PHONEME FIGURES OF SUSTAINED ORAL VOWELS BY TWO-DIMENSIONAL REPRESENTATION

(I) SOME BASIC OBSERVATIONS

TERUO FUKUMURA and KENZO HARA

Department of Electrical Engineering

(Received October 30, 1954)

Introduction

In a consideration of the rôle of speech in communication, one probably meets with the questions how the speech is really constructed and how informations are immanent in it. From the viewpoint of transmission engineering, these questions are connected with the intention to know how to design such relevant and delicate transmission systems that never slop any information, holding out persistently or rather actively emphasizing the very core of information. Then, this presses us to investigate the speech that comes into being with its full aspects. We are thus obliged to attempt to know acoustically, its physical construction, and subjectively, its corresponding quality; and further synthesically, to bind two up to realize the vivid reality of the speech. As for these themes, much has been done already about the vowel construction relating to its phonemic nature, and the so-called formant has been assigned according to each vowel. In spite of all its wide and detailed researches, we cannot find any study that touches the problem of finer quality such as timbre shadings. But, every individual has his own vowels that bear his own timbre. And a finer quality asks for a finer system. This is the reason why we take up again this problem. The reported here is but the entrance towards this end.

Samples and Measurements

One available clue to reveal the nature of a speech is to take out the steady part from its succession in time. Especially, in timbre study, the speech analysis of the steady state must be an indispensable way; where the fundamental qualities of speech can be well treated. In selecting voices for analysis, we endeavour to pick up some subjects who differ from each other in respect of their natural qualities. From this point, we select 2 males, 2 females and one child. Naturally their voices differ in pitch, loudness and timbre from each other. Since these subjects are all lays in phonological pronunciation, and furthermore we are under restrictions of the experimental apparatus which force the subjects to utter a single vowel sustained for several seconds, some proper training processes are beforehand needful in order to maintain the quality of voice constant in their voicing period. For instance, as the pitch of voice is raised, the articulatory effect of certain subject seems to become unnatural, and consequently the phoneme value of uttered sound tends to shift from what the subject intends to have. This details of procedure of measurement are seen in a previous report,[1] and the data readopted here are tabulated in Table 1.

[1] Y. Ochiai and T. Fukumura: Timbre Study of Vocalic Voices. Memoirs of the Faculty of Eng., Nagoya Univ. Vol. 5, No. 2. September, 1953.

TABLE 1. Illustration in Résumé of Subjects and Pitches Used.
(The pitches assigned for each subject are indicated by the sign *.)

Subject	Sex	Pitch in c.p.s.→ Age ↓	160	200	280	360	420	480	560
H.O.	child	12.7				*	*	*	*
Y.H.	(♀)	16.7				*	*	*	*
H.H.	(♀)	19.5		*	*	*	*	*	*
S.M.	(♂)	24.0	*	*	*				
T.F.	(♂)	28.6	*	*	*				

Formant Position

For the aim of frequency analysis, we used to obtain the oscillograms which show the energy or intesity distribution of vowels on the frequency continuum, and we call the part of energy or intensity concentration "formant." Thus visually representing the vowels in such a way, we see the formants vary their positions according to the kind of vowel, and we consider such moving formants as having some connection with the so-called phoneme values. But, if one would try to trace the envelope of pattern more minutely, he might find that the formant sometimes has two or more branches of peak, and that the shape of formant does not keep always uniformity even for the same subject, but varies delicately with change of pitch. How does the ear percept these auxiliary formant-peaks and further these delicate variation of the pattern? What rôle do these fine structures play in determining the vocal quality (and not phoneme quality)? Thus we have many questions of this kind to properly relate formant pattern to quality of voice. The examples are shown in Fig. 1 about the voices of Subjects S.M. and H.H., where the formant patterns are given for three fundamental pitches about three vowels "I", "A", "U"; binding them together with two other patterns of still different pitches in the case of Sub. S.M., also with four others in the case of Sub. H.H., and being averaged by juxtaposition method,[1] the mean formant structures are obtained as shown by the bottom figures in gray. The signs of F_1 and F_2 therin indicate the positions of so-called invariant formant which seems inherent to the individual.

To classify the vowels quantitatively with respect to their phoneme values, it is desirable to reduce the dimensions of pattern, and this necessarily leads us to determine the simple formant positions. Then we intend to locate the vowel positions by two formants, first and second, that is, in two-dimensional way. For this end, we must eliminate those auxiliary formant peaks from calculation which, standing on the positions correspond to F_1 and F_2, are seemed to contribute to the individual shading or vocal quality. Thus the calculating formula to determine the formant position is as follows;

$$f_i = \frac{\sum_j I_j p_j}{\sum_j I_j},$$

where f_i is the formant position, p_j is the frequency of No. j component, and I_j is the height of amplitude in mm above the base-line of the oscillogram of the concerned component. From the number which j can take, the components which are consistent with the invariant formants are omitted. This formula of weighting

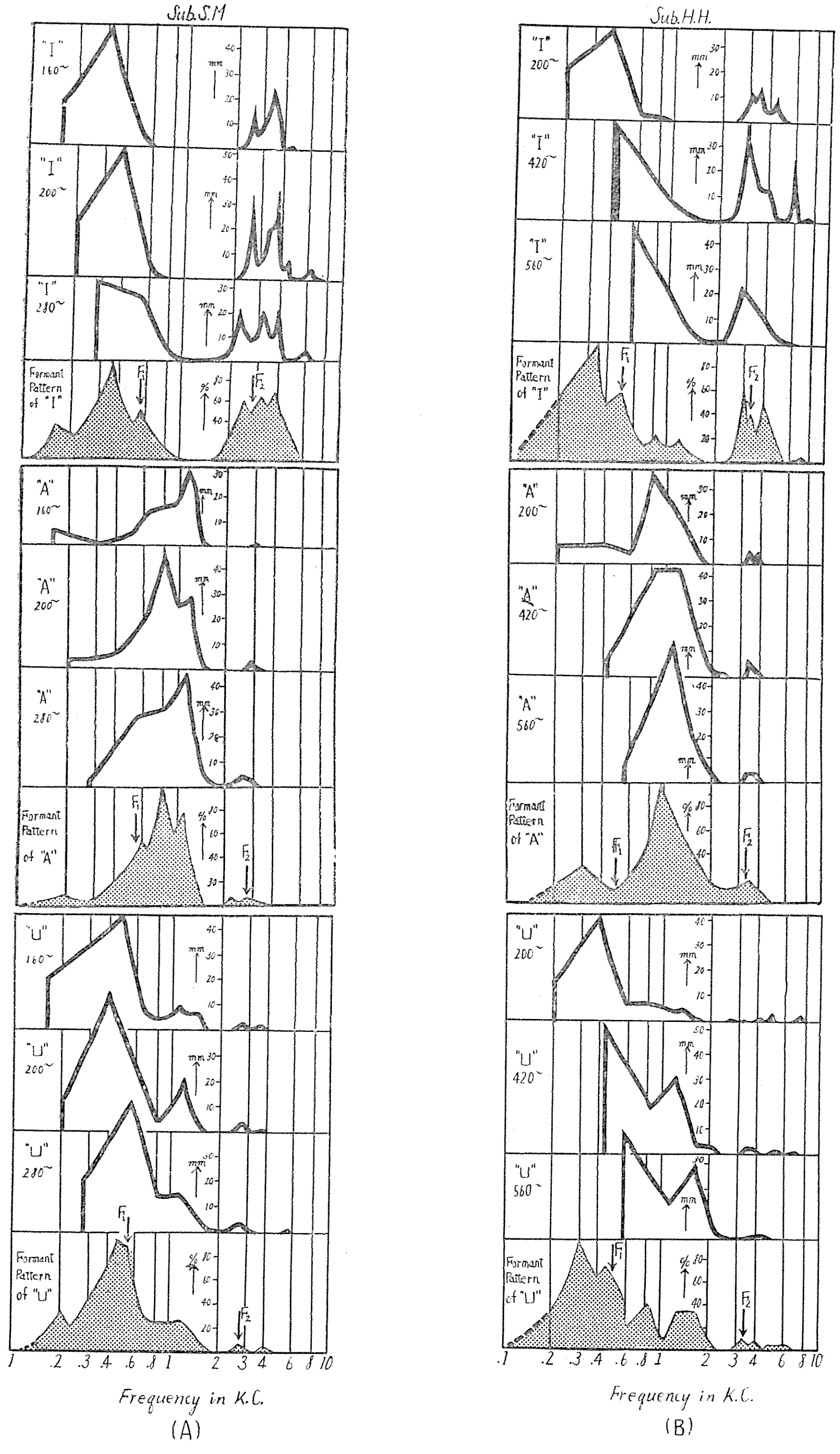

FIG. 1. Formant patterns of three vowels "I", "A", "U" at three different pitches (figures in white) and formant structures (figures in gray) for Subs. S.M. and H.H. In the process of juxtaposition, two (for Sub. S.M.) and four (for Sub. H.H.) other patterns of different pitches are lumped together.

method is same as the one adopted by Potter and Steinberg[2] except a limitation on its harmonics.*

Fluctuation of Phoneme Position

The formant position thus calculated is not always constant, but preferably tends to deviate over considerable wide range, even if uttered successively in time by the same subject. Fig. 2 tells us the character about this, where the abscissa

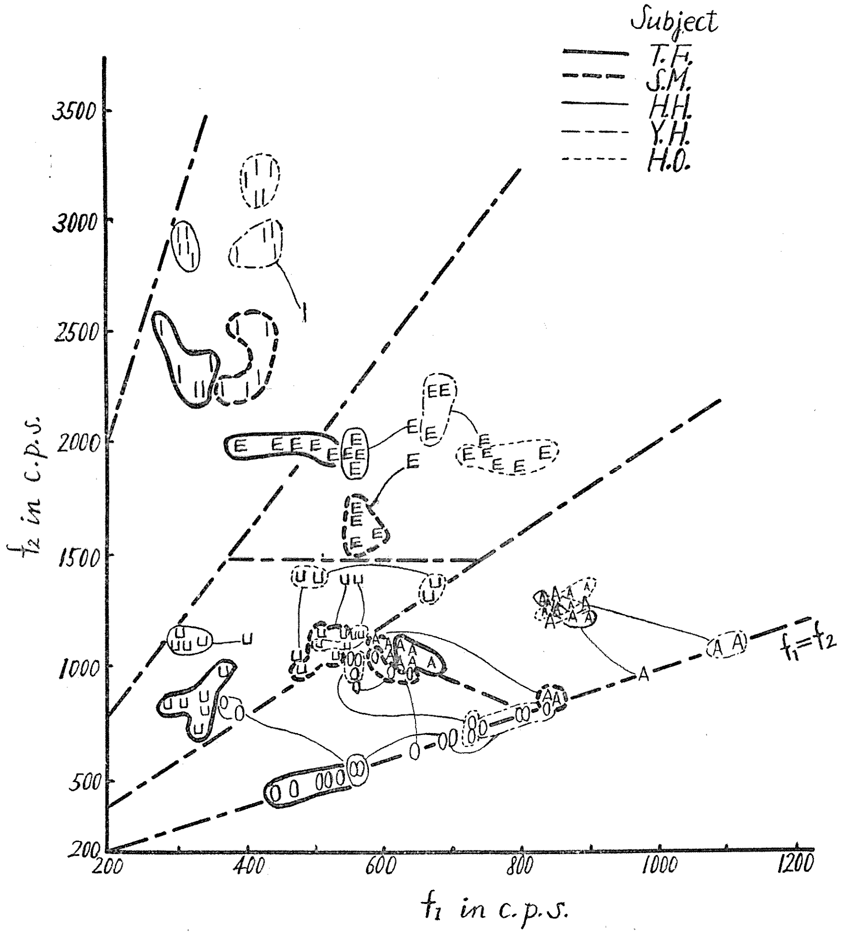

FIG. 2. Deviation of phoneme positions under the condition of constant uttering pitch 280~ with difference of subject.

---

[2] R. K. Potter and J. C. Steinberg: Toward the Specification of Speech. J.A.S.A. Vol. 22, No. 6, November 1950.

* For the most cases, only the harmonics which form the invariant formant F_2 are excluded from the calculation. Because, those which are participant in F_1 constitute, at the same time, the variant formant, and it is not yet clear how to treat them properly as auxiliary peak. It is also said that they contribute to determine the personal formant shape of any pattern. The problem of the position and shape of formant is left for further investigation.

stands for position of f_1 and ordinate for that of f_2 in c.p.s. respectively. The point, thus plotted in this f_1-f_2 plane, is tentatively called phoneme position, and the figure in f_1-f_2 plane, phoneme figure.* The ratio in the scaling of the former to the latter is 1 : 3; it is only for the sake of convenience. The result of five or more analyses of each vowel, uttered at the pitch of 280~, are presented for five subjects. Though the distinct trends of discrepancy is not seen, it is still noteworthy that the vowels "O" and "A" occassionally flop their phoneme positions onto the $f_1=f_2$ line, that is to say, two formants of these vowels are not always prominent simultaneously in their successions in time. Inspecting every subject's case, it is seen that the formant fluctuation of this kind has certain tendency, that is, Subs. T.F. and Y.H. have more chances to have one formant about the vowels concerned than the others. Then it may have some significance with respect to vocal quality.

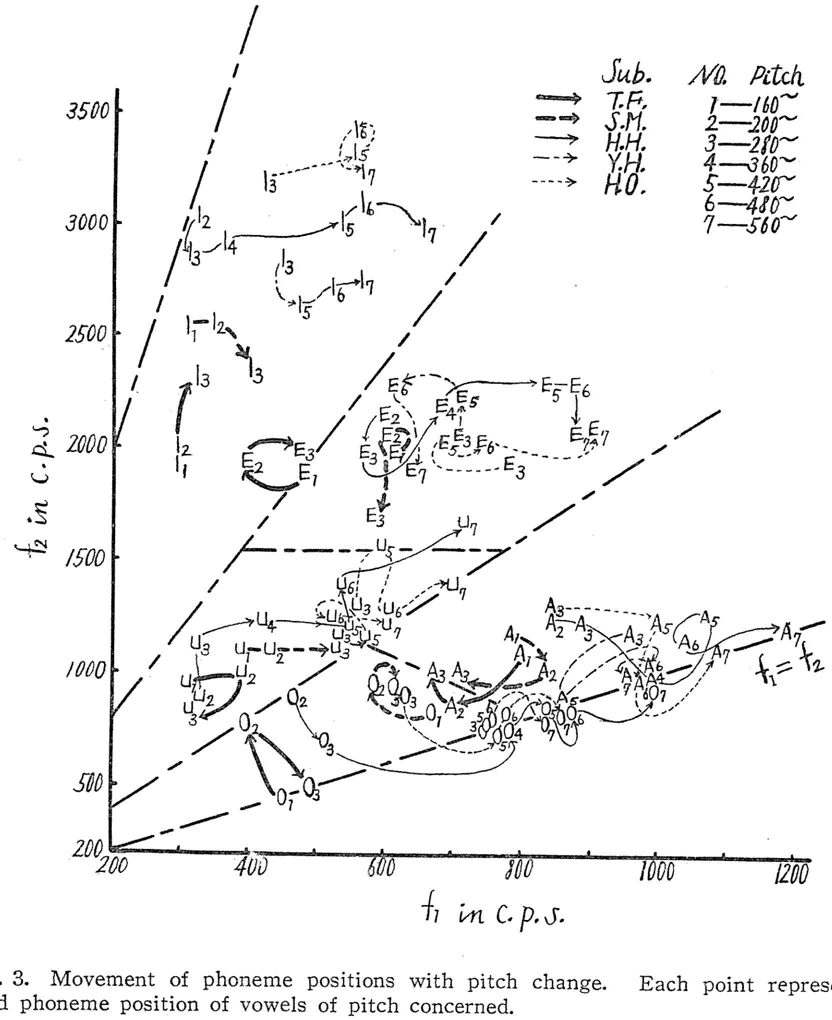

FIG. 3. Movement of phoneme positions with pitch change. Each point represents the averaged phoneme position of vowels of pitch concerned.

* The letters "I", "E", "A", "O", "U", which assign the phoneme positions in f_1-f_2 diagram, are temporally and tentatively used for convenience, and do not mean any phonetic sign.

Movement of Phoneme Position with Pitch

When one dimension of quality varies its value, how does the other behave? This is investigated about the effect of pitch change on the phoneme position proposed above, keeping the loudness of voice as constant as possible. Pitch conditions used for the subjects' vowel are shown in Table 1; and for every condition, from five to ten sheets of oscillogram are used to obtain the mean formant position; and the results are presented in Fig. 3, where the two axes represent the same as Fig. 2 (the legend of axis of f_1-f_2 diagram is consistent with each other through the Reports I and II), and the direction of curve, indicated by the arrows attached to it, shows that of pitch rise. The number suffixed to the letter of each vowel means the pitch at which the concerned vowel is uttered. In this figure we can find the fact that the phoneme position of vowel thus obtained shifts with the change of pitch. Among subjects and among vowels, some examples show no decidable trend of movement with pitch. But, inspecting individually, interesting modes of movement are found. One of them is seen, for example, in the case of vowels "E", "U" (Sub. T.F.), "O" (Sub. S.M.), "E", "A", "U" (Sub. Y.H.), where the curves of movement tend to close or alternate their directions according to the change of pitch. These cases mean the probable goodness of articulation or proficiency in uttering control of the subject concerned. In such cases, the so-called "changing-car phenomena"[1] are seen. On the other hand, in the case of Sub. H.O. moderately, and in the case of Sub. H.H. extremely, there are found some considerable upward transition of formant position as a whole. In the former case, the transition with pitch is not yet so evident in general, but is mainly found with a terminal pitch for each vowel, as shown by positions I_3, E_7, U_7, O_3 and A_3, and with other pitches phonemes keep their positions in the considerably fixed region. But, in the latter case, the amount of shift spreads over several hundred cycles extending from the male region up to the child region. Although we are not ready to discuss this phenomenon, it suggests the several points further to be investigated. One example; when the subject raises the uttering pitch, he tends to intensify the uttering breath and this effects on the shape and position of the formant.* Another; when the uttering pitch becomes extremely higher, the fundamental pitch or its neighbouring higher component of vowel happens to run beyond the proper region of lower formant assigned for the vowel concerned; especially in the case of vowels "I" and "U", where the lower formants used to occupy the lowest positions, there remains the problem how the proper position of formant is to be determined in such cases.

Relative Position among Subjects

For the convenience to inspect the relative localization on f_1-f_2 plane of each subject's voice, enveloping process of every point of a single subject with a curve is carried out graphically for each vowel, and shown in Fig. 4. Ignoring the singular parts of Sub. H.H., relative localization, from lower to higher position, is roughly in accordance with their statures. Their heights diminishes gradually in

* About the relation of the formant position and the intensity of uttering breath, some data will be published.

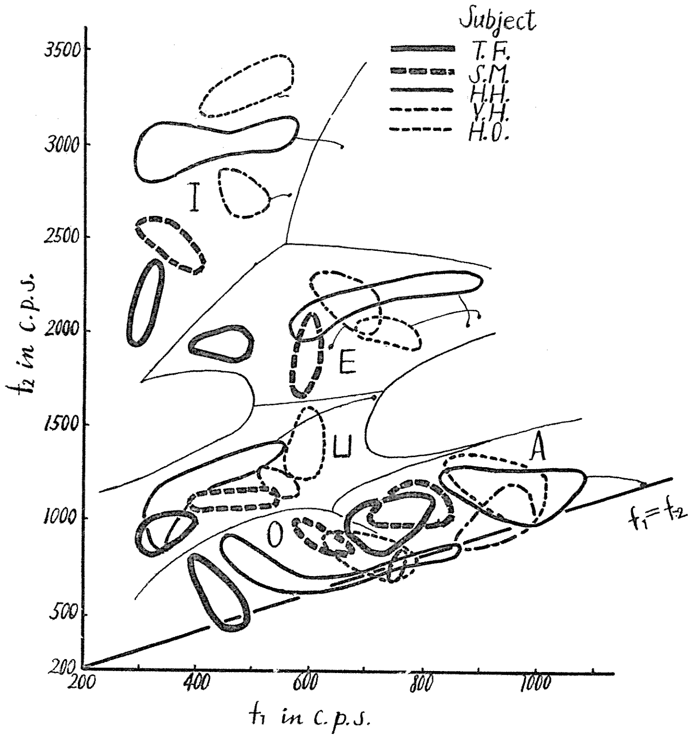

FIG. 4. Localization of phoneme positions with reference to five subjects.

the order of T.F.→S.M.→H.H.→Y.H.→H.O. Among them, S.M. and H.H. are almost comparably high. Assumably, the height relates to the size of oral cavity.