VOWEL QUALITY AND WAVE SHAPE

CELIA SCULLY

Many acoustic studies have been made of vowels and the frequencies F_1 and F_2 of the first and second formants are generally used as parameters for vowel quality. There are several objections to these measures however. Firstly, similar vowel sounds made by men, women and children have very different formant patterns. In the case of [æ] for example, a child's first formant can be about 50 % higher than a man's.[1] Some method of normalizing formant data seems essential. The ratio of F_2/F_1 seems to be a better measure for front vowels than the absolute values of F_2 and F_1, but this is not true of back vowels.[2] Some overlap occurs in the F_1 versus F_2 plot. For example a man's [*a*] and a child's [ɔ] can have the same values for F_1 and F_2; by changing the fundamental frequency but not the formant frequencies the vowel quality can be changed.[3] Secondly, it is not always easy to separate adjacent formants or indeed, for some back vowels, to find spectral peaks at all. This is because the resonances of the vocal tract do not always result in clearly separated peaks in the frequency spectrum. Thirdly, an F_1 versus F_2 plot takes no account of the relative strengths of the formants. In close front vowels F_2 is arbitrarily selected from the several strong high formants, of which F_2 is not always the strongest; and it has been shown that the relative amplitudes of the second and third formants can influence the perceived vowel quality.[4]

Although different speakers probably articulate similar vowels in a similar way, thus producing their characteristic formant patterns, it does not necessarily follow that perception of these sounds proceeds in the same way. Indeed, formantless complex sounds can be identified consistently as different vowels.[5] It may be that

---

[1] G. Fant: „A Note on Vocal Tract Size Factors and Non-Uniform F-Pattern Scalings". Royal Inst. of Tech. (Stockholm) *QPSR*-4/1966 pp. 22—30.

[2] G. E. Peterson: "Parameters of Vowel Quality". *J. of Speech and Hearing Res.* 4 (1961) pp. 10—29.

[3] R. K. Potter and J. C. Steinberg: "Toward the Specification of Speech". *J. of the Acous. Soc. of Amer.* 22 (1950) pp. 807—820.

[4] R. W. Peters: "Dimension of Quality for the Vowel [æ]". *J. of Speech and Hearing Res.* 6 (1963) pp. 239—248.

[5] A. Carpenter and J. Morton: "The Perception of Vowel Colour in Formantless Complex Sounds". *Land. and Speech* 5 (1962) pp. 205—214.

in speech perception the pitch of the voice producing a vowel sound is taken into account by the listener in assessing vowel quality, but in the experiment reported here the aim was to look for purely acoustic measures of vowel quality; ones not based on knowledge of the speaker's physiology.

Vowel sounds of Southern English were made by a few speakers imitating the experimenter. 62 of the sounds were made into 200 mS segments and presented in random order to 10 phonetically trained listeners, who were asked to identify each stimulus with one or more phonetic symbol, using English Received Pronunciation as a guide. [*i*], [æ], [*a*] and [*u*] were identified with nearly 100 % agreement. Formant frequencies were measured for the sounds and, in addition, detailed records of the wave shapes or oscillograms, were made, using an Ultra-Violet Recorder running at 100 ips and responding up to 7 000 Hz without significant distortion of the wave shape.

Fig. 1. shows the wave shapes for four sounds intended as [*i*] and identified as such by at least 9 out of the 10 listeners. Two prominent oscillations, or ripples, can be distinguished easily and their frequencies measured. The slower ripples have frequencies between 260 Hz and 350 Hz, corresponding to the first formants. The faster ripples have frequencies ranging between 3 030 Hz for a man and 3 640 Hz for the child. That is, near or a little above the frequency of F_3 for the men, between F_2 and F_3 for the woman and near or a little above F_2 for the child. The spread in ripple frequencies is very much smaller than the spread in F_2 frequencies, as table 1. shows.

Table 1.

Speaker	Ripple Frequencies			Formant Frequencies	
[*i*]	*F*. Low	*F*. High	F_H/F_L	F_2	F_2/F_1 (approx.)
RS (Child)	352 Hz	3640 Hz	10.3	3600 Hz	10.3
CS (Woman)	272	3270	12.0	2900	11.6
PADM (Man)	266	3180	11.9	2250	7.5
WRBA (Man)	273	3030	11.1	2200	7.3

Fig. 2. shows that, in some cases at least, this high frequency ripple is not merely the strongest of the high frequency formants, but is an amplified and damped wave resulting from two nearby prominent harmonics. Here the waveshapes of three [*i*] sounds are shown after high-pass filtering at 1 500 Hz. The two children's wave shapes show beating between two adjacent harmonics and the man's wave shape shows beating between harmonics separated by four times the fundamental frequency. The frequency agrees well with results for two-formant synthetic vowel sounds.[6,7]

Frequency and time measures are two aspects of the same thing; a formant shows as a frequency band in a spectrogram or as a ripple in the wave shape. But the wave shape, unlike the spectrogram, reveals combinational features of the formants.

For [æ] there is one prominent ripple at about 1 900 Hz. This is between F_1 and F_2 for the children, but between F_2 and F_3 for the adults. Back vowels showed ripples decreasing in frequency down to [*u*]. [ɔ] and [*a*] were well separated by their ripple frequencies although their F_1 and F_2 frequencies overlapped.

These results tend to support the view that front vowels are perceived as two-formant sounds, back vowels as one-formant sounds with [æ] or [*a*] at the boundary. Whether the measure of vowel quality should be absolute frequencies or ratios of frequencies, ripples in the wave shape could possibly be more appropriate than formants in the spectrogram as acoustic parameters of vowel quality.

DISCUSSION

Fujisaki:

It is quite true that the Sonagraph is far from an ideal tool for frequency analysis, but I believe one should not denounce the formant concept because of the incomplete instrumentation.

Maiwald:

A short remark to the phase discrimination in the human ear: In masking experiments with several harmonic components with fixed relative phase as masker, it can be shown that the threshold for an additional sinusoidal test-tone (with a frequency lying somewhere in between) is strongly dependent on the phase relation between the spectral components of the masker though the overall loudness of the complex sound does not change. Consequently there seems to be some evidence that the phase information is still present on a high processing level and might, in fact, influence the sound quality.

Merhaut:

What is the influence of phase of formants on the perception (when there is a strong influence on the shape)?

Slis:

Do you find with visual inspection of oscillograms of whispered vowels a differentiation between the different vowels comparable to the results with the normally spoken ones?

---

[6] P. Delattre, A. M. Liberman, F. S. Cooper and L. J. Gerstman: "An Experimental Study of the Acoustic Determinants of Vowel Color; Observations on one—and two-formant Vowels Synthesized from Spectrographic Patterns." *Word* 8 (1952) pp. 195—210.

[7] G. Fant and A. Risberg: "Auditory Matching of Vowels with Two Formant Synthetic Sounds". Royal Inst. of Tech. (Stockholm) *QSPR*-4/1963 pp. 7—11.

Scully:

Ad Merhaut, ad Fant: Phase relationships between the harmonics concerned in high frequency ripples probably are important in determining the amplitude of the resultant ripple. There is now a good deal of evidence to suggest that in auditory perception time measures, perhaps even of fine structure, may be possible. One method of processing the wave form is to make a time measure for half a ripple period at the point in time where the ripple amplitude is a maximum.

Ad Slis: Records for a few whispered vowels have been made. Their oscillograms show characteristic ripples although naturally there is no wave shape periodicity associated with a fundamental as there is for normal voiced vowel sounds. Although as yet no measurements have been made, visual inspection indicates that the ripple frequencies vary for different whispered vowels in the same sort of way that they do for normally spoken ones.

Ad Fujisaki: My objections to the use of formant patterns especially of F_1 and F_2 are not dependent upon the limitations of particular instruments. In any Fourier analysis, phase relationships between the harmonics are not shown.

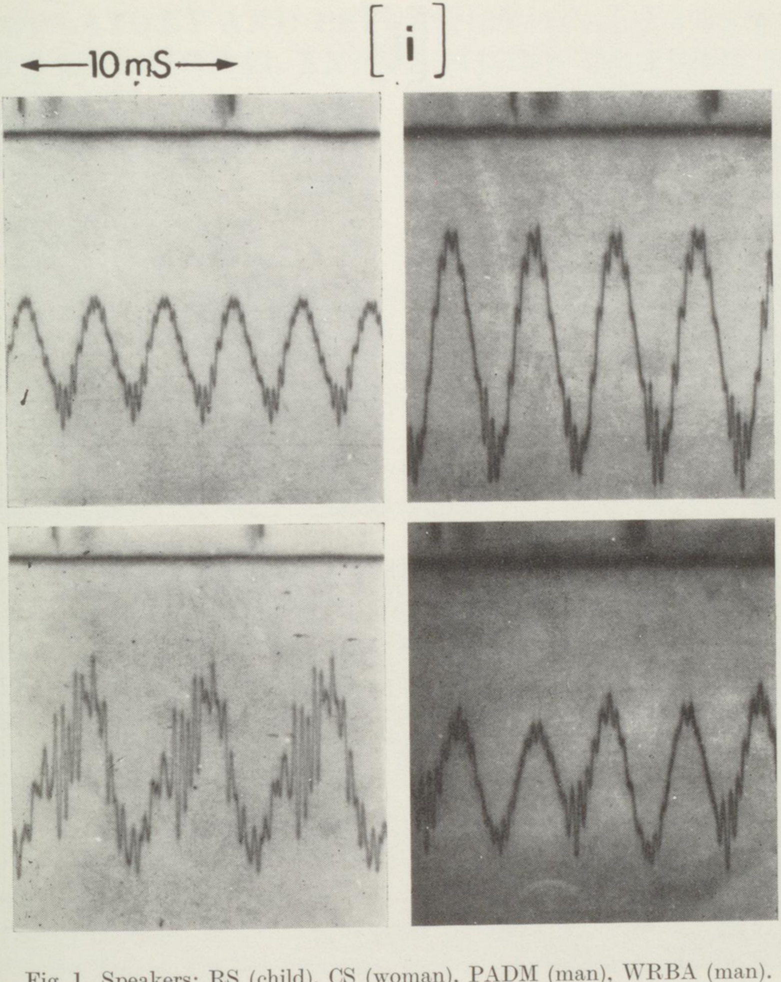

Fig. 1. Speakers: RS (child), CS (woman), PADM (man), WRBA (man).

[i]

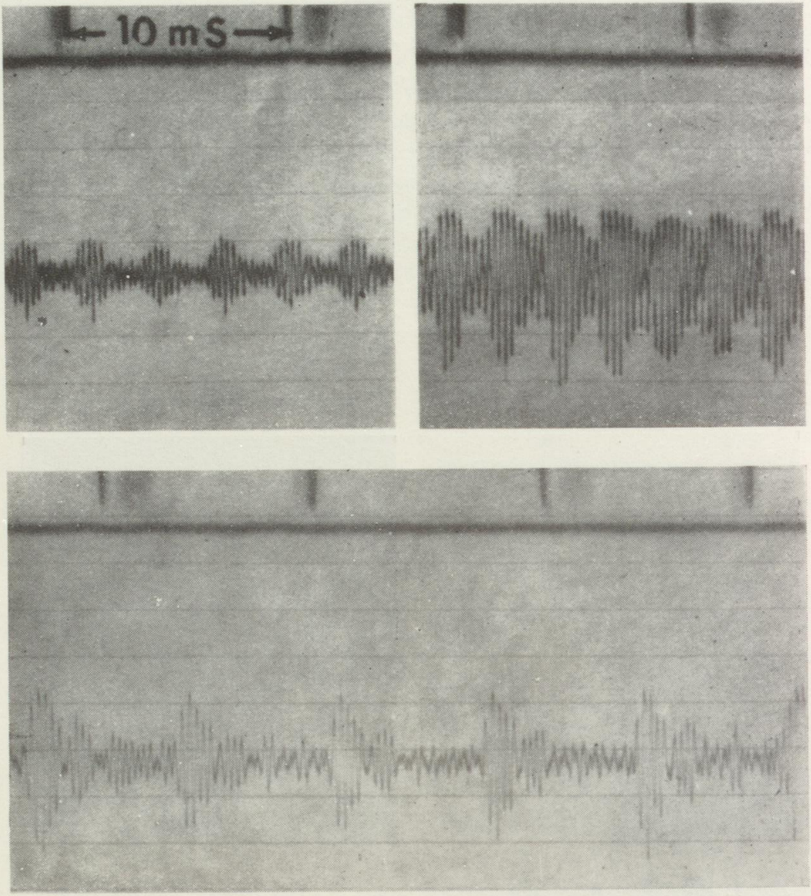

Fig. 2. Speakers: RS (child), Ch. S (child), WRBA (man).