STUDY ON FUNDAMENTAL QUALITIES OF VOCALIC TIMBRE BY ROTATIONAL SYNCHRONOUS DISTORTION

YOSHIYUKI OCHIAI and TERUO FUKUMURA

Department of Electrical Engineering

(Received May 31, 1956)

The observation and measurement of the fundamental qualities, i.e., the naturalness and articulation, of vocalic voices in the rotational synchronous distortion (RSD) are reported. Five sustained Japanese oral vowels uttered by four calling subjects were employed as timbre signals by being impressed on the distortion RSD, thereafter to be received and judged by four listening subjects. As a result of subjective judgement we obtained naturalness- and articulation-characteristic in general, and proceeded to study in greater detail both of these fundamental quality-characteristics by considering each individual vowel as well as each individual voice. This is the last and most important of all stages in a series of studies on speech qualities with the rotational synchronous distortion.

Introduction

It is needful in the transmission scheme of speech-communication engineering to pay just as much respect to the transmission of the voice-quality as to the transmission of the phoneme-quality, as we have already pointed out in previous papers.[1)2)3)4)] This is because, for example, the determination of the so-called transmission standard in telephone engineering is based exclusively upon the concept of *articulation-transmission* without any regard to *naturalness-transmission*, i.e., the transmission of vocal quality of telephone users. This imperfection in transmission design for speech communication seems to come from want of essential knowledge of speech quality as an object of transmission. In this circumstance we have to insist upon the necessity of quailty-study. To lay stress on the fact that quality-study is most indispensable in the field of transmission engineering, this small treatise is prepared wherein the leading idea that the concept of naturalness is quite different from that of articulation is effectively illustrated and clearly verified by utilization of the distortion RSD. A general articulation test in RSD and other measurements have already been carried out and reported in detail.[5)6)7)8)] For the direct purpose of showing the irrelevancy of naturalness and articulation, it is sufficient to refer to the so-called pitched timbres (vocalic voices) as timbre signals. This experiment on actual measurement of naturalness and articulation in RSD was originally executed in our Laboratory throughout the summer of 1954.

Procedure

We give a brief outline here of the experiment on timbre-quality measurement. The circuit used is shown in Figs. 1 (*a*), (*b*). The transmission characteristic of the parts of the equipment is similar to that described previously,[6)] with the exception of the recording microphone. In this experiment MR-103 type condenser

microphone* was employed for recording instead of the velocity microphone. As timbre signals, we used (1) five sustained oral vowels pronounced by four male subjects with an intensity level of about *mezzo forte* at a pitch of 140 cycles ; (2)

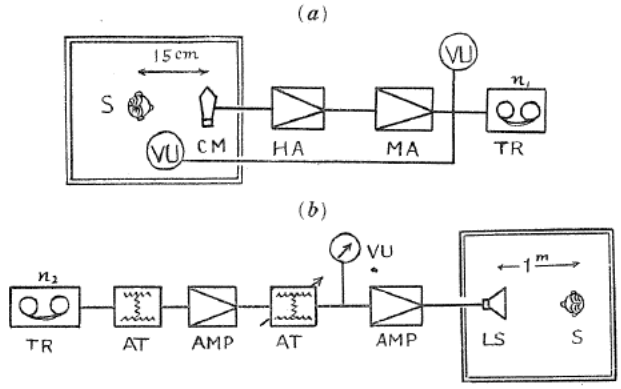

FIG. 1. Block-diagram of experimental circuits employed for timbre-quality measurement. (*a*): recording system, (b): reproducing system.

five sustained oral vowels by five subjects (four males and one female) uttered at a pitch of 240 cycles on the same level of about *mezzo forte*. As timbre signals, five vocalic voices of four subjects (twenty vocalic timbres), according to the process of level-matching, are recorded five times on one tape and are re-arranged and randomized in such a manner that they finally appear in reproduction quite at random with respect to the kind of vowel and voice. We must add one more point about the recording process. The timbre signals of vocalics for recording are to be prepared to make them fit for timbre judgement in pure sense.[1][2] We have cut down both the beginning and the end of each signal leaving only the quite stable part of about four seconds for the test and thus denying to timbre judgment any auxiliary help which might come from the perception of unsteady parts of voice signals, viz., the build-up and decaying parts of the presented voices. Each timbre signal is separated by a pause-interval of about four seconds, as shown in Fig. 2. Four listening subjects, seating themselves in a sound-proof room at a distance of one meter from a loudspeaker, are requested to identify not only the vowel of the timbre signal but the caller's voice to which they are listening. The discrimination and indentification of phonemes as well as voices is the task imposed

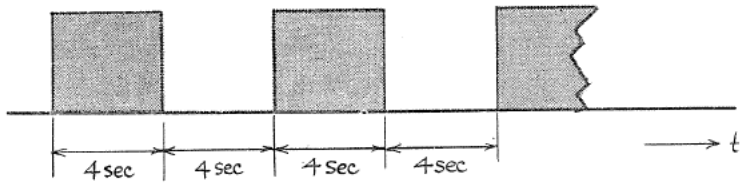

FIG. 2. Presentation of timbre signals.

* This was successfully developed in the Institute of Telecommunication of Japanese Telegraph- and Telephone-Public Corporation.

upon the listeners. The observations are repeated three times for each logatome-tape. This kind of judgement is the severest judgement of timbre because the listeners are forced to make judgement by appealing only to the difference in timbre structure. Other clues for judgement which proved to be effective in nuance (voice quality) judgement such as difference in levels and in envelopes were excluded in

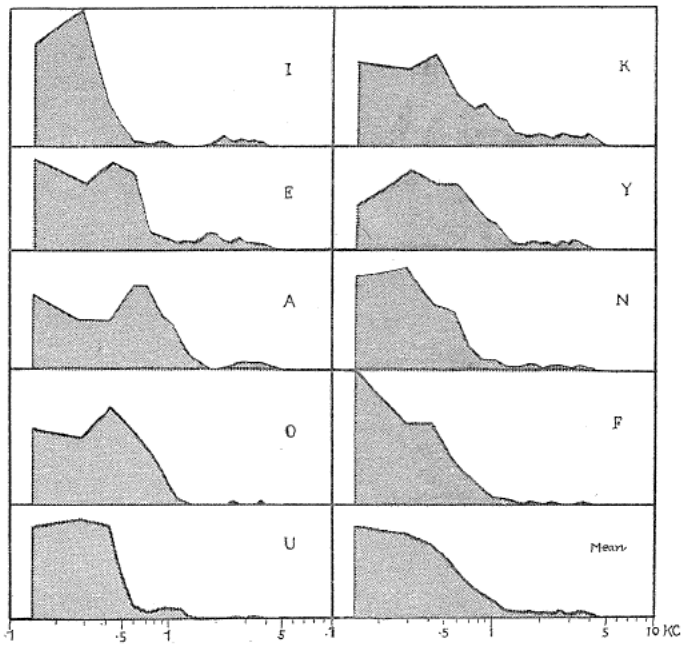

FIG. 3. Presentation of five phoneme patterns (I, E, A, O, U) and four vocal patterns (K, Y, N, F) and their mean pattern as timbre signals.

our experiment. Under these circumstances the listeners must necessarily form their judgement, relying exclusively upon the discrimination of the timbre-pattern in static sense, that is, the structural pattern of timbre in question. We show in Fig. 3 the vocal patterns of four speakers (K, Y, N, F) and also the phoneme patterns of five Japanese vowels (A, I, U, E, O). The former is obtained by the so-called *devocalizing* process and the latter by the *depersoning* process. As for the methods of pattern-obtaining, we have already discussed this in greater detail.[4)9)] As a measure of distortion of synchronous disorder, we use the term "*rotational ratio*" (r.r.) which is defined as a ratio of the reproducing speed n_2 to the recording speed n_1 or $r.r.=n_2:n_1$. We use three conditions of distortion in the speed-up region such as $r.r.=1.14$, 1.21, 1.36, and also three conditions in the speed-down region such as $r.r.=0.96$, 0.87, 0.74. In addition to these six points of condition, it is necessary to add one more point, i.e., the normal condition $r.r.=1$ and thus, under the conditions to be tested, there is a total of seven points. To change speed, we employ the source-frequency control method (FCM) for five distorted conditions $r.r.=1.21$, 1.13, 0.96, 0.87, with the exception of $r.r.=1.36$ where the capstan changing method (CCM) is used.[6)]

Results

We have already discussed in detail the nature of this distortion RSD which can be interpreted as some kind of transitional distortion in frequency dimension.[51] When the timbre patterns represented in Fig. 3 are shifted in frequency domain by this distortion, what changes produced in timbre quality can be perceived? It is this question that our experiment must answer. When we refer to timbre qualities, we mean, of course, phonemic as well as vocal values. Judgement as to phonemic value corresponds to that of articulation, and judgement as to vocal value corresponds to that of naturalness. Thus we can show naturalness characteristics as the result of judgement on vocal quality, and articulation characteristics as phoneme quality judgement.

Naturalness and articulation characteristics in general

The representation of fundamental characteristics of naturalness and articulation for seven conditions of distortion is done in such a manner that the speed-up region of the distortion is placed on the side opposite the speed-down region and the normal state of condition is the center. In Fig. 4 we show the naturalness and articulation characteristics for two sets of timbre signal, i.e., 140~-pitch group of timbre and 240~-pitch group. The former contains 20 kinds of timbre (5 vowels by 4 voices), and the latter 25 kinds of timbre (5 vowels by 5 voices). The four listeners are the same for both pitch group sets. Both voice groups are repeated five times and recorded on one tape. Measurements are completed by reproducing the same logatome-tape three times. Thus, each one vowel of each separate voice in both pitch cases is brought to the attention of all four listeners a total of fifteen

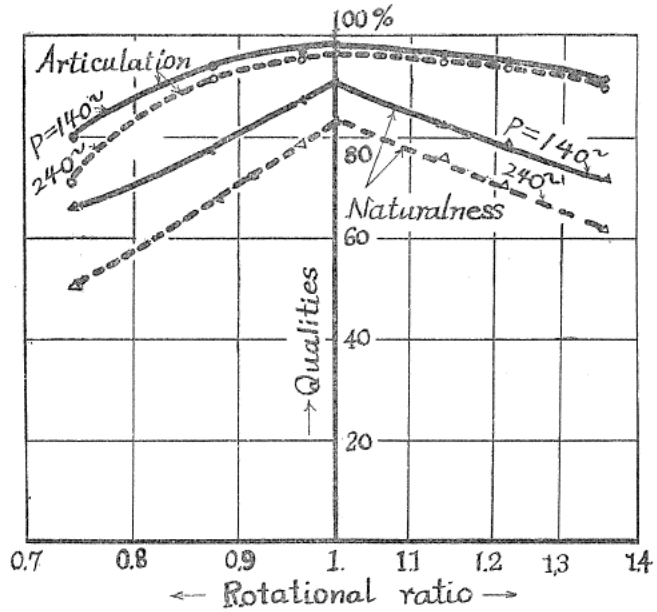

FIG. 4. Naturalness and articulation characteristics vs. rotational ratio of synchronous distortion. Solid-line curves are 140~-pitched timbres, and broken-line curves 240~-pitched timbres.

times. The quality measurements at every stage are the result of mean values of 1200 observations for 140~-pitch group and 1500 observations for 240~-pitch group. It goes without saying that the articulation characteristics shown here are the general or averaged articulation characteristics without reference to the kinds of vowels, and in the same manner that the naturalness characteristics are the general or averaged naturalness characteristics notwithstanding the kinds of voices. We shall hereafter abbreviate ARTICULATION CHARACTERISTIC IN GENERAL as AC(G) and NATURALNESS CHARACTERISTIC IN GENERAL as NC(G).

By inspecting Fig. 4, we can note the one clear-cut feature that naturalness characteristics have a tendency toward a nearly straight line which is quite different in shape from articulation characteristics. The trend of characteristic-shape of naturalness and articulation seems to be irrelevant to the pitch of the timbre signal, to the voice content (number of voices) and to the kind of vowel. It is important to point out that the transitional distortion of the constant interval type has a less vigorous effect upon the articulation than on naturalness where the same distortion becomes very serious. This empirical fact is most significant. But why and how in fact does this quality phenomenon take place? We must make a more positive approach to this problem. Let us therefore proceed to a detailed study.

Articulation and naturalness characteristics for individual vowels

For a more detailed study, we need only one example. Let us take the case of a 140~-pitched timbre. We show in Fig. 5 the fundamental quality characteristics for individual vowels; Fig. 5 (*a*) shows articulation characteristics per vowel (AC/V_I), and Fig. 5 (*b*) naturalness characteristics per vowel (NC/V_I). Each one of the points in these figures was determined by the mean result of 240 observations per condition. When we interpret the quality phenomena AC/V_I and NC/V_I, we must refer to the phoneme patterns given in Fig. 5 (*c*). The two most important and even conspicuous facts which we find on closer inspection of these characteristics are:

(1) Those vowels of relatively dull and round characteristics in NC/V_I correspond to vowels which have relatively sharp and pointed characteristics in AC/V_I. Here we mean the vowels "A" and "O". Contrarily, those characteristics of vowels "I", "E", "U" in NC/V_I which are typically pointed and sharp correspond to the characteristics which have flat and round typicality in AC/V_I. As to the irregularity and out-of-ordinary characteristic of "A"-vowel in AC/V_I in speed-up region, we shall discuss this later.

(2) In a general and rough sense the vowels of higher naturalness in NC/V_I correspond to lower articulation in AC/V_I. In the same way the vowels of higher articulation in AC/V_I correspond to lower naturalness in NC/V_I. By utilizing such outstanding features, we can classify these five vowels into two groups, one of "A" and "O", and the other of "I", "E" and "U". For an interpretation of this quality-phenomenon, we must remember that vowels "A" and "O" have the nature of single or nearly single formant.* On the contrary, vowels "I", "E", "U" (*uu*) have double or multiple formants. We can infer then that double-formant vowels with a relatively longer formant-interval are stronger in articulation and weaker

* We imply here that the interval of a double formant is so short that the double formant is taken as a nearly single formant.

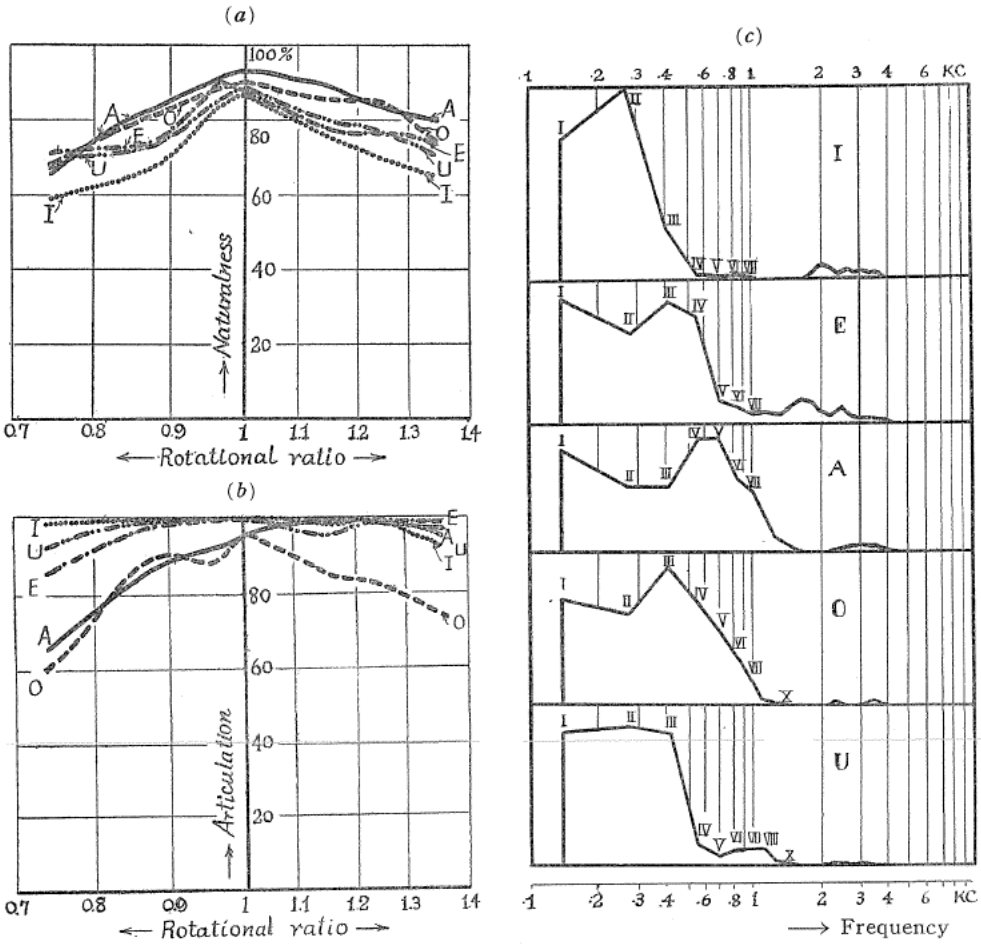

FIG. 5. Quality reponse characteristics per vowel, obtained from a result of 240 observations per condition. (*a*): Articulation characteristics per vowel (AC/V_I), (*b*): Naturalness characteristics per vowel (NC/V_I), (*c*): Phonemic patterns of five vowels obtained from mean results of 5-6 times measurements for each phoneme.

in naturalness when under the influence of the distortion RSD. Single-formant vowels or double-formant vowels with a relatively shorter formant-interval are more sensitive in articulation and less sensitive in naturalness. It is relatively easy to interpret the quality phenomena from the phoneme patterns which are comparatively well known. It is for this reason that we began the quality analysis with the study of quality characteristics per vowel. Now we go on to the study of another quality of a more delicate nature.

Articulation and naturalness characteristics for individual voices

We show in Figs. 6 (*a*), (*b*) the articulation and naturalness characteristics per voice (AC/V_c and NC/V_c). For the convenience of reference, in Fig. 6 (*c*) we added the figures of vocal patterns for voices of four speakers. By inspecting Figs. 6 (*a*),

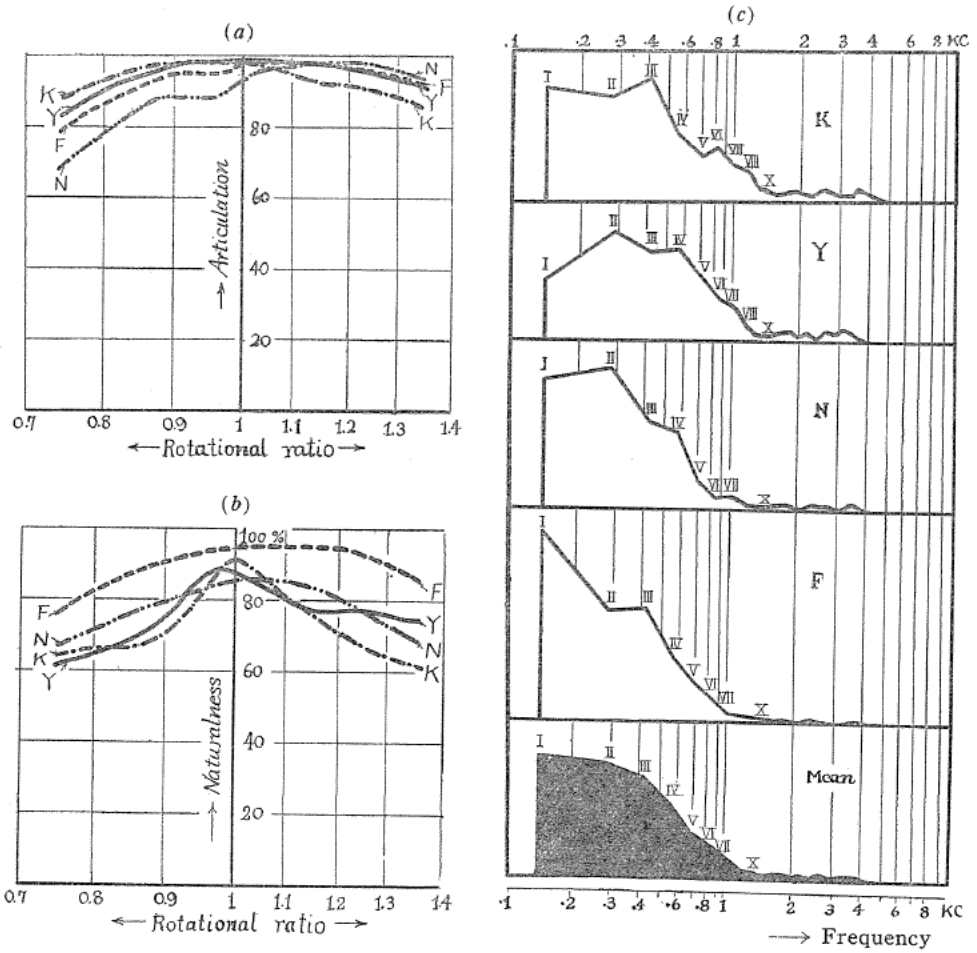

FIG. 6. Quality response characteristics per voice, obtained from a result of 300 observations per condition. (*a*): AC/V_c, (*b*): NC/V_c, (*c*) Individual vocal patterns and mean pattern of four speakers, obtained from the mean result of 5–6 measurements for each voice.

(*b*) we can see that the similar phenomena as observed in Figs. 5 (*a*), (*b*) are prevailing as to voice quality, viz., voices with fuller over-structures* become more sensitive in naturalness response to distortion than those with poorer over-structures. Therefore, we can classify these four voices into two groups, one of "K" and "Y", the other of "N" and "F". Voices "F" and "N" reveal the dull and round characteristics in naturalness response; voices "K" and "Y" show the more sharp and pointed characteristics in naturalness response. Referring to the vocal-pattern representation, we know that voices "K" and "Y" are rich in harmonics and therefore their patterns extend widely and strongly in higher frequency domain. Meanwhile "N" and especially "F" are poor in harmonics and their patterns remain in a range relatively more narrow. Timbres with more complicated structures become highly sensitive in naturalness response to the distortion RSD. The cor-

* Structures extended in upper frequency domain over the point of vocal glen.

respondency of higher naturalness in naturalness response with lower articulation in articulation response holds good here only in rough sense, but not exactly. For example, voice "F" with the highest naturalness in NC/V_c does not show the lowest articulation in AC/V_c. We can find that voices "K", "Y" with the lowest naturalness show the highest articulation only in the speed-down region. The same phenomenon can not be found in speed-up region. It is rather meaningful to point out that in characteristic AC/V_c voices with higher articulation in speed-down region must be subject to lower articulation in speed-up region and *vice versa*.* The articulation characteristics of individual voices in down-region are set in perfect order by gradation "N", "F", "Y", "K". Despite all the small deviations of these characteristics, the reversibility of the order with respect to up and down regions is quite remarkable in that it reflects some essential *unsymmetricity* in articulation characteristics. Of the four voices, voice "N" is most similar to the averaged mean pattern. It is for this reason that the articulation characteristic of "N"-voice and the articulation characteristic in general are most alike.

Common and connected observations on two types of characteristics per vowel and per voice

We summarize here our observation of two types of quality characteristic with respect to individual voices and individual vowels. As a common trend of characteristics without regard to difference in quality, we can state:

(1) That quality characteristics are generally nonsymmetric with respect to speed-up (U) and speed-down (D) regions; that quality loss in D-region is generally greater than in U-region.

(2) Generally speaking, naturalness characteristics are more or less symmetrical as compared with articulation characteristics.

(3) Also in a general sense as concerns our representation, naturalness characteristics show some sharp pointed forms and articulation has some round convex forms. Nevertheless there are relatively pointed forms in a group of round-form characteristics, and there are also relatively round-form curves in a group of pointed characteristics.

We are able to touch upon special meanings in our comparative study of pattern forms and to allude to facts connected with our observations. Included in these observations, but of less importance, we list the following facts:

(4) The particular representations of quality characteristic are NC/V_c and AC/V_I where the characteristics become more divergent accordingly as the distortion increases in both directions. This is most reasonable because naturalness characteristics are affected by voice-kind, and articulation characteristics are susceptible to phoneme-kind.

(5) The representation of characteristics by AC/V_c and NC/V_I is not so significant in the sense described in (4). Nevertheless, it has some importance for the following reason. It reflects some information on the interdependence of the phoneme and the voice. We should read something from these characteristics: voice has some effect upon phoneme quality and phoneme has some effect upon voice quality. Generally speaking, neither the deviation of articulation-character-

* This phenomenon concerning male and female voices has already been found—see Reference (6).

istics due to the variance of voice nor the deviation of naturalness-characteristics due to the variance of phoneme is large, and therefore these characteristics do not diverge even if the magnitude of distortion be increased.

Special observation

In U-region of characteristic AC/V_1 we can find the special phenomenon in "A"-vowel previously referred to. The optimal articulation of "A" takes place in the condition existing at about $r.r.=1.2$, and not in the normal condition at $r.r.=1.0$. This seems to imply a mispronunciation of "A"-vowel. By inspection of phoneme pattern of "A"-vowel, it was ascertained that its formant peak of 600–700 cycles is too low to be rightly understood as "A". For a perfect understanding of "A"-vowel in Japanese pronunciation, its formant peak must be situated within the frequency range of 900–1000 cycles[4)9)] which means a required speed-up distortion of at least 20% in transition. By utilizing distortion of RSD, we are able to detect with greatest sensitivity any minute defect in pronunciation and in voicing provided, of course, that the listening subjects are experts and their judgement reliable. As to the listeners in our experiment, they were quite physically fit and therefore we can infer that there was a mispronunciation of "A"-vowel as was detected through the shifting phenomenon of optimal articulation.

In any study of the fundamental qualities of speech and voice, the importance of the rôle of the transitional type distortion RSD cannot be overestimated; but we must not overlook the equally important study of the distortion of band elimination (BED).[10)] The importance of both studies lies in the fact that through the quality study of the former we attain an important concept of quality, i.e., the *quality in band position*, and through the quality study of the latter we attain the concept of *quality distribution in frequency dimension*. Not only by the cutting down of frequency band but also by the shifting of band position are both the phoneme and vocal qualities effectively influenced. It is only by studying these influences that we can grasp the essential character and behavior of phoneme and vocal qualities. Without the aid of these distortions, our knowledge of quality covers only the surface of things. We cannot attain the true concept of quality without basic data on quality-response characteristics which is supplied only by experiments on distorted conditions. For example, for the first time we understand the essentially weak character of naturalness in the distortion of position-shift by studying the rapid drop in linear form in naturalness-quality responses. This is in notable contrast to the articulation quality which drops reluctantly. It is also through this distortion that we have come to know that the voice pitch has little influence upon articulation; on the contrary, it has a considerable influence upon naturalness. We can say with some confidence that there is a general tendency in this distortion for the vowels with higher naturalness to become poorer in articulation quality and in the same manner for voices with higher naturalness to become poorer in articulation quality. It is possible that there may be some exceptional voices having a higher articulation quality notwithstanding a higher naturalness. These facts are important from a scientific standpoint and in addition they are also important when viewed from their practical application.

Acknowledgement

The authors take this opportunity to express their appreciation to the Institute of Telecommunication of Japanese Telegraph- and Telephone-Public Corporation for encouragement and use of equipment in this study. We are also indebted to our listening subjects: Miss Mitsue Sato, Messrs. Yoshiji Ono, Kenji Shimizu, Yutaka Sugimori for their active participation in this original experiment and in the timbre drill as well.

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 (Part I). M.F.E., Nagoya Univ., Vol. 6, No. 2, 1954.
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 N. Izumitachi: Timbre Study on Mishearing Phenomena of Speech Phones in Rotational Synchronous Distortion. M.F.E., Nagoya Univ., Vol. 7, No. 1, 1955.
8) Y. Ochiai, S. Saito and Y. Watanabe: Allowance Problem in Rotational Synchronous Distortion as a Study on Timbre Discrimination by Infinitesimal Position-Shift in the So-Called Timbre Space. M.F.E., Nagoya Univ., Vol. 7, No. 2, 1955.
9) Y. Ochiai und T. Fukumura: Beiträge zur Erkenntnis der Klangfarbestruktur bei Vokalischen Klangbildern. M.F.E., Nagoya Univ., unter der Presse.
10) 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., in press.