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

A note on vocal tract size factors and non-uniform F-pattern scalings

Fant, G.

journal: STL-QPSR
volume: 7
number: 4
year: 1966
pages: 022-030

B. A NOTE ON VOCAL TRACT SIZE FACTORS AND NON-UNIFORM F-PATTERN SCALINGS

G. Fant

Introduction

The common concept of physiologically induced differences in formant patterns comparing males and females is that the average female formant frequencies are related to those of the male by a simple scale factor inversely proportional to the overall vocal tract length. Thus on the average the female F-pattern (F_1 F_2 F_3 etc) is said to be scaled to about 20 % higher frequencies than the average male F-pattern. Children have even higher formant frequencies than grown up females and it is also well known that the individual size of the vocal cavities and thus of the F-pattern scale factor may vary appreciably among each age and sex category. A parallel reasoning is inherent in the common concept of perceptual invariance. Within certain limits vowels retain their phonemic identity if formant frequency ratios are preserved as can be judged from playing a tape or grammophone record at a somewhat higher or lower speed than normal.

The purpose of this report is to point out that the simple scale factor rule has important limitations. A range of typical and substantial deviations from this rule is concealed if the data are averaged over the whole vowel system of a language. Actually the female to male relations are typically different in the three groups of (1) rounded back vowels, (2) very open unrounded vowels, and (3) close front vowels. The main physiological determinant of the specific deviations from the average rule is that the ratio of pharynx length to mouth cavity length is greater for males than for females and that the laryngeal cavities are more developed in males.

Experimental data

In the years 1946-1947 I collected a material on formant frequencies and formant amplitudes of sustained vowels uttered by 7 male and 7 female Swedish subjects of a fairly homogeneous dialectal background. These data were originally published in an internal report of the Ericsson Telephone Company (2). A few years later the Peterson and Barney-study of American English vowels was published (7). Their investigation at Bell Telephone Laboratories was concerned with formant data sampled within test words of an h+vowel+d structure. It is the most extensive collection of formant data ever published, based on 33 men, 28 women, and 15 children. This American English study and my Swedish study are the only reliable sources of data on both formant frequencies and amplitudes.

At an information theory symposium in London, 1952, I gave a brief review of the general conformity of the Peterson and Barney-data with my own data [3], especially with respect to the average female to male relations in various vowel categories. This correlation was also mentioned in my monograph on acoustic analysis and synthesis of speech [4] in connection with a review of the material on Swedish vowels and consonants.

Here follows a more detailed attempt of physiological-acoustic interpretation of how female F-patterns differ from those of men.

Fig. II-B-1 illustrates the relation of female to male formant frequencies of Swedish vowels and comparable American English vowels selected according to an approximate F-pattern match.

The basic data are included in the followi ng tabulation:

American English (Peterson and Barney, 1952) [7] (A)

Swedish (Fant, 1959) [4] (S)

$$k_n = \left(\frac{F_n, \text{ female}}{F_n, \text{ male}} - 1\right) 100 \text{ per cent}$$

The Swedish vowels are denoted by a technical alphabet with index 1 for long vowels and 2 for short vowels and 3 for long pre-r allophones. The American English vowels are denoted by their IPA symbols.

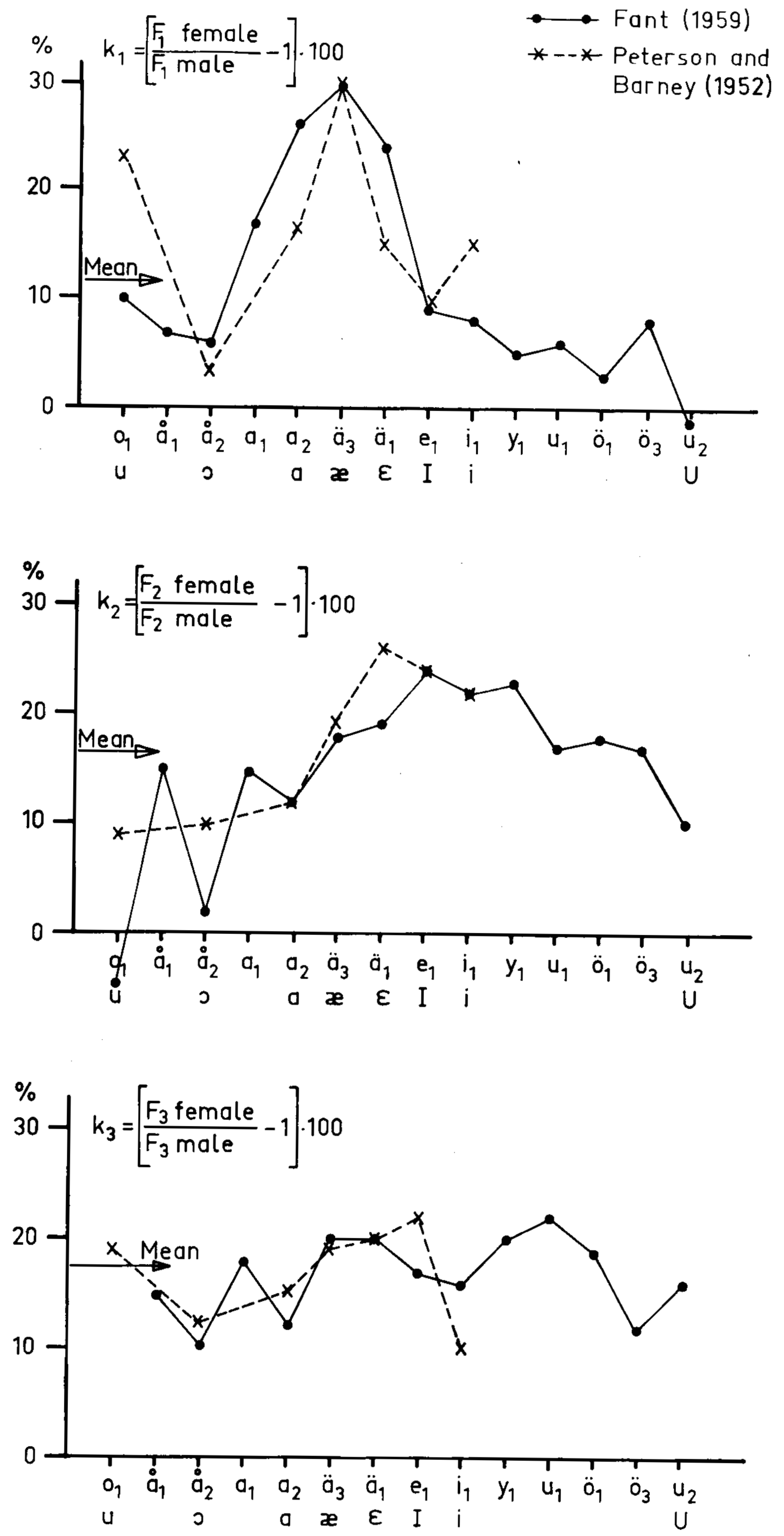

Fig. II-B-1. The average female/male formant ratio of F_1, F_2, and F_3. American English and Swedish data.

(A)	(S)	F_1 c/s	k_1 c/s	F_2 c/s	k_2 c/s	F_3 c/s	k_3 c/s
	o_1	310	10	710	-5	2230	
u		300	23	870	9	2240	19
	$å_1$	402	7	708	5	2460	15
	$å_2$	487	6	825	2	2560	10
ɔ		570	3.5	840	9.5	2410	12
	a_1	582	17	940	15	2480	18
	a_2	680	26	1070	12	2520	12
ɑ		730	16.5	1090	12	2440	15
	$ä_3$	606	30	1550	18	2450	20
æ		660	30	1720	19	2410	18.5
	$ä_1$	438	24	1795	19	2385	20
ɛ		530	15	1840	26.5	2480	20
	e_1	334	9	2050	24	2510	17
I		390	10	1990	24.5	2550	20
	i_1	256	8	2066	22	2960	16
i		270	15	2290	22	3010	10
	y_1	257	5	1928	24	2420	20
	$ʉ_1$	283	6	1633	17	2140	22
	$ö_1$	363	3	1690	18	2200	19
	u_2	416	-1	1070	10	2315	16
U		440	7	1020	14	2240	19.5
	$ö_3$	525	8	1103	17	2430	12
ʌ		640	19	1190	17.5	2390	16.5
ɚ		490	2	1350	21.5	1690	16

It can be seen that English and Swedish data correlate well. The first formant "sex factor" k_1 displays a pronounced maximum of 30 % in the very open back and front vowels [a] and [æ] and a minimum of the order of 5 % in the rounded, half open back vowels [o] and [ɔ]. Close front vowels and especially close rounded front vowels also show a low k_1.

The second formant "sex factor" k_2 is significantly lower in back vowels than in front vowels. A certain irregularity in the back vowel region is ascribed to the inherent spread of the original data. The k_3-data does not show equally promiment vowel class dependency. An exception is the relatively low k_3 of [ɔ]. The k_n-average of the entire ensemble of Swedish vowels is 11.5 % in F_1, 16.5 % in F_2, and 17,5 % in F_3.

Similar trends are found in a comparison of the American English male and children data, as can be seen from Fig. II-B-2. The children's formants are of about 35-40 % higher than those of the male group. The extreme range within any of the formants and any of the vowels are the k_1 = 20 % of [ɔ] and the k_1 = 54 % of [æ]. The k_n-curves of the children are fairly parallel to those of the women and about 20 % higher. It can accordingly be concluded that while children's F-patterns are scaled fairly proportionally relative to females this is not the case when the females are compared with the men.

Physiological interpretation

The major anatomical constraints on vowel articulation that can be correlated with these findings are the relatively greater pharynx length and more pronounced laryngeal cavities of grown up males compared with females and children, see Ref. (1) pp. 188-193. These authors state that in a girl of eight years the length of the mouth cavity, between the incisor and the pharynx wall, is 30 % smaller than that of a grown up male whilst the length of the girls' pharynges is 56 % shorter than that of the male. Chiba and Kajiyama (1) also exemplify the overall vocal tract length in relative numbers as 1.0 for males, 0,87 for females, 0.80 for a boy of nine and 0.70 for the girl of eight.

Now, in a first attempt to explain the systematic trends in Fig. II-B-1 it is first of all apparent that formants produced with a typical double Helmholtz resonator configuration as e.g. F_1 and F_2 of rounded back vowels are less critically dependent on the overall vocal tract length than other formants. A smaller overall length could be compensated for by more narrow lip opening and a more narrow tongue hump passage. Thus the low k_1 and k_2 of these vowels as will be discussed in more detail at the end of this article.

In other vowels the vocal cavities behave more like standing wave resonators. If the tract approximated a uniform tube as in the neutral vowel a reduction in length by 18 % would cause all formants to rise by the same amount. The American English ʌ conforms fairly well with this pattern in terms of the observed k_1, k_2, and k_3.

It is known from my earlier work (5)(6) to what extent a small reduction in length of any of the major vocal tract cavities or constrictions affects the F-pattern. The percentage relative increase

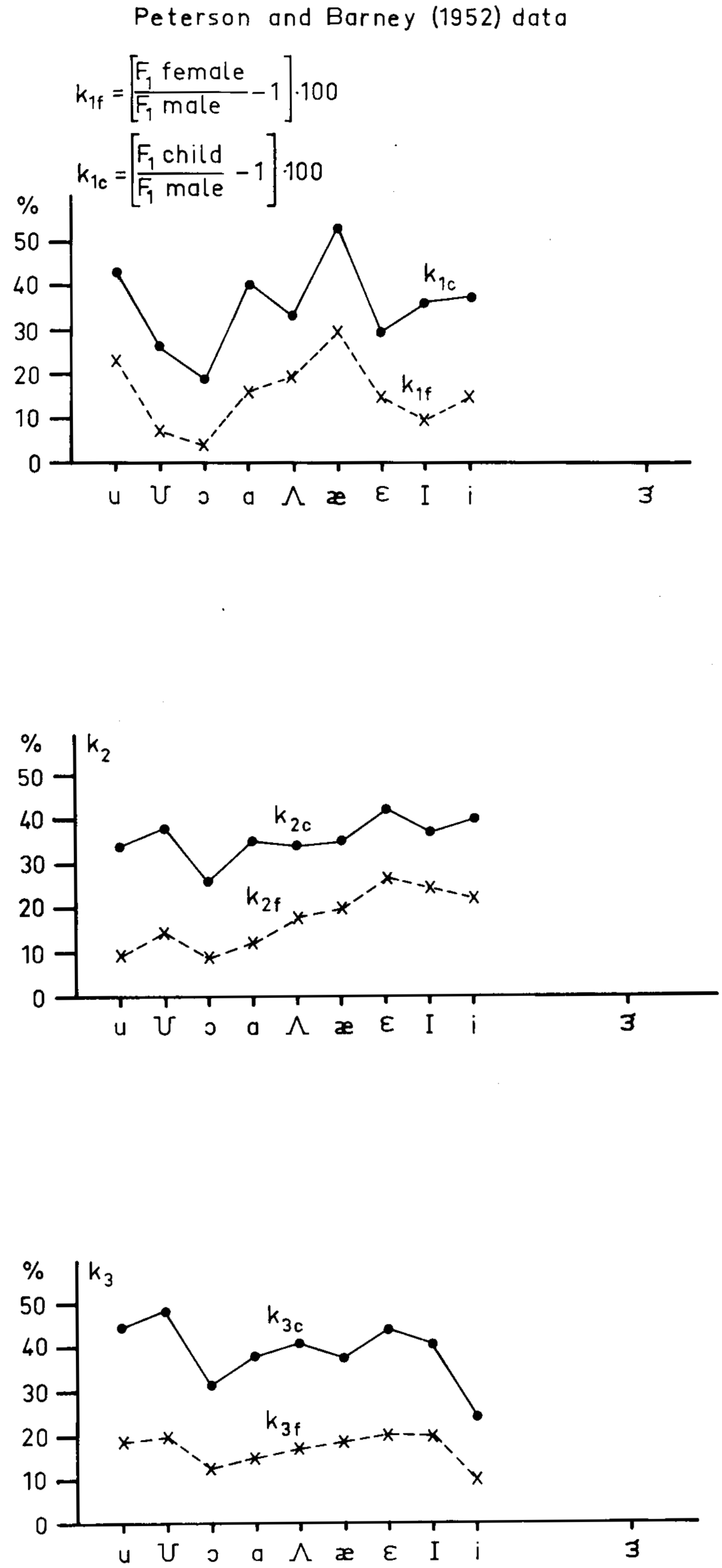

Fig. II-B-2. The average child/male formant ratio. American English data.

in formant frequencies associated with the removal of one-half centimeter of the pharynx of the vowel [i] is, accordingly to Table 2.33-4 of p. 120 in Ref. [5], 3.5 % increase in F_1, 4.7 % increase in F_2, and 0.5 % increase in F_3. Similarly a removal of a-half centimeter section of the frontal mouth cavity of [i] results in 1.3 % increase in F_1, 0.2 % increase in F_2, and 6.1 % increase in F_3.

Inversely, the observed k_2 = 22 % of [i] of the Swedish data above would correspond to a $\frac{22}{2 \cdot 4.7}$ = 2.3 cm shorter pharynx for the females compared to males. Similarly, noting the close association of F_3 with the mouth cavity and working from k_3 = 16 % of [i] we arrive in a $16/2 \cdot 6.1 = 1.3$ cm shorter female moth cavity. The extremely small contributions of mouth cavity length to F_2 and of pharynx cavity length to F_3 has been neglected in the calculations above.

In my older work I accordingly referred to F_2 of [i] as pharynx dependent and F_3 of [i] as mouth cavity dependent. These "formant-cavity affiliations" are apparent in a two-tube simplified model of the vowel [i] comprising a wide back tube and a narrow and somewhat shorter front tube. F_2 is associated with a half wave-length resonance of the back tube and F_3 with a half wave-length resonance of the narrow front tube. The lengths of these tubes matching the average Swedish male [i] would be:

$$\ell_b = \frac{c}{2F_2} = \frac{35300}{2 \cdot 2070} = 8.5 \text{ cm back tube}$$

and

$$\ell_f = \frac{c}{2F_3} = \frac{35300}{2 \cdot 2960} = 6 \text{ cm front tube}$$

In the next degree of approximation the mouth is conceived of as a symmetrical double horn with the two connecting narrow throats in the center of the mouth. Here again F_3 is associated with the basic "one-quarter wave-length" resonance of the front horn which is the same as the half wave-length resonance of the entire mouth.

It should be stressed that the estimates above of a 2.3 cm shorter female pharynx and a 1.3 cm shorter mouth cavity are approximate only since the coefficients were derived from a Russian articulation and are valid for small perturbations only.

For the normal male the total mouth cavity length from the incisors to the back pharynx wall is of the order of 8 cm. This measure is appropriately 2 cm greater than the front cavity length of the vowel [i] above. The total pharynx length from the soft palate to the level of the glottis is of the order of 8.5 cm. According to the calculations above the corresponding female measures would be 8 - 1.3 = 6.7 cm and 8.5 - 2.3 = 6.2 cm.

How do these inferred data compare to actual physiological data? From a material collected by the radiologist Paul Edholm of Karolinska Sjukhuset, Stockholm, the details of which have not been published I have collected the following measurements. A male articulating the vowel [i] had a pharynx length from glottis to the soft palate of 9.1 cm, and a mouth length from the incisors to the back pharynx wall of 8.25 cm. The corresponding female measures were 7.0 cm and 7.0 cm respectively. Thus the female pharynx was 2.1 cm shorter than the male pharynx and the female mouth 1.25 cm shorter than the male mouth, which conforms well with the previously inferred measures. Furthermore a detailed mapping of the vocal tract area functions involved provided the basis of calculating the vocal resonances of the subject's [i] which came out as follows:

Observed and calculated data for the vowel

	F_1		F_2		F_3	
	obs.	calc.	obs.	calc.	obs.	calc.
male	225	156	2060	2060	2960	3290
female	240	247	2600	2650	3550	2900

There is a specially good agreement in terms of F_2. The ratio of male to female pharynx lengths 9.1/7.0 = 1.30 is close to the observed 2650/2060 = 1.26 ratio and calculated ratio 1.28 of female F_2 to male F_2. The observed female F_3 divided by the observed male F_3 is 3550/2960 = 1.20 and the corresponding calculated ratio 3290/2900 = 1.14 are to be compared with the ratio of male mouth to female mouth lengths 8.25/7 = 1.18. In terms of absolute measures projected on the simple closed tube model referred to above a male pharynx length of 9.1 cm should provide an F_2 of 35300/2·9.1 = 1940 c/s which is only 6 % below the value of the complete calculation.

Chiba and Kajiyama [1] do not provide directly comparable measures. However, their data on the mouth and pharynx length of a girl of eight compared with a grown up male are expressed as a mouth cavity scale factor 0.77 and a pharynx scale factor of 0.64 which reflect the same tendency.

Now as a second step returning to the average Swedish male and female data and given the particular $\Delta\ell_p = 2.3$ cm shorter female pharynx and $\Delta\ell_m = 1.3$ cm shorter female mouth, what is their differential effect on the formant patterns of other vowels than [i] ? The tabulated data on p. 120 of Ref. [5] provides the necessary basis for such a check.

$$k_n = \Delta\ell_m \left(\frac{dF_n}{F_n}\right)_m + \Delta\ell_p\left(\frac{dF_n}{F_n}\right)_p$$

For the Russian vowel [e] corresponding to the Swedish $/\ddot{a}_1/$ we get

$k_1 = 20\ \%$, $k_2 = 20\ \%$, $k_3 = 19\ \%$

which is close to the observed

$k_1 = 24\ \%$, $k_2 = 19\ \%$, $k_3 = 20\ \%$

Inserting the perturbation coefficients of Russian [ɑ] for use with Swedish [a] we calculate

$k_1 = 28.5\ \%$, $k_2 = 23\ \%$, $k_3 = 9.5\ \%$

which is to be compared with the observed

$k_1 = 26\ \%$, $k_2 = 12\ \%$, $k_3 = 12\ \%$

The relatively high k_1 is predicted as well as a low k_3 but k_2 comes out to be high.

For the rounded back vowel [ɔ] the simple length perturbation leads to much higher k-values, 27 %, 24 %, and 19 % compared with the observed 6 %, 2 %, and 10 % respectively. Also the k_1 of [i] comes out at 19.5 % compared with the $k_1 = 8\ \%$ observed.

The vocal tract filtering may be approximated by a simple lumped element (Helmholtz) resonator for F1 of [i] and double Helmholtz resonator for F1 and F2 of [u], [o], and [ɔ]. The general formula for such a resonator of volume V, orifice length ℓ, and orifice area A is

$$F_r = \frac{c}{2\pi} \sqrt{\frac{A}{\ell \cdot V}}$$

If all dimensions are linearly scaled by one and the same factor k the resonance frequency comes out as F_r/k, which is the same effect as in standing wave resonances. If we assume that the length of the resonator neck (main tongue hump constriction and lip constriction) is the same for two speaker categories whilst all other dimensions (cavity length, cavity lateral width, and orifice effective diameter) are scaled by the same factor k we arrive at a resonance frequency of $F_r/(k)^{1/2}$. This scaling conforms with the observed k_1 of close front vowels and of rounded back vowels which are smaller than the average. The above physiological explanation is only one variant of many possible combinations. The low k_1 of the above discussed vowel category could also be the result of an attempt to tune F_1 closer to the male reference by decreasing the orifice areas A by more than k^2.

Summarizing these results:

(1) The scale factor relating average female formant frequencies to those of men is a function of the particular class of vowels. The American English vowel data display the same vowel category dependency of this factor as the Swedish data.

(2) The female to male scale factor is of the order of 18 % averaged over the whole vowel system. The typical deviations from this rule are:
  (a) The first and second formants of rounded back vowels have relatively low scale factors;
  (b) This is also the case with the first formant scale factor of any close or highly rounded vowel, i.e. high front vowels;
  (c) Very open front or back vowels display a first formant "sex factor" k_1 which is substantially higher than the average.

(3) These findings conform with anatomical constraints of the average female vocal tract. The particular scaling from male to female tract reduces the pharynx length more than the length of the mouth. Other differences may also contribute*, e.g. the relatively small female laryngeal cavities. More detailed anatomical studies and calculations are needed.

(4) The scaling of children's data from female data comes closer to a simple factor independent of vowel class.

---

*It has been suggested by Sven Öhman that a proportionally larger female mouth opening is a factor to consider.

These female/male departures from a uniform scaling are of some interest when attempting to normalize formant data. It is not within the scope of the present article to discuss their perceptual implications. They may not have a very crucial importance for the phonemic identity of perceived vowels in connected speech but are undoubtedly of interest as speaker category determinants.

References:

(1) Chiba, T. and Kajiyama, M.: The Vowel - Its Nature and Structure (Tokyo 1941).

(2) Fant, G.: "Analys av de svenska vokalljuden", L M Ericsson protokoll H/P 1035 (1948).

(3) Fant, G.: "Discussion of Paper Read by G.E. Peterson at the 1952 Symposium on the Applications of Communication Theory", Communication Theory (London 1953), pp. 421-424.

(4) Fant, G.: "Acoustic Analysis and Synthesis of Speech with Applications to Swedish", Ericsson Technics 15 (1959), pp. 3-108.

(5) Fant, G.: Acoustic Theory of Speech Production ('s-Gravenhage 1960).

(6) Fant, G.: "Formants and Cavities", Proc. 5th int. Congr. phon. Sci., Münster 1964 (Basel 1965).

(7) 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.