STUDIES ON QUALITIES OF SPEECH AND VOICE BY TIMBRE DISTORTION

YOSHIYUKI OCHIAI AND TERUO FUKUMURA

Department of Electrical Engineering

(Received October 31, 1952)

Epitome—*This is some study of trial concerning what is called transmission quality, that is, quality study of speech sound by distortion. As distortion, only one is selected; of very singular nature so as to give an unexampled effect as well to its timbres as to its pitches. As speech sounds, we selected also only one; Japanese speech phones of open-syllable type based upon its classical syllabary. Confining thus the type of distortions and the kind of languages respectively to only one, we have contrived, in return, to widen a field of vision on quality and refine at the same time our observation on it. Not having given up, therefore, the articulation study as traditional one, a field of mishearing phenomena was developed further as a sort of measure concerning the informations on subjective timbre. We have moreover provided the means by which we could enter far into the study of another kind of timbre treatment basing upon the aspect of naturalness. It comes in contact with the tolerance problem of distortion. In addition to these, we have made a new departure by carrying a pitch- and loudness-experiment of speech sounds, and covering thus a wider range of quality subjects. As to the distortion of this kind, the consideration and measurement on pitch-attribute are indispensable. We have ascertained it with an actual example of distorted pitches a part of which is determinable only by subjective comparison method. We were also enabled to deepen the idea of "pitchedness" by examining in full the vowels in phonated and whispered utterance, the voiced- and unvoiced-consonants as well in phonated as in whispered condition. In the treatment of mishearing, we have made some trial of sound classification which would enable us to contribute something to a sort of psychological specification of speech sounds. In short, this specific distortion can remind us of the very rôle of voice element in speech quality, that is, the timbre problem viewed from the side of naturalness, the pitch problem of speaking voices, the sonority problem of consonant phones; these were all neglected perfectly as yet without making themselves fall under our notice.*

CONTENTS

Introduction

It concerns us much to study the transmission qualities of communication systems which hold therein usually several kinds of distortion even in various degrees. The articulation quality, for instance, was already thoroughly investigated and reported as regards the ordinary distortion systems with which very often we fall in, such as, attenuation, band restriction, and some kinds of noisy circumstances. But, there still remain some specific distortions quite untouched. It is so with the synchronous distortion in carrier communication systems. The quality studies of this distortion will bring us, consequently, some useful informations for the carrier communication engineering. It lies beyond our comprehension that we have done without enough knowledge of quality about the system with which we could for the first time provide us the sheer communication circuit in its true meaning, that is, "channel circuit" in the wide band of frequency domain. But we shall go still further now. Reflecting that this kind of distortion is quite unique in itself, we must make it serve as indispensable means of a wider and more profound study in the nature of voice and speech. We should like to call this distortion "timbre distortion." Because, it is only one that gives rise to such an unexampled effect on the most internal structure of timbre as will be explained later. It is true however that the carrier synchronous distortion, by now, looks in no way a new as such. But, we are nevertheless quite fascinated in making a systematic study of it, being encouraged in the belief that a new contribution to the science of speech and voice, even if scarce and poor, will come back into the real field of the engineering of speech and voice with some multiplied validity.

General Consideration

The purpose of our present study is to make clear the nature of speech and voice as possible by making use of the synchronous distortion in carrier system. We can state also that, it is our proper purpose to clarify the character of the distortion, basing upon the transmission phenomena of speech sounds impressed. Therefore the objects of our observation are nothing but the speech sounds transmitted through the distortion, in other words, the transmission effect of the distortion by way of speech- and voice-quality: that is an indivisible effect of the transmitting means upon the transmitted things. Here ought to come into question anyway the psychological treatment of transmission effect of the distortion, besides its physical study. For covering a wider range of subjects under these circumstances, we must try to set orientation to the purpose of our quality study, without adhering too closely in vain to accustomed, old procedures. Let us think a little now about the nature of the distortion presented here.

The carrier synchronous distortion (we denote it C.S.D. hereafter) does not exert any influences upon the construction in time of speech current: speech is not affected thereby in the length of time duration of its sounds; needless to examine the speech, therefore, on the aspect of "quantity" in phonological meaning. However, this distortion can produce an important and significant effect on timbre problem. Well-known as it is, the effect of this distortion on complex sound is such that each component of it is displaced up or down by a constant frequency width which is equal to the shifting frequency owing to deviation of synchronism: the sounds that have harmonic structure in timbre or that can have some definite pitches, must subject thereby to vigorous changes; they become to lose their harmonic relation in timbre on the one hand, and give rise at the same time also to the gradual degradations of pitch on the other. Thus, comes into our questions at first the effect on the phonated (or voiced) vowels which have harmonic timbre-structure and also definite pitch sense. Next comes to our observations the effect on the voiced consonants which can not have any sense of pitch but bear nevertheless some traces of vocal element. For comparison's sake, we can employ en passant the whispered vowels which are originally of some sibilant character, not holding therein at all the harmonic structure in timbre. By way of investigation of the singular influences of this distortion upon the speech sounds of human voice, we cannot help bringing into focus at last the question of the voice itself in the most general and most basic sense. This very side of problem forces us to confront with an intimate relation between timbre and pitch, and between timbre and loudness, covering thus a wider surface of quality problem, and bidding farewell to the old, one-sided treatment of it.

Next we consider a little about a type of speech sounds used here. As objects of our observations, we adopted the unit-syllable sounds based on the so-called Japanese syllabary, that is, the phone type combined uniquely by CV construction (where C: consonant, V: vowel), including therein the lacked case of C (that is the case of pure vowels). The phone combination of *open-syllable* type seems particular, especially viewed from European close-syllable type of *CVC*. But, by utilising this open-syllable type, we take it easy to bring to light, not only the individual trend of separate component articulations (as an analysis of unit-syllable articulation), but also the most intimate question concerning the mutual relation between them. Because; the more the influences exerted by the distortion become different by degrees, according to the kind of phones and the grade of distortion, the more it becomes important to make known the correlational connection between two component articulations. This is one of the great advantages produced by adopting the simple phone system of Japanese syllabary.

In connection with the utilisation of Japanese sound system, it might be useful perhaps to lay stress on the next point. The Japanese phone system being exclusively employed in our studies, the results obtained can nevertheless persist in its general adequacy. Referring to the real kinds and the practical details of phones themselves, they might have to do with the particularities of languages used. But the universal characteristics of language phones and even the general tendencies of human voices are almost the same, notwithstanding the variety of languages and the diversity of voices. We put an end to the discussion in saying with reserve: the detailed data of quality phenomena may take part in the particularities, we are able to insist upon its general appropriateness in spite of the type of languages used.

Acoustical Attributes in Quality Studies

Before going too far into the detailed studies of quality problem, we are obliged to consider a little about the acoustical attribute of speech and voice. Because, it is possible, only the distorted state of them (speech and voice) can suggest us rightly and acutely the acoustical conditions under which they are essentially to be studied even when they have no distortion. Especially we feel it keenly in the pitch problem, but we will try here to think over the so-called three attributes of acoustics, beginning with the consideration in the most generalised manner.

We must first give the timbre item: in the study of distortion of speech sounds, the timbre problem must take thereby the top seat. Such being also the case with us, the quality studies are, for the most part, occupied by the timbre studies. Our timbre researches are assorted here in two divisions. The first: articulation quality on speech phones; The second: description about mishearing phenomena, as an indispensable observation on subjective timbre, that means, timbre received through our sense, not the timbre as an objective existence (or physical substance). As to the pitch studies of human voices, we have just set to work. To state the facts in due order, we shall show in the next the steps of thinking through which we hitted upon the idea of subjective experiment as to the pitch determination. This distortion, as mentioned above, acts on the complex sound in such a way as to produce the same displacement of each component, the harmonic relation in the timbre structure coming almost to be lost, and bringing also some obscurity to the pitch sense. This relation is illustrated in Fig. 1. It looks quite beside the question in the regular voices without distortion where the subjective difference-tones, if possible to appear, come to coincide perfectly with the fundamental component of sounds. But it is not so with the distorted voices, because the subjective difference-tones come out of position of the lowest frequency of all components. For this reason, we have tried to execute the subjective measurements of pitch in cases where there are at least the possibilities of occurrence of *multi-pitched* or *unpitched* phenomena.

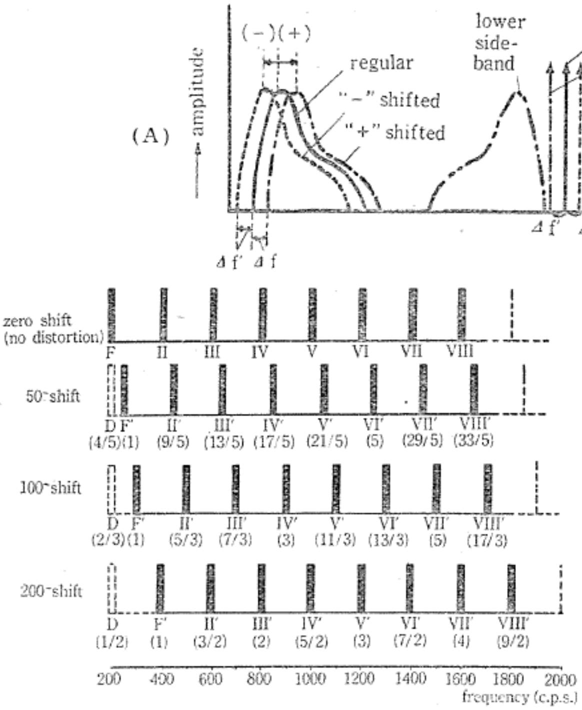


FIG. 1. Diagram for explanation of the distortion effect. (A): general displacement of voice range due to the shifting of carriers, taking place in either sides up or down according to the direction of carrier shifting. (B): effect on the internal structure of timbre, resulting at the same time in multi-pitched phenomena.

We do not think that the multi-pitched phenomenon is quite irrelevant to the deterioration of timbre. It is also under the same distorted circumstances that we come to be conscious of a vague knowledge of it, which in an ordinary case without distortion displays itself so latently that it does never fall under our eyes.

We can not neglect also the loudness effect of the speech sounds connecting with the timbre problem. To see how the dinginess of timbre by this distortion an interest in the change of loudness can bring, a series of experiment has been also executed by way of the procedure of loudness matching.

Theoretically it is in fact a most difficult problem to decide to the point the true relation among the acoustical attributes of speech phones. But we are forced to do with the actual problem concerning the determination of the tolerance (or allowance), or sufferable limit of this distortion, the experiment of which is also originally of subjective nature. The timbre must play an important rôle even in this problem, but the judgment thereby are not necessarily restricted only to the one, leaving rooms for co-oporation of other elements. As an all-inclusive study of quality problem, we have made investigation about the "torelance problem" of this distortion.

Studies on Articulation Quality

Here we call in question exclusively the quality study based upon the articulation aspect. That means, here we are confronted chiefly by the clearness of phonemical property of speech sounds. As test signals, we employ accordingly only the male voices without touching the problem of voice quality.

Outline of experiment

To realise the pure C.S.D. as possible, the intervening of any other distortions must be avoided, and we have endeavoured therefore to enlarge the transmission band of the distortion system under test as much as possible. It is for this purpose that we employ the velocity microphone and the high-quality speaker instead of the ordinary carbon microphone and telephone receiver. As the modulator and demodulator, we use an ordinary ring type set up by balanced dry rectifiers. The deviation from synchronous condition in carrier system is obtained very smoothly and continuously even to a large extent by adjustment of the detuning condenser set up in the demodulating oscillator. The regular frequency of carriers is 10 K.C. The circuit diagram of the distortion system employed here and its transmission characteristics in case of no distortion are shown in Fig. 2.

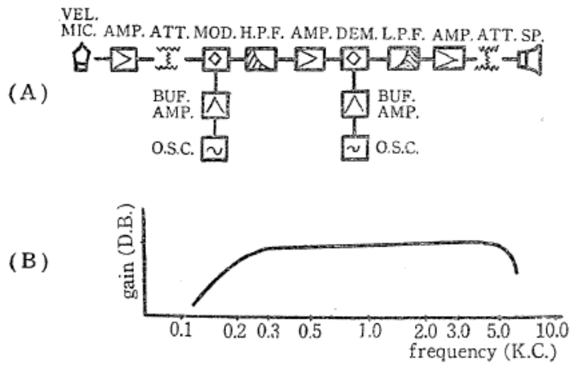


FIG. 2. (A): circuit diagram of distortion system. (B): its transmission characteristic.

The speech sounds tested here are indicated in Table 1, where 100 sounds are

TABLE 1. List of Unit-Syllable Sounds Used for Logatome, According to the Phone System of Japanese Syllabary

		Direct vowels					Glided vowels		
Series-consonants	( )	A	I	U	E	O	jA	jU	jO
	(k)	kA	kI	kU	kE	kO	kjA	kjU	kjO
	(s)	sA	sI	sU	sE	sO	sjA	sjU	sjO
	(t)	tA	tI	tsU	tE	tO	tjA	tjU	tjO
	(n)	nA	nI	nU	nE	nO	njA	njU	njO
	(h)	hA	hI	hU	hE	hO	hjA	hjU	hjO
	(m)	mA	mI	mU	mE	mO	mjA	mjU	mjO
	(j)	(jA)		(jU)		(jO)			
	(r)	rA	rI	rU	rE	rO	rjA	rjU	rjO
	(w)	wA							
	(p)	pA	pI	pU	pE	pO	pjA	pjU	pjO
	(g)	gA	gI	gU	gE	gO	gjA	gjU	gjO
	(z)	zA	zI	zU	zE	zO	zjA	zjU	zjO
	(d)	dA			dE	dO			
	(b)	bA	bI	bU	bE	bO	bjA	bjU	bjO
Simple sound-system (64 units)							Contracted sound-system (36 units)		

counted in total, including therein 64 sounds of simple character (*Chokuon* in Japanese) and 36 sounds of contracted character (*Yōon* in Japanese).

For utterance of speech sounds, two young male subjects* are employed. They have no defect in voicing and no dialect in pronunciation. They are also requested to play the part of listener. Their hearing are also quite sound. In the experiment, they are instructed to utter their voices in two manners: phonated (or voiced) utterance and whispered (or unvoiced) utterance. The distance between talker's mouth and pick-up microphone is about 50 cm. in phonated utterance, and about 30 cm. in whispered utterance.

The characteristics of articulation are obtained after the convenient periods of learning in order to avoid the deviation due to the so-called practice effect. The effect of tiredness must be also evitable to the end of obtaining the fine structure of characteristics depending exclusively on the pure distortion.

General characteristics of articulation quality

By measuring the unit-syllable articulation of Japanese phone system against two directions of the deviation of distortion, we can obtain the quality characteristics in which we try to detect the quality difference due to the difference of utterance. In order to explain the content of these characteristics, we take out the component articulations, that is, consonant- and vowel-articulation of which the unit-syllable articulation is composed.

In Fig. 3 are represented all these relations together.

Taking a view of these figures in detail, we can find out the following points distinguished by the general traits: (1) articulation quality in whispered utterance is lower generally than in phonated case, though the pick-up distance in whispered case is adjusted smaller than another case; (2) articulation qualities are not generally equal in two directions of deviation shifting. In the domain of negative shifting, the quality yields generally to that in the domain of positive shifting. We shall see it again later on; (3) as one of the general and striking traits of phonated voices, we can remark sudden depressions (like "pits") in the shoulder part of the quality characteristics, in comparison with which we can not find out the similars in case of whispered utterance. Therefore, the characteristics in

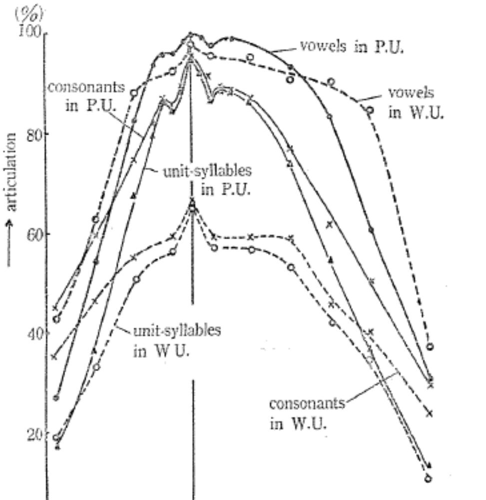


FIG. 3. Articulation characteristics by the distortion obtained in two modes of utterance, representing each component articulations.

* Articulation and mishearing studies were made chiefly by the hands of Mr. Tatsuo Nagase and Joji Kato, young engineers in the Bureau of Telecommunication of Tōkai-District, having engaged themselves in the laboratory study as special students in our Faculty.

phonated case look like somewhat "triple-peaks" type, and the characteristics in whispered case bear some resemblance to the "chestnut" form.

The "pit" phenomena in voiced articulation seems to have some concerns with the definite pitch of phonated voices (refer to the pitch study which will be noted later). Very interesting and worthy of attention as it is, it contains some original effects not reported as yet*, and we will try to describe them more in detail.

Influences due to the directions of deviation

It is to be noticed that the quality characteristics of articulation are not well-poised in two domains of distortion. It can be most evidently revealed by the representation of Fig. 4, in which the characteristics in two domains are turned up one over the another, taking the position of zero shifting as center. We can look thereby the degree of unsymmetry: the shiftings of carrier deviation that suffice to give 75% quality are respectively 485 c.p.s. in positive domain and 235 c.p.s. in negative domain; similars for 50% articulation are 750 c.p.s. in positive, and 470 c.p.s. in negative; the degree of discrepancy becomes more and more serious as the distortion increases. The speech sounds are more heavily affected by negative shifting of distortion than positive shifting of it.

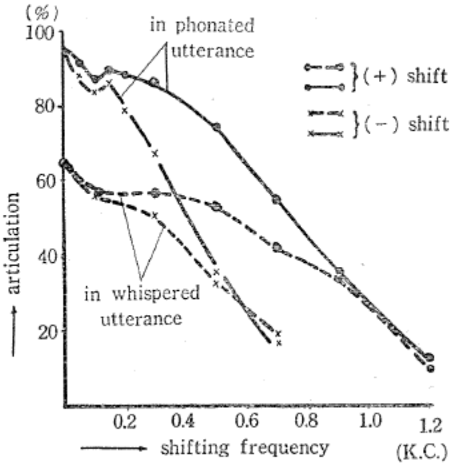


FIG. 4. Representation of unsymmetry of quality characteristics due to the direction of distortion

Influences due to the modes of utterance

So long as the distortion remains scarce, the speech sound uttered in phonated condition is able to show generally its higher qualities than in whispered case: but according as the distortion increases, the quality in phonated case runs lower inversely than in whispered case. Thus, the affairs go by contraries. It demands us to make a closer investigation and to give, if possible, a full explanation for this striking phenomenon. The elemental procedure of quality analysis represented in Fig. 3 will serve the purpose. Denote here with the following notations. $S=CV$: articulations of speech sounds in phonated utterance; $\underset{0}{S}=C'\underset{0}{V}$: similars in whispered condition. Then, S is greater than $\underset{0}{S}$ at the null distortion, $S-\underset{0}{S}$ is therefore positive. But the value $S-\underset{0}{S}$ will become by and by smaller, according to the increase of distortion. It will be quite so with the consonant articulations C and C', so far as the regions of distortion are restricted within the limits experimented here. As to the vowel articulations, the state of affairs undergoes a striking change. So long as the distortion is maintained within a definite extent, phonated vowel-quality can surpass the whispered vowel-quality. But, in course of time, the relation will go upside down: according as the distortion rises beyond a certain extent, the phonated vowel comes to give place to the whispered vowel in its quality response. It is indeed worthy of notice.

* *For example, refer to* H. Fletcher: Speech and Hearing. Fig. 146, p. 294, D. Van Nostrand Co. 1929.

Change of contribution of component qualities

The conspicuous phenomena that happened between the phonated (V) and the whispered vowels ($\underset{0}{V}$), also come about, but in a slighter degree this time, between the two component articulations (V and C) of unit-syllable quality in natural voicing. Because, it is likely to be a common phenomenon that happens almost always between sounds with pitches and without pitches. It owes in fact to the adoption of syllabary sounds of open-syllable type on the one hand, and to the trial of two modes of utterance on the other, that we can attain to the confirmation without much difficulty. To affirm more precisely these relations, we adopt the representation of Fig. 5, where the abscissa and ordinate denote respectively the consonant- and vowel-articulation, giving as parameter two modes of utterance and also two regions of distortion. The characteristics in phonated case are distinguished by becoming to intersect the bisector line of the rectangular co-ordinate system. This means: with the augmentation of distortion, the components which make a contribution to the quality loss of unit-syllable articulation appear in alternated way; so long as the distortion remains weak, the consonant part has predominantly to do with the quality change, but when the distortion becomes strong surpassing beyond a certain extent, the vowel part comes by turns to play the title rôle of quality comportment. It is to be noted that this effect is prevailed only in the distortion C.S.D. Man can not find it at all in any other distortions.

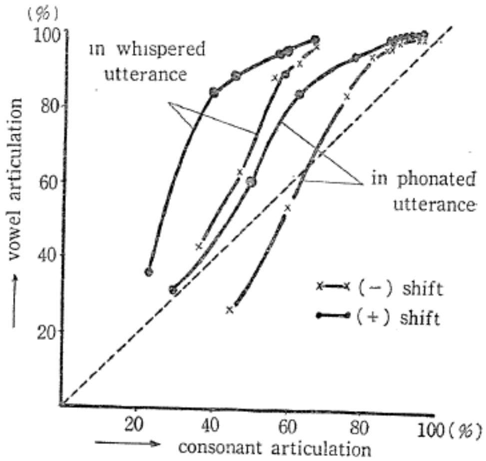
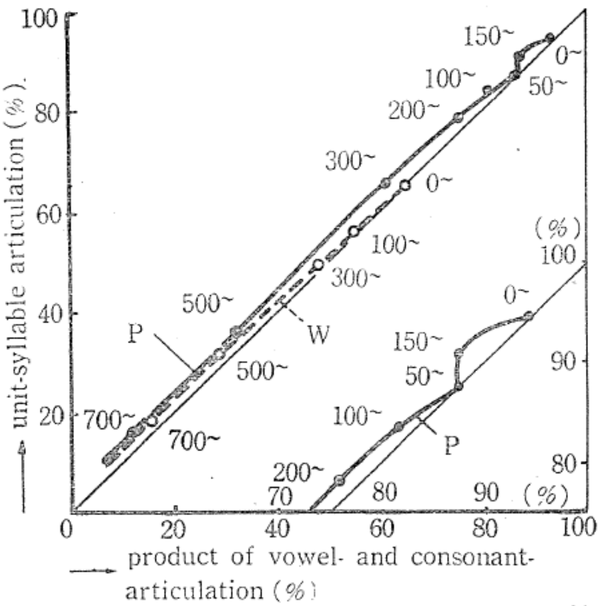


FIG. 5. Correspondence between two component articulations; consonant- and vowel-articulations.

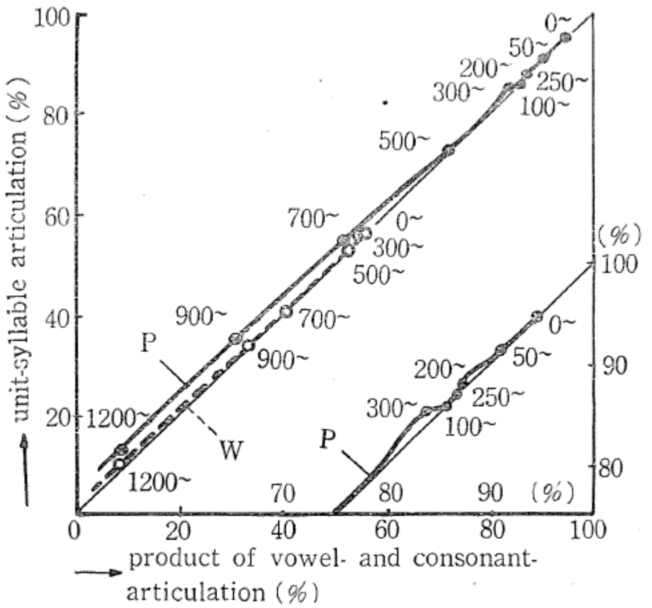
Mutual action of component qualities

As mentioned above, the vowels and consonants carry themselves quite differently and even characteristically in its quality response to the distortion. And, on the other hand, as Japanese speech sounds are reducible (almost) always to the unitary construction of CV, well-balanced in respect to the occurrence of each phone, it ought to come into question to see how it goes in reality between such component qualities: in the total course of their responses to the distortion, are they quite independent of each another or not? A mutual influence of component qualities will exist or not? This is likely one of the most fundamentals that do touch the vital point of quality problem.

For the purpose of inquiring into this question, it affords perhaps an aid to examine the linearity of correspondence between the unit-syllable articulation and the product of two component qualities. If C and V remain as quite independent events, so S will be just equal to the product of them, notwithstanding the degrees of distortion. For any other distortion than this, we can see the relation holding remarkably good, except the extreme regions of big distortion. But with this distortion C.S.D., the affair is quite another, especially in phonated case, as Fig. 6 shows. That means:



(A) "−" shifting domain of distortion, P: phonated, W: whispered.



(B) "+" shifting domain of distortion, P: phonated, W: whispered.

FIG. 6. Characteristics representing the correlational connection between vowel- and consonant-quality which compose the unit-syllable articulation; therein are shown in magnified manner some parts of characteristics of pulsatory nature at their starts.

the characteristics thereby come out of a bsiector line almost from the beginning. It is very curious and no less than instructive to find out the departure from the straight line at the start is of a little pulsative nature, as shown in more expanded way, in the same figure. Suffice it to say that this pulsatory departure from linearity (that means the pulsatory tendencies of correlation between component qualities) has some concerns with the "pit" effect of quality characteristics stated above.

Detailed characteristics of articulation quality

The "quality analysis" on articulation aspect has necessarily to come to do with the minute and particular phases of individual phone characters of speech sounds. Provided not being touched the problem of vocal characters (that is, notwithstanding the kind of voice qualities), there are so many kinds of interesting phone phenomena in our ordinary pronunciations, that the individual quality study of each phone will promise us a good yield. But we have not enough time to make an exhaustive study on it. We will give here only an outline, confining our treatment nearly to the group- or series-analysis at the utmost.

Before going into the detailed description, and for the purpose of preliminary preparation of quality analysis, we ought to explain in the beginning the manner of classification of phones used here conveniently. As to the vowel study, we classify roughly Japanese vowels into two divisions; direct vowels ("*Chokuboin*" in Japanese) and glided vowels (or contracted vowels, "*Yōboin*" in Japanese); the formers are ordinary vowels, the latters are similars but initiated by gliding or transitory parts that designate itself by notation "j" or "w." It is to be noted; what we want to call here in question is the difference of hearing on the stationary parts between direct vowel and glided one, not the adequacy of the classification itself. Referring to the consonants, we can not go into the study of sheer individual sounds. We restrict our sights only to the group- and series-classification, according to the traditional knowledge of Japanese syllabary.

Individual articulation characteristics of direct vowels

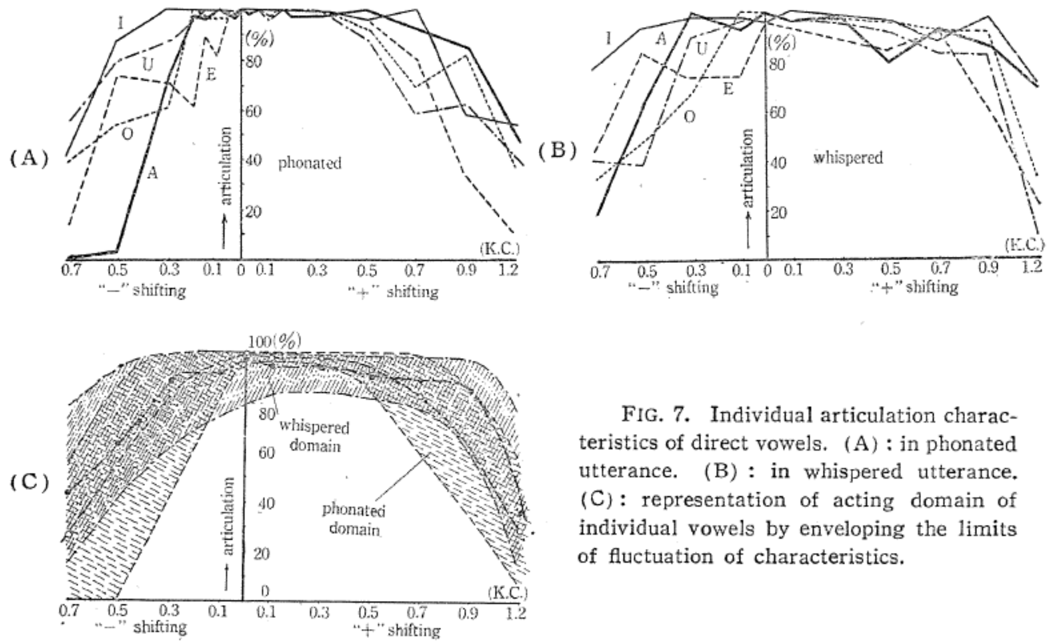


FIG. 7. Individual articulation characteristics of direct vowels. (A): in phonated utterance. (B): in whispered utterance. (C): representation of acting domain of individual vowels by enveloping the limits of fluctuation of characteristics.

The articulation characteristics of individual direct vowels are given in the Figs. 7 (A) and (B), shown separately as to the modes of utterance. Worth mentioning; the most scanty vowel in volume level is the most endurable in distortion; it is likely so with the vowel "I," and the reverse will be almost applicable to the vowel "A." Fig. 7 (C) is made by enveloping the limits of fluctuation of individual characteristics coming perhaps from the smallness of sampling. We can see therein the same relation as observed in the general articulation characteristics between phonated and whispered vowels holding also good with the enveloping domains represented here.

Difference between direct and glided vowels

Examine this time also the characters of the individual qualities but in the meaning of bearing comparison of direct vowels with glided ones. As direct vowels, we take three vowels; "A," "O," "U," and as corresponding glided ones, also three sounds of "jA," "jO," "jU."

Their individual characteristics are obtained also here as to two modes of utterance and also concerning two directions of distortion. Fig. 8 (A) and (B) show them. We can detect the most conspicuous characters as follows.

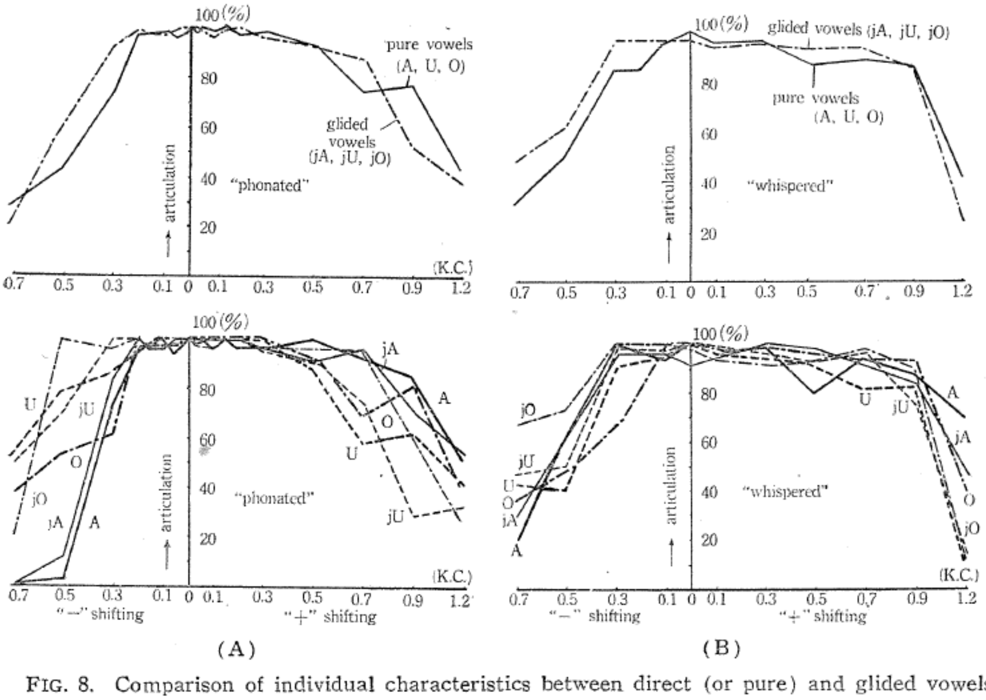


FIG. 8. Comparison of individual characteristics between direct (or pure) and glided vowels. (A): in phonated condition. (B): in whispered condition.

(1) As general tendencies, for biggest distortion merely in the domain of "+" shifting, the direct vowels seem to stand out the distortion slightly more than the glided ones.

(2) On the remaining domains of distortion, we see as the most remarkable characters the superiority of glided vowels to the direct ones. Speaking minutely,

especially in " – " domain, it is the pair of "O" and "jO," that the difference between sounds of every pair becomes most large.

(3) The degree of discrepancy of "O" and "jO" is so much that we can almost imagine as if they were the sounds not having any resemblances at all between them. If we try to search among direct vowels the very sound whose character is most akin to that of "jO," we can pick up the vowel "I" corresponding to it. Fig. 9 can show this interesting relation. Especially in the phonated condition, "jO" is nearly equal to "I" in its characteristics, and in the whispered condition, "jO" is slightly inferior to "I" in either domain.

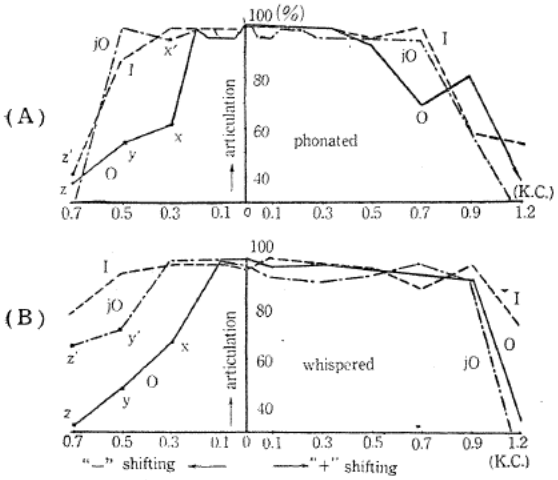


FIG. 9. Representing the resemblance of characteristics between jO and I. (A): in phonated. (B): in whispered condition.

But even here, it is not altered that "I" is more akin to "jO" than any other vowel.

(4) We can infer finally: the initial, transitory part of the phone "jO" is playing definitively the important rôle in perception of its stationary part. It corresponds perhaps to one of the rarest cases that suffice to verify the existence of mutual effect of component qualities on which we have touched above.

Quality characteristics of series-consonants

Now we ought to look more minutely the quality phenomena of consonants by entering into the detailed study of its series-element. By the name of series-consonants we mean the consonant sounds arranged in every series-lines in the table of Japanese syllabary. It is useful to remark here: by this classification, there happen some cases in our syllabary where the series-consonant does not represent precisely the consonant unit as phoneme; for example, it happens in "s," "t" and "h" series. But to our present purpose, this traditional classification will be enough serviceable. Because, the question before us, is not to find most precisely the individual characteristics or phoneme consonants in the most strict sense, but to find more distinctly the influences of characteristics due to modes of utterance and directions of distortion shifting.

We can present here in Fig. 10 the quality characteristics of series-consonant defined above.

It will be interesting here to compare any pure sound with its corresponding impure one, for instance, "k" series phone with "g" series one, "s" series one to "z" series one, in the same tone "p" with "b," "t" with "d" and so on. As general characters; in the phonated utterance, we can find almost some resemblances in form between sounds that compose such pairs; in the whispered utterance, the strong contrasts are thereby to be found without any resemblances at all. To describe more in detail under the condition of phonated utterance, the impure sounds form

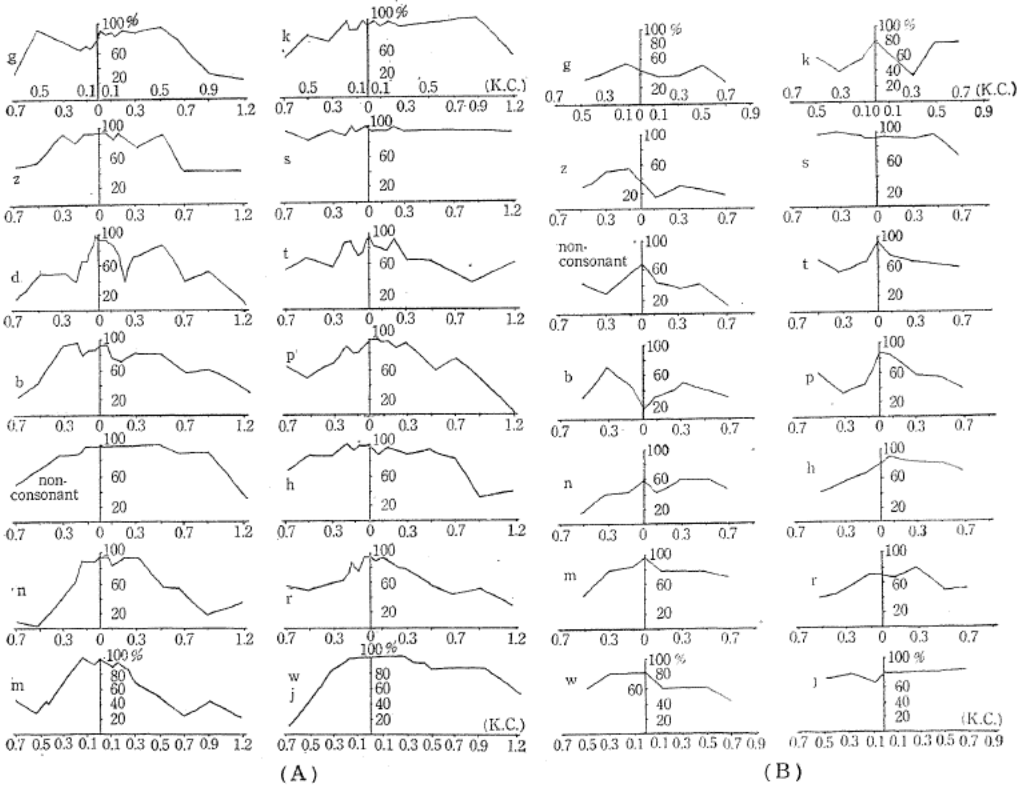


FIG. 10. Quality characteristics of series-consonant in two modes of utterance. (A): in phonated utterance. (B): in whispered utterance.

mountainous characteristics with indentations rough and coarse in shape but few in number; the pure ones in the same condition give relatively wide characteristics of tablelanded form with indentations small and fine in shape but numerous in number. It seems to be one of the principal properties that the characteristics of the so-called voiced consonants in phonated utterance are quite abundant in indentation. The characteristics of the corresponding sounds in whispered utterance are nearly short of them.

On what the jagged characteristics of sounds in phonated case depend in fact? That is the question on which everything seems to remain to be answered soon.

Group characteristics of consonant quality

Next we make an advance in preparing further classification of series-consonants. Because, we have seen above the pure sounds making a high contrast in their characteristics with the impure ones. For becoming aware of further informations on properties of consonant, we will try further classification:

1) *Pure* consonant group: here are included the sounds of what we call voiceless consonants or surds that give the impression of pureness, limpidness, and somewhat sereinity. We count therein such series-sounds as "k," "s," "t," "p," "h" etc.

2) *Impure* consonant group: to which belong the sounds of the so-called voiced consonants, such as "g," "z," "d," "b" series, the impressions of which are of impure, dull and somewhat dingy nature.

3) *Semi-vowelic* group : the phones such as those of "m," "n," "j," "w" series ones are included here, and their timbres are rather akin to a kind of vowels that can not have after all a definite sense of pitches, giving however somewhat harmonious impressions.

There remain still some phones that stand nearly in a midway : for example, Japanese "r" phone seems to lie between the impure and the semi-vowelic, rather nearer a little to the category of the impure ; finally the phone corresponding to the zero-consonant case is also needful to our study as a kind of consonant timbre, as will be understood later in the mishearing phenomena. These phones are however exceptional ones. The principal categories are the three above ; the pure, impure and harmonious. Thence we exhibit in Fig. 11 (A), (B), (C) the three group characteristics

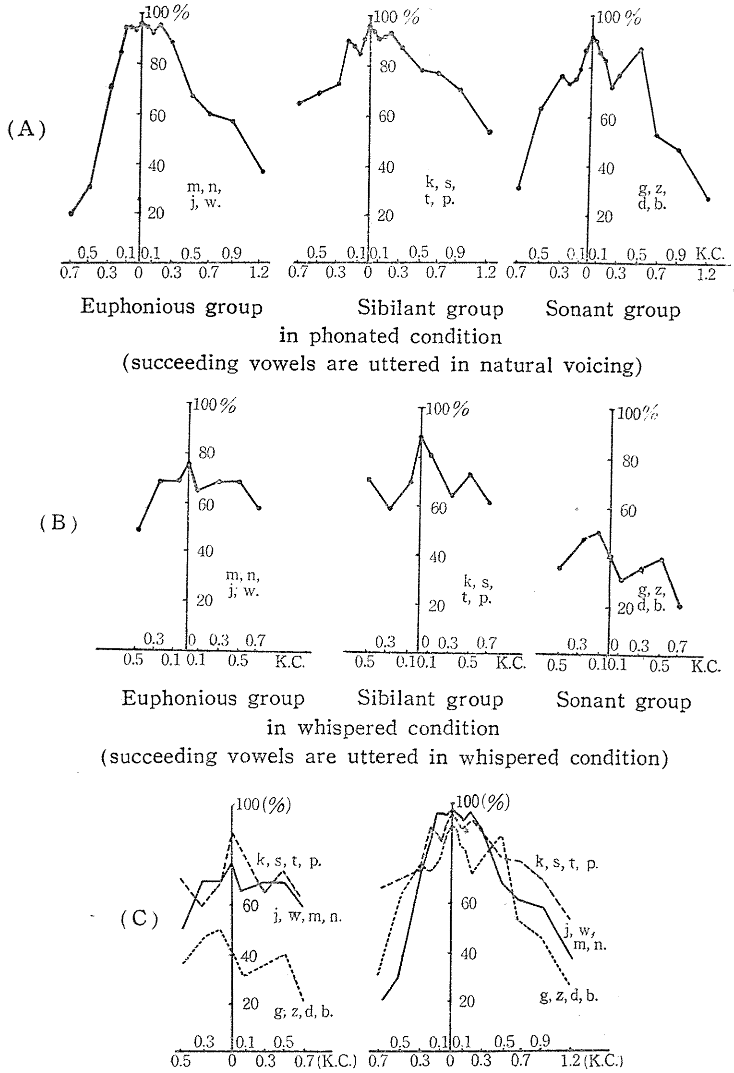


FIG. 11. Group characteristics of series-consonant quality. (A) : in phonated utterance. (B) : in whispered utterance. (C) : these characteristics are superposed in such a way that they can show their relative positions each other in two modes of utterance.

about them, showing also there the difference due to two modes of utterance. These characteristics can give a considerable differentiation from each other, in proportion to the variety of group-sounds on the one hand, and moreover to the difference of the modes of utterance on the other. Some conspicuous facts therein are to be mentioned;

1) In phonated utterance, all three groups come to experience similarly the maximum qualities at null distortion. In whispered utterance, however, the same tendency is observed only in cases of pure group and harmonious group, and the relation of impure group is quite another; in this case the pure and the impure groups come to contrast one with another.

2) The consonant sounds in phonated utterance is generally distinguished by having a shape something like the "triple-peaks" form, more or less akin to that of voiced vowels. The minute forms of three groups are not the same; they are alike in one point that they have therein some undulations, irreguralities of which are not quite the same among them. In phonated case, the two groups of pure and impure sounds do not give at all the antipodal relation as shown in case of whispering.

3) In whispered case, that shape something like the "chestnut" as we have seen it characterising the general sound quality in whispering, is only observed in some degree in the harmonious group.

4) In figure (C) of the same graph the characteristics are superposed and represented separately according to the modes of utterance. We can point out several interesting features worthy of mention. But we can not stay here too long.

Studies on Mishearing Phenomena

The "quality analysis" with which we have tried to keep touch in mid course of articulation treatment as a kind of timbre study, could not be brought to a conclusion without penetrating into the problem of mishearing phenomena. It is because: so long as we concern the commentary work of articulation characteristics by the aid of component articulations which belong after all to the same dimension, we cannot attain to the essential part of the problem.

We will expose here one example which is able to show concretely where the problem in fact exists. Fig. 12 presents the articulation characteristics of vowels and consonants respectively for the case of phonated and whispered utterance. As the graphs show, the vowel characteristics for both utterances cross each another, and the cosonant articulations at the same conditions do not come to cross. As quality characteristics, it must be an important event whether they have an intersection or not. It seems to lie the essential difference between quality phenomena of the crossed characteristics, and the characteristics that have no intersection look not have any

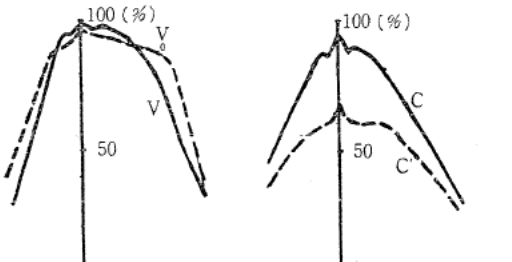


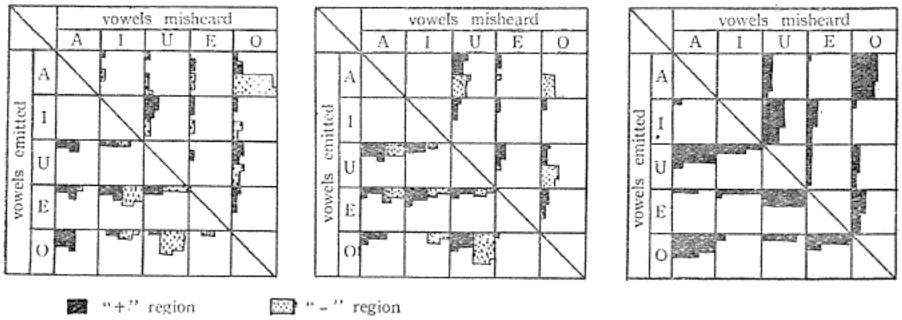
FIG. 12. Representation of general articulation characteristics of vowels (A) and consonants (B) by two modes of utterance.

V: vowel articulation in phonated utterance.
$\underset{0}{V}$: vowel articulation in whispered utterance.
C: consonant articulation in phonated utterance.
C': consonant articulation in whispered utterance.

fundamental differentiations in their quality phenomena. It is needful to be verified. Sure enough, be this the case? Will it be proved true? By studying the mishearing phenomena, we will try to answer it.

Mishearing of direct vowels

As we have seen before, the quality characteristics of individual vowels do not bear good resemblances among themselves. To itemize the detail of its quality loss, we must go into the troublesome studies on the mishearing phenomena of individual sounds upon which is based in fact the veritable cause of quality degradation. We show next the data on mishearing of five direct vowels in Table 2 where the numbers of occurrence of mishearing phenomena happened in the whole domain of each dis-



(A) (B) (C)

FIG. 13. Representation of mishearing of direct vowels, corresponding to each of Table 2.

TABLE 2

(A): in phonated utterance

Vowels pronounced	A (ア)	I (イ)	U (ウ)	E (エ)	O (オ)	Number of sounds observed
	Vowels misheard					
A (ア)	╲	1 3	1 6	3 4	8 47	168 140
I (イ)	2 0	╲	3 4	4 3	8 7	144 120
U (ウ)	12 0	5 2	╲	2 0	8 11	144 120
E (エ)	11 1	12 28	10 5	╲	7 0	156 130
O (オ)	21 0	2 8	3 27	1 1	╲	156 130

(B): in whispered utterance

Vowels pronounced	A (ア)	I (イ)	U (ウ)	E (エ)	O (オ)	Number of sounds observed
	Vowels misheard					
A (ア)	╲		26 34	7 2	1 31	224 336
I (イ)		╲	13 0	7 0	6 0	192 288
U (ウ)	29 29	17 4	╲	17 1	11 33	192 288
E (エ)	13 23	34 13	14 15	╲	15 3	208 312
O (オ)	14 1	0 16	33 68	1 0	╲	208 312

(C)

	A (ア) (%)	I (イ) (%)	U (ウ) (%)	E (エ) (%)	O (オ) (%)
A (ア)	╲	0.5	18	0.5	56
I (イ)	2	╲	46	16	1.5
U (ウ)	42	16	╲	9	16
E (エ)	6	11	40	╲	25
O (オ)	46	1.5	8	23	╲

tortion are summed up respectively and tabulated. Corresponding diagrams are shown in Fig. 13. We add there, for reference, the data on mishearing of five Japanese vowels voiced in phonated condition which were obtained by another procedure.* (Table 2 (C), Fig. 13 (C)). Discarding the minute influences due to the difference of distortion direction and thus attaching importance to the mere difference of utterance, we can make Table 3 and Fig. 14, where the numbers of phenomena are summed up together through two domains of distortion. By these tables and figures we can denote some remarkable points as follows:

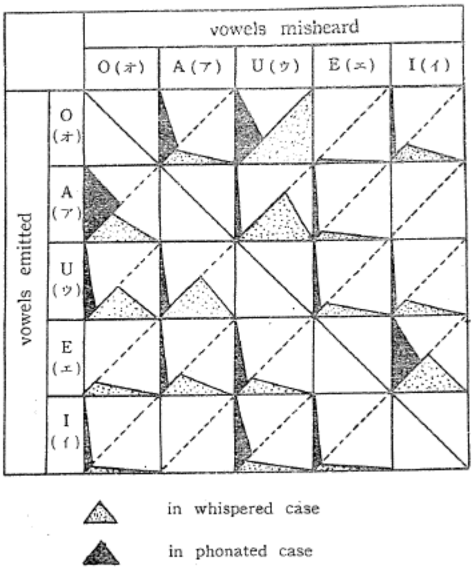


FIG. 14. Mishearing phenomena represented by stressing on the difference of modes of utterance.

TABLE 3

		Vowels misheard					Totals
		O (オ)	A (ア)	U (ウ)	E (エ)	I (イ)	
Vowels pronounced	O(オ)		21 15	30 101	2 1	10 16	63 133
	A(ア)	55 32		7 60	7 9	4 0	73 101
	U(ウ)	19 44	12 58		2 18	7 21	40 141
	E(エ)	7 18	12 36	15 29		40 47	74 130
	I(イ)	10 6	2 0	17 13	7 7		36 26

(1) Observing the phenomena of mishearing, the influences due to the modes of utterance are not so conspicuous as expected.

(2) In vowels whispered, the resemblances between "A" and "U," and moreover "O" and "U" become more remarkable than in vowels phonated.

(3) In vowels phonated, the resemblance between "A" and "O" happens in such a particular way that "A" can approach to "O" only by way of "−" distortion, and that "O" can approach to "A" only by way of "+" distortion.

(4) Generally speaking, the reciprocity of mishearing holds good roughly in some degree. By investigating more precisely, the mishearing of phonated vowels seems to have tendency to hold off the perfect reciprocity somewhat more evidently than whispered ones. In brief: from the essential judgement of the nature of influences of the distortion C.S.D., the vowels seems to subject to a great difference according to the modes of its utterance. At a glance, the phenomena take such appearances. Because, phonated vowels come to bear the impress of a dingy shade by this distortion. On the other hand, the whispered vowels remain quite free from it. Moreover, the general characteristics with their intersection seem also to support it. But as the results of precise observation of mishearing phenomena, we are able to reach a conclusion that there shall be no essential differences between phonated and whispered vowels. The facts that some variations of the degree of resemblance as the modes

* Y. Ochiai: Studies on Telephone Articulation Based upon the Observation about Distortion Phenomena of Japanese Speech-Sounds 1943.

of utterance change, are not essentially important, because the vanishing effect of volume difference among whispered vowels seems to play thereby an important part in the discrimination of vowels.

Mishearing of glided vowels

We have above mentioned about a great difference of quality characteristics between direct vowels and glided vowels, indicating thereby as its striking example the principal difference between sounds of a pair of "O" and "jO" in the "−" domain of distortion. We must try to examine here this interesting example from the side of mishearing phenomena. Referring back to Fig. 9, we will take out there 5 conditions X, Y, Z, and X′, Z′ in phonated case, and also 5 ones x, y, z, and x′, z′ in whispered case. We pass to the details of mishearing, and we can show its content, as given in Table 4. This table can explain pretty clearly the fact that the gliding-into action of the vowel "jO" can reduce the effect of mishearing "jO" → "jU" (that means after all O → U) that results in an increase of quality loss of direct vowel. The case where the direct vowel "I" as stationary timbre in the domain of "−" distortion can be heard amiss to "U," is very rare in phonated utterance and is quite null in whispered utterance, as shown in Fig. 13. This example can point out the existence of the rare case where the judgement on stationary part of the glided vowel "jO" is not determined mainly by its own stationary part, but rather by the powerful influence of its transitory part.

TABLE 4*

(A)

	X	Y	Z
direct	O→U (11)	O→U (12)	O→U (2) O→I (8) O→E (1) O→jU (1) O→jO (1)
	X′		Z′
glided	jO→jU (1)		jO→jU (4) jO→U (2) jO→O (4) jO→I (4)

(B)

	x	y	z
direct	O→U (16) O→I (3)	O→U (23) O→I (6)	O→U (28) O→I (7) O→A (1)
		y′	z′
glided		jO→jU (5) jO→U (1) jO→O (5) jO→I (1)	jO→jU (3) jO→U (3) jO→O (7) jO→I (1) jO→A (1)

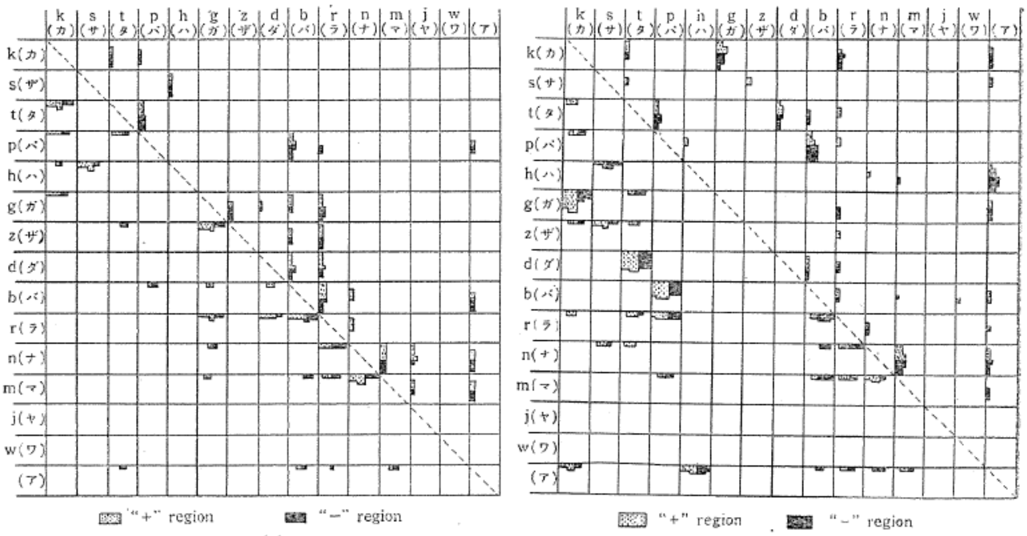
General tendencies of mishearing of series-consonants

In the study of mishearing of consonants, we take also the series-consonant sounds as the limit of our analysis. This means, we make an observation of mishearing among the series-consonant sounds. Thereby we take 15 series-sounds according to Japanese syllabary, including naturally pure vowels as one series-sound that lacks in consonant. Table 5 records the number of occurrence of mishearing of total 15 series-consonants (arranged in the first vertical column) that are heard amiss to any one of 15 sounds (arranged in the top line). The upper number in each column shows the record in the domain of "+" distortion, and the lower that of "−" distortion. And Fig. 15 is its diagrammatic representation. From the observation of this table and figure, we can denote some remarks:

* In this table, the numbers in parenchesisses denote the number of occurrences of mishearing.

TABLE 5 (A). Mishearing of Series-Consonants in Phonated Utterance

	k (カ)	s (サ)	t (タ)	p (パ)	h (ハ)	g (ガ)	z (ザ)	d (ダ)	b (バ)	r (ラ)	n (ナ)	m (マ)	j (ヤ)	w (ワ)	(ア)	Totals	Number of sounds observed
k (カ)			5	4												9	192
			9	6	2	1			1			1			2	22	144
s (サ)			2		8		1					1			1	13	192
			1		7											8	144
t (タ)	26			19	2		1	2	3	1					1	55	192
	6			20	1	1		3	3							34	144
p (パ)	10	1	7		3		1		13		1				3	39	192
	3		3		3				12	4					4	29	144
h (ハ)	2	26	1	1											1	31	192
	1	3		2					1						1	8	144
g (ガ)	9		3				4	4	8	9	2	1			3	43	192
	4		1	1			8	1	3	16						34	144
z (ザ)			1	1	1	26		2	6	8			1			46	192
	1	1	3			6			2	7			2			22	144
d (ダ)			2						11	11	2					26	79
			1	1			1		12	5						20	54
b (バ)	1		1	4	1	4		4		20	7	3	1		3	49	192
				2			1			11		2	1		8	25	144
r (ラ)			1		1	12	2	10	20		7			1		53	192
			1			5	1	3	22					1	1	34	144
n (ナ)						2			2	9		20	13		6	52	192
	1		1			3			2	8		19	2		8	44	144
m (マ)		1				4			1	5	33		7	1	6	57	192
									6	5	9		2	1	7	30	144
j (ヤ)	1										2	1				4	72
	1					1				1	2	1				6	54
w (ワ)													1		1	2	24
											1		2			3	18
(ア)				1	1			1	3	1		2				9	120
				2	1				3	2		3				11	90
Totals	49	28	23	30	17	44	9	23	68	68	53	28	23	2	25	422	
	19	4	20	34	14	21	11	7	66	55	12	26	9	2	31	329	



(A) (B)

FIG. 15. Representation of mishearing phenomena of series-consonant. (A): in phonated condition. (B): in whispered condition.

TABLE 5 (B). Mishearing of Series-Consonants in Whispered Utterance

	k (カ)	s (サ)	t (タ)	p (パ)	h (ハ)	g (ガ)	z (ザ)	d (ダ)	b (バ)	r (ラ)	n (ナ)	m (マ)	j (ヤ)	w (ワ)	(ア)	Totals	Number of sounds observed
k (カ)	\		9 3	3 2	1 	25 16	 1	1 2	1 1	4 11					6 9	50 45	128 96
s (サ)	1 	\	4 		1 1		4 1				2 1	2 			5 	19 3	128 96
t (タ)	8 1	 1	\	10 14		1 	1 1	16 12	4 6	5 1					1 	46 36	128 96
p (パ)	5 5		3 2	\	5 1	 1			33 40	5 		 1			3 2	54 52	128 96
h (ハ)	1 	14 12		2 	\			1 			3 	1 4	 1		15 25	37 42	128 96
g (ガ)	65 28		4 7	5 3		\		 1		1 8					4 8	79 55	128 96
z (ザ)	4 2	26 5	7 3		1 1	1 	\			4 	1 			 1		47 12	48 36
d (ダ)	5 6	2 	65 38	5 5		1 3		\	8 7	4 2	1 				1 	92 62	128 96
b (バ)	2 		1 1	58 38	4 	1 2			\	6 2		2 1		 2	5 3	79 49	128 96
r (ラ)	6 		8 2	19 20	2 	6 3	 1	2 	4 17	\	4 2	 2			2 1	53 48	128 96
n (ナ)		6 2	7 1	1 1	3 1	1 2		 2	1 6	7 10	\	28 28	1 		8 11	63 64	128 96
m (マ)	 1	1 		3 5	1 2	1 			3 10	6 3	15 2	\		1 1	2 6	33 30	128 96
j (ヤ)		 1											\	 1	 1	 3	48 36
w (ワ)					1 							1 		\		3 1	16 12
(ア)	13 3	1 	 1		25 16	1 	 1		2 2	5 4	3 4	5 2		2 	\	57 33	80 60
Totals	113 46	50 21	108 59	106 88	44 22	38 27	5 5	20 17	56 89	47 41	29 9	39 38	1 1	3 5	52 66	706 530	

(1) As regards the mishearing of series-consonants, the relation of its reciprocity is generally established in spite of the modes of utterance.

(2) The phenomena of mishearing of series-consonants show the considerable difference with respect to the difference of utterance.

(3) For example, in phonated utterance, the mishearing effect occurs mainly among the consonants of the same group; any consonant of pure group is misheard only to another sound of the same group, and any consonant of impure group is heard also amiss to another sound of the same group, we can call this type of mishearing: *mishearing within group.*

(4) In whispered utterance, on the contrary, we see the mishearing of another type: consonants of pure group are heard amiss to the corresponding consonants of impure group, in such a manner: s⇄z, p⇄b, k⇄g and so on. We can call this type of mishearing: *mishearing between groups.*

(5) The existence of such tendencies of quite different type of mishearing means: the nature of consoants due to the difference of utterance is quite essential. Referring back once again to Fig. 12 (B) and judging from the forms of quality characteristics, there seems to lie no essential difference between the consonants phonated and the consonants whispered. But the minute observation of the mishearing effect of consonants suffices to show that there exists in fact the essential difference between them. We know, the troublesome statistical labours on mishearing investigation were to be

rewarded sufficiently by making aware of its certain merit for developing the subtle side of timbre study.

Mishearing of series-consonants by group treatment

TABLE 6

	Phonated				Whispered			
	O. G	I. G	II. G	III. G	O. G	I. G	II. G	III. G
h (ハ)	5	92	3	0	50	37	2	11
s (サ)	5	85	5	5	23	31	23	23
k (カ)	7	84	7*	2	16	19	65*	0
t (タ)	1	83	14*	2	1	41	58*	0
p (パ)	10	44	44*	2	5	20	74*	1
b (バ)	15	12	54*	19	7	80	9*	4
z (ザ)	0	12	84*	4	0	90	9*	1
g (ガ)	4	22	70*	4	10	82	8*	0
d (ダ)	0	7	90*	3	1	82	16*	1
r (ラ)	1	3	86	10	3	56	33	8
m (マ)	15	1	13	71*	12	20	22	46*
n (ナ)	14	2	10	74*	18	19	10	53*
j (ヤ)	0	20	10	70(*)	33	33	0	33(*)
w (ワ)	20	0	0	80(*)	0	50	0	50(*)
(ア)		25	35	40*		66	7	27*

TABLE 7

	Phonated			Whispered		
	I (%)	II (%)	III (%)	I (%)	II (%)	III (%)
h (ハ)	97	3	0	74	4	22
s (サ)	90	5	5	40	30	30
k (カ)	91	8	1	22	78	0
t (タ)	84	14	2	42	58	0
p (パ)	47	47	6	21	78	1
b (バ)	17	61	22	86	14	0
z (ザ)	12	84	4	90	9	1
g (ガ)	23	73	4	91	9	0
d (ダ)	9	87	4	84	11	5
r (ラ)	4	86	10	58	34	8
m (マ)	2	15	83	23	25	52
n (ナ)	2	12	86	23	13	64
j (ヤ)	20	10	70	50	0	50
w (ワ)	0	0	100	50	0	50
(ア)	25	35	40	66	7	27

For carrying out the effective treatment of mishearing of consonants, here we try to classify the series-consonants of Japanese syllabary, conforming nearly to the manner as mentioned above:

O. 0 Group series of pure vowels, lacking in its consonant parts, designated by the sign " ,"
I. 1st Group "s," "k," "t," "p," "h" series-consonants,
II. 2nd Group "z," "g," "d," "b," ("r") series-consonants,
III. 3rd Group "m," "n," "j," "w," ("r") series-consonants.

The "r" series-sound in Japanese syllabary stands in such a midway between the impure group and the semi-vowelic, that it is sometimes to be counted in the 2nd group and sometimes counted in the 3rd group: under such circumstances when "r" is counted in 2nd we mark it with the sign *, and when counted in 3rd we mark it with the sign (*). In accordance with this understanding, we can treat the mishearing phenomena in the following manner shown in Table 6.

The diagrammatical expression for this could be done well with the pyramidal space representation. But for the sake of simplicity, we choose here three-groups method, taking No. zero group off the list of Table 6. Thus we can reach Table 7, where the numbers are recorded in percentage. The exclusion of No. zero group can not influence on the general tendency, because it is merely a partial effect concerning mainly the mishearing between "h" and " " (pure vowel) due to the difference of modes of utterance.

Triangular representation of series-consonants

By the treatment of three-groups classification method, we can give here a triangular representation for it, determining thus the positions of individual series-consonants in the surface of equilateral triangle. Fig. 16 shows it. First we will explain a little about this representation. The three tops of this equilateral triangle represent respectively the three poles by three-groups classification. The percentages of mishearing of individual series-consonants are scaled respectively on the three perpendicular lines drawn from each top to each corresponding base line. By this method, we can orientate 15 consonantal sounds on the plane of a triangle. Such

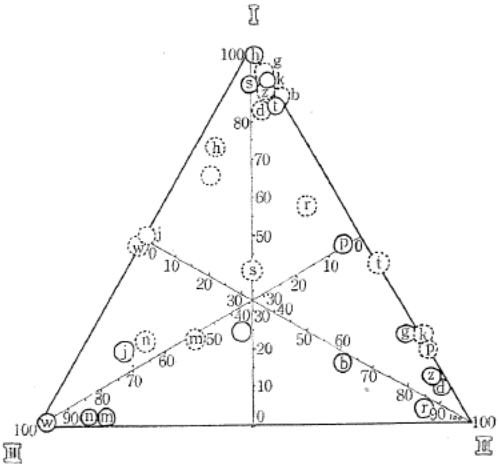


FIG. 16. Triangular representation of series-consonants in two modes of utterance. In this figure the circles in full line denote the consonants in phonated utterance, and the circles in broken line those in whispered case. The circles without letters therein denote the initial sounds of pure vowel, that is, the sounds of non-consonant.

configulation represents the positions of the sounds, so to speak, in a psychological space imagined as to the quality of timbre.

Next we try to trace the positions of the sounds. As three poles are the representatives of ideal group sounds, deduced from the actual phonated case, it is very reasonable in phonated condition to see the pure group sounds "h," "k," "t," "s" quite closing up in the region of the pole I, the semi-vowelic group sounds "j," "w," "m," "n" crowding in the neighbourhood of the pole III, and the impure group sounds "g," "z," "d," "b," "r" dispersing more or less in the near region of the pole II. That the sounds of each grouping are represented gathering respectively in the region of each pole corresponding to the grouping, is based on the nature of mishearing in phonated case, that is, the character of the mishearing within group. It is to be noted moreover that "the non-consonantal sound (that is, the pure vowel) goes almost near to the center of the triangle. The sound lacking in consonants means its neutrality in character. It is also reasonable that the sound neutral in charater can take position in the neighbourhood of the center, that is, the neutral position of this triangular representation. On the contrary, the more the sounds lie near the pole, the more the mishearing within group come about evidently with these sounds. Such is the example of the sounds "h" and "w." And the sounds "p" and "b" reveal themselves as a little intermediate ones.

In the whispered condition, we see, by this figure, the situation undergoing a sudden change. In general, all sounds seem to be attracted towards the pole I. The devocalization seems to mean the general increase of sibilant tendencies. Now look more minutely and partly. We can observe there still essential phenomena going on. For example, the impure group "z," "g," "b," "d" are pulled strongly into the region of the pole I, and the pure group "t," "k," "p" are attracted but less vigorously to the region of the pole II: that means; it happens in this case a quite different type of mishearing; *mishearing between groups*; sonant (impure) sounds heard amiss to corresponding elements of sibilant (pure) ones; sounds of sonant group advance in mass to the *sibilant* pole I, and an amount of sibilant sounds heard amiss to corresponding elements of sonant group; but as the reciprocity of this mishearing is not so perfect, that the sibilant sounds can not reach in the direct vicinity of the *sonant* pole II. The sounds "s" and "h" of sibilant group advance to the mid-region; this does not mean directly the effect of attraction of the *euphonious* pole III, but seems rather to mean the tendency of neutralization of these sounds; that is the reason the "h" and "s" sounds approach to the mid-position of *neuter* region, and the "non-consonant" sound in whispered case is attracted, owing to its devocalization, to the "h" position in phonated case. In this way, we know that the confusion of sounds happen very frequently between aspiratory sound "h" and devocalized pure vowelic sound, exclusively in case of whispered utterance. By the aid of this triangular representation, we can very easily express the influence of modes of utterance upon the series-consonants by the displacement of their configuration.

Studies on Pitch-Attribute

In the preceding, we have considerd the speech sounds exclusively on the clearness aspect, and the timbre studies performed above have been accordingly based upon the articulation judgment as a measure of transmission quality. This is a traditional way. We followed it also as possible. But there remains another field quite untouched till now in quality problem. We mean here the pitch study of voices. The reason why man did not yet set his hand to such interesting problem is nothing but that the voices without distortion can not show clearly where the problem in fact exists. The voices coming through the distortion of this type get aware for the first time of the true validity of pitch study. We had not to neglect this aspect even in any other distortion. Because, we think, the clues of sheer solution of timbre problem can not be found in any other than this inlet of pitch study.

Outline of experiment

A sustained vowel "I" which is voiced by a male at the pitch of 200 c.p.s., being recorded and driven in the endless manner by magnetic tape-recorder, is sent into the circuit through the distortion C.S.D., while referential pure-tone frequencies of which are previously assigned to the purpose are led to the circuit at the same time: thus, a referential pure tone and a distorted sustained vowel are alternately fed to the speaker before subjects by aid of electronic switch. These are illustrated schematically in Fig. 17.

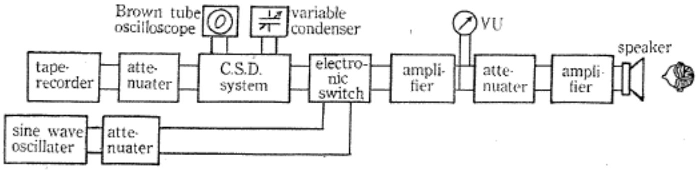


FIG. 17. Schematic diagram for pitch measurement; Brown-tube oscilloscope is inserted for the purpose of supervision of carrier frequency.

Procedure of subjective determination of distorted pitch

The scheme of presentation of two different signals in time is shown in Fig. 18, where by P is designated pure tone, and by V distorted vowel. The amounts of distortion (shifting frequency) are determinated by a variable air-condenser added to the tuning circuit of demodulator, and are varied by an operator according to the previously designed tables. The frequency intervals of distorting step are 3 and 6 c.p.s., and number of steps is eleven. Therefore, the vowel is distorted (or shifted) over the range of 30 and 60 c.p.s. The mid-point of eleven distorting steps is obtained previously by subjects' control of condenser. For nine pure tones which differ in pitches from each other, the mid-points of step are thus determined and other ten steps are scattered at random around their mid-points respectively.

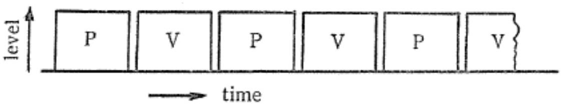


FIG. 18. Arrangement in time of pure tone and distorted vowel presented as signals, where the former is denoted by P, the latter by V. Intervals between them are very short, about 0.01 sec.

Two young subjects, both males, are instructed to judge whether the pitches of either sounds, alternately and different in time, are matched to each other or not, and to record on the score sheets whenever they judge either to be "matched." Every subject sits two rounds for each pure tone. In each round, every step is presented six times. Then one step is observed twenty-four times in total by both subjects. The pitch matching between a pure tone and a distorted complex sound is so much hard, that each distorting step is kept constant until the judgments of subjects are completed. The duration time and loudness of reproduced sounds are adjusted in advance by subjects so that the listening is as easy as possible. Duration time practically selected is from 1.5 to 3 seconds. Pitches of pure tone used, the corresponding mid-points of step and frequency intervals of step are tabulated in Table 8.

TABLE 8. Experimental Condition of Pitch Measurement

Pitch of pure tone (c.p.s.)	200	210	160	180	190	200	210	220	240	190	200
Mid-point of step (c.p.s.)	−200	−170	−120*	−45	−22.5	0	30	60	135*	170	200
Interval of step (c.p.s.)	3	6	15	6	6	3	6	6	15	6	6

Results of pitch degradation of vowelic timbre

Representing by the ordinate the number of cases where the subjects complete their judges of being "matched," and also by the abscissa, original point of which is identified with mid-point of step, the distorting step in c.p.s., such a histogram as shown in Fig. 19 is obtained for every measurement corresponding to each pure tone, where the data for both the observers are brought together. The histogram changes its form according to the position of mid-point (null shift) on the continuum of shifting frequency. To represent these aspects quantitatively, means and standard deviations of the histogram are calculated. These results are tabulated in Table 9 and illustrated in Fig. 20. Some data, those for the considerable amount of distortion,

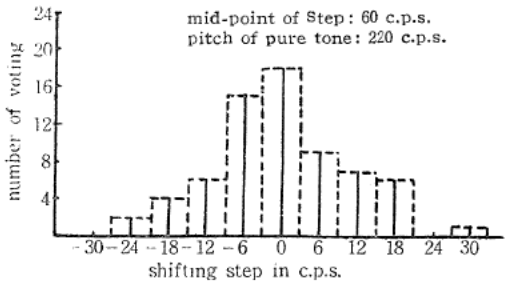


FIG. 19. Histogram obtained from pitch matching between pure tone and distorted vowel "I." The ordinate shows the number of votings, or of cases in which the subjects judge "matched." The abscissa represents the shifting steps of C.S.D., and its original point corresponds to the mid-point of step and is for the distortion of 60 c.p.s.

TABLE 9. Values of Mean and Deviation of Distortion in c.p.s. and Values of Index

Pitch of pure tone (c.p.s.)	200	210	180	190	200	210	220	190	200
Mean distortion (c.p.s.)	−202.6	−168.2	−53.5	−33.9	−0.5	33.0	54.0	158.9	202.0
Standard deviation (c.p.s.)	4.9	9.8	7.1	6.9	3.1	9.0	11.1	7.8	3.5
Value of index	21.2	6.0	7.1	7.2	21.6	8.2	6.1	7.7	13.6

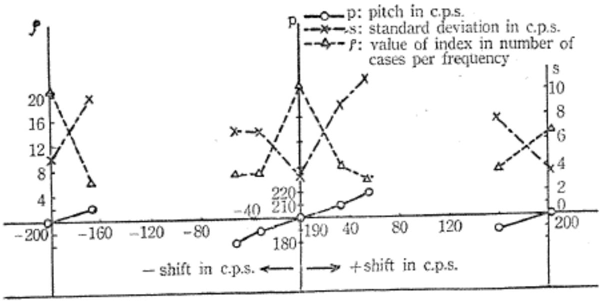


FIG. 20. These curves are obtained from the histograms of pitch matching between pure tone and distorted vowel "I." Full line is to show the mean values of distorted pitch, chain line is for the standard deviation of histograms, and broken line is for the value of index for representation of intensiveness of pitch.

being affixed with asterisk in Table 8, are omitted, because so much of distortion makes pitch matching with pure tone very hard and the votings of subject run low and become quite at-random.

Mean value of distorted pitch

As the full line in Fig. 20 shows, the pitch of distorted vowel changes with the shift in the same direction of distortion in the vicinity of null shift. The gradient of pitch change being nearly 15° in both sides of shifting, we know that the pitch changes of vowel due to the uniform shift of its components are always less than the shifting values of distortion. These phenomena reappear in both directions but at intervals of an octave shift leaving the middle regions between them almost undeterminable, for which it becomes very hard to grasp definite pitch.

Standard deviation of pitch matching

It is easily supposed that the size of deviation is subject to difficulty of pitch matching, and this is really shown in chain line in Fig. 20. Namely, the increases of the deviation with the amount of distortion shifting show the rise of difficulty of pitch matching, and this is no less than the fact that the more the distortion increases, the more the pitch-attribute of vowel is to be lost.

Index for representation of intensiveness of pitch-attribute

Owing solely to this distortion C.S.D., we can demolish the timbre structure continuously from its original state, where all the components of sound are in perfect union, and, taking the probably subjective difference-tone as the fundamental, we can also recover it cyclically at every octave shift. This fact induces us to prepare some indices for representation of subjective impression under such physical circumstances. Changes of pitch mentioned above are, so far, conceivable as one of them, but there still remain a clue for finer quality analysis. That is to say, it can be seen partially in the transition of deviation shown in Fig. 20, that the contents of pitch matching performed at every step of distortion are not uniform. And, this suggests the idea of *intensiveness* of pitch, besides the absolute value on its height. The values of standard deviation may look first adequate for the end of representation of its intensiveness, but it is not yet sufficient by reason of the following explanation.

If a sound at any distorting step has the definite pitch, votings in pitch matching must be done for every signal presented at this step of distortion, that is, the number of votings at this step must be twenty-four in our case. (This is really obtained at the vicinity of null distortion.) This leads us to think of the number of voting or size of histogram, besides its deviation, in treatment of the intensiveness of pitch. Because, the small deviation of histogram does not necessarily indicate the definite pitch of sound, without taking the size of histogram into consideration on the view-point above. To illustrate concretely, we select two examples from the data and show them in Fig. 21, where the curve of full line is the envelope of histogram for the experiment, in which the mid-point of distorting step is -170 c.p.s., and the broken line is that for the experiment, in which the mid-point is $+30$ c.p.s., and pitches of referential pure tones are 210 c.p.s. in both the cases. Values of deviation are in the same order; 9.8 c.p.s. for the former and 9.0 c.p.s. for the latter. But the number of votings in total is different from each another; 59 and 74 for either cases respectively. This means that the votings concentrate more seriously around the modal point of histogram in the latter case than in the former and indicate that the distorted sound in the latter has more intensive pitch than the former.

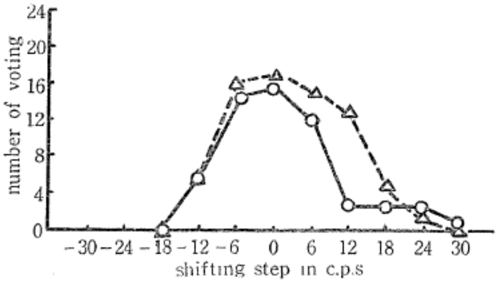


FIG. 21. Comparison of two histograms. Original point of the abscissa corresponds to the distortion of -170 c.p.s. for full line curve, and to that of 30 c.p.s. for the broken one.

Thus taking into account at the same time both the deviation and the number of voting, we come to give an index for intensiveness of pitch:

$$\rho=\frac{N}{s}$$

where N is total number of voting (or area of histogram), s is standard deviation of histogram and ρ is to represent intensiveness of pitch-attribute. This index is calculated for every case and tabulated in Table 9 and illustrated also by broken line in Fig. 20. As shown in this figure, the index varies in its values reasonably with the amount of distorting shift. This is also in accordance with our intuitive experiences.

Studies on Loudness-Attribute.

On another phonic attribute "loudness," the character of this specific distortion is also studied, that is, the change of loudness of distorted sound is measured subjectively by means of loudness balance.

Procedure of experiment

As shown schematically in Fig. 22, a sustained vowel "A" voiced by a young male at the pitch of 200 c.p.s., is sent into the line by way of endlessly driven tape-recorder, from which the one signal coming directly without distortion is fed in one

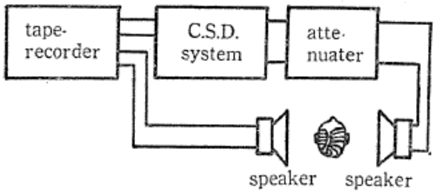
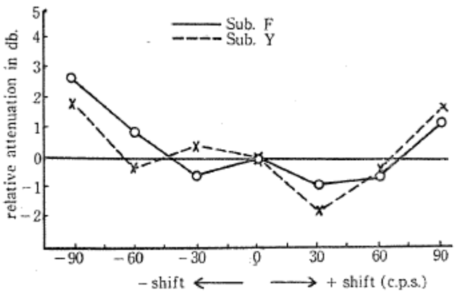


FIG. 22. Schematical diagram of loudness balance. The attenuater is to be controlled by the subject until he judges that both the distorted and non-distorted vowels are balanced completely.

speaker and the another coming through the distortion system fed to another speaker. The subjects situated at the mid-point between two speakers, control the attenuater inserted until the loudness of two sounds come to be balanced. Seven differently distorted sounds are balanced in loudness in one round, and every subject is engaged five rounds. The amounts of distortion used here are ±90, ±60, ±30, and 0 c.p.s.

Results of measurement

In Fig. 23, the ordinate shows the mean reading of attenuater (in db) relative to that of attenuater in case of the sound of null shift, while the ordinate in Fig. 24 represents the unbiased variances (in db²) at each shifting step. In either case of subjects, the mean value of attenuation tends to increase according as the distortion gets larger. But this tendency, when tested statistically, is not concluded significant under such a condition where the considerably much fluctuation takes place as indicated in the values of variance.



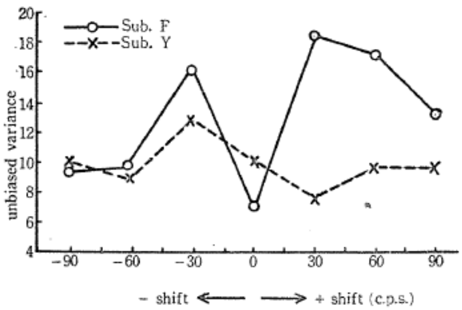


FIG. 23 (*left*). Mean values of attenuation relative to that of null-distortion step, obtained from the measurement of loudness balance between the distorted and non-distorted vowel "A."

FIG. 24 (*right*). Unbiased variance of relative attenuation in db² obtained from the measurement of loudness balance between the distorted and non-distorted vowel "A."

Variance also does not show the statistically significant changes over the whole region of distortion, especially in the case of subject Y, and the absolute values of variance are not so small. . This means that the loudness-attribute is not easily seized under such a condition as this.

Tolerance or Allowance Problem

Not a more important problem in practical field of transmission quality could be found anywhere else excepting the question as to the tolerance of distortion. In the tolerance problem, it is requested to decide pertinently to what extent the distortion is allowable for communication purpose. It concerns surely a matter of degree. We can call in question whatever distortions they may be. To establish to the point the tolerable limits of the degrees of distortion, it can not be based on the physical authorities, but must be solely founded on the psychological ones. We know thus the determination of allowance must conform therefore to the experiment of sheer subjective nature.

Then, as to the articulation experiments as have been executed before, are they serviceable for this purpose? Articulation quality based upon the clearness aspect, truth to say, can not answer it. Because: to bring forth more than ten per cent of articulation loss in the distortion C.S.D., it is needful to prepare the shifting of one hundred cycle or more, the effect of which on speech timbre is already formidable enough. We ought to search accordingly another kind of quality grounded upon any other aspect. Thus, we come to know at last this tolerance problem coming from the practical field originally is nothing less than the most intresting and instructive one that does touch seriously the pertinent study of timbre quality of another kind.

Outline of experiment

To discriminate the distorted sound from the non-distorted one, the subject needs two referential criteria of judgment, such as the "non-distorted" qualities and the "distorted." Though the distorted sound quality of "impurity" is characteristic enough at the not-small degree of distortion, non-distorted sound quality judged on the aspect of "natural timbre," is rather complicated nature and comes to be conceived differently by each subject. Then it remains left to individual attitudes, to catch hold of the timbre property with which the assigned sounds are endowed in their natural states. That is to say, every subject is requested to judge whether or not B sound is different from A sound, where A is non-distorted sound and B distorted one. And both the sounds are presented before subjects in alternated succession in time, being solely left to the discernment of subjects.

Allowance for continuous speech

Procedure of subjective measurement

As shown schematically in Fig. 25, the continuous speech is picked up by a velocity microphone and sent to the line, which includes the distorting system (C.S.D.), and finally reproduced from a high-quality speaker.

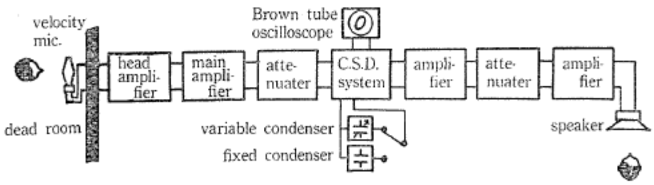


FIG. 25. Schematic diagram for the measurement of allowance. One of two condensers, which is variable, is to present the distorted signal and another, which is fixed, is for the non-distorted one. Synchronism of carrier frequencies is supervised by Brown-tube oscilloscope.

Sentences which two talkers have to read are selected arbitrarily from the handy essays. Four listeners, each young male and normal in hearing, sitting at one meter distance from the speaker and listening to the reproduced sounds by it, are requested to make a discrimination of the "distorted" speech from the "non-distorted." The speeches conditioned by such two regulations are conducted to the subjects' ears, the "nondistorted" thereby always preceding the "distorted," as shown in Fig. 26, where A and B represent two conditions respectively. The transitions in time between these conditions are instantaneously carried out, and the subjects recognize these transitions with the aid of feeble clicking sound by switching. The procedure of arrangements of speech in time is performed by an experimenter who operates the hand-switch for marking speech in time on the one hand, and controls at the same time the variable condenser for changing the degree of distortion on the other, conforming to the previously designed table.

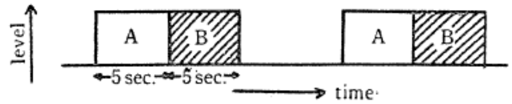


FIG. 26. Signal arrangement in time for the measurement of allowance. A is to represent the "non-distorted" and B the "distorted."

Every subject is engaged in measurement of five rounds for one talker. As the suitable number of steps of distortion used in one round, are allotted thirteen steps for negative and positive directions together, involving therein the step of non-distortion naturally. The frequency interval between two neighbouring steps is 3 c.p.s., and every step is repeated five times in one round. Then each subject listens to sixty-five pairs of A and B per round. To talk about the steps more minutely, every subject observes twenty-five pairs of A and B for each step in all the rounds of one talker. Then the total number of observations for each step amounts to two hundred times over all talkers and subjects.

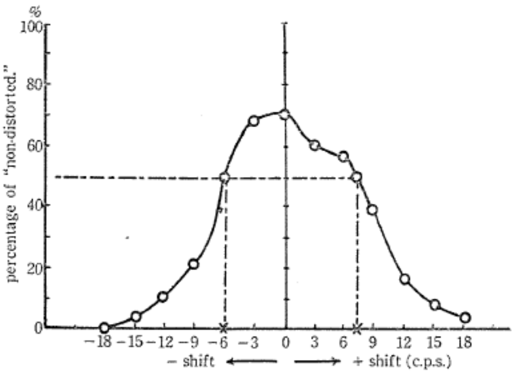
The instruction given to the subjects is as follows; "Preceding signal A passes into signal B with clicking sound. Signal A is designed always as non-distorted. When you come to judge that there is no difference between A and B, record sign ○ on the score sheets, and sign × when you judge that there is difference."

By way of some preliminary experiments, every subject can accept and follow this instruction.

Results of measurement

Rejecting some data which show the apparently significant difference from others, by Thomson's testing method, the results are readjusted and illustrated in Fig. 27.

FIG. 27. The results of measurement of allowance for male continuous speech. The signs × on the abscissa, corresponding to 50% of ordinate, represent the values of allowance.



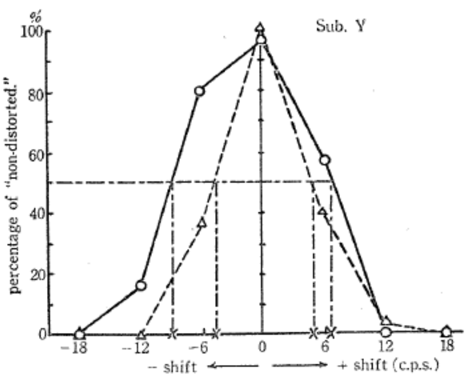
This is the mean result of that of every combination of talker and subject. In this figure, the ordinates stand for the percentage of cases where the subjects judge either "There is no difference" or "Sound B undergoes no distortion," and abscissa for the shifting of distortion in c.p.s. Taking the shifting amount corresponding to 50% of judgments as the value of allowance, we see that 7.2 c.p.s. for positive shift and 6 c.p.s. for negative are the values wanted for. These values of allowance are only the sample mean, and there is a scope for infering the population parameter. But a distribution function, to which the distribution of the point thus determined must follow, is not yet cleared up definitely, thence the authors leave off further discussions.

The percentage of judgments at null distortion does not perfectly amount to 100%. This is likely to be caused by the fact that the continuous speech keeps varying in their attributes or contents, and it becomes hard therefore to get hold of its natural colour definitely. It must be very hard for subjects to image to himself an unvariable natural quality of voice for continuous speech which is doomed to vary incessantly. But this deterioration in percentages can be improved more or less by widening the frequency interval of steps used as will be shown hereafter.

Comparison of allowances for various sounds

Vowel and musical sound in sustained condition

The allowances for sustained vowel "A" voiced by a male at the pitch of 200 c.p.s., and that for the sustained musical sound of violin played at the pitch of 400 c.p.s., are measured respectively. The procedures of experiment are the same as before, except that at first each sound is reproduced here, and finally that the frequency interval



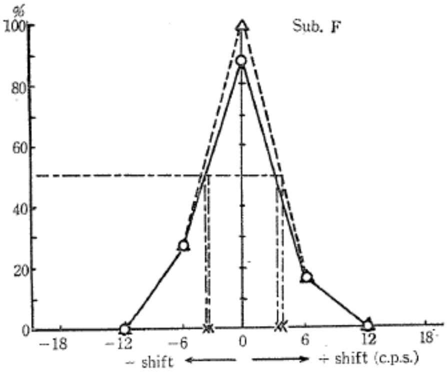


FIG. 28 (*left*). Comparison of values of allowance between sustained vowel "A" and violin for subject Y, where ○ denotes violin and △ vowel.

FIG. 29 (*right*). Comparison of the results of allowance between violin and vowel "A," under the sustained condition, for subject F, where sign ○ denotes violin and △ vowel.

TABLE 10. Values of Allowances for Both the Sustained Vowel "A" and Violin in c.p.s.

Subject	Direction of shift	Values of allowance	
		Vowel	Violin
Y	+	5.4	6.6
	−	5.8	8.7
F	+	3.6	3.1
	−	4.2	3.8

between neighbouring steps of distortion is here 6 c.p.s. The results of measurement are shown in Fig. 28 for subject Y, and in Fig. 29 for subject F. Representation of the co-ordinates is the same as before. The difference of values of allowances between two sounds can not be concluded, for the deviation between individuals is too large. The values of allowance obtained from the graphs are tabulated in Table 10.

Continuous speeches voiced by male and female

In this case, the continuous speeches are also reproduced from the tape-recorder. The talkers are a young male, whose vocal range belongs nearly to barytone, and a young female, whose vocal range is in alt. And this time, for every step of distortion, are observed measurements of ten times for each observer and for each talker. Other procedures and conditions of experiment are just the same as before. The result is shown in Fig. 30 for subject F. The result of another subject Y, is quite the same. In this case, the percentages at null shift are considerably improved comparing to the result of the same kind of experiment described above. The values of allowance obtained from the graphs are tabulated in Table 11, where the values of female are somewhat smaller than those of male. Comparing with the case of sustained vowel, there are some differences for subject F, but scarcely for subject Y.

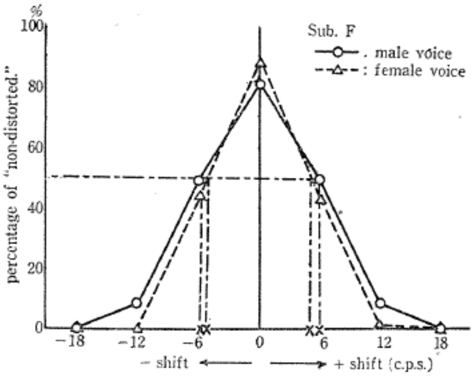


FIG. 30. Comparison of the results of allowance measurement between male and female continuous speech for subject F.

TABLE 11. Values of Allowance for the Continuous Speech Both of Male and Female in c.p.s.

Subject	Direction of shift	Value of allowance	
		Male	Female
Y	+	6.0	5.2
	−	5.5	5.5
F	+	5.7	5.0
	−	5.7	5.1

Continuous music

From violin sonata "La Folia" by A. Corelli, some of the representative modes of playing are picked up and arranged, and reproduced from the tape-recorder. Other methods of experiment are the same as in the preceding. One of the results is illustrated for reference in Fig. 31 for subject F, and the values of allowance, taken from the graphs, are tabulated in Table 12. The values shown there are markedly larger than that for continuous speech. As the difference of allowances between sustained vocal and musical sounds are not likely of decisive nature, it

TABLE 12. Values of Allowance for the Continuous Music of Violin in c.p.s.

Subject	Direction of shift	Value of allowance
Y	+	10.8
	−	10.5
F	+	9.6
	−	9.6

may be suggested here that in alterations especially as to pitch and loudness, the continuous music is more striking than the continuous speech, and in addition to this, as to the familiarity with signals used, the music may be of lower grade than speech, for the subjects employed in this case.

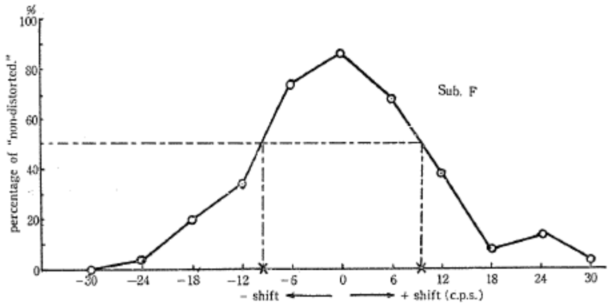


FIG. 31. Allowance of C.S.D. for continuous music in case of subject F, values of which correspond to the shifting frequencies in c.p.s., marked by the sign × on the abscissa.

SUMMARY

For convenience's sake, we will denote in itemizing some relatively novel facts brought to light in our studies.

In articulation study

1. Speaking in the most general sense, *phones with pitched timbre* are higher at the start in quality response than phones with unpitched timbre, but the formers come to differ more heavily from the influences of bigger distortion (C.S.D.) than the latters.
2. Vowels in phonated utterance (so-called voiced vowels) are quite the same with vowels in whispered utterance (whispered vowels).
3. The relation of the vowels with the consonants in natural utterance is also quite similar.
4. Though anyone of all consonants does not extend itself to show explicitly its pitchedness measurable, yet there seems to lie among consonants certain stages of sonority, the activeness of which has some concerns with voice element in speech phones.
5. For example, we can trace an apparent difference in quality characteristics between semi-vowelic and impure groups and also a similar between impure and pure groups.
6. The phenomena described in 1, 2, and 3 items are only perceived in this particular distortion (C.S.D.). Any other distortion, even the synchronous distortion in recording system (R.S.D.) can not bring about them.
7. By adopting a logatome of open-syllable type, we can make a verification with facility about the existence of correlational connection between component qualities in the rôle of perception of speech phones.
8. Beyond the extent that the (physical) degradation of consonants themselves take part directly in the hearing loss (quality loss) of them, the effect of vigorous degradation of vowels come to exert some subjective influences upon the hearing of consonants.

9. As to the reverse relation, the intervening action of consonant upon the hearing of vowels is too rare to be perceived evidently. But in the observation of glided vowel "jO," we can remark that the transitory part (part "j") of it acts in such a way as to safeguard the hearing of its stationary part (part "O") against the destructive power of distortion. The similars are found with any glided vowels, but in slighter degrees.

10. Correlations between consonant and vowel qualities are scarcely perceived only in the biggest regions of distortion of any other types than this (C.S.D.). So far as this distortion (C.S.D.) is concerned, the correlational connection is established at the start of the distortion.

11. Articulation characteristics in case of phonated utterance can have "pit" effect on its shoulder. In case of whispering, there seems to be none.

12. Item 11 looks to have some concerns with item 10. Correlational connection associated with the "pit" effect reveals itself in somewhat pulsatory movement.

In mishearing study

13. Quality analysis is not completed without going into the study of mishearing phenomena. The number of articulation and the characteristic of it remain to show the mere outline of quality, not showing its sheer content. Mishearing study can touch more or less the content of quality.

14. Judging from the observation of mishearing, the differences between phonated and whispered vowels are not essential. It is possible only thereby to deduce that the degree of affinity of vowels undergoes some changes.

15. The difference of characters of consonants in phonated and whispered utterances seems to be essential, judging from the observation of mishearing.

16. As to the consonants in phonated condition, there dominates the mishearing within group, and mishearing between groups is predominate with the consonants in whispered condition.

17. The essential properties of consonants reflected in the mishearing phenomena, are represented to the point by triangular or pyramidal figures.

In pitch study

18. When the sustained vowels come to lose their harmonic structures, resulting from the fact that every component of them is uniformly shifted on the frequency continuum by distortion (C.S.D.), their pitches, so long as they are subjectively measurable, begin to rise or fall according to the directions of distortion.

19. Pitch-attribute of distorted vowel becomes by and by less intensive with the gradual increase of distortion, and finally it comes to bear the impress of unpitched and somewhat noisy appearances.

20. Pitch change caused by the distortion, is less than the shifting frequency of distortion, so far as it is subjectively measurable.

In loudness study

21. The variation of loudness-attribute under this distorting condition is not so noticeable as it is concluded significantly in statistical sense.

In allowance problem

22. As values of allowance of distortion (C.S.D.), 7.2 c.p.s. for positive shift and 6 c.p.s. for negative shift, are obtained as a mean of a sample of considerably large size.

23. Difference of allowances between vocal and musical (of violin) sounds, both are sustained, is not detected conclusively.

24. Continuous speech of a female presents somewhat smaller values of allowance than that of a male.

25. For the continuous music (of violin), considerably larger values than that for continuous speech are obtained.

Acknowledgement—The authors wish to express appreciation of assistance and advice given by Mr. K. Tsuchiya and Mr. H. Mori in the Faculty of Engineering, and also wish to thank to Mr. S. Nomura in the Faculty of Literature of Nagoya University for his useful, advicing discussion during the whole course of this work ; and finally the direct aids in the experiment of Mr. S. Saito and Mr. T. Yamashita in our Laboratory of Voice- and Speech-Engineering, and the laboursome assistances and studies offered by Mr. Tatsuo Nagase and Mr. Joji Kato in the Bureau of Telecommunication of Tōkai-District who have achieved an important part in our systematic studies on quality problem are also gratefully and profoundly appreciated.