Dept. for Speech, Music and Hearing

Quarterly Progress and Status Report

Formant frequency tracking

Fujisaki, H. and Risberg, A.

journal: STL-QPSR
volume: 1
number: 1
year: 1960
pages: 006-009

C. FORMANT FREQUENCY TRACKING

The scope of this project is to gain experience on various schemes of formant tracking, i.e., automatic extraction of formant frequency data from connected speech.

During the past year we have made some preliminary studies on methods of counting zero-crossings and of moment weighting. The moment-weighting method was engineered simply by carrying out the analog division between the differentiated speech wave and the undifferentiated speech wave.

The recent work during the last three months has been concerned with more detailed studies of zero-counting techniques and with a method of anti-resonance filtering. When the zero of an anti-resonance circuit coincides in frequency and bandwidth with a formant pole there is perfect cancellation providing the residues from other poles may be neglected. This procedure may be regarded as the time-domain equivalent of the pole-zero spectral-matching method developed at M.I.T.

Our results are negative in so far as all the methods mentioned above have serious inherent limitations, at least within the experimental conditions of our present studies. Our main conclusions are as follows:

(1) Zero-crossing counting.

a. If the frequency of the damped oscillation corresponding to a single formant is integrated over a period longer than that of a voice fundamental period the measure will coincide with the frequency of the most intense harmonic within the formant. This well-known effect will cause objectionable jitter effects in resonance vocoders unless an excessively long time constant is chosen for the frequency parameter smoothing.

b. An initial transposition of the speech band to higher frequencies will make possible the use of smoothing time constants much smaller than that of a voice fundamental period. However, the presence of the discontinuity at the time position of the initial voice excitation and also the frequent absence of oscillation in a part of the cycle will complicate the measuring procedure by the need of an intricate sampling system.

c. The presence of insufficiently suppressed residues from other formants might distort the measurement owing to averaging effects. If two frequency components are less than one octave apart this effect becomes negligible as long as the unwanted signal is more than 5 dB below the level of the signal to be measured. As the frequency ratio increases the effect is more pronounced. In the case of a frequency ratio of $F_2/F_1 = 10$ and an amplitude ratio of $A_2/A_1 = 10$ dB there results an error of $0.15(F_2-F_1)$. A prefiltering continuously controlled by the measured positions of other formants is needed [1] in order to get the best results from the method. A supplementary method which has been tried is to restrict the time constant of the flipflop in the frequency counter circuit so as to avoid synchronization on interfering oscillations from high-frequency formants.

(2) Moment weighting.

a. Moment-weighting and zero-crossing counting techniques are known to provide equal results on random noise signals. In general both methods are dependent on a prefiltering.

b. Our experiments on the use of the simple moment-weighting technique of analog division of two speech signals differing by a prefiltering of 6 dB/octave and subjected to linear rectification and short-time smoothing, have shown that the function is very sensitive to the particular prefiltering and to the degree of asymmetry of the formant. In case of reasonably symmetric formants the systems tend to follow the leading harmonic.

c. The moment-weighting procedure may from a systematic point of view be regarded as a simplified instance of interpolation within a band of filters connected to a maximum selector. Such systems have been tried with some success in England [2] and will be evaluated in coming phases of our work.

(3) Anti-resonance filtering.

A pilot study on anti-resonance methods of formant tracking has recently been completed. The ideal system would be one in which the speech wave is passed through a series of anti-resonance circuits in cascade, one for each formant. By proper adjustment of

Fig. I-4 Spectrogram with superimposed F_1 - values originating from a minimum selection of the outputs of a anti-resonance curcuit spaced with 100 c/s intervals.

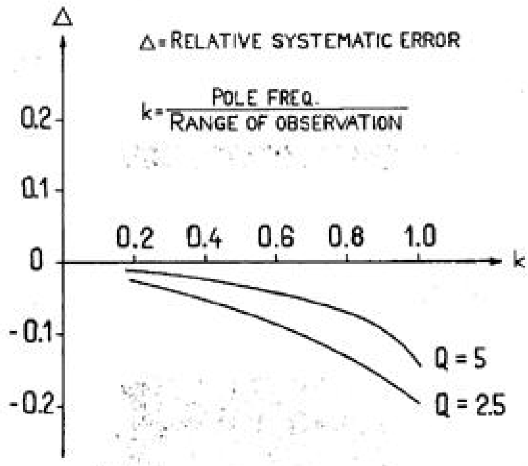

Fig. I-5 Systematic error in the tracking of a one-formant vowel owing to prefiltering with a LP filter.

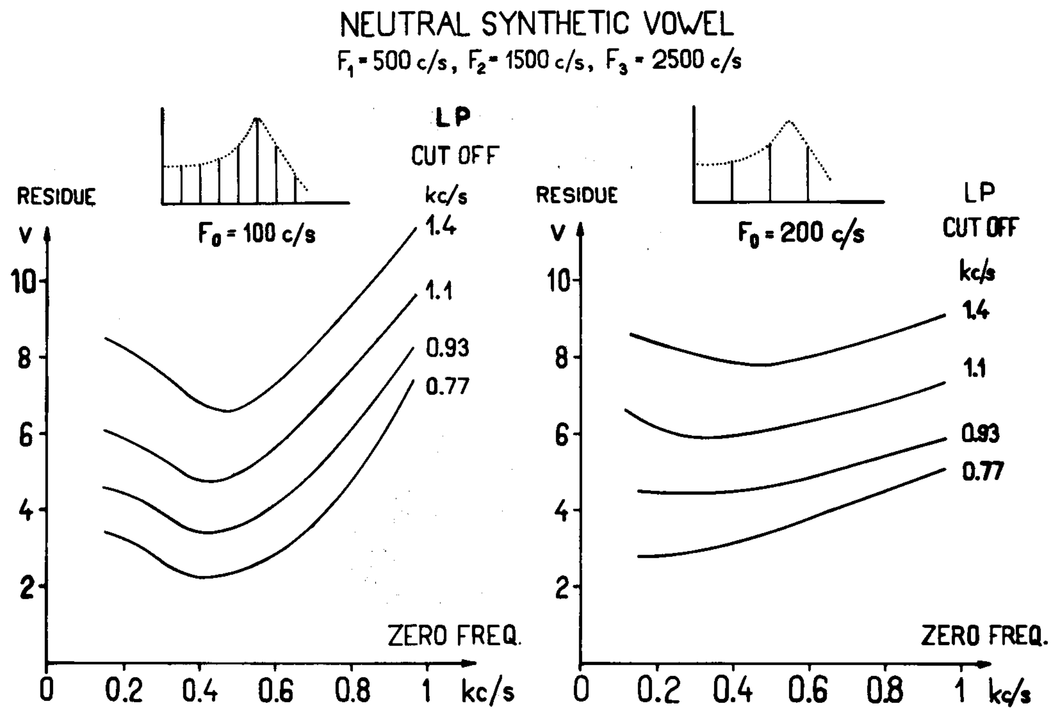

Fig. I-6 Anti-resonance analysis of the F_1 - range of a neutral vowel at two different F_0 - pitches and varying prefiltering LP cut off frequency.

the frequency and bandwidth of each anti-resonance it would be possible to cancel the damped oscillations of all formants and there would remain merely a voice excitation function. This is actually the method utilized in the voice-source studies, section I-A of this report. In formant-tracking systems, however, it is necessary to adopt a matching criterion that is independent of a human operator.

Our efforts have concentrated on the theoretical and experimental study of tracking the first formant after it has been isolated by a prefiltering stage. The theoretical approach was to calculate the intensity output of the anti-resonance circuit as a continuous function of its center frequency setting. The bandwidth of the zero was assumed to coincide with that of the formant pole but an exact match is not critical. The experimental set up consisted of measuring the linear rectified and smoothed output of an anti-resonance circuit varied in steps of 100 c/s between successive recordings thus simulating a bank of parallel channels of anti-resonance filters. The F_1-frequency was selected as the channel carrying minimum intensity. A fairly promising result was obtained in a test with a piece of connected human speech as can be seen in Fig. I-4, where the data are superimposed upon a spectrogram. A theoretical analysis of the anti-resonance method provides the following results:

a. There is a systematic error when the low-pass prefiltering range is narrowed as may be seen from Fig. I-5. This error is approximately inversely proportional to the Q of the formant and amounts to 10 % when the cutoff of the prefilter is 10 % above the formant frequency assuming a Q of 5 and further a very low F_0 and a negligible influence from higher formants.

b. Systematic errors of greater importance occur when the voice fundamental frequency, F_0, is high and especially when the formant peak falls halfway between two harmonics. This is illustrated in Fig. I-6, which pertains to the experimental analysis of a standard neutral vowel [ɜ] (F_1=500, F_2=1500, F_3=2500 c/s).

At the pitch of F_0=200 c/s the selectivity is very bad and a minimum is obtained only with the widest possible prefilter.

The fundamental weakness of the approach reported on above is that only a limited part of the spectrum is allowed to contribute to the minimum response whereas a frequency shift of a formant affects the entire spectrum. The sensitivity of the method is upset by incomplete cancellation of residues from other formants and from the vocal-cord wave. A cascaded system of several anti-resonances, one for each formant, would do better but is more complicated for automatic tracking systems. The supplementary usage of phase information might provide greater accuracy.*

H. Fujisaki, A. Risberg

* Such a system is that of Lawrence (personal communication) who works with a system of two self-adjusting anti-resonance circuits which continuously follow the first two formants of speech. His criterion for self-adjustment is the relative phase between the output and the input of the anti-resonance circuit.

(1) Chang, S.-H.: "Two schemes of speech compression system", J.Acoust. Soc.Am. 28, 565-572 (1956).

(2) Holmes, J.N.: "A method of tracking formants which remains effective in the frequency regions common to two formants", Res. Rep. JU8-2, Joint Speech Research Unit, B.P.O., Dec. 1958; and

Holmes, J.N., Kelly, L.C.: "Apparatus for segmenting the formant frequency regions of a speech signal", Res. Rep. JU9-4, Joint Speech Research Unit, B.P.O., Aug. 1959.