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

Auditory matching of vowels with two formant synthetic sound

Fant, G. and Risberg, A.

journal: STL-QPSR
volume: 4
number: 4
year: 1963
pages: 007-011

KTH Computer Science and Communication

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

II. SPEECH PERCEPTION

A. AUDITORY MATCHING OF VOWELS WITH TWO FORMANT SYNTHETIC SOUNDS

It has been suggested (1)(2)(6)(7) that vowels may be regarded as essentially two-formant stimuli as far as their spectral characteristics are concerned. The first formant of the vowel would be one of the constituents and a weighted average of the second and higher formant frequencies would constitute an effective "mean" pitch of the upper part of the vowel spectrum. A further simplification holds for back vowels with F_2 close to F_1. The main energy of back vowels is concentrated to F_1 and F_2 and these vowels can be approximated by a single formant placed close to the mean of F_1 and F_2 in the case of [u] closer to F_1. Even front vowels can be associated with a single formant corresponding to the "mean pitch" of the second and higher formants. The very extreme is to ask subjects to associate pure tones of various frequencies with vowels (ref. (2) page 87). Such tests provide results which are consistent with the general theory of vowel perception.

Our means of interpreting data from spectrographic analysis would gain from a knowledge of how the ear weights the several formants of front vowels. Frequently the third formant of [i] has a higher intensity level than the second formant and a fourth, fifth, and even a sixth formant may appear in the spectra of some front vowels produced by a male voice with well developed timbre. What is the effective mean of such a group of higher formants? Is it possible to derive the rules for the particular weighting function applied by the auditory system? It is felt that an application of such rules to a female and a male sample, pertaining to a vowel of one and the same phonetic quality, would provide center of gravity measures that differed less than, for instance, the F_2-frequency.

A general interest in these matters was apparent in the 1950's. (3)(4)(5)(8) Attention was devoted to the possible invariance of vowel quality with certain F_2/F_1-ratios and F_2/F_3- versus F_1/F_3-ratios and other formulas.

Vowels can be synthesized with great naturalness and exactness on the basis of a set of F_1- F_2- F_3- F_4- etc. values and other pertinent information but we still make tremendous over-simplifications, if we attempt to compare the phonetic similarities and dissimilarities in terms of a F_2- versus F_1-plot only. The vowels [i] and [y], for instance, would fall almost on the same point in the vowel diagram. One suggestion for over-coming this difficulty when plotting vowel data was to calculate an effective F_2 called F_2' from the particular F_1 F_2 and F_3 (ref. (2), page 80). This formula improves the relative crowdedness in the area of the high front vowels but it is rather arbitrary and does not provide significant measures for back vowels.

One possible technique that should be tried would be to compare the spectra from an analyzer approximating the function of the inner ear organs with that of a conventional analyzer. Our 20-channel LUCIA spectrograph used for speech training with hard of hearing subjects performs a semi-auditory analysis but is not specifically designed as a cochlear model. The LUCIA display is less sensitive to typical male-female differences than the Sona-Graph.

No single experiments can be expected to provide all data pertinent for defining the auditory weighting function. Here follows a report on a pilot experiment we made several years ago based on two-formant synthetic vowels which are varied to provide a best match with human vowels. Other types of stimuli which will be used in coming experiments should have more selective formants in contrast to the very gradual slopes offered by the single resonant circuits utilized for the spectrum shaping in the previous experiments.

The two-formant synthesizer offers a fairly natural reproduction of the main characteristics of back vowels with F_2 close to F_1. In case of front vowels there may be several higher formants of equal strenght building up a sort of pass-band region of equal ripple type as is typical of a horn resonator. A single resonance curve substituting such a group is not very natural and the same can be said of a narrow spectral band shaped with a band-pass filter with very steep skirts. The latter may, however, provide suitable conditions for specifying a lumped energy distribution on the frequency scale. Both these two-formant approximations should be of interest for matching purposes.

Typical spectral envelopes of the two-formant vowels utilized in our investigation are shown in Fig. II-1. Six subjects were asked to sustain each of 10 Swedish vowels. They were then set to the task of matching tape-recordings of their own vowels with synthetic sounds, F_0 and F_1 being preset from measurements of their spoken sample. The difference in second and first formant amplitude, L_2-L_1, was also preset according to an average frequency dependency rule. The only variable was thus F_2 of the synthesis which is denoted by F_2' in Table II-1. Each subject made on the average 2 matches. A series of repeated matchings carried out by one of the subjects showed that the reproducibility was generally within less than 100-200 c/s except for the high front vowel [i], where the spread could be very large.

The general tendency was to match F_2' closer to F_1 than F_2 of the vowel [u] and closer to F_2 than F_1 for [o]. The vowels [ɑ], [ø], [œ], [ʉ] were matched with F_2 quite close to F_2 of the human vowels, whereas [æ] was matched closer to F_3 than to F_2. For the vowel [e] a F_2'-frequency just above F_3 was preferred and in case of [i] most subjects attempted to place F_2' above F_4.

This very high location and the considerable spread encountered is ascribed to the short-comings of the synthesis.

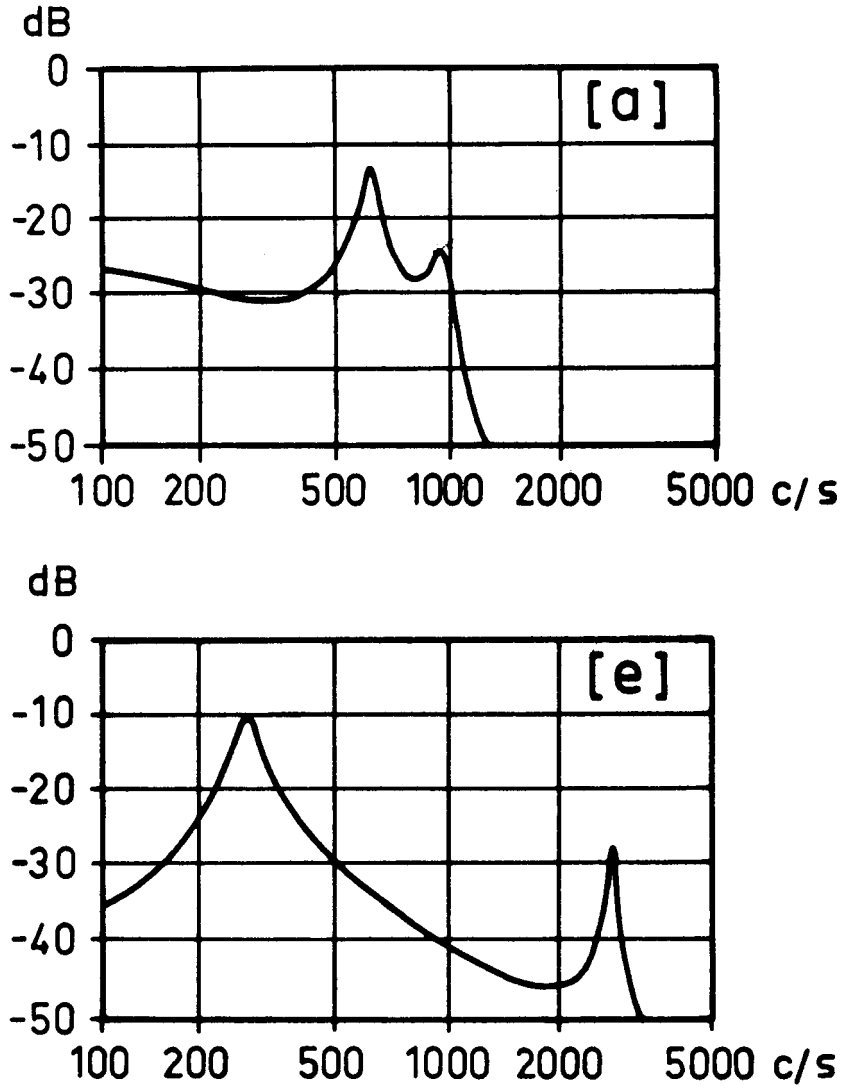

Fig. II-1. Spectrum envelopes of two-formant synthetic vowels [e] and [ɑ]. Parallel synthesis with phase inversion.

Under conditions of high $L_2'-L_1'$ the F2 appears as a small peak only on the skirt of the F_1-resonance curve.

The data are summarized in the vowel diagram of Fig. II-2. The representation of vowels by their two-formant approximations from the matching experiments has apparently eliminated the crowdedness in the [i] [e] [y] [ø] region of the vowel diagram thus sharpening the apparent distinctions. It remains, however, to test other synthesis methods and to gain deeper insight in the invariance criteria, e.g. by investigating the finite trading relation between formant frequency and formant intensity (3) and between F_o and apparent F_1' (8).

TABLE II-1.

(Mean of 10 values in all 6 subjects)

	Formant frequencies of natural vowels			F_2' of synthetic $L_2'-L_1'$ approximation	
Sustained vowels	F_1 c/s	F_2 c/s	F_3 c/s	F_2' c/s	$L_2'-L_1'$ dB
[u]	335	620		440	14
[o]	400	700		635	13
[ɑ]	620	920		940	12
[æ]	635	1810	2525	2320	14
[e]	350	2270	2775	2950	17
[i]	240	2210	3100	4700	22
[y]	240	2060	2720	2400	22
[ʉ]	235	1770	2280	1670	21
[ø]	350	1770	2260	1820	18
[œ]	480	1250	2300	1310	14

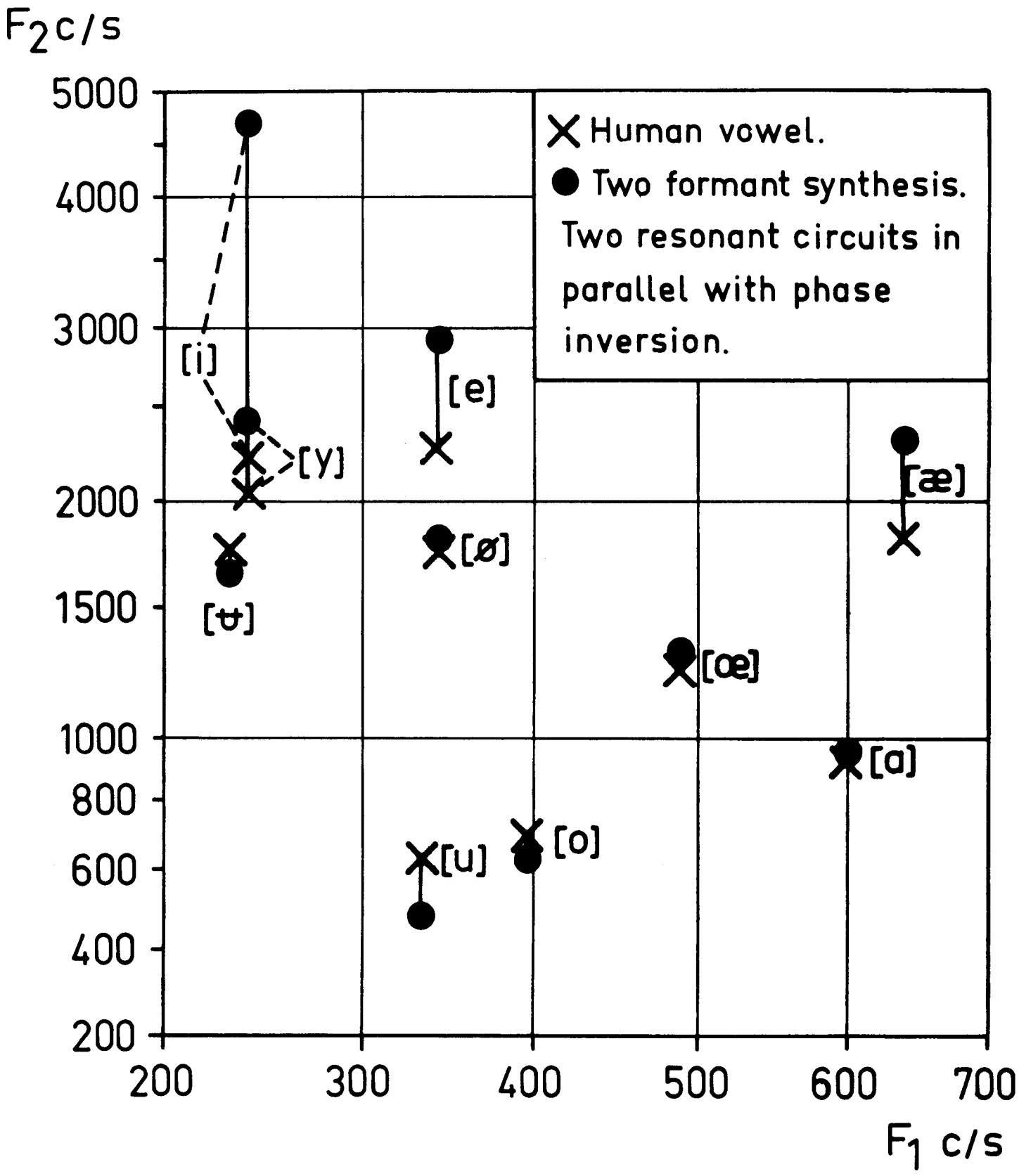

Fig. II-2. F_2 versus F_1 plot of Swedish sustained vowels and F_2' versus F_1 of two-formant synthetic approximations, $L_2 - L_1$ ranging from 12 dB in [ɑ], to 22 dB in [y] and [i]. The very high F_2 of [i] may probably be ascribed to the particular synthesis technique.

References:

(1) Fant, G.: "Modern instruments and methods for acoustic studies of speech", Acta Polytechnica Scandinavica Ph 1 (246/1958).

(2) Fant, G.: "Acoustic analysis and synthesis of speech with applications to Swedish", Ericsson Technics, 15, No. 1 (1959), pp. 3-108.

(3) Miller, R.L.: "Auditory tests with synthetic vowels", J.Acoust.Soc.Am., 25 (1953), pp. 114-121.

(4) Potter, R.K. and Steinberg, J.C.: "Toward the specification of speech", J.Acoust.Soc.Am., 22 (1950), pp. 807-820.

(5) Peterson, G.E. and Barney, H.L.: "Control methods used in a study of the vowels", J.Acoust.Soc.Am., 24 (1952), pp. 175-184.

(6) Delattre, P., Liberman, A.M., and Cooper, F.S.: "Voyelles synthétiques a deux formantes et voyelles cardinales", Maître Phonétique, 96 (1951), pp. 30-36.

(7) Delattre, P., Liberman, A.M., Cooper, F.S., and Gerstman, L.J.: "An experimental study of the acoustic determinants of vowel color", Word, 8 (1952), pp. 195-210.

(8) Fant, G.: "Discussion of paper read by G.E. Peterson at the 1952 Symposium on the applications of communication theory", publ. in Communication Theory, ed. by W. Jackson, London 1953, pp. 421-424.