TIMBRE STUDY OF VOCALIC VOICES VIEWED FROM SUBJECTIVE PHONAL ASPECT

PART II (b)—SIMPLE DESCRIPTION OF CONFUSION PHENOMENA

YOSHIYUKI OCHIAI, TERUO FUKUMURA and AKIRA HATTORI

Department of Electrical Engineering

(Received October 10, 1956)

In this present work we describe in detail the trends of both vocal and phonemic confusion in response to distortions of band-eliminating nature. For tracing the finest phases and modes of confusion-characteristics as functions of distortion, it is important above all to have a full knowledge of the referential state of timbre-confusion in the normal condidion (no-cutting) and we take this as the point of beginning for our study.

Simple Description of Confusion Phenomena

Timbre Confusion in Normal Condition

First, let us consider phonemic confusion. Fortunately we have some data[5] on Japanese language vowel confusion, shown in Table 1, where similarity of any two phones is arranged in proper order in all ten combinations of ${}_5C_2$. The data contained in Table 1 is the result of observations on a total of 60,000 sounds, averaging a confusion of 1.72%. In Table 1, we make use of very slight distortions in the testing system, such as +5 db, −5 db, LC below 0.3 kc, HC above 4 kc, in order to reduce the number of observations. Based on Table 1, we get a configuration of phoneme viewed from the similarity-relation as shown in Fig. 2, where the thickness of combination bar between phonemes is nearly proportional to the similarity of the phonemes. This means that five Japanese vowels can be put into two groups, the "E" and "I" group, the "A" and "O" group, "U" standing between these two groups. This configuration is considered reasonable because, in ordinary oral Japanese vowels, "A" and "O" are of a nearly single-formant nature and vowels "I" and "E" are clearly of double-formant nature. Vowel "U" is of double-formant which is a special characteristic of the Japanese "U" standing in contrast to the pronunciation of "U" in both English and German. Through intermediary of "U", the other two groups can be brought into contact. If the size of the statistical object is reduced, *i.e.*, if the number of observations is lowered, then this configuration becomes lost. We feel sure that such configuration is available as the basis of a standard of reference for the state of confusion, upon which basis the various states of confusion in distorted conditions must be considered. As for vocal confusion, we must be careful to think in the following way: There may be an infinitely greater variety of vocal patterns than phoneme patterns even when we are concerned only with the Japanese vocal type. It might be difficult to identify some

TABLE 1. Data of the Similarity-Index Based upon the Confusion of Five Japanese Oral Vowels in the Referential Condition of No-Distortion

Phoneme	No. of sounds observed	Per cent of confusion occurrence (%)
A	15,600	1.60
I	7,200	1.80
U	14,400	1.66
E	7,800	1.78
O	15,000	1.74
Total	60,000	Mean 1.72

Combination of phonemes	Similarity-index	Order of similarity
A—O	1.03	I (almost 1)
U—A	0.60	II (almost 1/2)
U—I	0.57	
E—U	0.49	
O—E	0.48	
O—U	0.24	III (almost 1/4)
E—I	0.22	
E—A	0.06	IV (below 1/10)
O—I	0.03	
I—A	0.01	

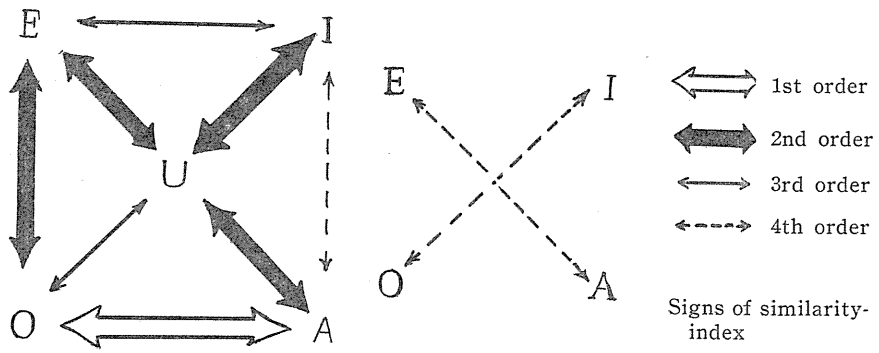

FIG. 2. Phoneme configuration of five Japanese oral vowels arranged in the consideration of the order of their similarity.

configuration as consistently holding good for vocal patterns, but when we restrict our consideration to the circle of vocal patterns presented here, the situation is simplified. It will be sufficient only to recall what was stated under "Commentary on Timbre Pattern." Considered from the viewpoints of *wideness and narrowness* in pattern extension and *evenness and oddness* in harmonics structure, these given voices can be put into two classes: "F" and "S"; "C", "H" and "Y". By imitating the *X*-configuration as in the case of phoneme, we can conveniently select one of two methods of grouping, *viz.*, the group composed of the male-voices "F", "S"; the female-voices "H", "Y" through the intermediary of the indeterminate voice "C" of boy-subject from the viewpoint of sex-voice-quality or the group composed of the evenness-dominant "F", "S" and the oddness-dominant "H", "C" through the intermediary of characterless voice "Y" from the viewpoint of evenness and oddness in harmonics structure. The former is shown in Fig. 3 and the latter in Fig. 4. Though we think the latter grouping is more reasonable, we employ here the former merely because it is more easily acceptable from the viewpoint of sex in voice classification, a point which we shall consider more in detail later.

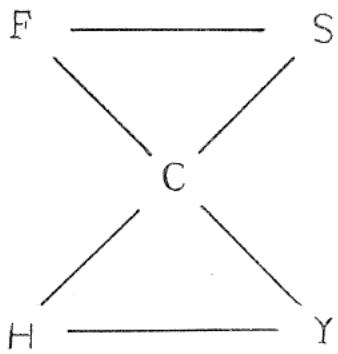

FIG. 3. One configuration of vocal timbres.

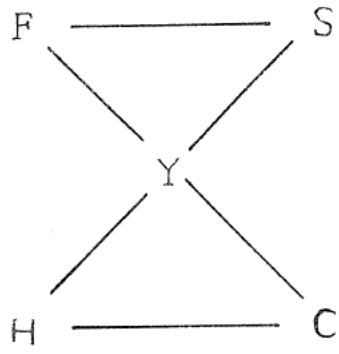

FIG. 4. Another configuration of vocal timbres.

Graphical Representation of Timbre Confusion as a Whole—Confusion-Modes of Phonemic and Vocal Confusions

After having established the reference standard for timbre confusion, we can now proceed to discuss how timbre confusion actually appears as a phenomenon when distortions are applied to timbre signals. Our present measurements in confusion are based upon a total of 400 observations by four listeners for each phone as well as for each voice. This number of observations is not sufficient to establish the so-called *reciprocity* principle of confusion upon which "similarity" is defined even for a distortionless condition. But this matters little. For our purpose, it is of prime importance that we detect clear variations in the confusion relation caused by the distortion and, moreover, that through the method of these observations we can secure evidence on how the two types of confusion, phonemic and vocal, differ from each other. In Fig. 5 we give a summary of graphical representations by this method where the thickness of the sign of bondage arrow means the approximate intensity and direction of confusion. The numerals inscribed on or under the bondage arrows mean the percentages of confusion which actually happened in the direction given by bondage arrows between the two timbres concerned. By attentive inspection of these confusion phenomena as a whole, we can point out the following four trends:

(1) As for the general difference between two types of confusion, vocal confusion is characterized by a quite gradual variation shown in its response to the distortion change; on the contrary, phonemic confusion shows a very rapid and abrupt variation due to the change of distortion.

(2) As for the difference of confusion due to the difference of direction in any combination of two timbres, vocal confusion generally shows small difference, a trend which still maintains even when the distortion becomes great; contrarily, phonemic confusion shows large differences when the distortion applied becomes greater and attains to some regions.

(3) Notwithstanding the small difference due to direction in vocal confusion, it is a notable fact that the general confusion trend is almost always found in the arrangement of confusion with respect to its dominancy. This we shall hereafter refer to as *mode of confusion*. The mode of vocal confusion is found (a) in peripheral clockwise movement from "F" to "H", such as F→S→Y→H, and (b)

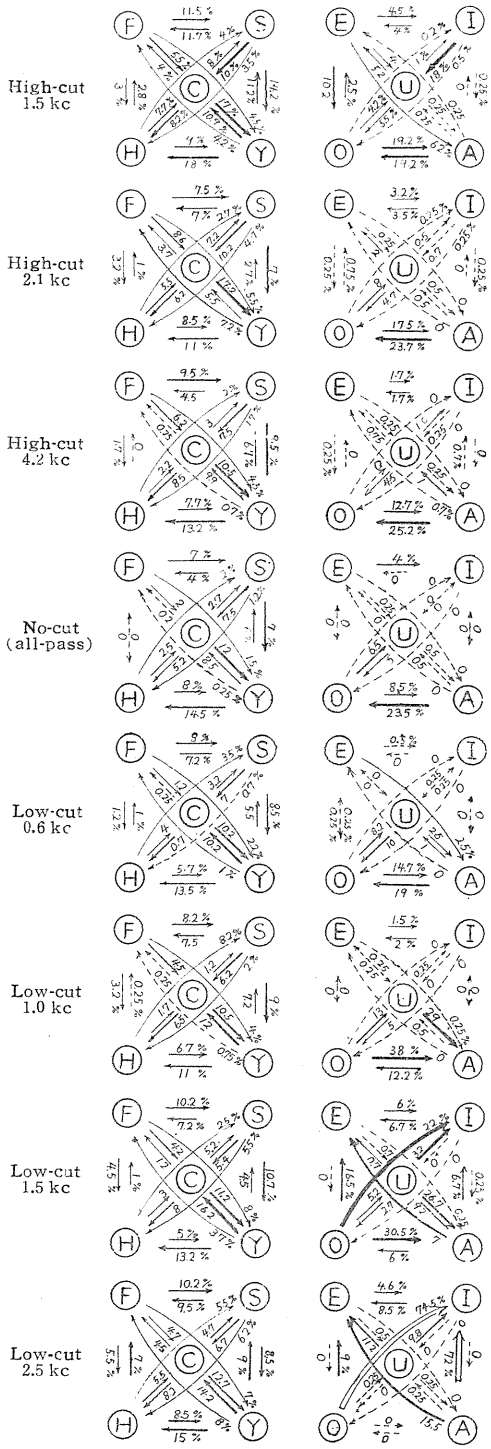

FIG. 5. Graphical representation of phonemic confusion as well as vocal confusion.

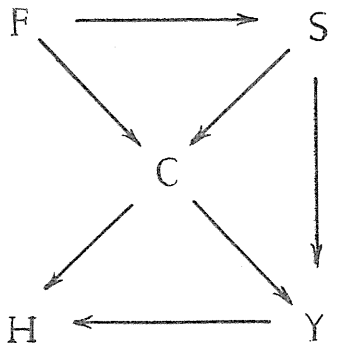

FIG. 6. Characteristic confusion-mode in vocal confusion.

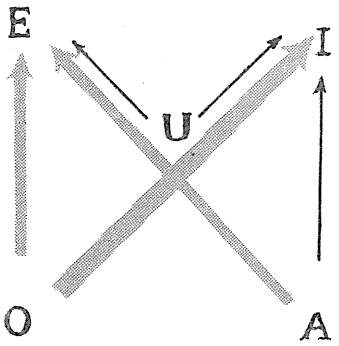

FIG. 7. Change of confusion-mode in the condition of 2.5 kc cut-off in LCD.

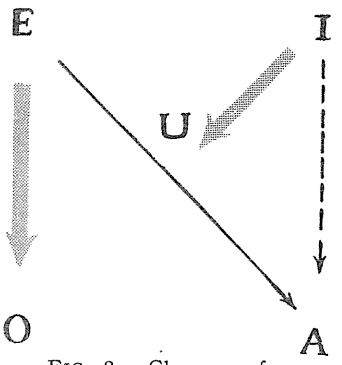

FIG. 8. Change of confusion mode in the condition of 1.5 kc cut-off in HCD in addition to the persistent confusion between "A" and "O".

in diagonal movements such as F→C→Y, and S→C→H, as clearly illustrated in Fig. 6, and (*c*) further in the state that there is no bondage worthy of notice between voices "F" and "H". Only when the distortion becomes great, does some partial change in these three points of configuration take place. For example, in LCD a reversed relation in dominancy takes place which is C←Y (1 kc, 1.5 kc, 2.5 kc in LCD) and some new bondage appears between "F" and "H" in LCD 2.5 kc and in HCD 1.5 kc also.

(4) We can describe phonemic confusion mode roughly as follows: When the distortion in low-cut direction becomes serious, the mode of confusion undergoes unique changes showing only dominant confusions which go from lower to higher phonemes in the given configuration, as, for example, in LCD of 2.5 kc cut-off we have the clear typical confusion-mode shown in Fig. 7 on one hand; on the other hand, in HCD we can find the reversed tendency showing the mode of confusion from higher to lower phoneme such as shown imperfectly in HCD of 1.5 kc cut-off, Fig. 8. This is reasonable because by cutting down the lower part of the phonemic patterns in question, these phonemes might be confused with those phonemes which keep their essentially inherent and important qualities chiefly in the upper part of their patterns. A reversed phenomenon might take place, *viz.*, by eliminating the upper part of the patterns, the phonemes characterized by the higher-part quality become confused with the phonemes which have their most important qualities mainly in the lower-part of the pattern. This is the essential mode of confusion which belongs to the timbre pattern with *concentrated distribution* of quality. In timbre pattern with *dispersed distribution* of quality, this type of confusion-mode does not occur, as is clearly shown in examples of vocal confusion. For a closer study of the detailed phases for these two types of confusion, we must use different methods of observation. Let us now begin with phonemic confusion and then go on to vocal confusion in our study.

Phonemic-Confusion Characteristics

The best understanding of articulation-quality characteristics *vs.* distortion is surely attained through a complete interpretation of separate articulation-loss characteristics for individual phonemes, and this interpretation is in conformity with our seeking after minute movement of characteristics of every possible outgoing confusion of individual phoneme. We illustrate all outgoing movements in Fig. 9 where, for the purpose of comparative examination, each phonemic pattern of five vowels is shown side-by-side. In this figure the heavy solid-line curves represent articulation-loss characteristics, and curves given by dotted-, broken-, and one and two chain-lines and another full line mean outgoing confusion characteristics which fully describe the constituent parts of articualtion-loss phenomena. By the vowels beside the confusion curves we can judge toward what phonemes the outgoing confusions are directed. Through a closer inspection of these confusion characteristics in this representation, we can find a most significant fact which might not be recognized in expressions such as given in Fig. 5. As is clearly found in LCD cases, and less clearly in HCD cases merely because of its insufficient cutting, there is a conspicuous trend (the cause of which we are not now in a position to give) in the outgoing confusion phenomena. We will explain by quoting actual examples. For instance, incoming confusions to phoneme "A" being markedly found in the outgoings from both "O" and from "U" have almost the same place of about

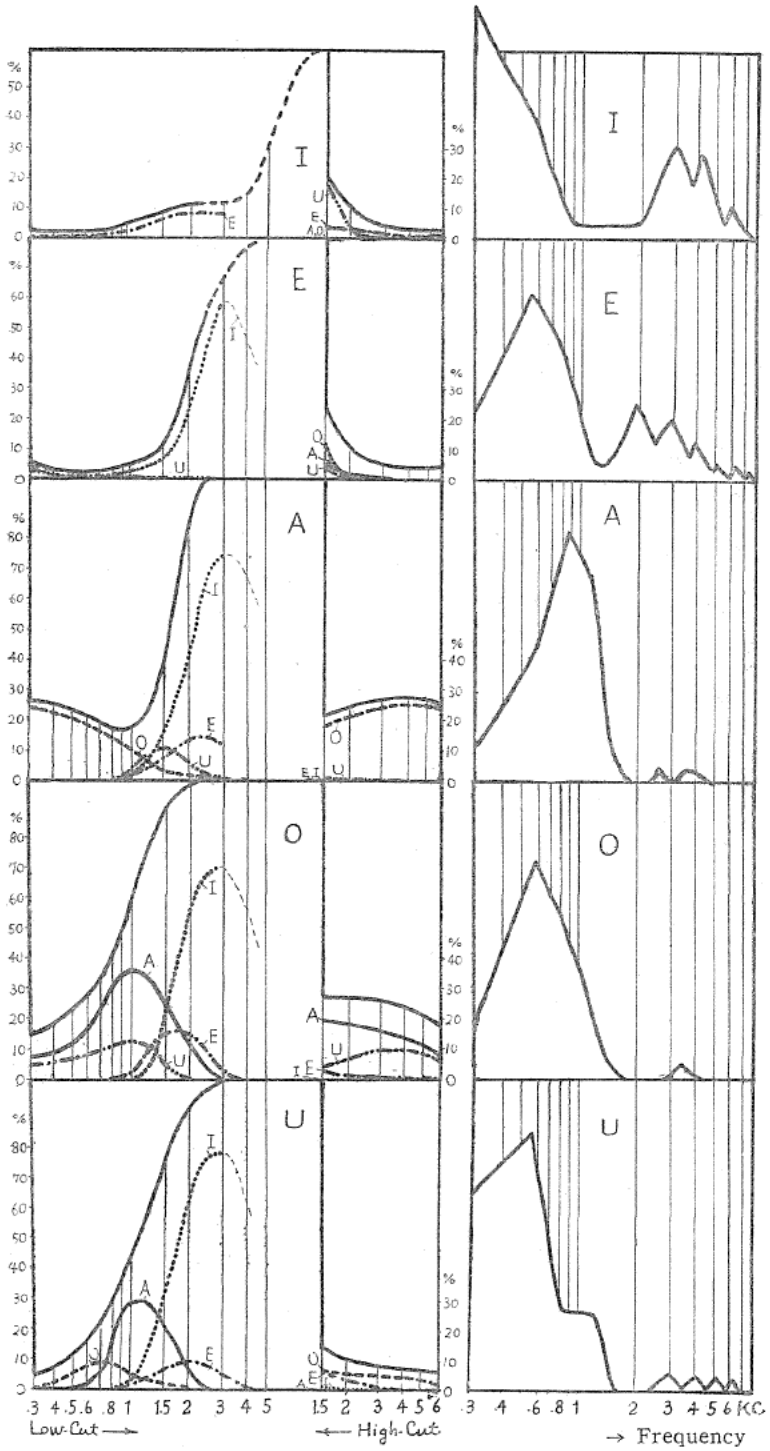

FIG. 9. Minute characteristics of phonemic confusion.

1–1.2 kc in LCD as their maximum-occurrence point. We can thus conclude that the most probable region of maximum happening of incoming confusions to "A" is nearly 1–1.2 kc. As for incoming confusions to phoneme "E", a similar phenomenon will be found in the outgoing confusions from phonemes "A", "O" and "U", but with a slighter degree of markedness. Incoming confusions to phoneme "I" seem by far the most conspicuous, but in this preliminary experiment we lack sufficient cuttings for more definite conclusion.* Nevertheless, we can say in short that incoming confusions, from whatever phonemes they may be coming, have a nearly constant place as their maximum happening. This trend is extraordinarily clear and sharp. As for incoming confusions to "O", the trend must be found in the outgoing confusion from "U" and "A". In this latter, however, there is some irregularity due to out-of-ordinary articulation-loss characteristics. Because there was some mispronunciation of phoneme "A" by one particular subject, the confusion from "A" to "O" was noticeably most violent at the beginning point of the distortionless condition.

Vocal-Confusion Characteristics

We show detailed characteristics of vocal confusions of each of all five voices with naturalness-loss characteristics in Fig. 10 where for the convenience of comparative observation the vocal patterns of the five voices are supplementarily placed side-by-side. By the subject voices inscribed at the side of confusion-curves we are able to see the destinations of outgoing confusions of each of the voices in question, that is, we know the details of outgoings which come from each of the five voices and at the same time we know the details of incomings with which all outgoings end. In this way we can easily trace out the finest movements of confusion-characteristics as functions of the distortions of either HC or LC. But in vocal confusion we cannot find such a conspicuous trend as perceived in phonemic confusion because here the changes in characteristics are gradual and relatively monotonous; there are no abrupt changes. There is also no frequent intersection of characteristics in either of the two directions of distortion, meaning there are no essential complications in this type of confusion. This is a general tendency which should be noted. Closer inspection of the relatively changeless characteristics of vocal confusion reveal the following facts:

(1) Judged from their individual vocal patterns, voices "F" and "H" show two extremes between which there is little connection, and in each of these two extremes we see the smallest quality-loss in the referential condition. Contrarily, for voice "Y" with a characterless pattern, the quality-loss in this condition is greatest; consequently, as a whole, the confusions can very easily take place in the order Y→H, Y→C, Y→S, Y→F.

(2) Generally speaking, the order of activeness in which confusion actually occurs is not altered by the application of distortion excepting only where the cutting distortion becomes very large, approaching to or extending beyond the vocal glen.

(3) There are many cases in which no reciprocity in confusion is found, for example, between large (max.) outgoing confusion from "F" to "S" and small (min.) outgoing confusion from "S" to "F", and also between heavy confusion outgoing from "Y" and incoming to "H" and slight confusion from "H" to "Y".

* In succeeding experiments we could fully verify this point.

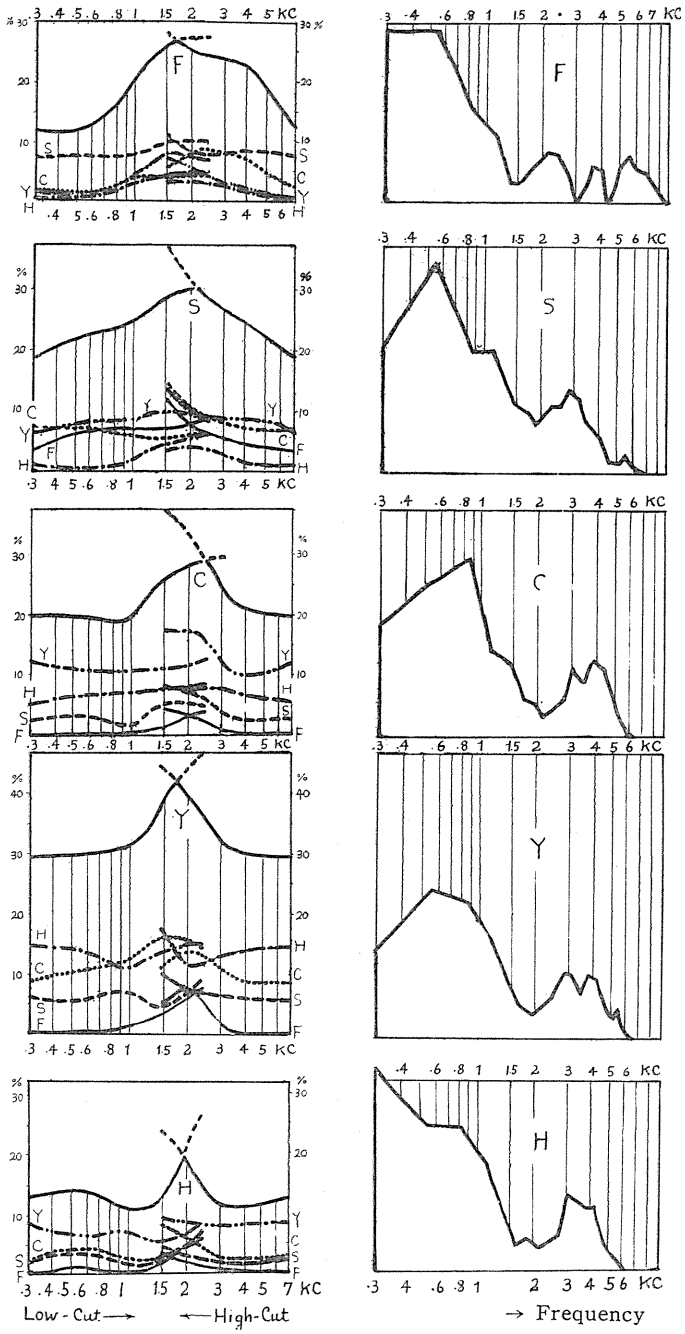

FIG. 10. Minute characteristics of vocal confusion.

It is interesting and most important to see how and where the reciprocal relation holds good in the course of the change in confusion characteristic toward the increase of distortion, but we will postpone this subject for the present. The great affinity between "Y" and "H", *i.e.*, the large confusibility between "Y" and "H", both female voices, is evidenced in Fig. 11 where it is apparent that there is very little difference between patterns of "Y" and "H" in the examination of pattern from the viewpoint of *frequency structure* only. In spite of large differences of patterns in such a sense, there are some cases where considerable affinity between voices exists, for instance, between "F" and "S". On this point, is it sufficient to stress the similarity of dominance in evenness in *harmonics structure* already mentioned or must we recall the fact that these two voices are both male? To approach this question it seems pertinent to consider, for the moment, another question: "Is there any element of sex-voice-quality perception?"

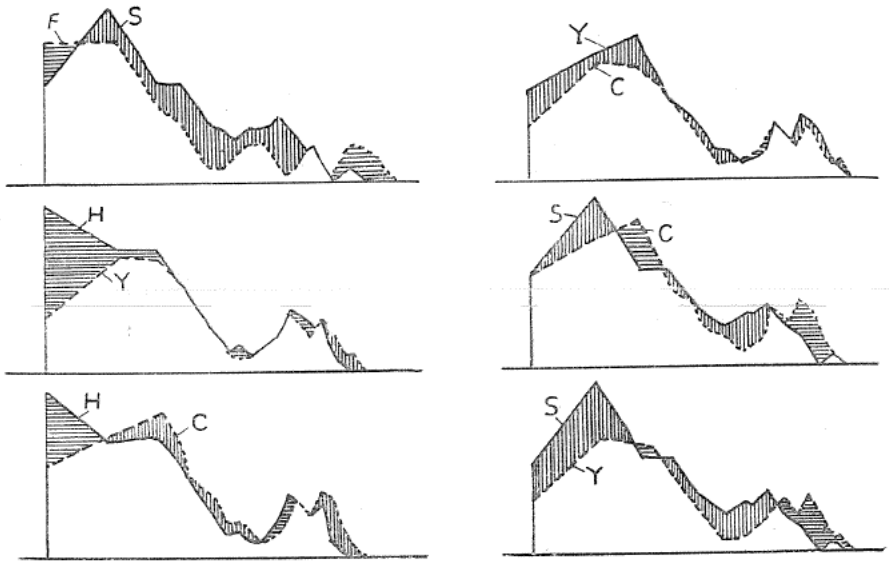

FIG. 11. Comparative study of vocal patterns, taking into consideration the pattern-difference (by shaded area) in any combinations of two vocal patterns.

Special Observation on Confusion—Judgement of Sex-Voice-Quality

For solution of a problem of the kind with which we are now confronted, it is better to look upon confusion phenomena as a whole and not try to arbitrarily pick out only that part of the phenomena which seems most convenient for ready conclusion. Consequently we continue to treat confusion phenomena collectively as we have from the beginning.

We first classify our five voices into three groups: Male-voice (M) composed of "F" and "S"; female-voice (W) composed of "H" and "Y", and an indeterminate voice (K)* composed of "C" only. Then we proceed to transform our origi-

* We call the voice of boy-subject before mutation "indeterminate" because his voice with 280-cycle pitch, while having some resemblance to a female voice in its limpidness and sonorousness, as yet it has no resemblance to the male-voice.

nal sex-voice-quality subject into one of "Confusion between Groups," leaving, for the time being, "Confusion within Groups" out of our discussion.

Our next step is to trace the movement of the characteristics of confusion between groups as functions of distortion-changes in HC and LC with the purpose in mind of discovering how the confusion between groups *M* and *W* is revealed, and, further, to see how the indeterminate group *K* moves in confusion response. This is well expressed by the so-called triangular representation in Fig. 12 which is based upon the percentage numbers tabulated in Table 2. In this representation

TABLE 2. Data of the Confusion-between-Groups Tabulated for the Purpose of Showing the Confusion Movement in Sex-Voice-Quality Aspect

		NC (all pass)		LC (0.6 kc)		LC (1.0 kc)		LC (1.5 kc)		LC (2.5 kc)		HC (4.2 kc)		HC (2.1 kc)		HC (1.5 kc)	
		No. of conf. obs.	Per cent of conf.	No. of conf. obs.	Per cent of conf.	No. of conf. obs.	Per cent of conf.	No. of conf. obs.	Per cent of conf.	No. of conf. obs.	Per cent of conf.	No. of conf. obs.	Per cent of conf.	No. of conf. obs.	Per cent of conf.	No. of conf. obs.	Per cent of conf.
W (♀)	M	33	4.1	44	5.5	38	4.7	47	5.8	121	15.1	39	5	73	9.1	89	11.2
	K	44	5.5	57	7.1	54	6.7	77	9.6	79	9.8	48	6	77	9.6	74	9.2
	W		(90.5)		(87.4)		(88.6)		(84.5)		(75.1)		(89)		(81.3)		(79.6)
M (♂)	W	42	5.2	50	6.2	73	9.1	113	14.1	109	13.6	70	8.7	89	11.1	113	14.1
	K	38	4.9	33	4.1	43	5.4	40	5.0	46	5.7	84	7.2	75	9.4	82	12.5
	M		(89.9)		(89.7)		(85.5)		(80.9)		(80.7)		(84.1)		(79.5)		(73.4)
K (△)	W	68	17	68	17	68	17	77	16.7	84	21.0	77	19.2	94	23.5	101	25.2
	M	12	3	14	3.5	6	1.5	27	6.7	37	9.2	12	3.0	44	11.0	49	12.2
	K		(80)		(79.5)		(81.5)		(76.6)		(69.8)		(77.8)		(65.5)		(62.6)

Key: ♀ female, ♂ male, △ indeterminate.

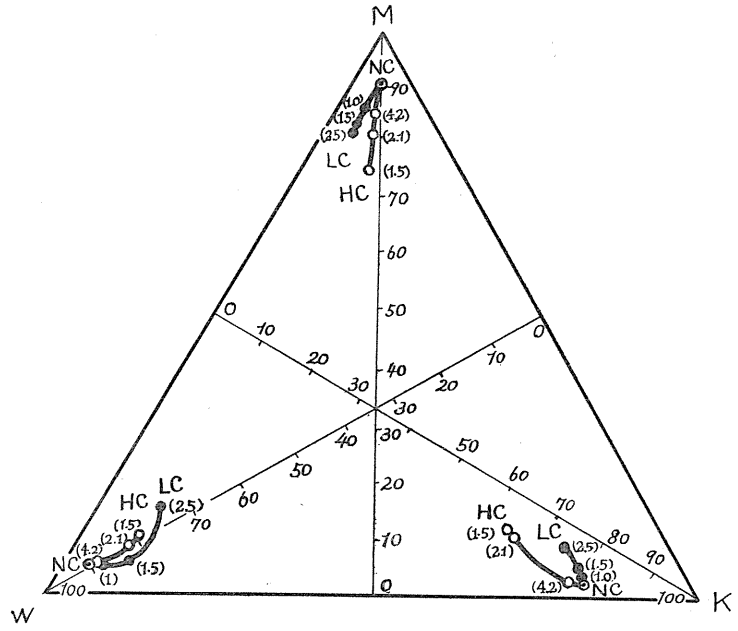

FIG. 12. Triangular representation of confusion movement viewed from sex-voice quality.

the parameters of distortion (HC and LC) are inscribed on the characteristics in heavy curves which move according to change of distortion. The important points in this figure are:

(1) For referential condition of no-cut (NC) *M* and *W* are both on their neutral lines, and that means all outgoings from *M* end as two incomings directed to *W* and *K* with a percentage almost comparable; and in a similar way, outgoings from *W* end as two incomings to *M* and *K* with almost equal percentage of confusion. As for *K*, *K* finds itself in a position attracted to *W* out of its neutral line meaning that the outgoing from *K* is directed only to *W* and not to *M*. This reminds us of the fact that the voice of the boy-soprano is akin to the female voice in its timbre effect.

(2) By lower-cutting (LC) *M* moves to *W* only, and that means all outgoings from *M* almost end as incoming to *W*, and also that *K* moves to *M* only which means outgoings from *K* almost end as incoming to *M*; meanwhile *W* behaves in another way by lower cutting. First, with slighter distortion, *W* moves toward *K*, but accordingly as the distortion increases, *W* begins to move so as to cross the neutral line, and finally with heavier distortion *W* moves toward *M* only.

(3) By higher-cutting (HC) the situation is considerably changed, for instance *M* and *W* almost follow the neutral lines, implying no biased confusion for this distortion. As for *K*, primarily with slighter distortion in this direction, *K* moves in parallel with the neutral line, and with heavier distortion *K* begins to move gradually toward *M*.

In short, in the original state without distortion, the boy's voice being similar to a female's, its confusion consequently moves only in the direction $K \rightarrow W$, and there is no noticeable similarity and hence no apparent route between *K* and *M*. Only by losing the lower parts of their own patterns, can *K* be confused with *M*, and *M* be confused with *W*, this latter being a sort of sex-voice confusion. As for *W*, cutting-down the pattern beyond the vocal glen only brings about sex-voice confusion in direction $W \rightarrow M$. On the contrary, by cutting-down the higher part of the patterns only, little is contributed to sex-voice confusion so long as the cutting distortion does not extend downward below the glen, as proved in our present experiment. This HC distortion has some influence only upon the indeterminate group *K*, bringing about a confusion toward *M* more slowly-paced than in LCD where one-sided confusion of $K \rightarrow M$ occurs from beginning to end with unmistakable suddenness and clearness.

In short, the most essential element indispensable for distinguishing the male voice from the female seems to be found chiefly in the lower part and not the higher part of their pattern. For the voice of boy-subject, the situation seems slightly different. Despite its similarity to female voice in the referential condition, this voice tends to approach exclusively toward male voice through the distortions in either direction, that is, under both HC and LC of distortion, the only difference is that the tendency of $K \rightarrow M$ is more rapid in LC than in HC.

Discussion

In our previous paper—Part I—on Timbre-Quality Study, we insisted upon the prime importance of the fact that naturalness quality differs quite clearly from articulation quality. In Part II now, this fact is clarified with more precision and lucidity by introducing the study on confusion presented here.

Actually, quality characteristics on whatever qualities they may be based are merely outlines depicted by mean values of various responses of all timbres concerned. A representation of such outline-figures can give us, so to speak, silhouettes or envelopes of quality phenomena by which we can infer the movement of phenomena as a whole; but beyond a description by the mean-value observation, neither can directly reveal an inward phase of quality phenomena where the actual and active real phenomenal appearance occurs.

Our confusion study in the aspect of dual-quality is of course composed of two elements, phonemic confusion and vocal confusion which, by comparing them with the so-called phonemic pattern and vocal pattern respectively we can, in some measure, make clear the essential differences in the two types of confusion.

For verification of the essential differences in quality between naturalness and articulation, it may suffice now to point out the clear difference in forms of their general and individual quality characteristics *vs.* certain distortions, as shown in publications listed in References (1) and (8).

For more positive verification we must secure indisputable evidence of confusion characteristics. Many conspicuous differences are found in the summarized representation of vocal and phonemic confusions. That the reciprocity in confusion between two timbres despite confusion direction (balanced confusibility) is seldom found in our referential condition of no-cutting, may be due either to the small number of observations or to the fact that the condition thereunder was not perfectly free from distortion. This aspect of reciprocity in confusion is of vital importance to the study of vocal confusion especially where reciprocity between outgoing and incoming confusions is recurrent during the whole course of outgoing and incoming confusion-characteristics. But we cannot dwell too long on this point only.

We have already aluded to the two types of quality-distribution, *concentrated distribution* and *dispersed distribution.* Both of these types are inferred from the results of confusion study. We have now something important to add here. The usual way to define the so-called *importance* is by describing the quality-density distributed per band by the aid of mathematically differentiating the curves of quality-characteristics (*vs.* BED) with regard to frequency band. This consideration is seemingly without merit and the process seems purposeless. We must reflect here that the process useful for physical quantity has nothing to do with the psychological quantity in that the magnitude of quality does not belong to physical quantity. We are therefore inclined to feel that deduction of the so-called *importance* through mathematical differentiation of quality-characteristics is irrelevant. Conversely, it is reasonable and meaningful to infer *importance* from the study of confusion phenomena because confusion itself involves a sort of differential problem of timbre-sensitivity. We shall reflect here on our scheme of interpretation and follow the steps of our thinking. With the help of confusion study, we have explained quality-loss characteristics, and for an explanation of confusion itself, we resorted to the study of timbre patterns. Those steps stem directly from the framework of our quality theory on timbre aspect. We consider our subtle approach to the problem on shading or nuance (color) is, after all, attained by the study of timbre-quality with which we became familiar through the intermediary of timbre pattern because timbre can be understood as a problem of some sort of construction. From these reflections it is logical that as confusion might be explained by

timbre-discrimination, so timbre discrimination might be interpreted as pattern-contradistinction. In the nature of things, the problem of sex-voice-confusion discussed here is of secondary importance. According to the procedure of our experiment, sex-voice-quality is but a pattern difference insofar as our concern is only with uttering voices. It remains for us to examine sex-voice quality from the listeners' side, that is, to see how the sex-voice differences influence the voice-quality judgement of listeners. This requires group test: Sex-voice-quality testing by listeners of both sexes. There is great value in such testing from the standpoint of telephone perfection.

Speaking from a more radical viewpoint, this subject of sex-voice-quality itself involves the difficult question of the delicate nature of an inner structure in harmonics timbre-construction. We cannot give an immediate answer here. This is, in part, because of the extraordinarily intricate nature of the subject and, in part, because, while most suitable for examining pattern from the standpoint of pattern-extension in frequency dimension, the band-eliminating distortion used in our experiments seems no longer suitable for the pattern-study in an aspect of harmonics-structure. What should be noted in this statement is that the failure to attain our highest goal of solving the harmonics-structure problem is not the fault of the confusion-treatment presented here but it is directly attributable to the unfitness of the distortion adopted.

Summary

After reviewing and summarizing the results obtained in these confusion experiments, we want to list the following points of importance.

The phase of phonemic confusion is quite different from that of vocal confusion. The former is characterized by a tendency toward abrupt and violent confusion changes, and the latter by a preservative inclination toward vocal configuration in normal condition with only gradual progressive confusion changes.

In Phonemic Confusion: Generally speaking, by lower cutting process, the phonemes having their most effective formant in the lower position tend to be confused with the phonemes having their most effective formant in a high position; by higher cutting, phonemes with formant in high position are apt to create confusion with phonemes having their formant situated in low position. With the revelation of sharp characteristics in phonemic confusion, incoming confusion to some particular phoneme (from whatever other phonemes it has come as the result of outgoing confusion) subtly suggests an inherent position in the frequency region which characterizes the phoneme in question. From these relations we are led to the concept of *phonemic formant*. This is one of the most important gains of our confusion study.

In Vocal Confusion: The confusion modes in referential condition of no-distortion are nearly always maintained even in cutting conditions. There are two marked modes, peripheral and diagonal, as shown in fig. 6. These modes of confusion tend to be changed slightly only with very large distortion. From the characteristics of vocal confusion, we can deduce some representation of the distribution of *vocal formant*. Vocal formant can be considered as formant appearing in pairs irrespective of the kinds of phonemes, that is, a pair consists of high and

low frequency structures and this fact is closely connected with the so-called *invariant formants* mentioned in Reference (2).

Postscript

Our concern with confusions is deep and sincere and our confusion study is one of long-continued concentration.[5)6)7)] The present paper plays only a prelusive role to more complete studies to follow. There are many difficulties which cannot be easily overcome and which necessitate further consideration and experiment. Supported by the ideas expressed here in this paper, we describe confusion phenomena with some ease and confidence. There remain many facts to be re-examined by further study because there are some shortcomings in our experiments which must be eliminated in future studies; for example, insufficient cuttings and failing to adopt a level-recovering process in cutting condition.

We think the solving of confusion is one of the most important of all quality problems. It is only through confusion study that we can ever get into close relationship with the inward sphere of quality phenomena and acquire living contact with the core of quality theory. This is our conviction as well as our conclusion.

References

1) Y. Ochiai and T. Fukumura: 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. M.F.E., Nagoya Univ., Vol. 8, No. 1, May, 1956.
2) Y. Ochiai and T. Fukumura: Timbre Study of Vocalic Voices. M.F.E., Nagoya Univ., Vol. 5, No. 2, September, 1953.
3) Y. Ochiai und T. Fukumura: Beiträge zur Erkenntnis der Klangfarbestruktur bei vokalischen Klangbildern. M.F.E., Nagoya Univ., Vol. 8, No. 1, May, 1956.
4) Y. Ochiai, T. Yamashita et T. Fukumura: A la Recherche du Timbre des Voyelles. —Quelques Considérations Basées sur Expériences de la Structure de Timbre Sonore des Voyelles Soutenues. M.F.E., Nagoya Univ., IV-2, 1951.
5) Y. Ochiai: Nippon On-In no Hizumi-Gensho to Denwa-Meiryodo ni Kansuru Kenkyu (in Japanese).—Studies on Telephone Articulation with Reference to the Distortion Phenomena of Japanese Speech-Phones. Paper presented for Doctoral Dissertation at the Tohoku Imperial University in 1943.
6) Y. Ochiai and T. Fukumura: Studies on Qualities of Speech and Voice by Timbre Distortion. M.F.E., Nagoya Univ., Vol. 4, No. 2, November, 1952.
7) Y. Ochiai and N. Izumitachi: Timbre Study on Mishearing Phenomena of Speech-Phones in Rotational Synchronous Distortion (Report I: Mishearing of Vowelic Timbre). M.F.E., Nagoya Univ., Vol. 7, No. 1, May, 1955.
8) Y. Ochiai and T. Fukumura: Study on Fundamental Qualities of Vocalic Timbre by Rotational Synchronous Distortion. M.F.E., Nagoya Univ., Vol. 8, No. 1, May, 1956.