QUALITY STUDIES OF SPEECH TRANSMISSION SYSTEM ESPECIALLY WITH REFERENCE TO CHARACTERISTIC FORMS IN FREQUENCY RESPONSE

Yoshiyuki OCHIAI, Shuzo SAITO, Masanobu WATANABE and Teruo FUKUMURA

Department of Electrical Engineering

(Received October 31, 1954)

We have reported here on quality studies of the speech transmission systems especially with reference to their nominal forms of frequency-response characteristics: a kind of study on transmission merit, dealing exclusively with questions on over-all forms of transmission characteristics considered to be an objective measure in transmission rating. This study is divided into two sections; one, loudness balancing; the other, timbre estimation. These two sets of quality study are bound together more closely by a unitary viewpoint of quality transmission. We have based that part dealing with timbre study chiefly upon articulation measurement, but, at the same time we have ventured an experiment by cutting a path to the measurement of naturalness. At the beginning, in our "Introductory Consideration" we endeavored to make clear the fundamental motive which led us to attempt such research and at the end, under "General Discussion," we added a summarized view to give our purpose in entering into so wide a field of quality discussion.

CONTENTS

Introductory Consideration

Preface

It was, in fact, about the specific distortion due to the lack of synchronism in the carrier telephone transmission system, that the first attempt to study quality was made in a small systematic way, though experiments in the articulation measurements themselves were often made in our laboratory. The next study conforming to the same principle followed, dealing with distortion due to the lack of synchronism in revolving type of sound recording system. As for other types of distortion such as band-eliminating (BED) and level-attenuating (LAD), some quality studies were occasionally executed, but there was no summarized, collective study carried out. At any rate those distortions that were treated by us are naturally considered simple and pure in character.

The third study on quality will be reported here; it concerns the distortion caused by the deflection from evenness (flatness) as to the form of frequency-response characteristics. But when we think of what is meant by this problem, we cannot help shrinking before its inherent difficulties, first, because it is one of the most serious distortions, *i.e.*, a distortion of an extremely delicate and difficult nature and, secondly, such a problem necessarily connects with the discussion of so-called "fidelity" in the science of communication. It pushes us into the study of concepts of functional core of communication itself. We have so many problems to contend with; so many questions to be answered; we indeed face a difficult proposition. In facing such a situation we would do well to take a step backward to consider the essential nature of that proposition. Thus we must first discuss the nature of our problem.

Proposition

If we want to consider freely, that is, exempt from the influences of current opinion or conventional view, then we must place ourselves outside the idea that the flat response characteristic is one of the undistorted conditions or the non-flat response is nothing more than distortion. Speaking logically, even this seemingly undoubted principle must stand before the bar. We are required to give up the premise in traditional opinion. Or, to be more precise, we must begin with doubting the so-called "undisputed" premise. Following such a line of thinking, we are persuaded into posing the question in the following manner. "What is the most appropriate form of transmission characteristic of frequency response, for example?" That question concerns the expediency of the form of frequency response. Naturally no difficulty is reduced by such an argument. Difficulties in the experiment will rather increase because the scope of the experiment will be extended and the moves of the experimental processes will be infinitely augmented.

But, before setting out on this main attack, we must try some skirmishes. For example, by what measurement can we attain judgment on their expediency? It is utterly unknown what nature of quality is available for such research. Because, for the sound quality of speech we can name three attributes: loudness, pitch, and timbre; the last, as timbre-quality of speech, has two sides.

In short, at this stage, our research plan is as follows: Suppose that speech signals are given, and that the transmission systems are also given as to their forms of frequency responses; to seek out the differences in quality responses in given varying systems under these conditions, it is our present aim to reach our goal by way of inventing, or improving upon a method of quality measurements. Our subject for the time being, then, is to see what differences in quality responses are caused by such-and-such deformations in shape of transmission characteristic. Our present task, therefore, concerns solving this problem which can possibly answer the question whether the main problem on the expediency of characteristic form or on the optimum choice of characteristic form is, in fact, worthy of study. Further, in our study we have a subsidiary purpose to make some contribution to the selection of the most appropriate method concerning a quality judgment of this kind, and moreover, in greater detail, to the determination of the degree of precision which is required for the measurement of quality discrimination in the main and decisive problem.

Now that we call in question the form difference of frequency characteristics, we must give some little thought to those attributes of sound qualities which can really be considered important in our quality theory. By observing grossly the phenomena of transmission by a distortion of this sort, we can conclude that the loudness and the timbre are of considerable consequence, and that the pitch counts for little. We may possibly do away with pitch entirely. Then, provided that the loudness and timbre are seriously studied in the form, are they to be studied separately without any connection between them? Or, are they to be placed in a correlated position, bound together, making a complete quality theory which is summarized and synthesized by a unitary aspect of quality transmission? We must probably demand a fundamental and general consideration in the very beginning in order to make clear our views on quality treatment.

General consideration

Before going into the detailed study of quality measurement, it is important, even convenient, to contemplate first the significance and the value of the timbre study in speech communication. The articulation measurements are traditional and practical ways for quality appreciation of speech transmission system, and one cares little for a more fundamental base than this in which the articulation-judgment itself is essentially and principally rooted. This is justly the timbre attribute; without serious reference to the pitch element, the articulation-judgment has to do with the timbre element. The interpreting operation on articulation is carried on in the clearness aspect, relying naturally upon a sort of timbre-judgment. This paper deals mainly with the articulation quality to which the timbre-judgment is not limited. There is another timbre judgment which exists, a full report on which will be made later. Apart from the detailed discussion on timbre-judgment, we now have to proceed to a consideration of the most fundamental condition under which the timbre studies are to be performed. As we shall see, this consideration will provide us with a principal standpoint on the treatment of the timbre problem. This consideration is available for general treatment notwithstanding the type of timbre-judgment.

How is the timbre to be defined in our quality theory? In sober truth it puzzles us. We cannot answer immediately. And, on the other hand, we have not space enough to particularize. Insufficient explanation is likely to cause misunderstanding. It will not be long before we report on this in two other papers, "On Timbre Quality" and "Transmission of Quality." Here we simplify by questioning: How can the timbre be most conveniently considered? We think it depends sensibly on the circumstances under which we are forced to meet the problem. We will now pass on to the next step.

Timbre in pure sense

When we consider the timbre as depending on the pitch as well as on the loudness, we assume that such timbre is the pure timbre; more strictly, the timbre in pure sense or timbre in pure judgment. By thus setting up the conception of timbre in such a restricted way, we will proceed to meet the actual problem. When we are obliged to compare one sound with another in respect to their timbre in pure sense, we must set to work beforehand to match the two sets of sound in pitch as well as in level. For an exact study of timbre the pitch- and level-matching become an indispensable, preliminary condition. Speaking from this standpoint, it might be meaningless to compare one timbre with another when the two sounds concerned are found to be different either in pitch or in level. Let us explain further by taking another example. In the treatment of the timbre of human voice, the vowels that are to be placed under test must be voiced at a pitch and on a level conveniently prescribed in advance. Moreover, for example, if we are not successful in controlling the uttering level on the part of the talking subject, it is necessary, as the next best condition, that the sounds under test be well matched through transmission system at the reproduced level. In order to conform to such a definition of timbre, the method for the timbre study must follow the steps thus far mentioned. We have not enough space here to describe more closely the train of reasoning.

Problem of timbre comparison in transmission system

Aside from the abstract contemplation on quality, we must proceed to an actual case of timbre treatment. Naturally we deal with this problem in the SUB-

JECTIVE PHONAL ASPECT. We can measure the transmission faculty of transmission system by means of absolute or relative calibration. Even the physical data on speech sound can be obtained. But there is no connection between these three, in-so-far as they remain physical data.

Quality theory in speech communication must begin by closely connecting the three elements, speech as signal; transmission faculty of natural and technical system; hearing, and in addition, if possible, judging operation of the person. We admit the existence of quality only where and when the subjectiveness of the human being takes part in its phenomena. In communication, of course, we can find the quality of communication; communication has for its premise subjective mutual responses between COMMUNICANT and RECIPIENT, without which communication ceases to exist.

We define the SPEECH QUALITY as that which is to be expressed and understood. By speech quality thus we mean the speech which is assumed to be expressed and to be impressed subjectively. In the same manner we can also define the TRANSMISSION QUALITY or SYSTEM QUALITY by the expression of speech quality received over transmission system. Thus, according to the prescribed manner of symbolic method* we have as speech quality

$$\boldsymbol{H}(X) = \Xi,$$

and as system quality

$$\boldsymbol{H} \cdot \mathrm{D}(X) = \Xi'.$$

For a judgment of the adequacy of speech transmission system, we must resort to the subjective rating based upon the principle of direct or indirect comparison of transmission quality with speech quality as expressed by

$$\boldsymbol{H} \cdot \mathrm{D}(X) : \boldsymbol{H}(X).$$

For an establishment of method of quality measurement, it is necessary to find some reference type of distortionless transmission system in which the transmission quality is the same or similar to the speech quality itself.

In general we can pick up the phonal attributes of pitch and loudness for speech sound as we did for pure tone. Further, with speech sound we must rather over-stress the timbre attribute. Accordingly, for the subjective rating of transmission system, we must specify three elements of transmission effect in speech communication system. As a result we have three transmission-qualities, *viz.* quality of loudness-transmission, quality of pitch-transmission, quality of timbre-transmission.** According to distortion, we can question any one or all of these three qualities.

If we are forced to rate any two systems in respect to their transmission effect of timbre, we must examine the timbre qualities of the speech transmitted and reproduced by these two systems. We call this the timbre comparison between two transmission systems.

As long as the phonal attributes are interpreted in a purely physical way,

* In our symbolic description: **D** is the physical operation of distortion due to transmission system, $\boldsymbol{H}$ is the hearing or judging operation; X implies the speech signal in physical sense and Ξ the speech signal in subjective sense. The symbol: or / in our description implies some procedure of subjective comparison. For further information on this, refer to the paper "Transmission of Quality" *loc. cit.*

** We do not now touch upon the problem of time-quality (duration-quality), which will be called into question in the treatise "On timbre Quality."

they can remain quite independent of each other. But when we take them in the subjective way, they cease to be independent of each other. For example, loudness is relevant to pitch, and *vice versa*; and timbre more relevant to both, and *vice versa.* Timbre cannot be subjectively and must not be strictly, considered independent of loudness and pitch. Therefore, in touching the problem of timbre quality, we are unable to avoid or to neglect the treatment of speech sound as to its loudness and pitch.

Pure comparison of timbre

The lines of thinking so far mentioned lead to the distinctive view of the pure comparison of timbre. If we try to compare two speech-phones as to their timbre, we must do it after equating their loudness and pitch. And if we are forced to compare any one system with any other with reference to timbre transmission effect, we must also make timbre-comparison after finishing the matching of the two systems in both loudness and pitch. This is the very idea of strict comparison of timbre available for the positive detection of timbre difference, eliminating all differences caused by pitch and loudness.

Speech and system

In speech communication the study of speech signal and that of transmission system stand in such a strange relationship that we cannot describe it in a word. A profound study of speech cannot be achieved without making use of the advantages offered by transmission engineering, advantages given by interesting transmission-performance of various distortion systems. Even studies on both pitch and loudness of vowels and voices cannot be executed sufficiently by refusing the aid of transmission engineering in modern telecommunication. On the other hand, the study of transmission-system itself, after all, comes to nothing unless we give some regard to the actual speech signal, as we have shown in earlier reports.

We are duly convinced that the transmission plan inclusive of transmission system design must rest upon the knowledge of speech signal (speech quality), but for an understanding of speech itself, the study must be submitted to some restriction or to some accomodation for research in an exact ratio to the degree of precision and dexterity of the technique in transmission engineering. We should like to emphasize the fact that the system design in speech communication must have some deep-seated, rational ground, supported by serious inquiries into speech quality, not by faint and perfunctory opinions which are based on the simple empirical facts or merely on desirable expectations. In the succeeding section we illustrate this by an outline sketch of an actual example of system design, incomplete and rough though it may be.

Outline of Experimental Plan

In general

The first meaning of loudness study is naturally to obtain the condition of loudness matching and to make it available for an exact study of timbre in pure sense. But in general loudness study as one kind of quality transmission can have its own meaning. For example, there is an interesting problem which we actually meet in the field of broadcast engineering. The question arises when local broadcastings are compared with country-wide relay broadcastings through a trunk-line connection-system in which the transmission lines differ in frequency-response characteristics. The question here is reducible to the condition where the loudness effect must be kept constant irrespective of the differences in transmission characteristics of the intervening systems. These circumstances enable us to make a closer investigation of the effect of loudness transmission.* Apart from the practical application given by this example, the problem of loudness transmission is in itself quite interesting.

Choice of transmission characteristics

Choice of form

In order to study the effect of forms of over-all characteristics of frequency-response upon qualities of loudness and timbre, we must consider briefly how the types of characteristic forms are to be selected, because we have the widest latitude of choice. There are, in fact, innumerable conditions in which there are a variety and diversity of forms. We must choose the most typical and the most valuable.

It is suitable and even well-advised to adopt the so-called flat or approximately flat response because the flat response characteristics are convenient as a reference condition in such a comparison. It has one advantage that it can be reproduced with ease. This condition is well known as one of the so-called distortionless conditions in the science of transmission engineering.

With the choice of other characteristics for study, this time we should like to have some characteristics considerably deflected from flatness. In the present stage of study, it is above all most desirable to know how the differences in quality can actually be detected by the influence of form change. and less desirable to decide in detail the most preferable type of characteristic forms. Guided by these ideas, we have selected two types of characteristics, one, peak-form having resonance point at 1,000 c.p.s.; the other, valley-form with anti-resonance point at 1,000 c.p.s.

Why did we choose the resonance or anti-resonance point at 1,000 cycles? Furthermore, for what reason were the peak and valley types chosen? Is there any reason other than that they are the most deflective types from evenness? We will discuss this in the following paragraphs.

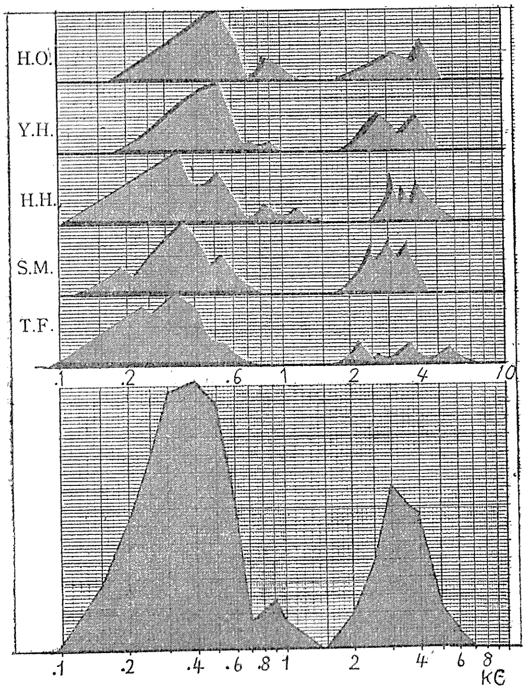
Results of vowel analysis

We reported the result of vowel- and voice-analysis in the paper titled "Timbre Study of Vocalic Voices" where five oral Japanese vowels** uttered by five normal Japanese subjects are obtained by a so-called DEPITCHING process, although so few as five subjects, in the strict sense, should not be taken as representative Japanese. Nevertheless, we can suppose that the average mean pattern of vowels for the five subjects gives some indication of a system design for timbre transmission. Thus, if we take a DEPERSONING process, details of which course are given in Fig. 1, we obtain the mean pattern of five vowels, as shown in Fig. 2. By these representations we know that the five vowels may be classified roughly into two or three divisions: a group of "I" and "E"; a group of "A" and "O"; with "U"

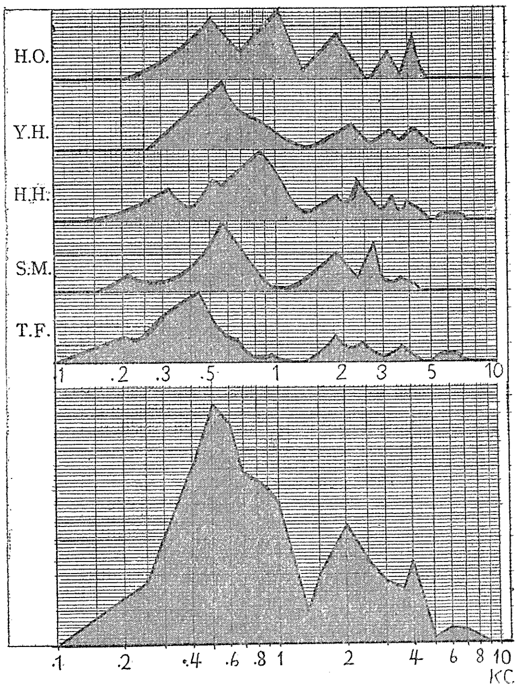
* We were encouraged on this point by an idea proposed by the discussion group in the Society of Science of Speech and Sound.

** Five oral vowels employed are: "A" (ah), "I" (ee), "U" (ōō), "E" (eh), "O" (oh).

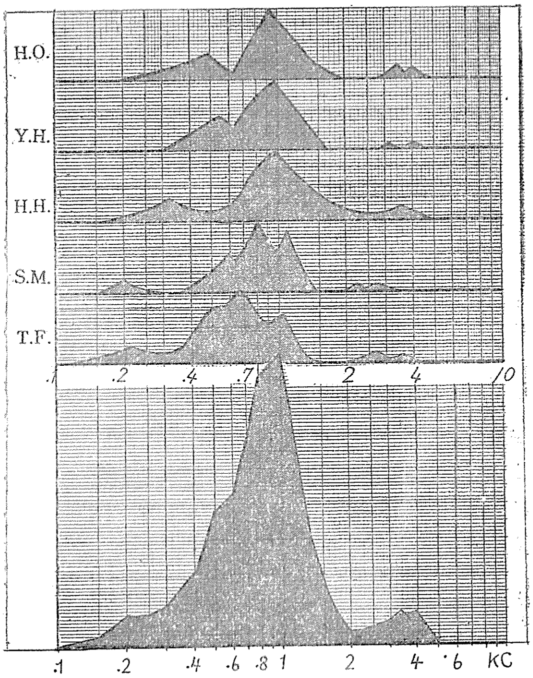
Vowel "I"



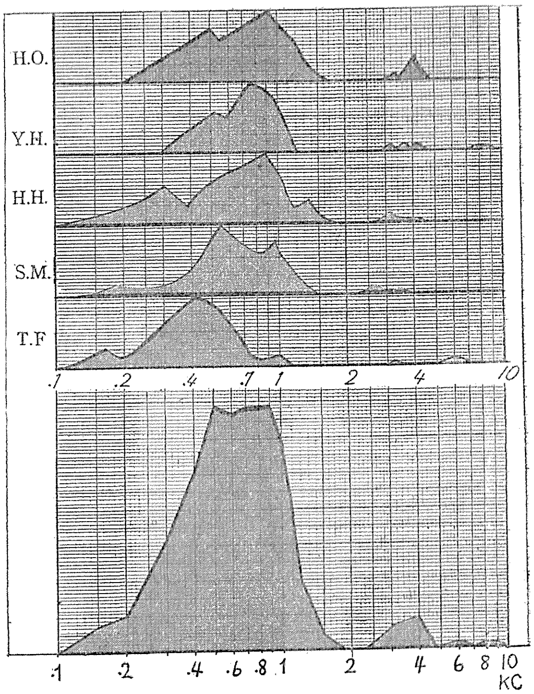
Vowel "E"



Vowel "A"



Vowel "O"



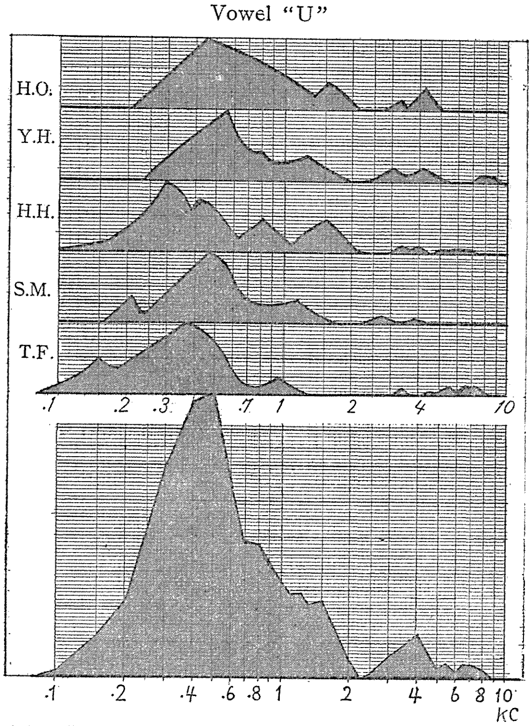


FIG. 1. Five Japanese vowels, averaged as to five voices.
Subs.: T. F. (♂), S. M. (♂), H. H. (♀), Y. H. (♀), H. O. (child).

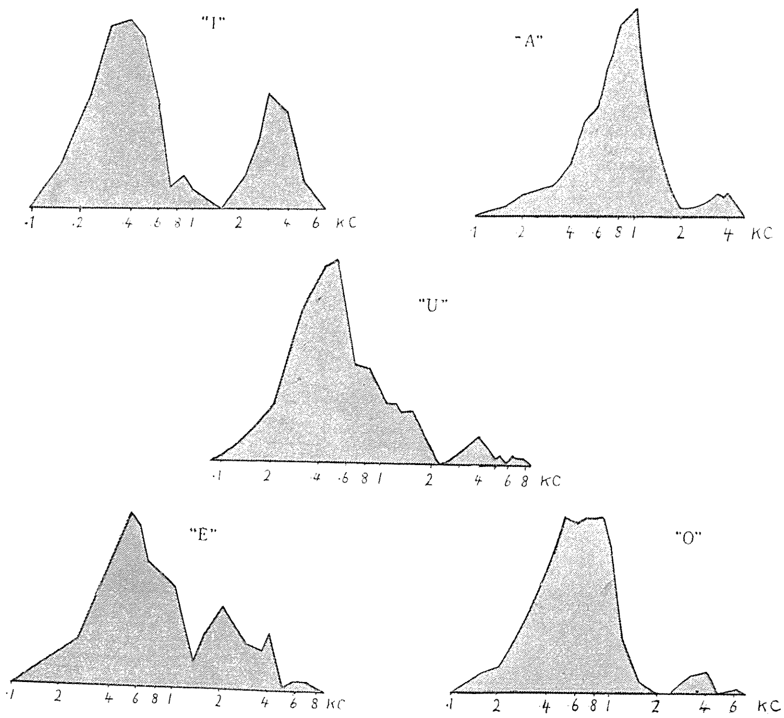


FIG. 2. Vowel patterns in convenient arrangement.

classified either in the third group, or attributed to either of the first two groups. The intermediate nature of "U" seems to have some resemblance to both "O" and "I". Therefore, taking "A" and "I" patterns as representative of the two groups, and further by modifying the detailed parts of the curves which are not important, we obtain the two modified patterns shown in Fig. 3, which we have named "A"- and "I"-conformer respectively. When we speak of only two cases of "A" and "I" conformers we do so because the most conspicuous types are the vowels "A" and "I" which stand in the greatest contrast. The question arises when do the frequency-responses approach these two conformer types which are deflected from standard evenness. It is no longer subdued to the principle of FAITHFUL REPRODUCTION; in other words the peak-type system works so as to emphasize the group of the vowel "A", and the valley type to emphasize the group of the vowel "I". The shapes in the pattern thus finally attained are shown in Fig. 4 (*a*), where

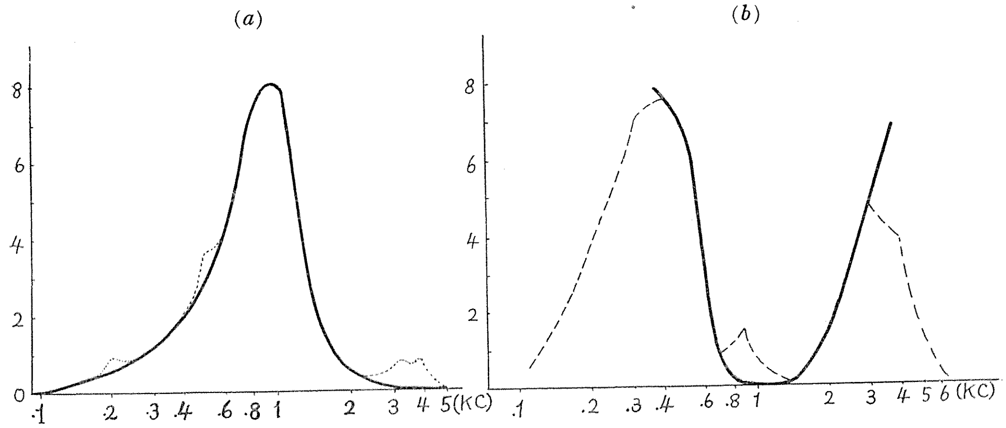


FIG. 3. Typical conformer, (*a*): "A"-vowel conformer, (*b*): "I"-vowel conformer.

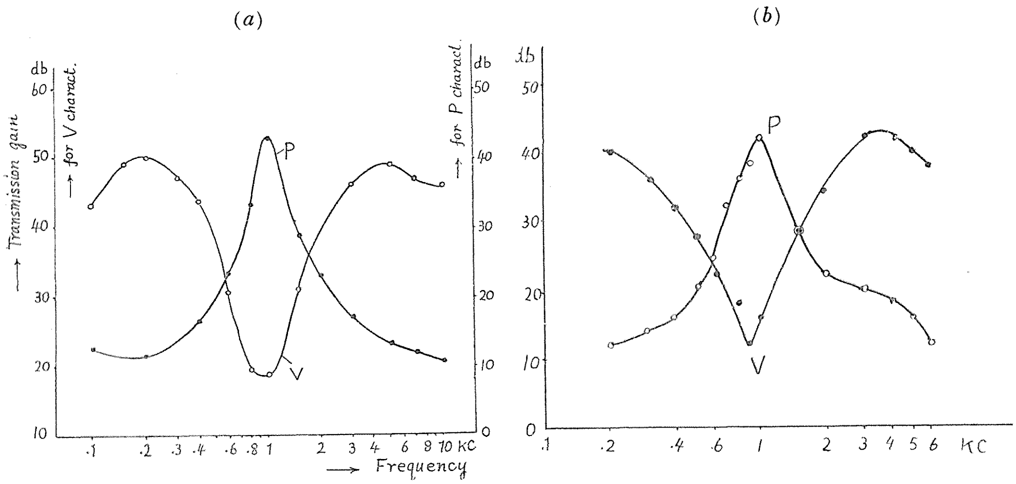


FIG. 4. (*a*) Exact forms of P and V characteristics measured in the electrical part of transmission. (*b*) Exact over-all forms of P and V characteristics measured including the acoustic transmission part due to loudspeaker.

the characteristics of peak and valley are emphasized still more than those of conformers. The reason is that the width of the transmission band grows gradually narrower, accordingly it is easier to make the comparison of peak and valley characteristics when the difference between these two characteristics and the flat is found great enough. The sharpness of these two characteristics are of the order of about 2 in number of selectivity. More in detail, P is in 2.1 and V in 1.8. Fig. 4 (b) gives the over-all characteristic forms of total electrical transmission system inclusive of the acoustical transmission part of the loudspeaker.

Choice of band

In view of the knowledge of the timbre structure of vowel and voice, it seems very important to set up the conceptions of band width as well as of band position concerning the so-called transmission band. Therefore, in choosing the transmission band of the systems in question, it is needful to take into consideration these two conceptions.

The number of combinations of width and position is so great that we cannot follow out an experiment that satisfies all the conditions in every conceivable case. For the purpose of actual experiment, we are compelled to reduce the number of combinations that can be investigated practically.

Here what we are trying to study is the combined case of constant position and variable band width. We have chosen the form of frequency characteristics those approximately the same as "A" and "I" emphasizers, the band position selected at 1,000 cycles as its center. The selections of band width must be carried out so that their mid-frequency of boundary product coincides with the 1,000 cycle region. But for the convenience of variable filters placed at our disposal, we have selected the cutting bands as shown in Table 1. In Fig. 5 we show the transmission characteristics of filters employed here.

Combination of distortions

By selecting the types of characteristics, and further, by choosing the band widths of transmission band, we finally obtain nine cases of distortion which are tabulated in Table 2 and illustrated in Fig. 6. These nine cases are obtained from a combination of Fig. 4 and 5. It is reasonable that these two types of distortion are so absolutely independent that the order of arrangement can be changed without bringing about any influence upon quality consideration. From the actual process of this experiment, we can judge that the distortion of the level-attenuation is a supplementary, subsidiary distortion.

TABLE 1. Data on Transmission Band

Band signature	Cut-frequency		Mid-frequency	Band width	
	F_1 (c.p.s.)	F_2 (c.p.s.)		Frequency difference (c.p.s.)	Frequency interval (P in O)
N	600～	2,000～	1,100～	1,400～	1.74
M	250～	3,200～	890～	2,950～	3.68
W	150～	5,000～	890～	4,850～	5.05
R	70～	10,000～	850～	9,930～	7.06

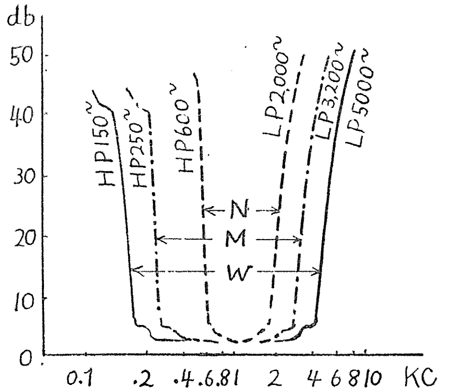


FIG. 5. Characteristics of HP and LP filters used, forming BP of the three transmission band widths.

TABLE 2. Combination of Distortions

$D_2 \rightarrow$ / $D_1 \downarrow$	*W*	*M*	*N*
P	*PW* (P_W)	*PM* (P_M)	*PN* (P_N)
F	*FW* (F_W)	*FM* (F_M)	*FN* (F_N)
V	*VW* (V_W)	*VM* (V_M)	*VN* (V_N)

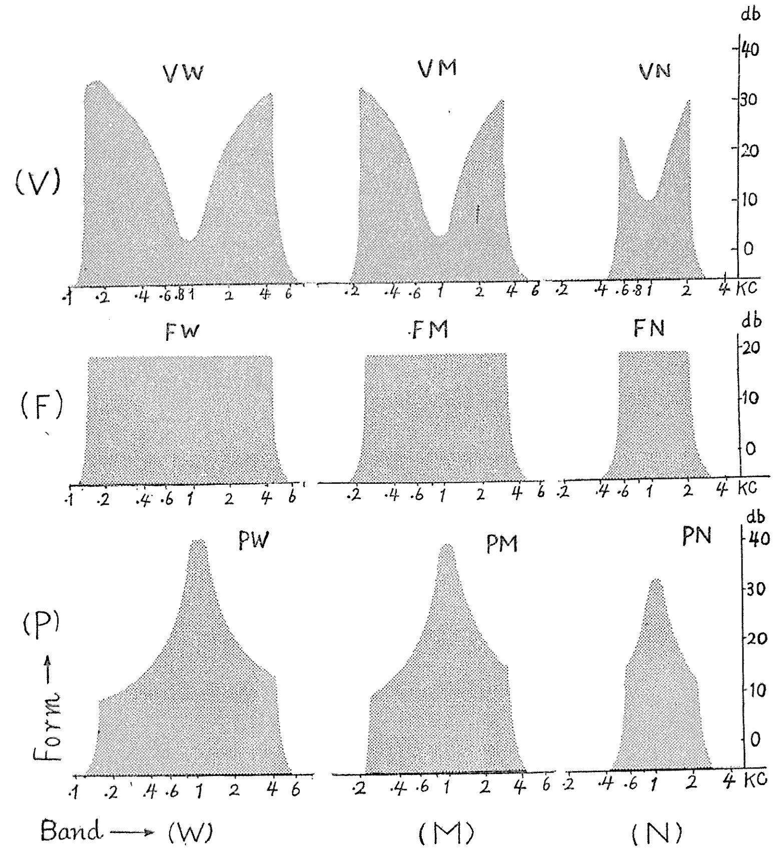


FIG. 6. Nine combinations of distortion derived from the three band restrictions and the three form characteristics.

Loudness Matching

This is one phase of quality study called LOUDNESS TRANSMISSION. At first we studied it passively merely as a preparation for timbre study. But we soon arrived at the conclusion that this field of loudness is applicable to an increasingly wide variety of phonal problems. It is for this reason that we made an effort to study it more closely and more particularly than is demanded for our immediate need. Needless to say, this part of our study is nevertheless far from a state of thoroughness and perfection, knowing as we do that there are so many interesting problems remaining untouched in this branch.

This is a summarized, final report of fragmentary studies published earlier in Research Report, some of which were re-examined and revised, in addition to which there are certain supplementary studies.

Introduction

In the nature of things, the measurement of loudness balancing of the pure tones with different pitches is not carried on without effort. The problems of loudness balancing between complex sounds of different timbres are necessarily difficult enough even when they happen to have the same pitch. Still more difficult might be the problem of loudness matching between two systems which are different in their transmission characteristics. This is justly called the problem of loudness transmission. By going further along this way, we are persuaded that the problem of loudness transmission is very interesting and important for its own sake. Because there is a scarcity of good studies touching on the loudness of actual complex speech-sounds, we believe such a study is vitally needed.

Fundamentals in loudness matching

The idea which underlies the loudness matching of transmission systems is reduced to the fundamental principle of loudness transmission which is:

$$\boldsymbol{H}_l \cdot \mathrm{D}(X) : \boldsymbol{H}_l(X).$$

If, instead of $\boldsymbol{H}_l(X)$, we can adopt the expression $\boldsymbol{H}_l \cdot \mathrm{D}_0(X)$ where D_0 implies some convenient reference system, we are able to evaluate the loudness effects of different transmission systems D_1, D_2, D_3, . . . , as shown by the following expressions

$$\begin{aligned}
&\boldsymbol{H}_l \cdot \mathrm{D}_1(X) : \boldsymbol{H}_l \cdot \mathrm{D}_0(X),\\
&\boldsymbol{H}_l \cdot \mathrm{D}_2(X) : \boldsymbol{H}_l \cdot \mathrm{D}_0(X),\\
&\boldsymbol{H}_l \cdot \mathrm{D}_3(X) : \boldsymbol{H}_l \cdot \mathrm{D}_0(X),\\
&\ldots\ldots\ldots\ldots\ldots,\\
&\ldots\ldots\ldots\ldots\ldots,
\end{aligned}$$

which enable us to attain the comparative method of loudness appreciation of systems. This is the practical process of loudness matching.

Some remarks on this problem

The study of the problem of loudness matching as derived from the fundamental idea of loudness transmission must have as its object the locating, by means of observations on speech signal, indications of transmission systems as to their faculty for reproducing loudness. Appreciation of the qualities of system is the main object; observation of speech signal is a subsidiary aim. Here comes into question primarily the faculties of carrying the speech-sounds as a complete whole. The problem of loudness balancing of each individual speech-phone is of secondary importance. Therefore, we usually used continuous speech as the signal, and isolated vowel-sounds were used in special cases only. The difference in band widths and the variety of over-all responses of transmission systems bring about some considerable changes not only in the loudness but also in the timbre of speech-sound, having hardly any influences on the pitch.

In resorting to subjective judgment, we must consider that the measurement of loudness is subject to fluctuation because of the differences in timbres, a point on which we must reflect further in connection with any study on the method of measurement. A comparison of operator control method (OCM) and subject control method (SCM) is carried out in this respect, as is also the duration time or the time of presentation of signal discussed for the same reason.

Transmission level and system level

When we speak of loudness matching of the three systems of peak, flat, and valley forms, we mean loudness equating by means of the adjustable attenuators with which each system is provided, called here SYSTEM LEVEL ADJUSTER (SLA). There is a further level shifter connected with the final reproduction level of speech signals which we call TRANSMISSION LEVEL SHIFTER (TLS). As this latter level is the final level of speech sound over the transmission system, we can conveniently call it speech level (SL) or voice level (VL). But this level has not the slightest connection with the uttering or pronouncing level itself.

Loudness matching by continuous speech

Experimental procedure

The circuit employed is diagrammatically shown in Fig. 7. The combination of distortions *W*, *M*, *N* and *P*, *V*, *F* can be selected freely by switch, SLA implying system level adjusters, in which the variables are inserted in *P* and *V*, and the fixed in *F* only. In the method of OCM, the SLA of *P* and *V* are manipulated by the hand of the operator, while in SCM, SLA is placed in a sound-proof room to be adjusted directly by listening in. The real circuits which come into use for making *P*-, *V*-, *F*-characteristic are shown in Fig. 8.

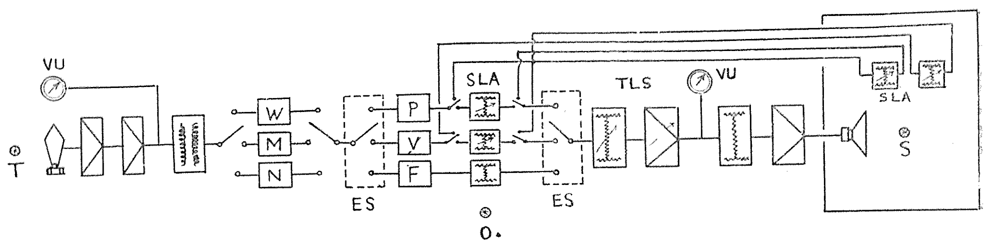
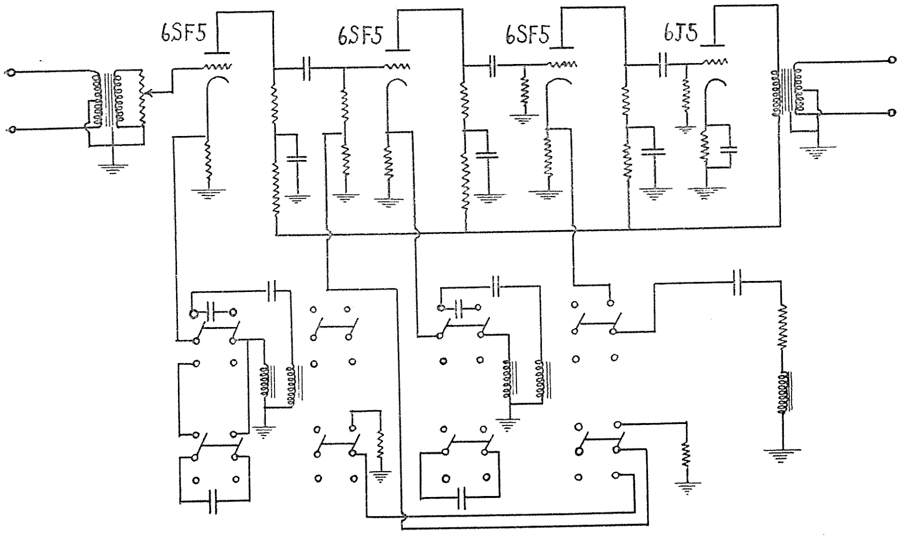


FIG. 7. Schematic diagram of subjective measurement of loudness matching of *P*-, *V*-, *F*-system. ES: Electronic switch. SLA: System level adjuster. TLS: Transmission level shifter. T: Talking subject. S: Listening subject. O: Operator.



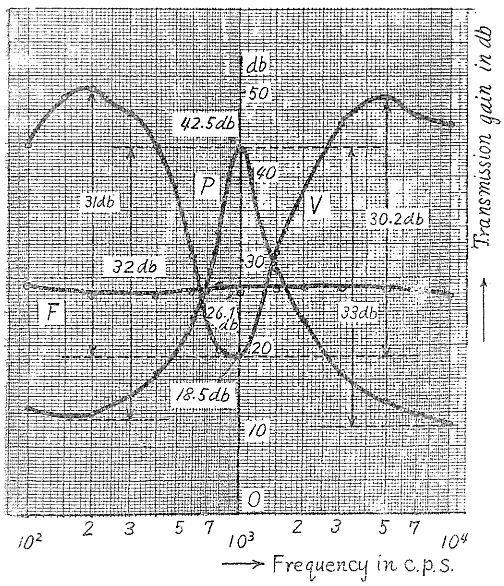
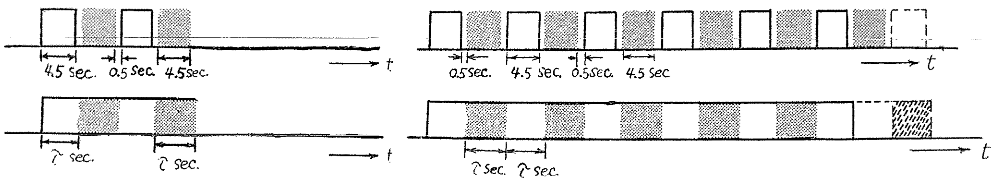


FIG. 8 (*a*) (*upper*). Circuit of *P*-, *V*-, *F*-characteristic. Switch thrown in upper side, circuit makes *P* charact eristic, switch thrown in lower side, makes *V* characteristic, switch off makes *F* characteristic.

FIG. 8 (*b*) (*lower*). Exact relation in level of relative position of *P*-, *V*-, and *F*-system.

For continuous speech, the living—not tape-recorded—voices of two different subjects were used, both reading some handy essay at an appropriate speed and at a distance of 50 cm from the microphone. The speaking subject is requested to keep his voice-level as constant as he can, by watching the VU meter in front of him. The speech signals are given to the listening subject who is seated at exactly 1 m from the loudspeaker, conforming to the pattern of presentation shown in Fig. 9 (*a* and *b*); (a) corresponds to the method OCM where the pair of conditions is repeated only twice during which the listener must complete his greater-or-smaller judgment; and (b) corresponds to the method SCM where the repetition of pairs



(*a*) Presentation of signal in the method of OCM.

(*b*) Presentation of signal in the method of SCM.

FIG. 9. Presentation of continuous speech for loudness matching of any pair of conditions. *uppers*: anterior test by mechanical switch with constant duration. *lowers*: posterior test by electronic switch with variable duration.

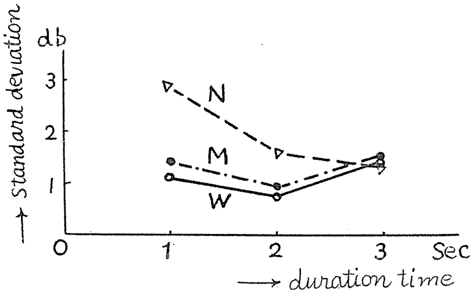


FIG. 10. Standard deviations of the results in loudness matching, related to duration time of signal presentation.

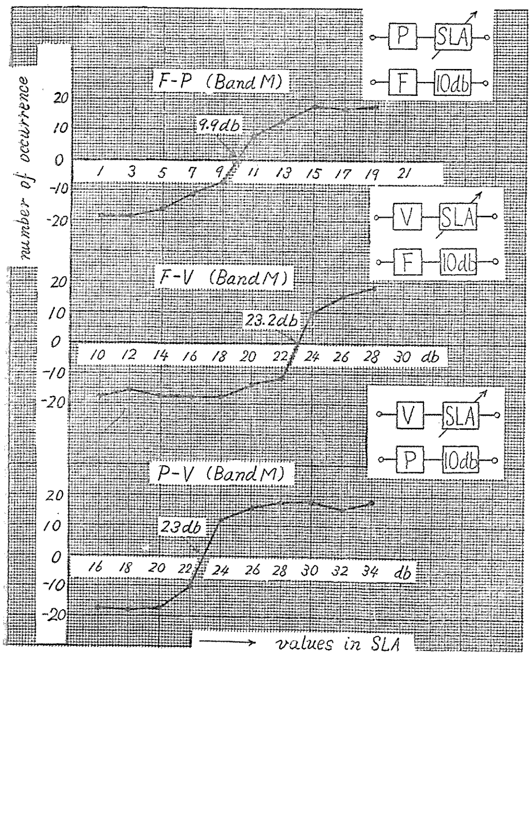
of condition is not limited; the number is left to the listener's discretion, until he feels he has finished both his judgment and adjustment; the upper patterns in Fig. 9 show the mechanical switch system (MSS) in anterior test where the duration of presentation is fixed at only 4.5 seconds, manipulation carried by the hand of the operator, the lower patterns show the electronic switch system (ESS) in posterior test in which time duration of presentation is variable and manipulation is automatic. The relation between the value of duration time and the deviation of measurements is given in Fig. 10. By this diagram we know that the suitable duration of presentation of continuous speech signal is 3 seconds or more. Any shorter duration than this, irrespective of the band width of transmission, is not suitable.

Comparison between OCM and SCM

To test the expediency of the two subjective measurements, OCM and SCM, in the loudness matching as to the sounds with different timbres, we investigated the three systems *P*, *V*, *F* in the intermediate band width *M*, at the medium transmission level *m*. The *m* level in these experiments is taken as the referential level which almost corresponds to 57 db re. 0.0002 dyn/cm^2. Three subjects are selected, all young males. They come into play in this manner: one, uttering; the other two, listening; they take their turns in these roles. Thus one round is six sittings in all. As to the judgment OCM, the greater-or-smaller judgment must be given by the listener, no equal judgment being permitted. Of course this is the so-called *AB* test. Type of switching is MSS. By method OCM, we test three combinations, *P-F*, *V-F*, and *P-V*. In each case the system level adjuster (SLA) which is inserted in each system is adjusted conveniently by the hand of the operator and all that the listener must answer is "greater" or "smaller" on the sounds of the conditions under test in a comparison with the sounds of the referential condition during the time of repetition of the pair of conditions. The result is plotted in Fig. 11, where the positive direction of ordinate represents the number of the occurrence of vote "greater", and the negative direction is the number of the occurrences of vote "smaller"; meanwhile the abscissa gives the values of SLA at each step for which the foregoing judgment is requested. The cross points of the curves in the abscissa are called the points of loudness-matching which points correspond advantageously to the parts of greatest steepness. Thus we have the results of OCM.

The method SCM in the same condition of M band and m level is also carried out. This time, the adjustment of SLA placed in a sound-proof room with listening subject, is left to the manipulating hand of the subject who continues to listen to the succession of pairs of stimulus until he believes that a matched condition is duly attained. The results obtained through a total of 18 observations are shown in Table 3.

By comparing the results of OCM and SCM, we can infer that the results remain practically unchanged because of the difference in measuring methods. As for the time needed in these measurements, SCM needs only about one-third the time of OCM. Hence we hereafter employ mainly the method SCM, exceptional cases apart. When we speak of "exceptional" cases, we mean those cases of loudness matching in a condition where there are very great differences in timbre. In order to account for the difference in timbre, we calculated the standard devia-



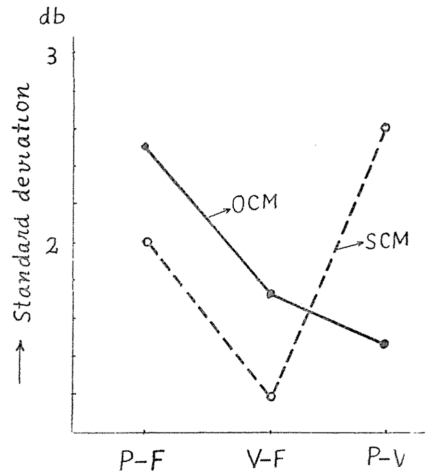


FIG. 11 (*left*). Data on loudness matching of systems by OCM in transmission band M. Method: OCM. Condition: On voice level m, in band width M by MSS (4.5 sec. duration).

FIG. 12 (*right*). Standard deviations in methods OCM and SCM.

TABLE 3. Results of Level Matching by Two Methods

Method→ / Condition ↓	SCM	OCM	Mean	Difference
P-F	9.4 db	9.9 db	9.65 db	0.5 db
V-F	23.4 db	23.2 db	23.30 db	0.2 db
P-V	22.2 db	22.9 db	22.55 db	0.7 db

TABLE 4. Standard Deviations in Two Methods

Condition→ / Method ↓	*P-F*	*V-F*	*P-V*
OCM	2.55 db	1.75 db	1.49 db
SCM	2.00 db	1.20 db	2.60 db

tions in our measurement as shown in Table 4 and plotted in Fig. 12. As for the variances, we can find considerable differences in the two methods. The facts of these phenomena are:

(1) for *P-F* and *V-F*, SCM is preferable to OCM;

(2) for *F-V*, variances are of lower grade, notwithstanding the difference in method;

(3) in the combination *P-V*, variances in SCM are greater by far than in OCM.

On the other hand, in a strict sense, introspection of the listening subject indicates the fact that the dissimilarity of timbre between two successive signal-pairs is brought out most remarkably in *P-V*, and the resemblance of the two timbres becomes most effective in the *F-V* combination.

If we suppose that the dissimilarities in timbre cause difficulties in loudness-judgment and, further, that by variances in loudness data we can infer the degree of difficulties in such subjective measuremnt, we can suggest that SCM is more desirable than OCM, at least where the timbre resemblance is called into being.

Comparative study of loudness matching and volume matching by continuous speech

As we have already pointed out, loudness matching is a purely subjective method for estimating transmission systems considering only their loudness effect. There is available another method rather of a physical nature, so to speak, based on an application of VU meter, hereafter called VU or volume matching. In a comparative study of VOLUME MATCHING (VM) and LOUDNESS MATCHING (LM) we employ as our signal both continuous speech-sounds (CSS) and isolated vowel-sounds (IVS) because we are disposed to study by volume matching the difference between CSS and IVS of speech signal. In the research where loudness matching was tried only in SCM, our studies deal entirely with:

(1) *F*-characteristic systems in different band widths,

(2) *P*- and *V*-characteristic systems in three different transmission bands.

Comparison between F-systems in different transmission bands

Circuit diagram for LM experiment is given in Fig. 13, indicating only SCM where three *F*-systems in bands *W, M, N* are studied in comparison with *R*-

system which, in the widest of all conceivable bands (75–10,000 c.p.s.) is practically the same as F-characteristic system. For the VM experiment we used the circuit shown in Fig. 14 where the operator O has charge of adjusting SLA to equate the two readings VU_2 and VU_3.

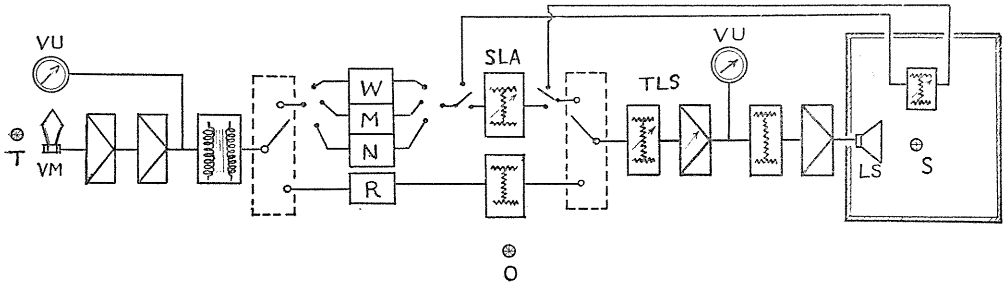


FIG. 13. Circuit diagram of LM in F system in different transmission bands.

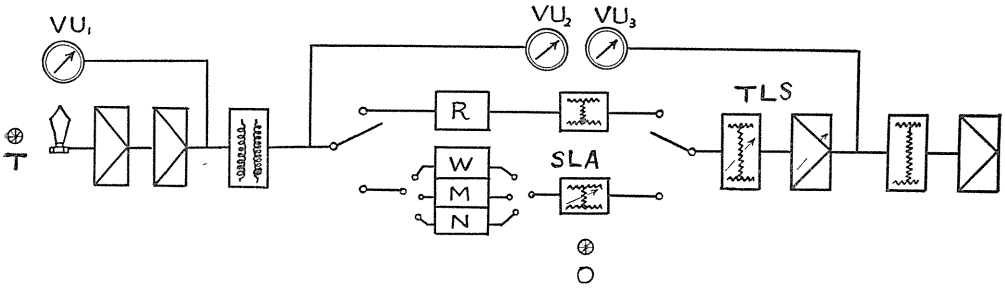


FIG. 14. Circuit diagram of VM in F system in different transmission bands.

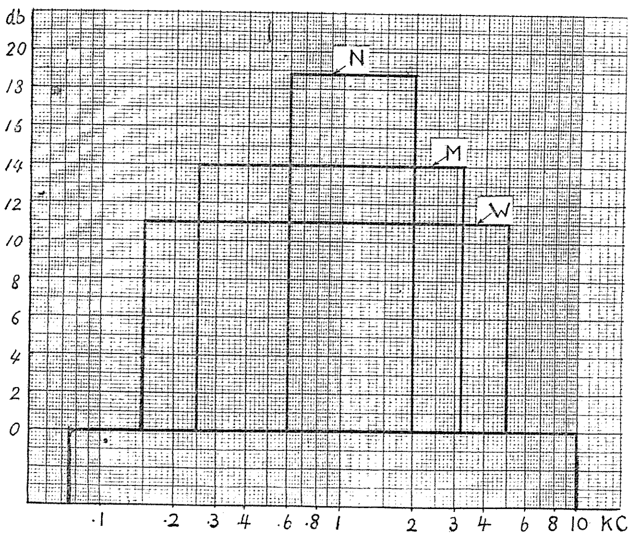
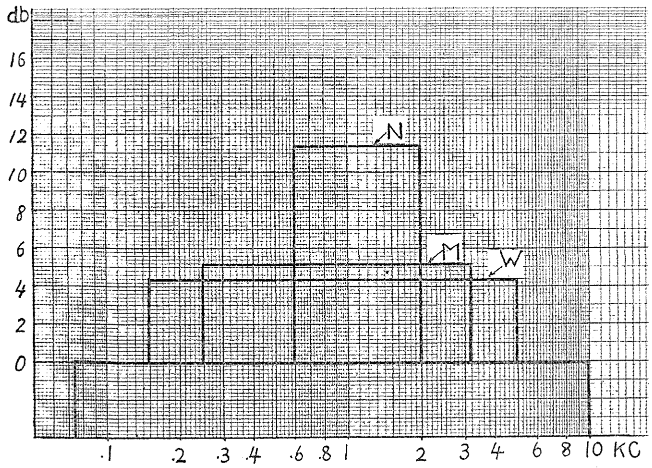


FIG. 15. Balanced positions of W, M, N band systems in loudness matching by continuous speech signal.



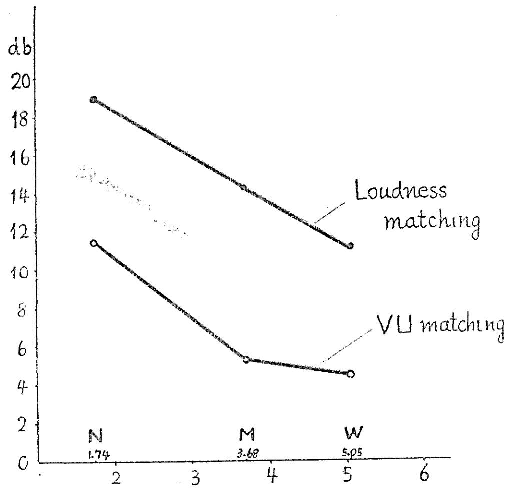


FIG. 16 (*upper*). Balanced positions of *W*, *M*, *N* band systems in VU-meter matching by continuous speech signal.

FIG. 17 (*lower*). Comparison of LM with VM concerning the level matching of flat-response *F* systems in different transmission bands by continuous speech.

The result of LM experimented by CSS at *m* transmission level is given in Fig. 15, and that of VM by CSS in Fig. 16. When we use another representation, we have Fig. 17 where the abscissa shows the transmission band width in *pitch in octave* and the ordinate indicates the system levels for matching both LM and VM. This figure shows the following tendencies:

(1) Incremental shifting of each system level needed for balancing caused by narrowing of the transmission band is generally greater in LM than in VM;

(2) Level shift for loudness matching stands inversely to transmission band width, becoming smaller as the width increases;

(3) No exact reversal proportion was found in the characteristics of VM in our experiment.

The aforementioned facts hold good for *F*-characteristic systems, so far as they were tested by continuous speech-sounds at medium transmission level.

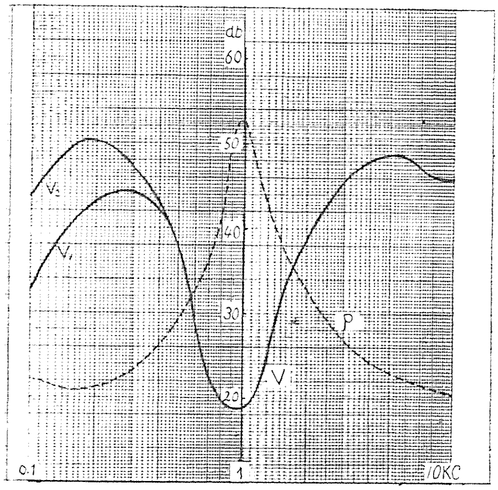


FIG. 18 (*a*). Exact form of *V* characteristics. V_1: in anterior test, V_2: in posterior test.

Comparison between LM and VM in both P- and V-systems in different bands

We will find the results of this experiment summarized briefly in Fig. 18, showing two sets of balanced positions of matching by both LM (in solid curve) and VM (in dotted curve) in three transmission bands *W*, *M*, *N*.

Experimental results:

(1) Loudness matching (LM), in large measure, produces different results from volume matching (VM).

(2) The difference between LM and VM is more conspicuous in *P*-systems and less in *V*-systems.

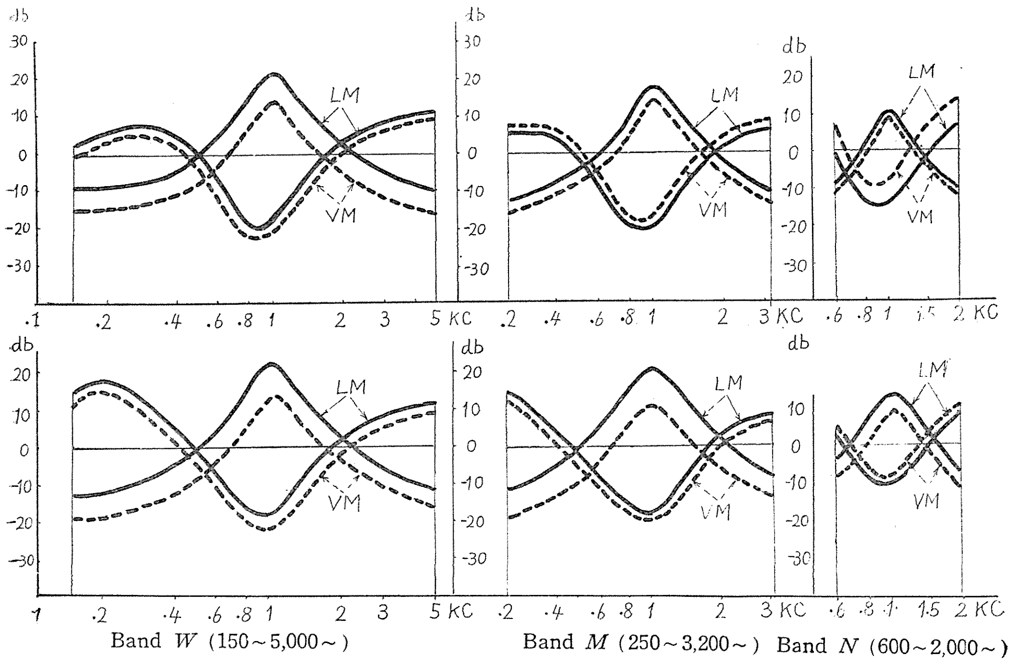


FIG. 18 (*b*). Comparison of LM with VM by means of continuous speech as to *P* and *V* systems in different transmission bands (*W*, *M*, *N*), on the transmission level *m*. Uppers are the results of anterior test; Lowers those of posterior test. Only the form of *V* is little different between these test cases.

(3) The difference in *P*-systems due to two different matching grows greater as the transmission band becomes wider.

(4) The greatest difference between LM and VM is of the order of about 10 db in these experiments.

Studies of VM and LM by isolated vowel sounds

The problem of level matching between different speech phones is important and interesting. We discussed it fully in our treatise on timbre analysis of vowel sounds. Here we present this problem from the aspect of volume matching only. There seem to be so many precautions needed to dispose of VU meter when the differences in transmission systems come into question. We require a study of two cases which may possibly be reduced to actual problems. Finally, we will bring the phenomena of IVS into relationship with that of CSS. In other words, we will illustrate the differences between continuous speech-sound method and isolated vowel-sound method.

Study of VM in F-systems in different bands

Circuit diagram used is the same as Fig. 14. The results are shown in Fig. 19, where the abscissa gives the positions of vowels, the intervals of which are chosen most convenient for our use, and the ordinate gives the shift of system level needed for volume matching, the parameter being taken as transmission band width. The shift level for matching is zero in the all-pass system *R*, and in proportion to the narrowing of the transmission band the attenuation to be decreased gradually increases when the vowel "I" is taken as the center. The arrangement of vowels is intended to show that *F*-system with its narrower band comes to resemble *P*-system, by a sharpening of the figures as shown in Fig. 19. Further, in this diagram we show that the matched levels obtained by CSS and designated as l_n, l_m, and l_w, correspond to *N*, *M*, and *W* bands respectively.

Study of VM in P- and V-system in different bands

Circuit used in this test is given in Fig. 20, and results shown in Fig. 21, illustrating the following facts: The reversal nature of *P* and *V* becomes more marked as the transmission band widens. The matching levels obtained by CSS are also

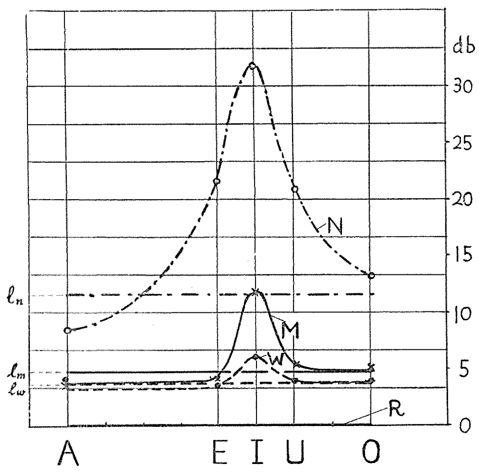


FIG. 19. Volume matching (VM) of isolated vowel sounds (IVS) in *F* systems in different transmission bands.

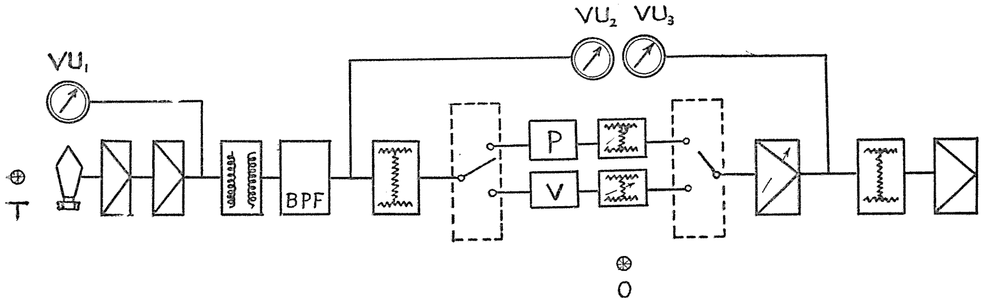


FIG. 20. Circuit diagram of volume matching (VM) of *P, V, F* systems with respectively *N, M, W* transmission bands by means of isolated vowel sounds (IVS).

designated here, given as l_{pn}, l_{pm} and l_{pw} and l_{vn}, l_{vm} and l_{vw}, respectively. These values come into close agreement at mid-position of "A" and "O" in *P*-characteristic systems, and approximately the same as "I" in *V*-characteristic systems, which is reasonable when we select *P*-system as almost the same as "A"-emphasizer, and *V*-system as "I"-emphasizer.

Comparison between LM and VM in F-systems in different bands

By employing the discrete three vowel sounds "A", "U", "I", we have tested the difference between LM and VM in *F*-systems in three different band widths. The circuits for the experiment were the same as shown in Figs. 13 and 14. Fig. 22 shows the result of our experiment. The ordinate expresses the shift of system level for the matching of different band systems. In these diagrams we see in vowel "I" almost no difference between LM and VM, but for vowels "A" and "U" there are considerable differences between LM and VM. These differences are greater accordingly as the band width becomes narrower. We can find the same tendency in all the five (Japanese) vowels. As the band width decreases the vowels have a greater effect on the shift of system level for matching.

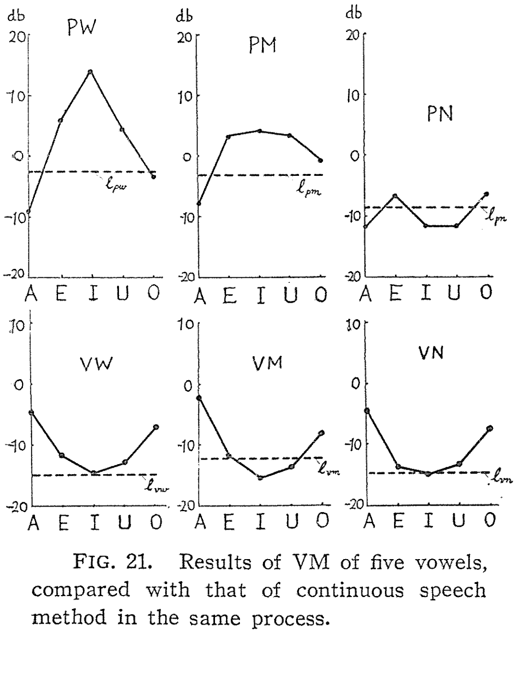


FIG. 21. Results of VM of five vowels, compared with that of continuous speech method in the same process.

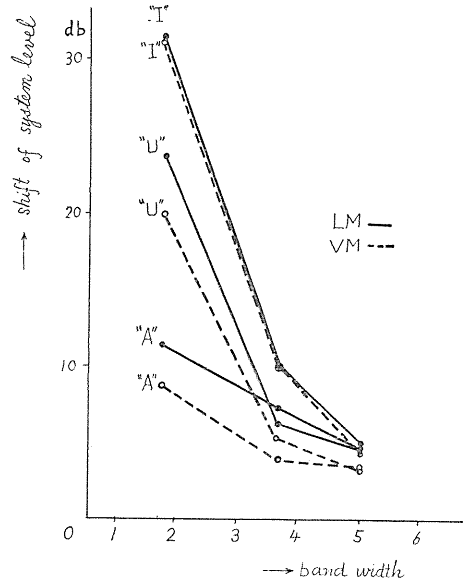


FIG. 22

Discussion

As the volume unit is not equal to the loudness unit, it is not strange that the result of loudness matching of the systems does differ from that of volume-matching. Also from the standpoint of theoretical quality transmission, it is more suitable to adopt loudness matching. It is needless to say that loudness-matching by continuous speech-sounds is preferable to that by isolated speech-sounds, in-so-far as it is necessary to consider the summarized effect of loudness transmission.

We must refer to the two cases of loudness-matching, *i.e.*, the matching between *F*-systems in different bands and the matching between systems of different forms but in the same band width. By combining these, we are in a good position to deal with loudness matching between any two systems irrespective of both form and band, setting aside all measurement difficulties. Measurement trouble happens where timbre sounds differ. Our experiment warns us to adopt, in such cases, a method of indirect rather than direct comparison, choosing and utilizing the intermediate state and thus avoiding a difficult comparison. For example, we can compare *P*-system with *V*-system by indirect process through a comparison of two combinations such as *P-F* and *V-F*. According to the degree of difficulty of measurement, we can make the best use of the so-called OCM with the expenditure of a much longer time needed for measurement.

We added some final remarks, redundant as they may be, on a procedure for obtaining the loudness-matching adopted here. We have expressed loudness matching in terms of the shifted value of system level required for loudness-matching. By expressing the level matching in terms of system level we have obtained the order of loudness by means of the relative magnitude of level of the system concerned as it relates to the reference level of the referential system. Naturally this means that there is no absolute measure for loudness. This procedure is based exclusively upon the idea of loudness transmission. The fact that the loudness evaluation of system is of the first importance must be kept in mind.

Quality Appreciation in Timbre Aspect

Introduction

The term "quality appreciation" in the heading, of course, means the appreciation of systems with respect to timbre quality. Compared with loudness quality, timbre is a still higher quality. Therefore, the methodology with loudness quality is not always applicable to timbre quality. For instance, in spite of having tried LOUDNESS MATCHING (or volume matching) as measure of quality appreciation in loudness transmission, we cannot easily introduce the concept of TIMBRE MATCHING.

The timbre appreciation introduced here is chiefly an appreciation of articulation aspect, a one-sided point of view, not all-inclusive but considered suitable for estimating the purely intellectual information as found in commercial telephone-conversation. It is difficult, therefore, to forecast whether or not the articulation quality is, in fact, fit for either treatment or detection of the delicate quality-difference owing to the change of characteristic form. We approached this problem by improving somewhat upon the process of articulation measurements.

Experimental Procedure

Circuits and characteristics

The circuit in this articulation experiment is shown in Fig. 23 where the speech signals are taped voices supplied by tape-recorder equipment. Circuit for speech recording is shown in Fig. 24. The transmission characteristics of velocity microphone (*a*), tape-recorder (*b*), and loudspeaker (*c*) used here are shown in Fig. 25.

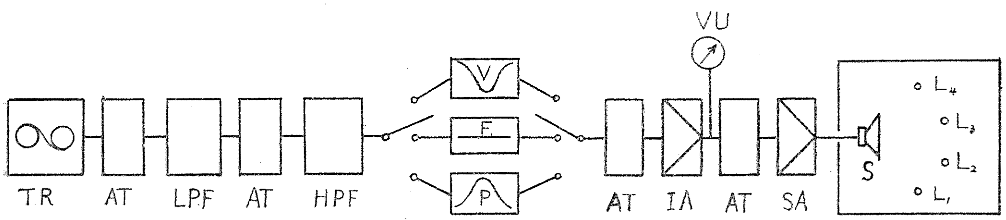


FIG. 23. Circuit diagram for articulation study of *P*, *V*, *F* systems in different bands at various transmission level.

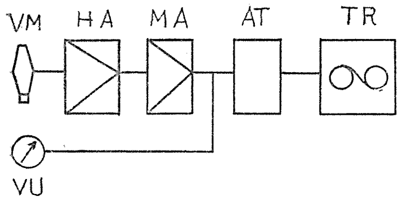


FIG. 24 (*upper*). Circuit for speech recording.

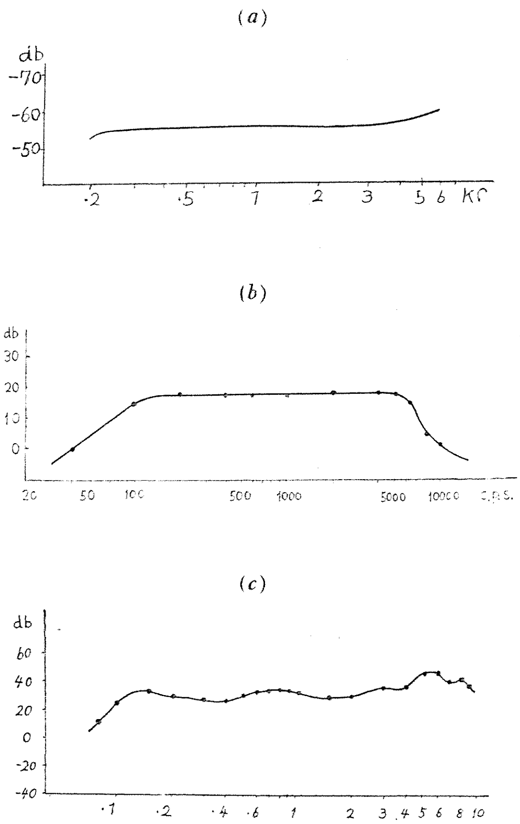


FIG. 25. (*a*) Frequency response characteristic of velocity microphone. (*b*) Frequency response characteristic of tape-recorder. (*c*) Frequency response characteristic of loudspeaker.

Experimental processes

As talking subjects, a young male and female* voices are selected. Loudness matching is traced from their continuous speech. The articulation test is made by means of articulation logatome composed of syllable-unit sounds. As listening subjects, four male students with normal hearing are employed.

Recording of speeches was made at the distance of about 50 cm in front of a velocity microphone. Level of uttering is about *mezzo forte*. Listening was carried on in front of a high-quality loudspeaker at the distance of about 1 m in the sound-proof room. Drilling for uttering and listening was carried on for about six months, considered a sufficient time for our experiments.

Condition crossing method

The important factor of this experiment consists of the finding at the medium transmission level m of the difference of articulation quality caused by changed conditions due to the combinations of three transmission bands N, M, W, with three characteristic forms, P, V, F. As the subjective measurement is generally susceptible to fluctuations in the so-called practice effect founded mainly on the subjective condition of listening crew, it is most important to reduce the practice effect to that exact extent which counts for little when compared with the deflection of quality responses due entirely to the change in testing conditions.

For that reason we adopted our crossed-condition method previously proposed by one of the authors. This method has now been improved and developed by S. Saito for the purpose of finer detection. To avoid any inevitable deviation caused by the practice effect, the testing conditions to be compared must not be so arranged that any one of the conditions is studied over a long interval of time, during which the study of any of the other conditions was completed, while the scores of subjective measurement were gradually increasing. A series of testing conditions must be studied successively after intervals so short that they do not permit of any perceptible increase in scores.

According to our classical syllabary, 100 syllable-unit sounds of CV construction are collected in one logatome as an almost complete whole of the Japanese language phone system. So that the listener may remain unaware of occurrence of only one phone at any one time in one logatome, we have developed a logatome system with 200 phones, divided into four divisions of 50 phones each, between each division the testing conditions were finely crossed and randomized. We think this method efficacious enough to be free from disturbances due to practice effect, and to bring about desirable results in sensitive detection of the influences of testing conditions.

A total of 800 observations were used to determine one point of articulation characteristics. For each uttering voice, and for each listening subject, five repetitions of the 200-phone logatome test were carried out, with four listening subjects scoring in 4,000 observations; 8,000 observations of two different voices thus determined every one point of articulation characteristics.

Quality characteristics of articulation response

By taking P, V, F as parameter, we can show the relation between articula-

* Regular female anouncer in CBC (Central Broadcasting Company at Nagoya).

tion responses and transmission bands (M, N, W) in Fig. 26 (*a*), (*b*), where the transmission level is always maintained at m (medium) level, (*a*) in this figure corresponds to anterior test, and (*b*) to posterior test. As a precaution, we made two tests, the first after about a 30-day interval. It goes without saying that the process of loudness matching is executed in each transmission band. By these quality characteristics, it is evident that the quality of P system is apt to subside to the lowest position under every condition. It is difficult to decide the superiority of either the V- or the F-system. Generally speaking, however, V is found slightly better in narrower bands and F in the wider band. However, the difference in the degree of this superiority is only of the order of one or two per cent in articulation, very small values but sufficiently to give confidence because the difference is brought out by the finely crossed-condition method.

It is to be noted that the accuracy of results of timbre comparison within the same band is more reliable than that between bands because the procedure of loudness matching and condition crossing within one band is more definite. Further, we must make a test of the articulation characteristics when the transmission level is varied. We show the results of such a test in Fig. 27. Here the transmission level is changed by spacing five steps in the dynamical range of 70 db. The transmission band is assigned to the band condition M. From Fig. 27 we see P-system, for the most part, runs lower and V-system higher, with an exception on the loudest level where V is very slightly smaller than F. Level matching and condition-crossing are tested on each level of P, V, F.

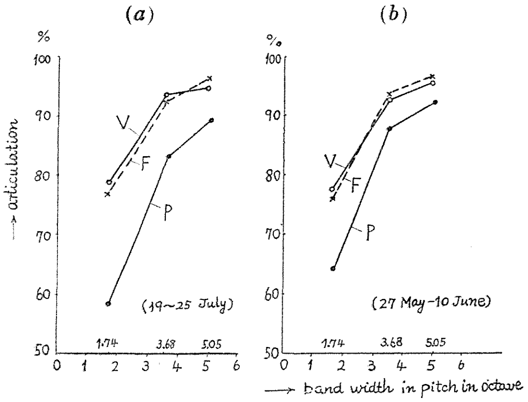


FIG. 26. Articulation characteristics of P, V, F systems in different bands (M, N, W) on the transmission level m.

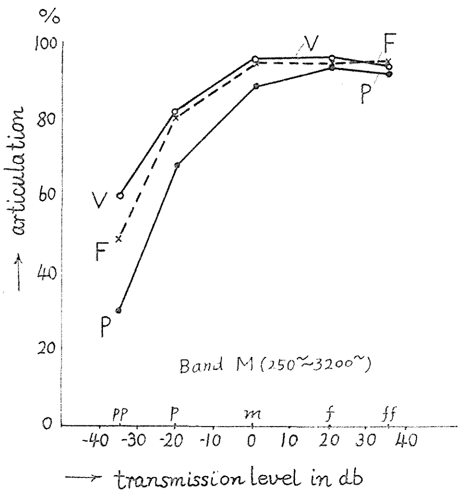


FIG. 27. Articulation characteristics related to the transmission levels in the same transmission band M. In the transmission level, zero level corresponds to the medium level m.

Practice effect

The so-called practice effect being considered as the effect of experiences under unfamiliar conditions, we must examine this effect separately under all three conditions, P-, F-, and V-form. This is given in Fig. 28, showing the development curve of practice phenomena obtained by a crew composed of four listening subjects under three conditions in the same intermediate transmission band; the abscissa means the total number of speech-phones already experienced accumulatively as an index of practice phenomena. The deviation is in general of the order of one or two per cent.

FIG. 28. Practice curve in articulation test with *F*, *V*, *P* systems.

Analysis of articulation

Finally, we have made an analytical study of articulation quality by which we can make clear not only a part of phonal character of speech-sounds but the characteristics of distortion. Syllable-unit articulation can be divided into two component-articulations, that is, vowel- and consonant-articulation. We prepared two sets of relations one of which was studied using the grade of distortion as parameter, as the direct relations of vowel- to consonant-articulation; the other studied as the

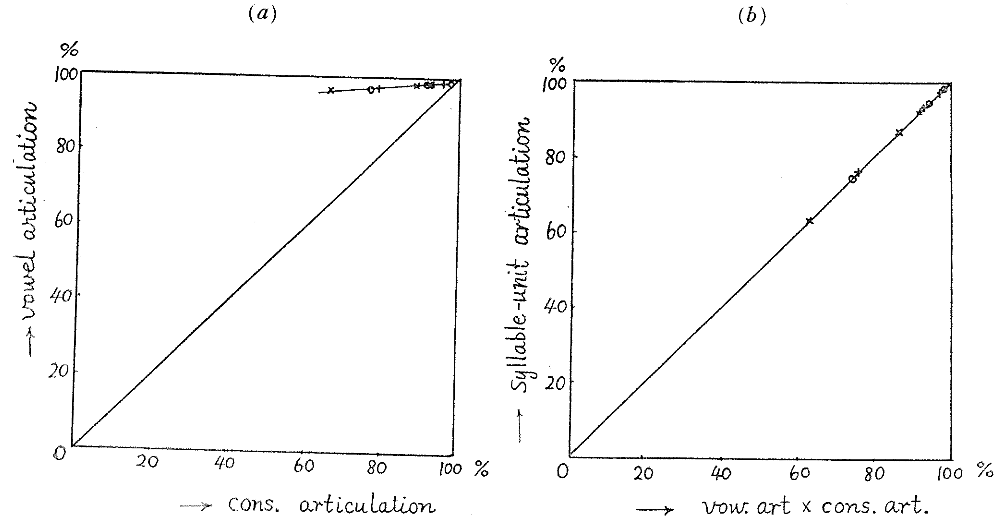


FIG. 29. Band effect at medium level.

relation between syllable-unit articulation and product of vowel- and consonant-articulation. By inspecting the former, we can see the distortion character by the proportion of contributions of each component-articulation to a complete syllable-unit articulation. By inspecting the latter relation, we can divine the degree of independency of each component-articulation, or measure of correlation between two component-articulations. Fig. 29 indicates the band width effect of transmission system upon component-articulations at medium transmission level, (*a*) corresponds to the former case, and (*b*) to the latter.

Timbre impressions of *P*-, *V*-, *F*-system

Timbre impressions of the three systems *P*, *V*, *F* are quite distinctive in character. The wider the transmission band becomes and the nearer the transmission level approaches medium level, the more distinct the character of timbre impression becomes. It is really difficult to describe the details of timbre impression in words. However, *P*-system usually gives an impression of some annoyance quite indistinguishable in spite of its full force. *V* and *F* are both easily heard, one being a little different in impression than the other. Only when the transmission band becomes narrower or the transmission level rises, does *F* begin to give an impression similar to *P*, differing nevertheless, in impression from *P*. The most interesting point we noticed in the course of these experiments is that despite some clearcut distinctness in timbre impressions between *F*- and *V*-system, we cannot sufficiently follow out the differences in articulation scores of *V* and *F* to account fully for such timbre differences. We here need another judgment on timbre other than the clearness aspect. Let us now go on to the next observation.

Test of guessing-at-condition

After a six-month experience with these tests, the listeners are sufficiently educated to discriminate between the timbre impressions of *P*, *F*, *V*. Therefore, we must consider some test for guessing-at-conditions. This certainly means a timbre judgment other than articulation. Of what quality this timbre judgment really consists is not yet certain. As there is wide range of thought, we have considered that the VOCAL QUALITY (and not the PHONEME QUALITY) is brought into play and therefore we investigated quite separately the effects of the male and female voices. In Table 5 we show the result of the guessing-at-condition for the male voice with the judgment based on 60 observations for each condition. Every outstanding features leading to the misjudgment of the three transmission systems in three transmission bands are ingeniously represented in Fig. 30. By this representation, we can know positively at what condition and to what extent each of these three conditions in each three transmission band is guessed at. *P* has fewer occasions to be heard amiss than any other, *F* is heard primarily as *V* and secondly as *P*, meanwhile *V* tends almost always to be confused with *F*. In Fig. 31 we show in (*a*) the detailed character of right guessing and in (*b*) misguessing. From the latter figures we know the degree of discrepancy in the reciprocity of misguessing measurement of 30 observations per condition. We show in the same manner in Table 6 the result of condition-guessing with the female voice, this time based on only 14 observations per condition. As the trend of reciprocity seems to hold good here, we cannot infer that the number of observations is too small. By picking up the results of right guessing on male and female voices we can draw the figures as shown in Fig. 32. From that fact we can point out that the linear trend of right guessing toward the increase in transmission band (expressed in *pitch in octave*) is found only in *F-F* judgment.

TABLE 5. Condition Guessing on Male Voice

W band		Judged condition F (%)	P (%)	V (%)	*M* band		Judged condition F (%)	P (%)	V (%)	*N* band		Judged condition F (%)	P (%)	V (%)
Given condition	*F*	91.7	0	8.3	Given condition	*F*	70.0	1.7	28.3	Given condition	*F*	61.4	18.2	20.4
	P	0	100	0		*P*	0	100	0		*P*	6.8	90.9	2.3
	V	6.7	0	93.3		*V*	35	0	65		*V*	20.4	2.3	77.3

TABLE 6. Condition Guessing on Female Voice

W band		Judged condition F (%)	P (%)	V (%)	*M* band		Judged condition F (%)	P (%)	V (%)	*N* band		Judged condition F (%)	P (%)	V (%)
Given condition	*F*	100	0	0	Given condition	*F*	92.9	0	7.1	Given condition	*F*	78.6	7.1	14.3
	P	0	92.9	7.1		*P*	0	100	0		*P*	0	92.9	7.1
	V	7.1	0	92.9		*V*	21.4	0	78.6		*V*	14.3	7.1	78.6

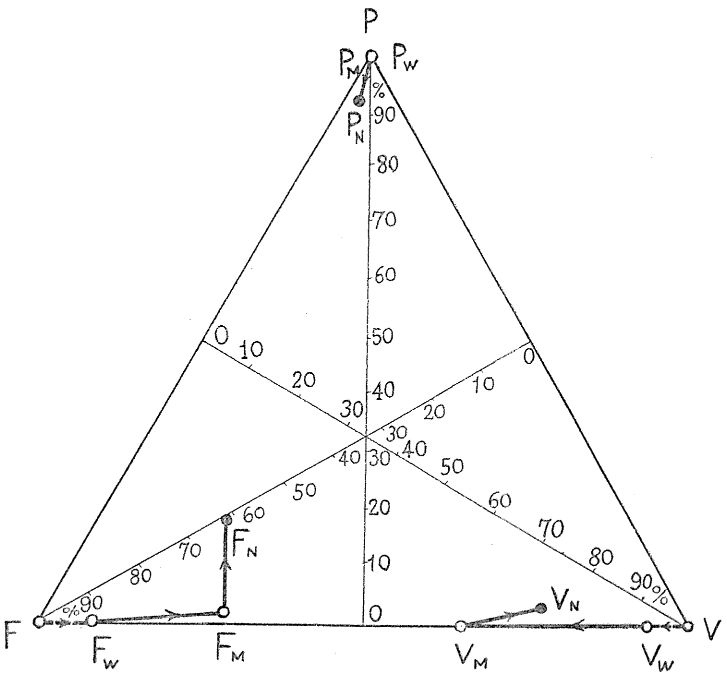


FIG. 30. Triangular representation of the result of condition-guessing in the case of male voice as a measure of timbre discrimination of *P*-, *V*-, *F*-system in different transmission bands.

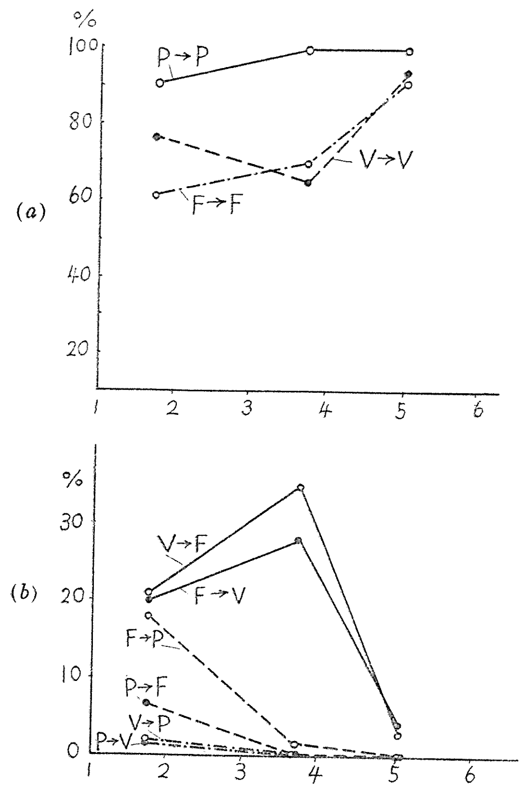


Fig. 31. Right judgment (*a*) and misjudgment (*b*), in condition-guessing in *P*-, *V*-, *F*-systems with a male voice.

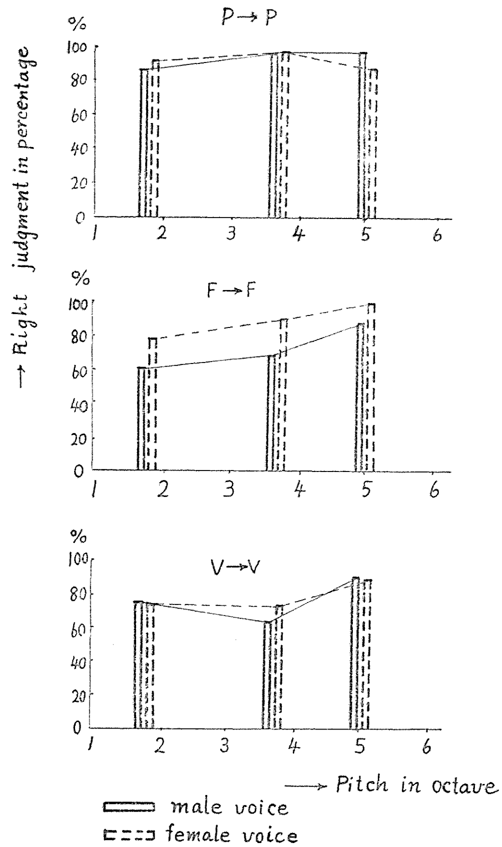


FIG. 32. Right judgment characteristics of *P*-, *V*-, *F*-systems with male voice, compared with that with female voice.

Some experiments on naturalness and articulation of vocalic voices

In addition to the guessing-at-condition test, we have tried some direct experiment on naturalness of voices by using only the sustained five vowels of five subjects. As we have launched the measuring method of naturalness, and we have already performed the naturalness experiment on several distortions, we will have occasion to give a report on it in detail. Here we describe briefly only the outline of an experiment of this kind. Fig. 33 shows the experimental block diagram, and Fig. 34 shows the way of presentation of signals where it is to be noted that both the beginning and end of a succession of time in waves of sustained vowels have been omitted, leaving stable parts only about four seconds, as recorded for greater simplicity on the tape-recorder following the processes of volume matching (instead of loudness matching) and pitch matching, leading thereafter to the listening subjects through the distortion circuits under test. Some of the results are illustrated in Table 7 where the calling pitches selected are 140~ and 240~, and the band conditions selected are only two: one, band unrestricted; the other, band-pass of 400–3,000 c.p.s. We do not intend to conclude anything from these data. We are content with the fact only that the naturalness of personal voice can be detected by a method so strict that it reflects nothing but the vocal quality in timbre construction in almost pure sense.

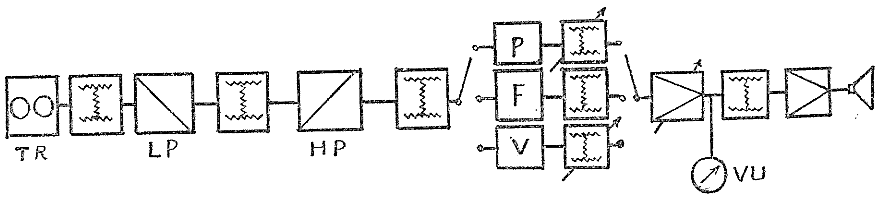


FIG. 33. Block diagram of the experiment of naturalness measurement.

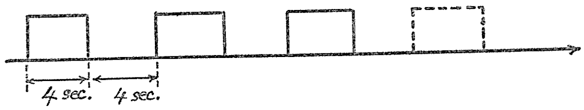


FIG. 34. Presentation of vowel signal in the naturalness experiment.

TABLE 7. Result of Naturalness and Articulation

Calling pitch	Band condition	Articulation in %			Naturalness in %		
		P	*V*	*F*	*P*	*V*	*F*
140 ~	Band unrestricted	88.0	96.0	98.9	84.8	90.7	91.4
140 ~	Band pass 400 ~ 3,000 c.p.s.	81.5	97.3	99.3	80.5	87.0	86.2
240 ~	Band unrestricted	85.9	89.6	96.5	68.3	81.1	83.5

Discussion

As we have already pointed out, there must be strict accuracy in-so-far as the differences in articulation quality are concerned in any comparison of articulation data between the conditions where the treatment of condition-crossing is used.

In this articulation test usually the taped voice was used. The characteristics of a tape-recorder are suitable for studying the quality in the pitch range below about 6 KC. Articulation data on this experiment was obtained by averaging the results on the male and female voices, the difference in which is beside the question. On the contrary, in the experiment of guessing-at-condition, we called into question for the first time the differences in vocal quality, and we investigated quite separately the characteristics of judgment on the male and female voices.

Finally, in any experiment on naturalness, an investigation is needed not only on the vocal difference due to sex, but on the pitch difference in uttered voices, because the vocal pattern may depend entirely on the pitches employed. A study of the performance of *P*-, *F*-, and *V*-system on naturalness is now in progress. All that we can say here is that quality studies of characteristic form are insufficient for tests from an articulation standpoint only. As an indispensable quality assessment on timbre, naturalness measurement in addition to articulation measurement must be reflected.

General Discussion

Now that we have discussed in detail the items at the end of each section, we here make only some final comments on quality transmission from a collective viewpoint. In order first to re-examine the steps in abstract thinking by referring to the results of experimental scrutiny, and, further, to show how to interpret the results of actual experiences by consulting the processes of rational thinking, we are surely in need of some summarized discussion which will aid us to focus such divergent considerations. When the problems we meet become too complex and too confused, we must demand for deeper study a higher view which can provide us with a broader field of vision.

If the so-called flat response of frequency characteristic is really the most excellent for communication, all discussions will come to naught, and even the proposition or suggestion of a discussion on such a problem will be equally senseless. It is a well-known fact that flat response as one of the distortionless conditions cannot be doubted from the PHYSICAL FIDELITY viewpoint, on which the traditional transmission-engineering and -science rests. It is, in fact, only from the standpoint of SUBJECTIVE FIDELITY that flat response can be doubted. To propose, is of first importance; to verify is the second step, and to verify we believe that only quality assessment will be of any avail. The ultimate reference is to quality reference, to which we resort. The merit of articulation quality cannot be over-rated. In our opinion the study of articulation is insufficient for quality study under these circumstances. But even if we adopt only articulation quality, we are still in a position to suggest with positiveness that such a quality study on frequency response is of considerable significance.

All sounds can be analyzed as signals of sound transmission. Sound components can be also synthesized into the original sound. Several transmission characteristics were formerly obtained in the field of transmission engineering as a physical measure of transmission rating. It depends entirely on whether the opinion is most deeply rooted in fundamental conviction or on a foregone conclusion as to what channel there is between analysis and synthesis. That idea is only true *objectively* (physically). It cannot be true *subjectively*. The most difficult and most fundamental thesis in quality theory is that analysis and synthesis cannot be communicative. As the quality response is the response of systems which include in themselves a subjective system, all conditions for distortionless transmission which arise from physical fidelity must be subject to verification from the subjective standpoint.

Although we could successfully detect the quality difference caused by the difference of forms such as *P, F, V,* the problem of determination of the optimum form of frequency response is a most difficult one. We cannot but think that the optimum form depends on one hand on the transmission band width and on the transmission level of the system on the other hand. When we resort to the method of cut-and-try, we will need to do many hours of labor. Instead of cut-and-try method in circuit design, we must use some rational method derived from the theory of quality transmission.

In the expression of

$$\boldsymbol{H}_t \cdot \mathrm{D}_1 \cdot \mathrm{D}_2(X) : \boldsymbol{H}_t(X) = \Gamma,$$

to obtain maximum quality, we must endeavor to find the form of R_f as D_1 for the given band B_w as D_2. But here we must admit two qualities; one, articulation, the other, naturalness. We must therefore make every effort to obtain two maximum forms. The expressions

$$\boldsymbol{H}_{t_1} \cdot \mathrm{D}_1 \cdot \mathrm{D}_2(X) : \boldsymbol{H}_{t_1}(X) = \Gamma_1,$$

and

$$\boldsymbol{H}_{t_2} \cdot \mathrm{D}_1 \cdot \mathrm{D}_2(X) : \boldsymbol{H}_{t_2}(X) = \Gamma_2,$$

must become our maxima. Here Γ_1 represents specific articulation, and Γ_2 specific naturalness.

The principle of conformer or emphasizer type of the circuit design is but a primitive step in the transmission scheme founded on the idea of quality transmission. The transmission performance of communication system must be matched with the speech quality. Therefore, the system design for speech communication must lie based on the knowledge of speech response characteristics. System is not essentially irrelevant to speech, that is, system is not to be designed so as to transmit its signal no matter what the nature of that signal is. System must be designed so as to express most efficiently the effect of speech quality.

Finally, we wish to add some words on the progress of this study. The original and general project of this research was begun five years ago. The experiment on loudness balancing was first carried out four years ago by Masanobu Watanabe, and Shuzo Saito, graduate student. The next experiment on articulation measurement was partially in charge of S. Saito, and Akinori Izumitachi, special student, with the assistance of many other students. The actual measurement of naturalness was originally and only recently carried out by Teruo Fukumura, who also took charge of final checking experiments for discussion. In addition to the executers of the experiments, Mr. Tohru Yamashita also took part in the discussion.

Acknowledgement

We express our deep appreciation to the Nippon Telegraph and Telephone Public Corporation for their assistance and the use of their equipment and to the Society of Science of Speech and Sound for their encouragement.

References

1) Y. Ochiai and T. Fukumura: Studies on Qualities of Speech and Voice by Timbre Distortion. Memoirs of the Faculty of Engineering (M.F.E.), Nagoya Univ. 4, 2, November, 1952.

2) Y. Ochiai and T. Fukumura: Timbre Study of Vocalic Voices. M.F.E., Nagoya Univ. 5, 2, September, 1953.

3) M. Watanabe: Denso-tokusei no Katachi to Onsei no Onryo-seigo tono Kankei ni tsuite. (On the Relation between the Nominal Forms of Transmission Characteristic and the Loudness Balance of Transmitted Speech Sounds.) (Thesis for the Degree Bachelor of Engineering, 1950, Nagoya University).

4) Y. Ochiai, S. Saito and M. Watanabe: Denso-onsei no Onryo-seigo ni tsuite (On Loudness

Matching of Transmitted Speech Sounds), 1st Report. M.F.E., Nagoya Univ. 4, 1, 1951. 2nd Report. M.F.E., Nagoya Univ. 4, 2, 1951.
5) Y. Ochiai and S. Saito: Loudness Matching. M.F.E., Nagoya Univ. 4, 1, 1952.
6) S. Saito and Y. Ochiai: Loudness Matching. M.F.E., Nagoya Univ. 4, 2, 1952.
7) Y. Ochiai: Transmission of Quality (M.F.E. in press).
8) Y. Ochiai and T. Yamashita: On Timbre Quality (unpublished).
9) Y. Ochiai: Reflection on the Basis of Transmission in Communication (unpublished).