 c/s.
K_0/K_1	1	2	Two alternative values of K_0/K_1 = 1/2 and $1/\sqrt{2}$ were adopted.
A_0	3	8	A_0 = +5, 0, -5, -10, -15, -20, -25, and -oo dB.
A_C	3	8	Same as A_0 above.
A_H	2	4	-5, -10, -20, and -oo dB.
A_N	2	4	Same as A_H above.
	Total 29 bits/sample		

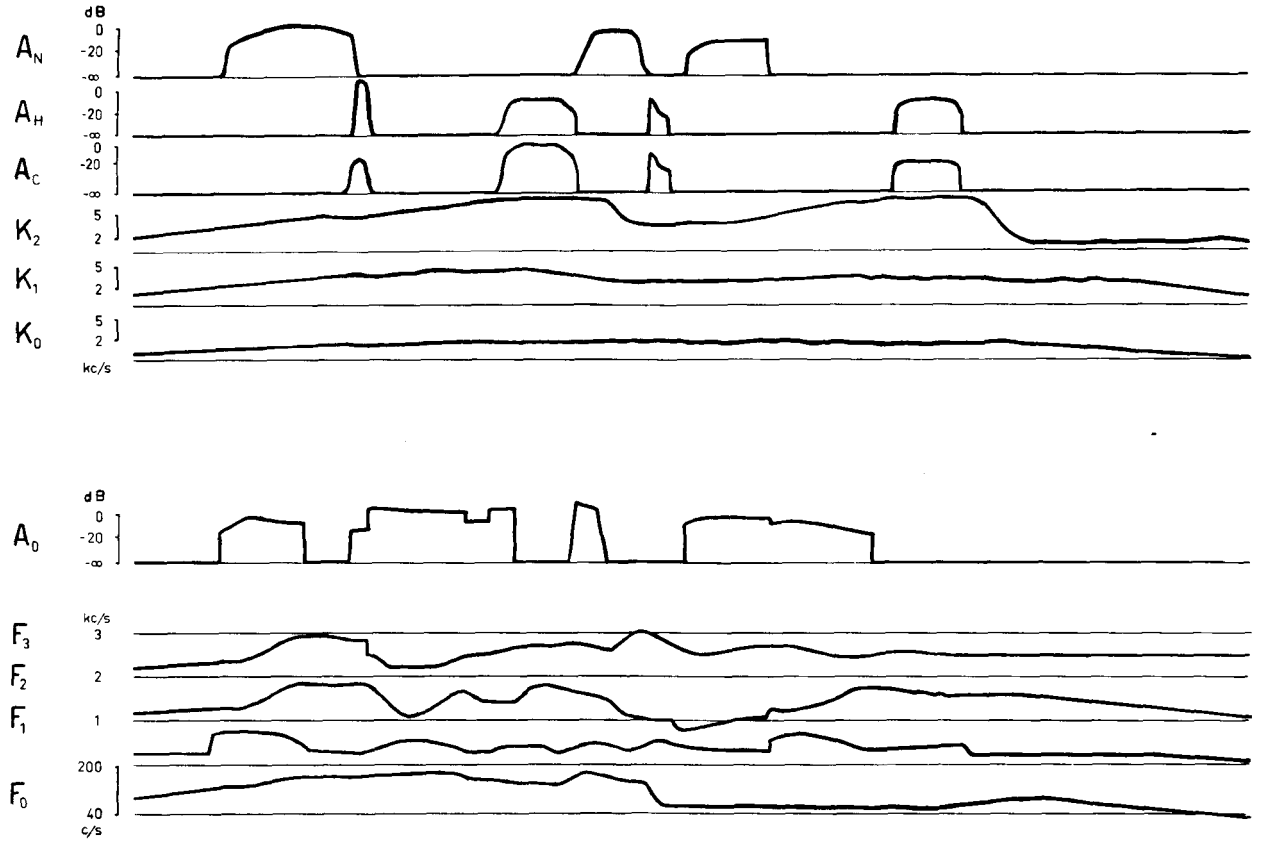

Fig. III-2 The time-variation of the 11 synthesis parameters within the sentence "I enjoy the simple life".

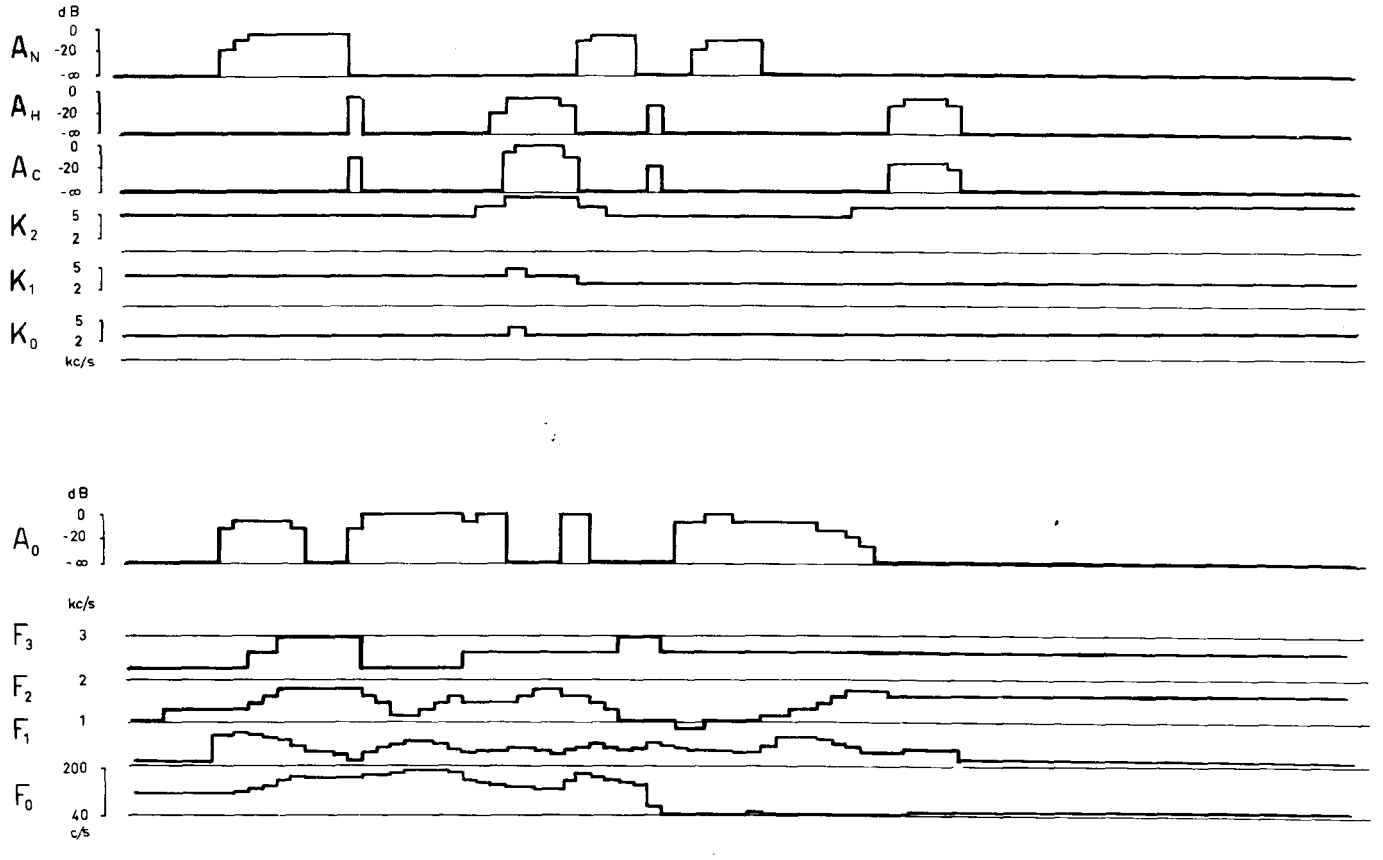

Fig. III-3 The time-variation of the 11 synthesis parameters within the sentence "I enjoy the simple life". The parameters have been sampled at a rate of 40 times per second and quantized.

Since it is theoretically possible to let the F_1 information occupy the same signal channel as the K_0 K_1 K_2 information (mutually exclusive variations) it would be possible to subtract 4 units from the number of bits 29 and conceive of the coding as requiring a channel of the capacity of transmitting 25 bits/sample and thus 1000 bits/second.

A special test was run on the effect of varying the cutoff frequency and thus the time constant of the smoothing filters in each control signal channel connecting the function generator output of a channel and the corresponding input control terminal of the synthesizer. It was found that a time constant of 10 msec was sufficient to smooth out the discontinuities of the control signals to the extent that the audible effects were eliminated. A time constant of 20 msec did not noticeably affect the quality of the analog or quantized speech. These low-pass smoothing experiments were made with our standard 3rd order minimum overshoot pass filters of the transform

$$H(s) = \frac{1.27\omega_I^3}{(s + 0.85\omega_I)[(s + 0.7\omega_I)^2 + \omega_I^2]}$$

In producing the analog speech the smoothing filters were set at a time constant of 10 msec corresponding to a low-pass cutoff frequency of 40 c/s.

Special tests on varying the quantal steps in F_0 showed that the 4 bits approximation with mixed absolute and differential code did provide an improvement compared with a 3 bits approximation, the difference not being very great. However, these tests were made on a very limited speech material and should thus be regarded as merely indicative of practical coding demands.

G. Fant, J. Mártony

(1) Holmes, J.N.: "Notes on Synthesis Work", STL-QPSR-1/1961, p. 10-12.

(2) Fant, G.: "Acoustic Analysis and Synthesis of Speech with Applications to Swedish", Ericsson Technics 1/1959, p. 3-108.