TIMBRE STUDY OF VOCALIC VOICES VIEWED FROM SUBJECTIVE PHONAL ASPECT

PART I. PRELIMINARY STUDIES ON NATURALNESS AND ARTICULATION QUALITIES ACTUALLY AND DIRECTLY MEASURED WITH RESPECT TO BAND-ELIMINATING DISTORTION

Yoshiyuki OCHIAI and Teruo FUKUMURA

Department of Electrical Engineering

(Received May 31, 1956)

Introduction

A few years ago, after much consideration,[1][2][3] we introduced a method of determining two different timbre patterns of vocalic voice, i.e., the *phonemic pattern* and the *vocal pattern* both of which were originally derived from a fundamental consideration of quality; but at that time we studied from a purely physical standpoint, and published a paper under the title "Timbre Study of Vocalic Voices."[4] In that paper we described some examples of timbre representation, selecting five vowels spoken by five calling subjects. In order to fully appreciate that study of representation of timbre-structure, we need some verification from the aspect of quality, i.e., there remains a subjective study which is the other side of a perfect timbre interpretation. To complete the timbre study of vocalic voices under the transmissional condition of band-limiting, we tried quality measurement of timbre signals impressed upon listening subjects by making use of the same examples previously given for a preliminary study of direct measurement of quality. This time we made not only an articulation measurement, but naturalness measurement as well. Our direct measurement of naturalness of vocalic voices commenced with this experiment which was for the first time carried out at our laboratory throughout the winter of 1953.

Experimental Procedure

Circuit of Experiment

We selected for the constituents of our subjective system, five callers with no defect of speech and four listeners of normal acuity of hearing. As an objective transmission system, we used band-eliminating distortion with low-pass and high-pass filters of known characteristics (60 db attenuation per octave). Experimental circuit is given in Fig. 1.

Presentation of Timbre-Signal

Taped voices are usually used as timbre signals in accordance with the scheme shown in Fig. 2 where each vocalic signal of four seconds duration is followed by a pause-interval of four seconds which separates the signals. In keeping with our policy of carefulness all through our timbre study, every caution was taken to carry

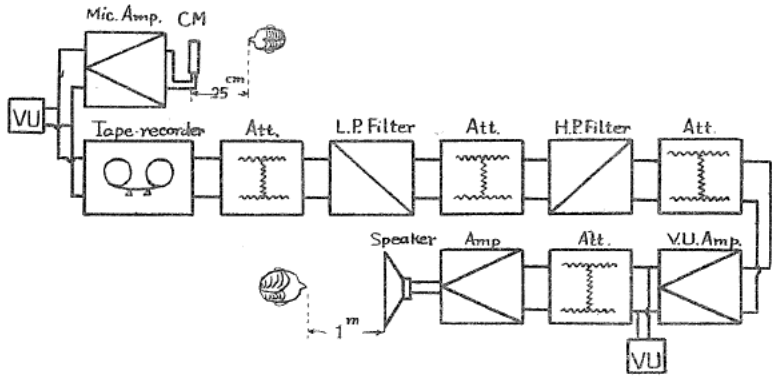

FIG 1. Experimental circuit. Calling and listening subjects are placed in a sound-proof room with the equipment of condenser microphone and loud speaker.

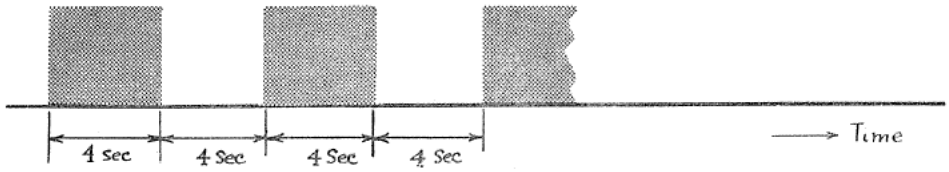

FIG. 2. Presentation of timbre signals.

out pitch and level matchings for each signal. As to the selection of timbre-signal, in each instance we took five oral vowels spoken with ordinary conversational pronunciation but in a sustained manner by five calling subjects whose vocal and vowel patterns had already been fully studied.[4] The five calling subjects consisted of two females, and three males, one of whom was an adolescent boy before mutation of voice. To find the common voicing pitch of the five callers, all of different ages and of both sexes, a 280-cycle pitch was conveniently selected.* The level of utterance is given to calling subjects, instructing them to maintain *mezzo forte*, and the level of reproduction of signals is kept constant at nearly 65 phons for the position of listening subject. In Fig. 3 (*a*) we show the vowel pattern for each of five vowels voiced by five subjects formed by averaging the timbre patterns for each vowel. In Fig. 3 (*b*) we show the vocal pattern for each of five subjects† formed by devocalizing the same five vowels for each subject.

List of Vowel-Kind Inclusive of Voice-Element

As we have already stated, we use five vowels as timbre signals uttered by five callers. From the viewpoint of timbre study which deals with the problem of not only the phoneme but also of the voice, we are forced to make ready a total of twenty-five timbres as our signals to be tested as shown in Table 1. Here "F", "S", "C", "Y", "H" mean the voices of five callers† and "A", "I", "U", "E", "O"

* A 280~-pitch was selected as the one most suitable alike for male and female subjects.

† Voices employed are; F: TF (♂), S: SM (♂), C: HO (boy), Y: YH (♀), H: HH (♀) —see Reference (4).

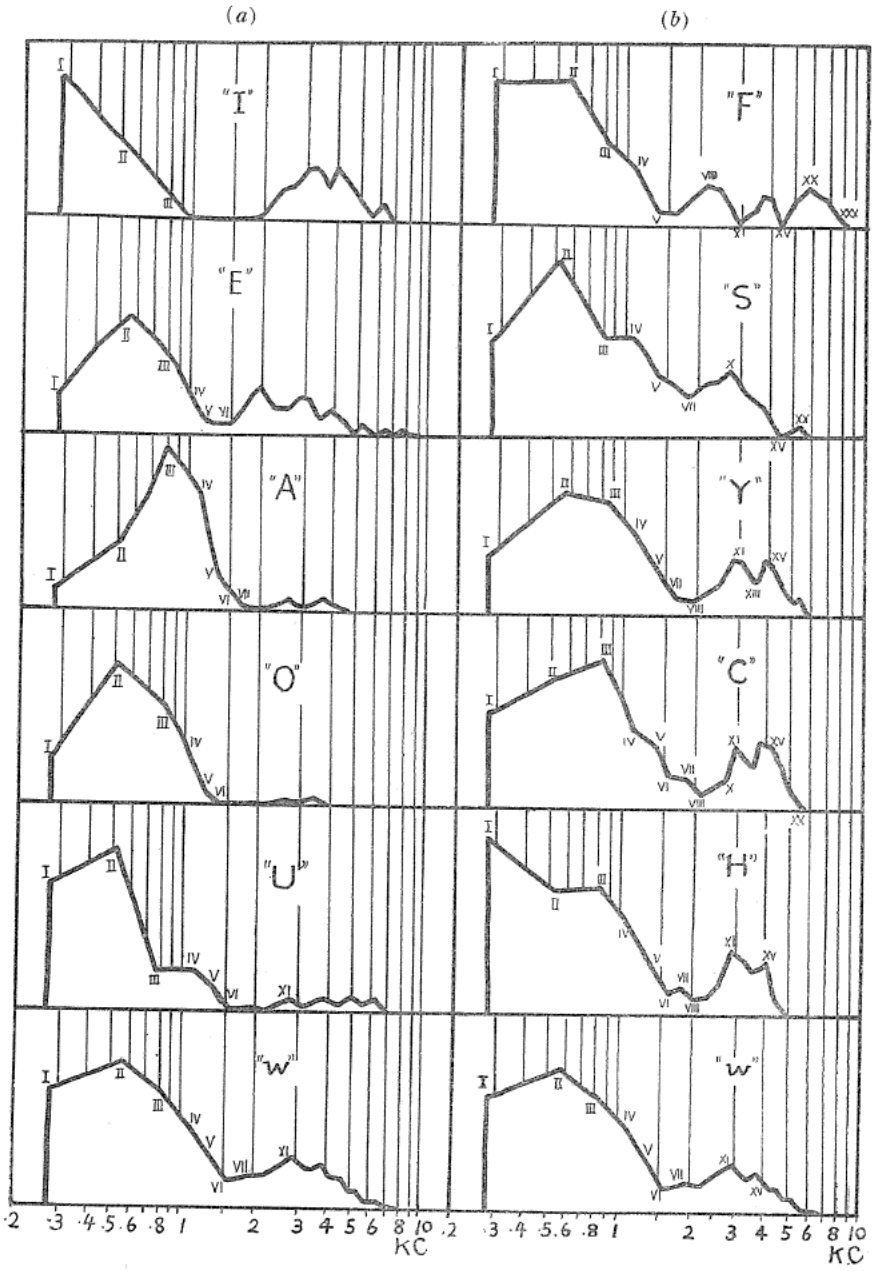

FIG. 3. Representation of timbre patterns. (*a*): phoneme patterns, (*b*): vocal patterns.

mean the phonemes of five vowels. In Table 2 we show a list of 100 phones for timbre drilling where timbres are arranged in five groups and each vowel is repeated four times in five different voices. This list is designed to be used as an exercise in vocal discrimination for each vowel group. There are four repetitions

TABLE 1. List of Timbre-Signals of Twenty-Five Kinds in Total

Vowel→ Voice↓	A	I	U	E	O
F	FA	FI	FU	FE	FO
S	SA	SI	SU	SE	SO
C	CA	CI	CU	CE	CO
Y	YA	YI	YU	YE	YO
H	HA	HI	HU	HE	HO

TABLE 2. Timbre-Logatome for Exercise in Naturalness Judgement

A		E		I		O		U	
1	FA	21	HE	41	FI	61	CO	81	YU
2	HA	22	YE	42	CI	62	YO	82	SU
3	CA	23	SE	43	SI	63	HO	83	FU
4	YA	24	CE	44	YI	64	SO	84	HU
5	SA	25	FE	45	HI	65	FO	85	CU
6	HA	26	FE	46	SI	66	YO	86	HU
7	YA	27	HE	47	YI	67	HO	87	CU
8	SA	28	CE	48	CI	68	FO	88	SU
9	CA	29	YE	49	HI	69	CO	89	FU
10	FA	30	SE	50	FI	70	SO	90	YU
11	YA	31	YE	51	HI	71	SO	91	SU
12	CA	32	CE	52	SI	72	CO	92	CU
13	FA	33	FE	53	FI	73	YO	93	FU
14	SA	34	SE	54	CI	74	FO	94	HU
15	HA	35	HE	55	YI	75	HO	95	YU
16	CA	36	CE	56	YI	76	HO	96	SU
17	SA	37	SE	57	FI	77	FO	97	YU
18	HA	38	HE	58	HI	78	SO	98	CU
19	FA	39	FE	59	SI	79	CO	99	FU
20	YA	40	YE	60	CI	80	YO	100	HU

TABLE 3. One Example of Timbre-Logatomes Recorded on Tape for Actual Measurements of Both Naturalness and Articulation with Reference to Vocalic Voices

1	SA	21	YI	41	FO	61	HI	81	CI
2	YI	22	CA	42	YA	62	SI	82	FO
3	FE	23	SE	43	CI	63	CE	83	HA
4	HO	24	FE	44	SE	64	FO	84	SU
5	FI	25	CO	45	FU	65	YI	85	YO
6	HO	26	HE	46	HA	66	HI	86	FA
7	SE	27	CO	47	YU	67	SO	87	CE
8	FU	28	FI	48	SI	68	YI	88	CO
9	YA	29	CE	49	CU	69	FE	89	SA
10	CO	30	SO	50	SU	70	HI	90	YU
11	SI	31	FA	51	FE	71	SI	91	SU
12	HU	32	HA	52	HU	72	HI	92	FA
13	SO	33	YA	53	YE	73	FE	93	HO
14	CU	34	HE	54	CA	74	YE	94	CE
15	FA	35	YO	55	HE	75	SE	95	SA
16	CU	36	SO	56	SU	76	CU	96	YO
17	HU	37	HE	57	CA	77	FI	97	FO
18	YE	38	FI	58	YA	78	CI	98	HO
19	CA	39	YI	59	HU	79	CI	99	YU
20	YU	40	SA	60	FU	80	YO	100	HA

of twenty-five combinations of timbres, making one hundred sounds in all. For a real measurement of naturalness and articulation data, we prepared other lists of timbres, an example of which is given in Table 3 where the timbres in Table 2 are re-arranged in such a way as to make a perfect randomization with reference to voice and vowel. The manner of randomization can be altered to avoid, if necessary, effect of the listener's memory. By using such timbre-logatome such measurements of qualities are usually repeated five times in one test-run.

Listening Subjects

We show the threshold characteristics of four young male listeners in Fig. 4 by which we can see that there are some slight deviations among these listeners, but such deviations have little influence on quality response. In Fig. 5 we show for a crew of four listeners the so-called practice-effect in naturalness experiment only under the referential no-limit band condition. For these experiments we use the timbre logatome shown in Table 1. We see in Fig. 5 a relatively large discrepancy between listeners in spite of a relatively small deviation in threshold values. At the very beginning of the experiment only one subject could be considered proficient in his measurement of naturalness, the other three showed relatively lower ablilities in the same period. Among these three of lower quality, one subject was marked by a relatively faster build-up toward proficiency than the other two. This crew was instructed to carry out actual measurements after drilling for about two months.

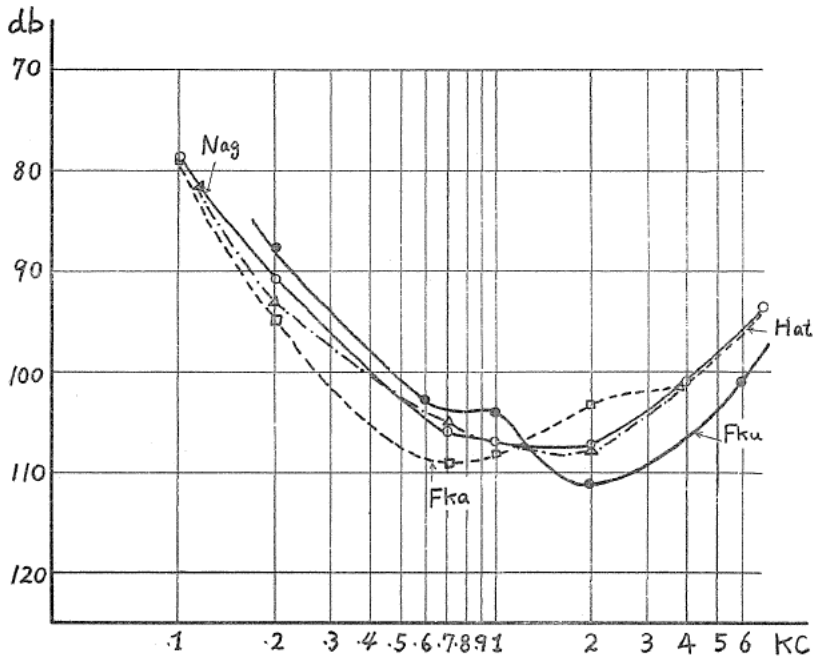

FIG. 4. Threshold characteristics of four listeners.

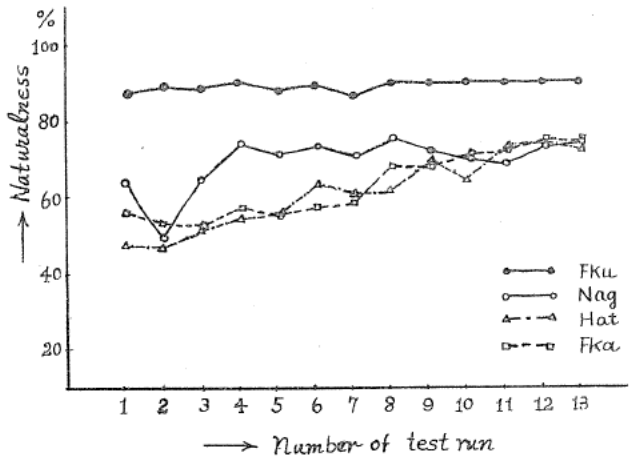

FIG. 5. Exercise characteristic curves of four listeners in naturalness quality measurement.

On Distortion

For distortion we use Low-Pass Filter as High-Cut Distortion (HCD) and High-Pass Filter as Low-Cut Distortion (LCD); the transmissional characteristics for both were quite the same as in our previous paper.[5 6 7] In cutting the band frequency for the present experiment, the power level of the transmitted voice-signals through the distortion is never recovered to the original level of the uncut signal. The band-cutting condition, it must be noted here, is processed without recovering its level.*

Experimental Result

General Description of Quality Response

Characteristics of both naturalness and articulation quality in response to the distortion of band-elimination are shown in Fig. 6. Each point is obtained by a total of 2000 observations per condition. By characteristics in Fig. 6 we see that with the exception of cases in very large distortion the naturalness characteristics $\Gamma(N)$ are nearly always in lower positions than articulation characteristics $\Gamma(A)$ the fact of which suggests that the crew even after practising for about 60 days was still less familiar with timbre discrimination in vocal aspect than in phonemic aspect. But what is most significant is that the relative positions of starting points of the two kinds of quality characteristics count for nothing insofar as the general tendency of characteristic forms is concerned. For example, first we find the important fact that the intersection of two naturalness characteristics, i.e., the characteristics in low-cut and in high-cut distortions seems to be always situated in a higher position in frequency dimension than the intersection of articulation characteristics. A clear intersection of articulation characteristics could not be obtained

* In suceeding experiments of this kind we used the level-recovering process in the case of band-cutting.

in the first stage of our experiment.* Why quality characteristics cross each other and what the point of intersection in quality theory means are two points which we will study in detail later.

Then, the second fact worthy of mention is gradient of characteristics. The slopes of $\Gamma(N)$ are usually dull; on the contrary, the slopes of $\Gamma(A)$ are usually steep. The articulation characteristics generally start with a relatively small gradient and usually end with a steeper gradient. From the beginning, the naturalness characteristics start with a relatively small gradient which continues unaltered to the larger distortion. Since the degree of exercise for naturalness is not the same as for articulation, it is meaningless to touch upon the intersection in different qualities. Therefore, here we had better deal exclusively with intersections in the same qualities.

Last, we must add some facts to give meaning to the question why we discuss such an intersection problem. We believe it is most important to interpret this

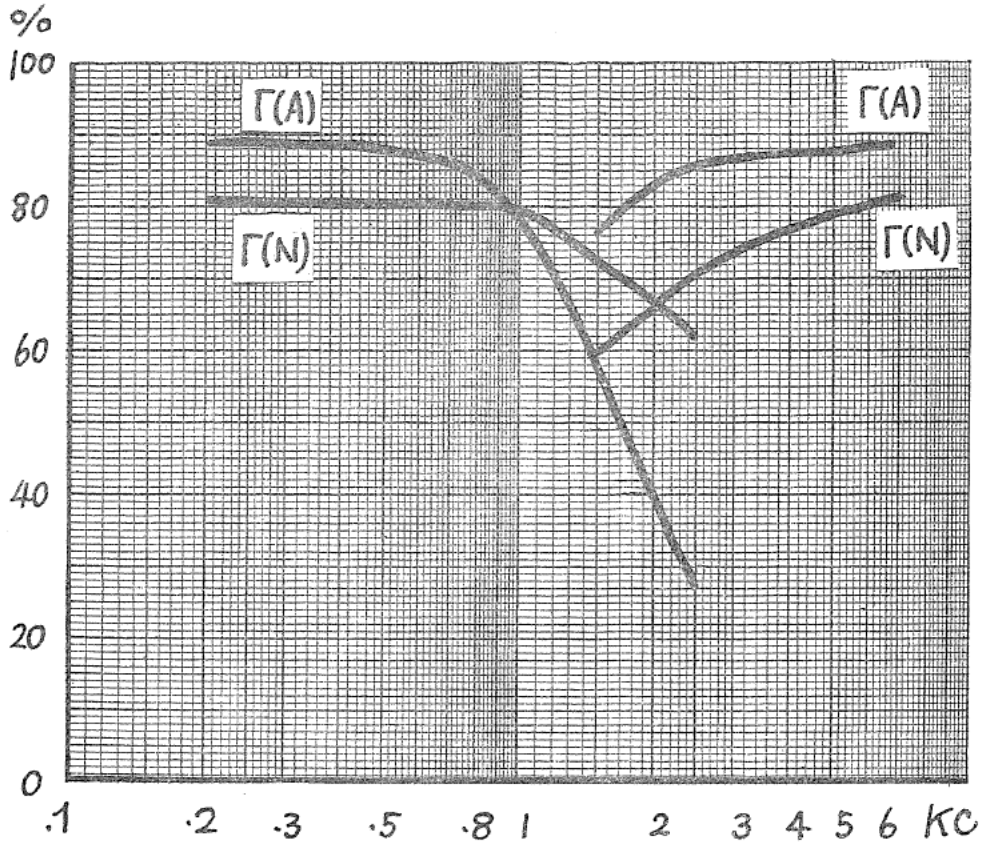

FIG. 6. General characteristics of naturalness and articulation in the band-eliminating distortion. $\Gamma(A)$: Articulation characteristics, $\Gamma(N)$: Naturalness characteristics.

* In succeeding experiments of this kind we did obtain the clear intersections of articulation characteristics.

quality phenomenon as is well presented in the articulation-index theory in American practice. But we should not discuss this problem from a general inspection of over-all characteristics only. We must postpone a fuller discussion and proceed rather to the problem of quality analysis.

Quality Analysis

To carry on a more detailed study of quality phenomena and not to be content with the study of general characteristics of quality which give information on averaged quality only, we are obliged to make an anlysis of quality. Quality analysis is necessary for an understanding of general characteristics because the analysis of quality can play the rôle of an individual study of quality, treating the quality phenomena with respect to each individual timbre.

Characteristic per Ear (Listener Effect)

We have to first detect quality characteristics for individual listeners which we will call hereafter *Characteristic per Ear** (C/E). This is seen in Figs. 7 (*a*), (*b*) where in (*a*) the naturalness characteristic (NC/E) is represented, and in (*b*) the articulation characteristic (AC/E) is represented. Each point of these characteristics was determined by the result of 500 observations per condition. In these figures we find at a glance that:

(1) The fluctuation of articulation characteristics due to the differences in listeners is relatively small in comparison with the deviation of naturalness characteristics. The deviation of AC/E is no more than about 10%, but that of NC/E averages 25%.

(2) Notwithstanding the deviation of C/E, the tendencies of all individual characteristics are very much alike. For example, in NC/E we see a marked similarity among the four listeners (Fku, Fka, Nag, Hat) with respect to the shape of characteristics.

(3) Insofar as the frequency of the intersection is concerned, in our quality study we make an issue of the point of intersection of characteristics, especially those of NC/E, by ignoring the absolute quality-value of the intersecting point, because the point of intersection is the result of a general tendency and reflects no effect caused by different listeners.

After summing up the various observations made in the treatment of characteristics per ear, we were able to complete the first step. Disregarding not only the unequal effects of exercises in articulation and naturalness but also the greatness of the deviations of characteristics in such quality measurements, we can proceed to the next stage of our study where we must analyse the qualities of the phoneme and the voice, the two most fundamental and essential elements of speech quality.

We therefore proceed to the study of *Characteristics per Voice* and *Characteristics per Vowel.* We summarize in Figs. 8 (*a*), (*b*), (*c*), (*d*) the quality analysis results which are so important. Fig. 8 (*a*) shows naturalness characteristics per voice (NC/V_o); Fig. 8 (*b*) articulation characteristics per voice (AC/V_o); Fig. 8 (*c*) naturalness characteristics per vowel (NC/V_1) and Fig. 8 (*d*) articulation characteristics per vowel (AC/V_1). Based upon 400 observations per condition, we ob-

* Per listener.

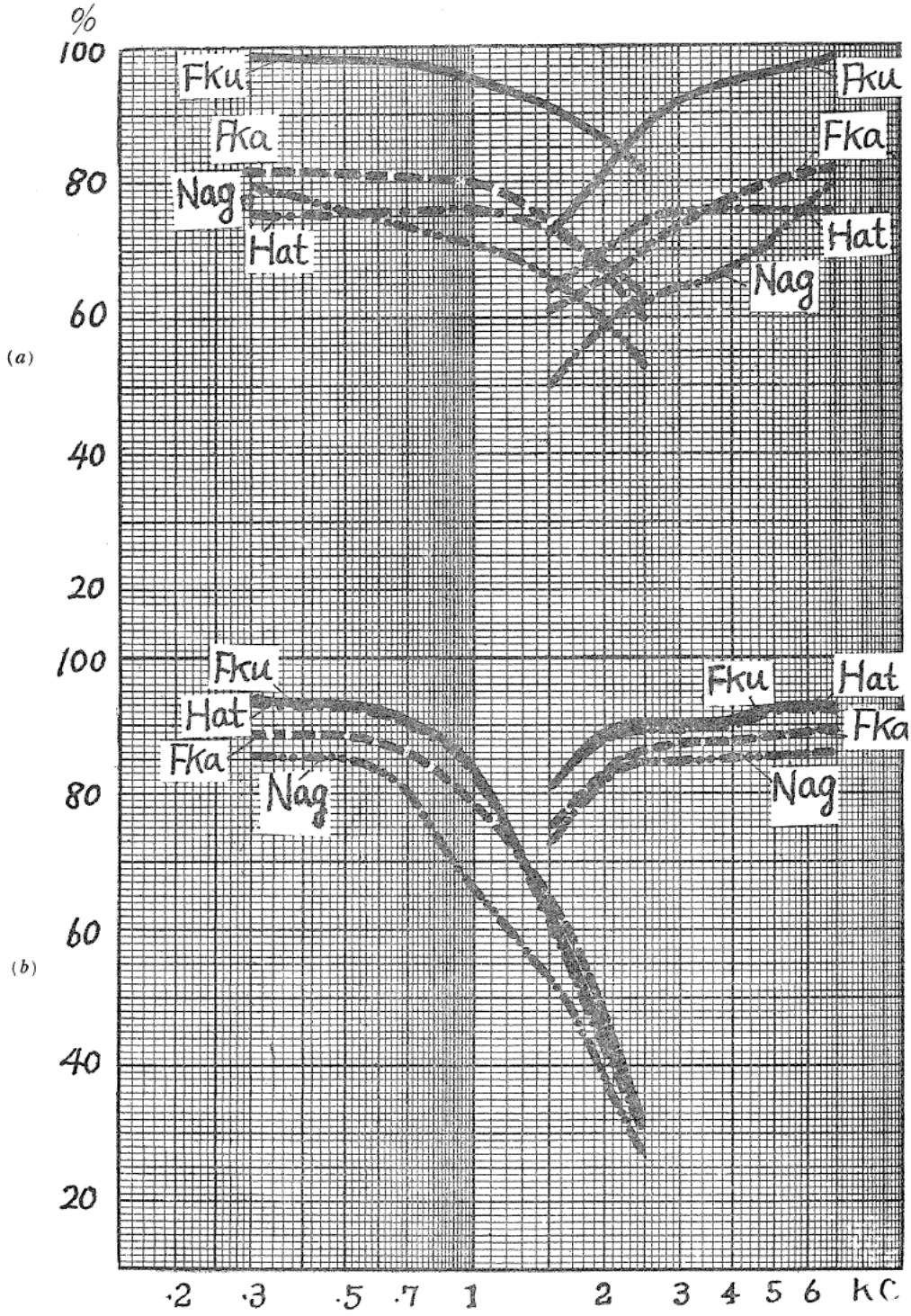

FIG. 7. Characteristics per ear. (*a*): Naturalness characteristic per ear (NC/E), (*b*): Articulation characteristic per ear (AC/E).

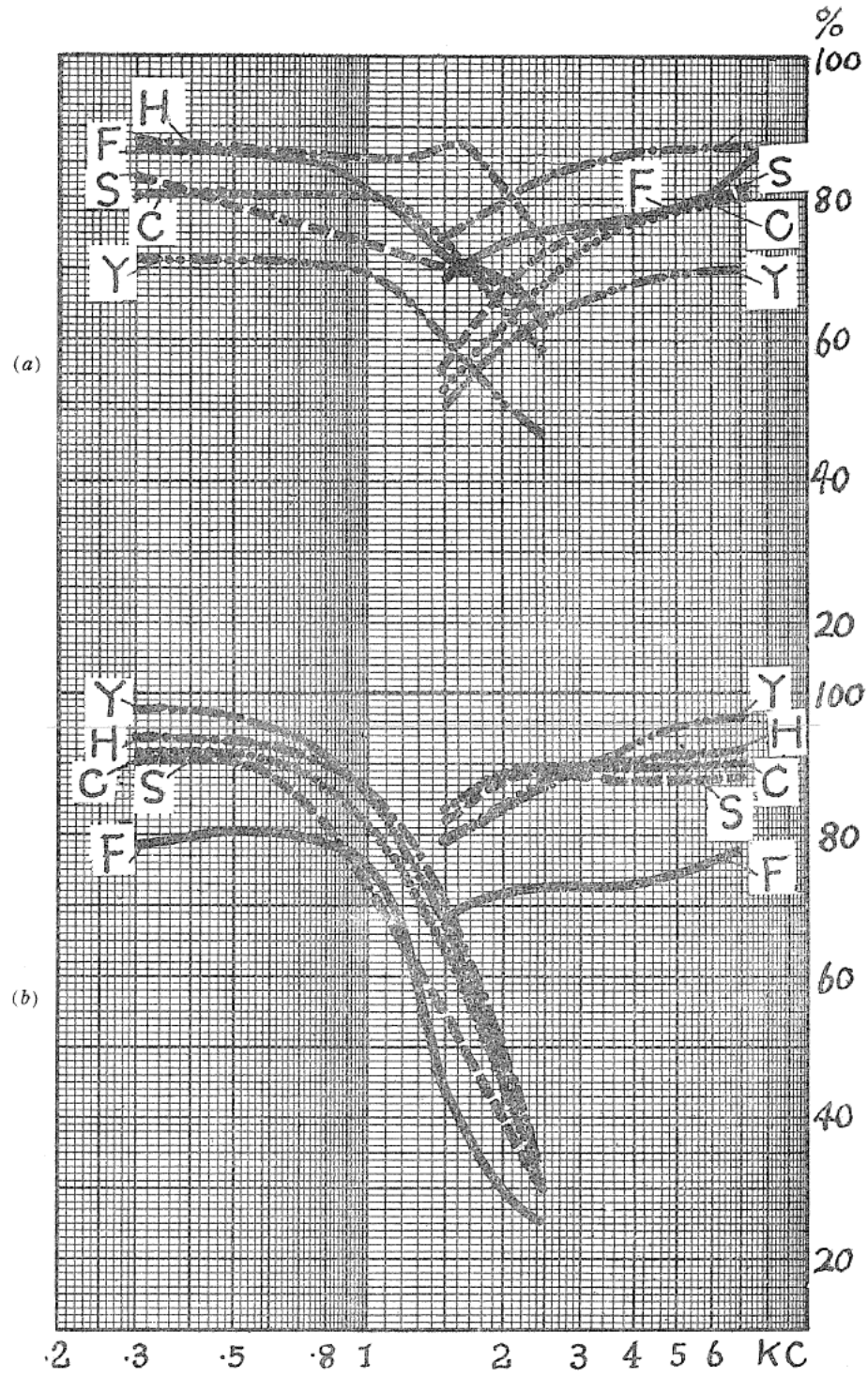

FIG. 8 (*a*), (*b*). Most important parts of quality analysis. (*a*): Naturalness characteristics per voice (NC/V_c), (*b*): Articulation characteristics per voice (AC/V_c).

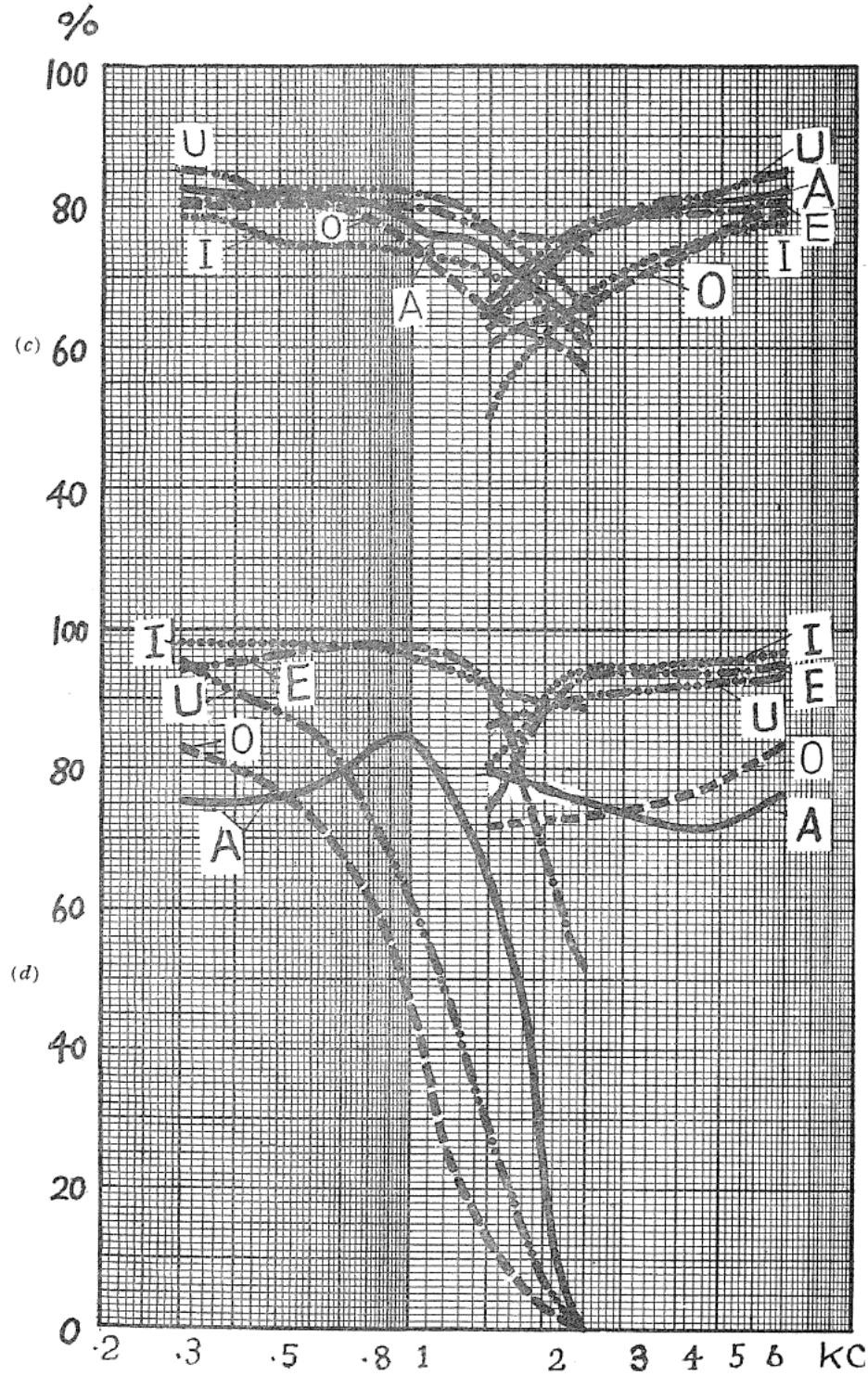

FIG. 8 (*c*), (*d*). Most important parts of quality analysis. (*c*): Naturalness characteristics per vowel (NC/V_1), (*d*): Articulation characteristics per vowel (AC/V_1).

tained these characteristics by averaging the mean value of four listeners. Of the four representations of quality analysis, the NC/V_c and AC/V_I are the most essential in the sense that there the corresponding analysis are done with reference to the direct item (or content) of the quality judgements. That is, naturalness characteristic of NC/V_c is analysed relative to the voice element upon which the naturalness-quality judgement must be based and, in the same manner, articulation characteristic of AC/V_I is analysed relative to the phoneme element on which the articulation-quality judgement must be founded. Nevertheless, the remaining two characteristics AC/V_c and NC/V_I are very important because the delicate interdependence between the phoneme and the voice is picked up from these two characteristics only. The former characteristic reveals the voice effect upon the phoneme-quality judgement, and the latter the vowel effect upon the voice-quality judgement. These four characteristics each in its turn are examined more closely.

Naturalness Characteristics per Voice (NC/V_c)

These characteristics corresponding to Fig. 8 (*a*) undergo relatively wide and large deviations due to individual voices. The characteristics of both female voices "H" and "Y" are quite similar, one obtained simply by the process of parallel displacement of the other, putting aside the question of the fine point of curves. The soprano-voice "C", the boy-subject, considering the whole tendency of characteristics, is somewhat similar to these female voices. As for the voices "F" and "S" of male subjects, the naturalness characteristics are complicated, crossing each other in their course.

Articulation Characteristics per Vowel (AC/V_I)

The characteristics AC/V_I corresponding to Fig. 8 (*d*) are subject to the widest and largest deviations effected by the difference of vowels. Strictly speaking, there are no two vowels having similar characteristics. Accordingly, for five vowels there are five pairs of characteristics, each pair quite different from the others. From the viewpoint of characteristic tendency, i.e., characteristic shape, we can separate five vowels into two groups, one the "I" and "E" group, the other "A", "O", "U" group. But here we must admit the unusual characteristic of the vowel "A" upon which we will touch again.

Articulation Characteristics per Voice (AC/V_c)

Referring to Fig. 8 (*b*) we can conclude a convergent tendency of characteristics in spite of the difference in voices, with the one exception of voice "F", the case of which will be discussed more thoroughly later for the same reason as alluded to for the "A" vowel. As for the female voices "Y" and "H", the naturalness characteristics are a little different, crossing each other in both pairs of characteristics. The female voice "H" is somewhat similar to the soprano-voice "C" of the boy-subject.

Naturalness Characteristics per Vowel (NC/V_I)

It is possible to find somewhat convergent characteristics in Fig. 8 (*c*), and to conclude that as a whole there is a relatively small vowel effect upon naturalness judgement. But on closer inspection, we can probably discover something important to suggest a delicate mutual relationship between the phoneme and the voice. We show only the following facts: (1) that as a whole the vowels "E" and "U"

stand in higher naturalness, and (2) that the vowel "I" stands in lower naturalness, and (3) that the vowels "A" and "O" appear to belong to an intermediate group. As for the characteristics of NC/V_l and AC/V_c we refer to the timbre patterns shown in Fig. 3 for a better understanding of characteristics.

What is Found by Connected Inspection

In running through these four characteristics, we must pick out the two cases of AC/V_l and NC/V_c as the most essential. It is to be noted that the most essential characteristics usually have the widest and largest deviations between individual characteristics. On the contrary, the non-essential characteristics NC/V_l and AC/V_c usually have small deviations in characteristics, reflecting some rather delicate relationship between phoneme and voice. By comparing NC/V_l with AC/V_l we can introduce the most interesting and even astonishing fact that although the articulation of vowels "U", "O", "A" come to zero at the cutting condition of LCD 2.5 kc, they can still maintain a naturalness quality of about 60% at that condition. In other words, by cutting down the frequency region to below 2.5 kc the phonemic quality of vowels "U", "O", "A" are completely lost; meanwhile the vocal qualities of these vowels have lost no more than 40%. Taking this comparison to be the most fundamental common fact on these two characteristics, we finally observe the following points:

(1) Naturalness characteristics are generally so situated that they bear nearly symmetrical features for the up- and down-pairs of characteristic, and have intersection-points as centers. On the contrary, articulation characteristics show hardly any symmetricity in this sense. They are rather far from a symmetrical representation. This point seems to suggest a great difference in quality distribution between naturalness and articulation.

(2) Judged from vowel effect, the vowels "I" and "E" seem to occupy simultaneously a high position in articulation and a low position in naturalness; but at the same time the vowels "A" and "O" seem to give a low value of articulation and a high value of naturalness. It is meaningful that vowels with higher articulation tend to show lower naturalness, and *vice versa*. This inclination, however, must be stated with some reservation until a closer examination through a great many examples has been made.

(3) From observations of voice effect, we can infer that in the same way voices with higher articulation tend to show relatively lower naturalness, and *vice versa*. But there is the one exception to this, i.e., in the case of voice "H" where both high articulation and high naturalness are found.

Remarks on Pronunciation and Voicing

Next we take note of the reason for low articulation in this experiment.

(1) Vowel articulation is in general too low seemingly to have been caused partly by the selection of 280~-pitch, too high for male subjects, resulting in a relatively poor vowel pronunciation, and partly by the condition of sustained state which has a somewhat inadequate influence upon vowel perception. Of all the vowels, the vowel "A" of the subject "F" was received the most poorly. This fact can be explained by a defect in pronunciation of subject F which can be verified by closer inspection of his individual vowel pattern. Detailed examination of mishearing phenomena will ascertain the reasonableness of our explanation.

(2) As described above, low quality of naturalness can in general be explained

by the lack of skill of the crew. On the other hand, this phenomenon can also be explained in part by the fact that a selection of the pitch 280~, which is a little too high for male subjects, forces their utterances to come out under stress and strain with no semblance of their at-home voices. This of course resulted in the lowering of scores in the total judgement on voice quality.

Further Studies on Quality Phenomena

We have discussed quality characteristics in general and individual quality characteristics of various sorts with respect to characteristic forms. In these studies we have pointed out certain remarkable characters of quality which concern macroscopic observations in distortions of band-limiting nature. To go into a more minute study on quality, we must proceed either with observations of a more microscopical nature or with the study viewed from the standpoint of other than characteristic slopes which we have called into question in the above discussion.

Balancing Point of Quality

When the distortion in question has two opposing directions, as is the case with the band-eliminating distortion, the quality characteristics thereof are usually represented by a pair of characteristics which intersect at some intermediate point in the distortion range concerned. We can define the crossover point of the quality characteristics in the distortion BED as a BALANCING POINT OF QUALITY, because the total of qualities distributed in the band region above the crossover point is just equal to those distributed in the band region below the crossover point. This is exactly what the characteristics mean in that situation. Of course we have two sorts of balancing points, one the balancing point of articulation, the other the

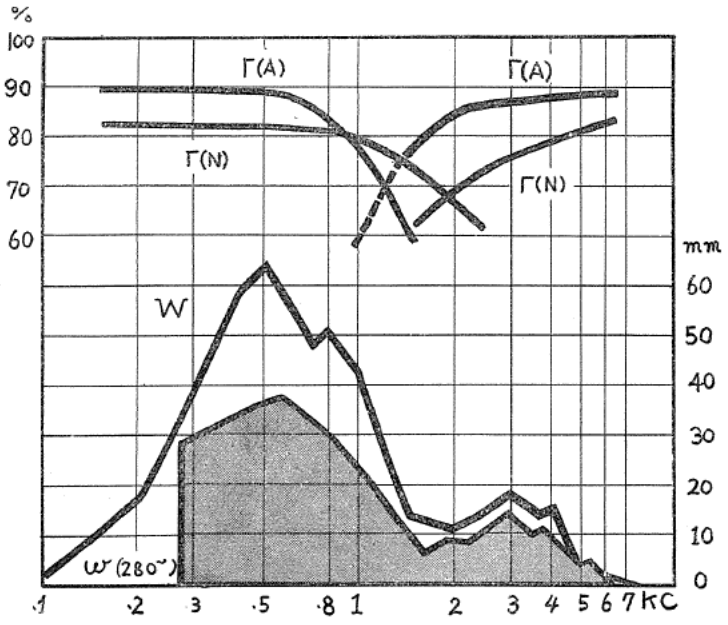

FIG. 9. Quality characteristics in general in comparison with timbre patterns of white-phone *W* and white-voice *w*.

balancing point of naturalness. Now we are confronted with the problem of interpreting the balancing points of quality. Are they actually the centers of the concentration of qualities concerned? Or do they correspond to the empty or most dilute point of quality? These are vitally important questions in timbre-quality theory. To answer them we must commence our study by examining the result of quality responses in connection with timbre patterns, as shown in Fig. 9, where the general articulation characteristics $\Gamma(A)$ and general naturalness characteristics $\Gamma(N)$ are shown first in a comparison with the timbre pattern of white-phone representation W, the most typical expression of this caller-group including both voices and vowels, and then in the timbre-pattern of the white-voice representation w which gives only a 280~-voice pattern. It is necessary to remember that the quality characteristics of both $\Gamma(A)$, $\Gamma(N)$ are the result of subjective or psychological estimation on qualities and that the timbre patterns W and w (280~) shown there are nothing but representations of timbre construction in a purely objective (or physical) way. By comparing the timbre patterns thus given in a purely objective way with the quality-characteristics thus measured in a subjective way, we immediately see that notwithstanding the differences of naturalness and articulation, the crossover points of quality-characteristics nearly coincide with the positions of the so-called *glen* in timbre patterns. This fact is significant because a glen does

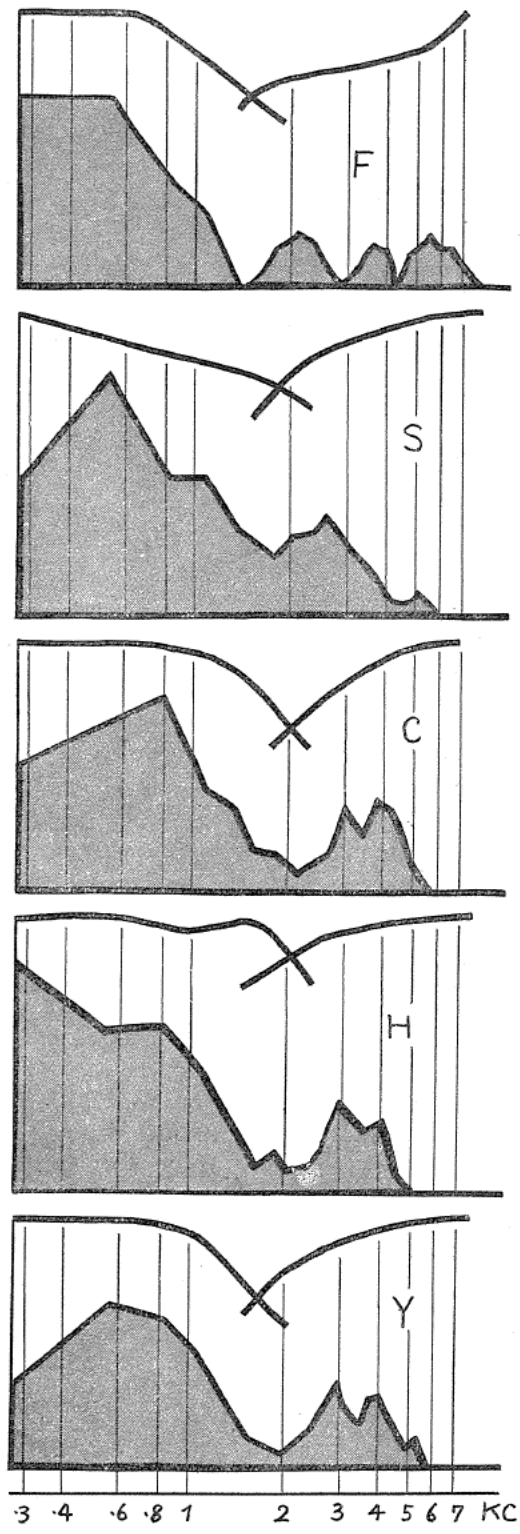

FIG. 10. Individual vocal patterns of five callers in comparison with each of five naturalness characteristics per voice.

divide the timbre pattern of frequency dimension into two sections, i.e., the upper section and the lower section. These two sections are not equal in either height or area so far as timbre pattern is concerned, but the qualities which justly correspond to these two pattern-sections are nearly balanced. This is the case with naturalness as well as with articulation.

Balancing Point of Naturalness Quality

Since an interpretation of quality intersection is important in our quality study, we think it needful to examine more strictly the phenomena of coincidence of quality intersection with the pattern glen. To avoid the inferences and presumptions which lack support by solid experimental results, we have studied this problem of the voice pattern restricted to naturalness quality only. This is shown in Fig. 10 where the individual vocal patterns of five callers are represented in a comparison with the individual naturalness characteristics per voice (NC/V_c). On closer inspection of these figures, we can confidently conclude that the intersections of naturalness-quality almost coincide with the timbre glens of vocal pattern. For example, in voice "F" where the glen is relatively low, the intersection of the quality characteristics of the voice "F" is also low; and the relatively high positions of the quality intersections of voices "H" and "C" always correspond exactly to the higher positions of their own vocal glens. As to articulation quality, the discussion could continue in the same manner. But now, as we are lacking a crossover point in articulation measurement in general, we had better postpone any discussion on the balancing point of articulation quality.

Allocation of Quality in Timbre Pattern

We can roughly calculate the magnitude of quality corresponding to several pattern sections. As the most simple procedure, in Fig. 11 we give an example of quality allocation in the two pattern sections, i.e., the upper and the lower sections. In Fig. 11 (*a*) we show the allocation of naturalness quality in the vocal patterns where the point of vocal glen, i.e., the balancing point for individual voices, is selected as the demarcation point of quality. In such selection of division, the qualities allocated to the two sections are balanced. In Fig. 11 (*b*) we give an example of the allocation of articulation-quality for individual vowels where the demarcation point is selected fixedly at the frequency of 1.5 kc notwithstanding the difference of vowels. Although such a representation is a very rough one, we can suggest a most interesting fact which is that the glen in a vowel pattern by no means represents the balancing point of articulation quality. For example, in both vowels "I" and "E" the upper formant part is more weighted in articulation quality than is the lower part; on the contrary, in the vowels "A" and "O", the articulation quality seems to be concentrated only in the lower part. As to the vowel "U" in Japanese pronunciation, this phoneme may belong rather to the latter group than to the former. For the balancing points of articulation of vowels "I" and "E", we will find the corresponding points in the frequency range above 2 kc where the articulation characteristics really come to intersect. By these inspections and considerations we can introduce a most important fact, viz., the balancing point of articulation tends to drop in the neighborhood of the most concentrated point of quality. This is decidedly not the case with naturalness quality. It is essential to point out that the allocation of naturalness by the method of demarcation of

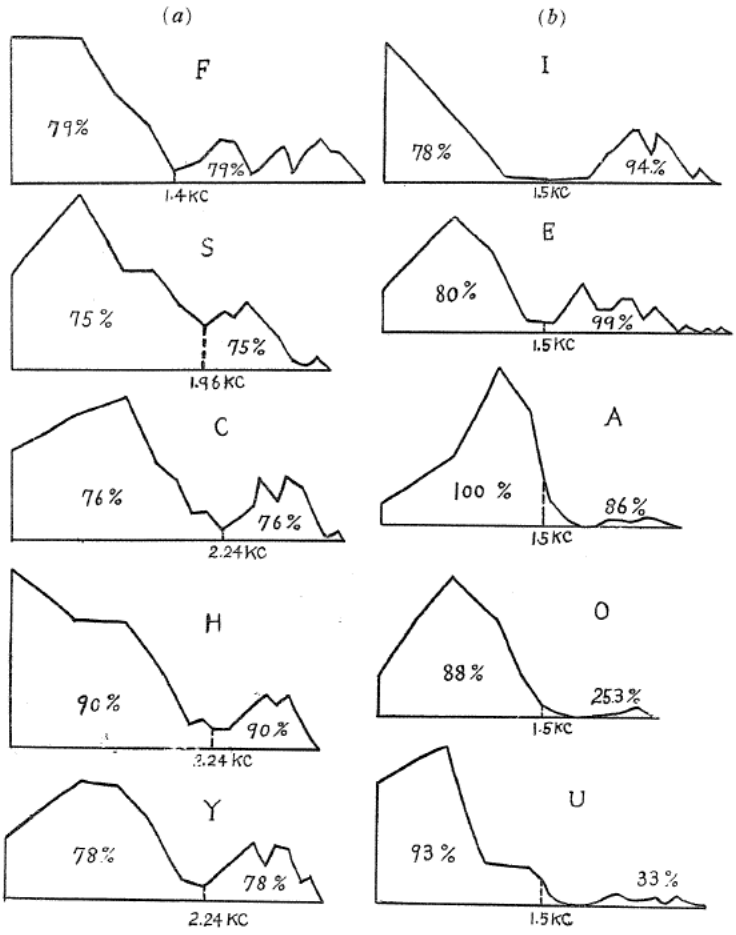

FIG. 11. Quality allocation in timbre patterns. (*a*) Naturalness allocation in vocal patterns, (*b*) Articulation allocation in phoneme patterns.

the vocal glen corresponds to the halving quality. The one-half quality thereby becomes anywhere between 75% and 80% by our calculation. The case for voice "H" is an exceptional one where we obtained too high values for half quality.

Discussion

The so-called formant theory in the traditional sense, confronts us now with the need for checking from a more generalised and wider viewpoint of timbre quality such as our researches on speech sounds have been conducted. We think, at least the primitive manner of thinking must be given up in which one used to assume in defining "formant" that it was a simple measure of the strongness in physical intensity of component partials. As there are at least two aspects of quality, the concept of formant must necessarily undergo new checkings. For example, we must at least distinguish between the two types of formant: one due

to phoneme quality, i.e., phonemic formant, the other due to vocal quality, i.e., vocal formant, a distinction often pointed out in our previous papers.[3)4)] There may be a further formant of another kind. At any rate it is sure that for verification of this distinction we need some subjective quality measurements wherein utilization of some distortions of certain types are indispensable.[5)6)7)] The present study was carried out for this reason.

Lastly, we wish to add one vital fact which is that for the most positive verification of formant theory it is not sufficient to discuss a conceptional problem on quality only. We must proceed further and devise new methods in measurement to detect extraordinarily delicate nuance-differences. To this end, we expect much from an auditory confusion study[8)9)10)11)12)] now in progress on which we will report more completely in the near future.

Acknowledgement

The main part of this study was contracted for by the Institute of Telecommunication of Nippon Telegraph and Telephone Public Corporation for whose cooperation we are indebted. We appreciate the cooperation also of Messrs. Masahiko Fukamachi, Kenji Nagayoshi and Akira Hattori as experimenters and at the same time as listening subjects and cooperation of the calling subjects Messrs. H. Ohnishi, S. Masumoto, Misses H. Higuchi, and Y. Hashimoto.

References

1) Y. Ochiai: Transmission of Quality. M.F.E., Nagoya Univ., Vol. 6, No. 2, 1954.
2) Y. Ochiai and T. Yamashita: On Timbre Quality (I Report). M.F.E., Nagoya Univ., Vol. 7, No. 1, 1955.
3) Y. Ochiai: Mémoire sur les Sons des Voix Humaines. M.F.E., Nagoya Univ., Vol. 4, No. 1, 1952.
4) Y. Ochiai and T. Fukumura: Timbre Study of Vocalic Voices. M.F.E., Nagoya Univ., Vol. 5, No. 2, 1953.
5) Y. Ochiai: General Consideration on Studies of Speech Qualities in Rotational Synchronous Distortion. M.F.E., Nagoya Univ., Vol. 7, No. 1, 1955.
6) Y. Ochiai, S. Saito and Y. Sakai: Articulation Study of Speech Quality in Rotational Synchronous Distortion. M.F.E., Nagoya Univ., Vol. 7, No. 1, 1955.
7) Y. Ochiai and T. Fukumura: Study on Fundamental Qualities of Vocalic Timbres by Rotational Synchronous Distortion. M.F.E., Nagoya Univ., in press.
8) Y. Ochiai and T. Fukumura: On the Fundamental Qualities of Speaking Voices. Report Read at Spring Meeting of Acoust. Soc. of Japan. May, 1956.
9) Y. Ochiai and T. Fukumura: On Vocal Confusion. Report Read at Spring Meeting of Acoust. Soc. of Japan. May, 1956.
10) Y. Ochiai and T. Fukumura: On Phonemic Confusion. Report Read at Spring Meeting of Acoust. Soc. of Japan. May, 1956.
11) Y. Ochiai and T. Fukumura: Some Significant Observations on Quality Phenomena of Vocalic Voices. Report Read at Spring Meeting of Acoust. Soc. of Japan. May, 1956.
12) Y. Ochiai: Study on Naturalness and Articulation of Vocalic Voices, Especially with Reference to Auditory Confusion Viewed from both Vocal and Phoneme Qualities. Reports I, II and III. Special Reports Presented for the Study under Contract with the Institute of Telecommunication of Japanese Telegraph- and Telephone-Public Corporation. Feb., 1954; Feb., 1955; Feb., 1956.