CHARACTERIZATION OF TRANSMISSIONAL DISTORTIONS IN SPEECH COMMUNICATION

YOSHIYUKI OCHIAI and TOHRU YAMASHITA

Department of Electrical Engineering

(Received October 13, 1954)

As an application of "quality-transmission" principle we tried here a sort of quality study, showing a method of characterization of transmission distortions in speech communication systems by means of two-dimensional representation of articulation-quality. We are interested in the marked difference of phonal characters between stationary vowels and transient consonants, of which articulation qualities may be differently and characteristically influenced by distortion systems according as the type of the distortions varies. This fact is of first importance in our present study.

Notation on Quality

By the prescribed method of symbolic description, we can define the transmission quality by the form

$$\boldsymbol{H}\cdot\mathbf{D}(X)=\Xi',$$

where $\boldsymbol{H}$ signifies the subjective operation of recipient side as to the sensation and judgment, and $\mathbf{D}$ means the transmission operation in physical sense, X the speech sound signal in physical sense, Ξ' the similar in subjective sense. In the same manner, we can define the speech quality by

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

And as an index of quality transmission which is our main subject, we have

$$\Gamma=\frac{\boldsymbol{H}\cdot\mathbf{D}(X)}{\boldsymbol{H}(X)}.$$

Suppose now that we are in need of the t_1-judgment, a judgment on phonal (or phonemic) quality, and further that we consider now exclusively the syllable- and sound-articulation. Naturally the sound construction of syllabic words in speech entirely depends on the kind of language. For an Europenan language we take as its main representative $X=CVC$, and for Japanese as its main representative $X=CV$, where C means consonant phone and V vowel phone.

Quality Product

If we suppose that the quality phenomena of speech sound of compound structure are to be treated as being quite independent of combination of its component

phones, we can assume the following formal operation:

$$H\cdot\mathrm{D}(X)=H\cdot\mathrm{D}(CV)=H\cdot\mathrm{D}(C)\times H\cdot\mathrm{D}(V),$$

and

$$H\cdot\mathrm{D}(X)=H\cdot\mathrm{D}(CVC)=H\cdot\mathrm{D}(C)\times H\cdot\mathrm{D}(V)\times H\cdot\mathrm{D}(C).$$

Consequently it goes for Japanese syllable-units

$$\Gamma(J)=\frac{H_{t_1}\cdot\mathrm{D}(X)}{H_{t_1}(X)}=\frac{H_{t_1}\cdot\mathrm{D}(CV)}{H_{t_1}(CV)}=\frac{H_{t_1}\cdot\mathrm{D}(C)}{H_{t_1}(C)}\cdot\frac{H_{t_1}\cdot\mathrm{D}(V)}{H_{t_1}(V)},$$

and for European syllable sounds

$$\Gamma(E)=\frac{H_{t_1}\cdot\mathrm{D}(X)}{H_{t_1}(X)}=\frac{H_{t_1}\cdot\mathrm{D}(CVC)}{H_{t_1}(CVC)}=\frac{H_{t_1}\cdot\mathrm{D}(C)}{H_{t_1}(C)}\cdot\frac{H_{t_1}\cdot\mathrm{D}(V)}{H_{t_1}(V)}\cdot\frac{H_{t_1}\cdot\mathrm{D}(C)}{H_{t_1}(C)}.$$

From these relations, we lead at once the following forms respectively

$$\Gamma_s(J)=\Gamma_c\Gamma_v,$$
$$\Gamma_s(E)=\Gamma_c\Gamma_v\Gamma_c,$$

which mean that the syllable quality is obtained by producting its component qualities. For Japanese case, the syllable articulation is equal to the product of consonant- and vowel-articulation. It is naturally supposed that the notation

$$\Gamma=\frac{H\cdot\mathrm{D}(X)}{H(X)}$$

signifies not only the conception of quality transmission but also even the quantified quality itself.

Distortion Characteristics in Quality Responses

When we interpret the "quality transmission" as a subjective appereciation concerning the transmission effect upon the signal sound, we shall be able to connect the distortion characteristics with the quality responses. Thus, if the phone system of signal sounds (X) is given, and the distortional function of transmission system (D) and further the judging operation of hearing subjects (H) are also given, then we can obtain the relation of the quality function (Γ) versus the variable (d) of the stated distortion (D) of pure and simple nature.

We have

$$\Gamma_s=F_s(d),\qquad \Gamma_v=F_v(d),\qquad \Gamma_c=F_c(d).$$

When the distortion is considered, as is usual, in two directions, we have

$$\Gamma_s(+)=F_s(+d),\qquad \Gamma_s(-)=F_s(-d),$$
$$\Gamma_v(+)=F_v(+d),\qquad \Gamma_v(-)=F_v(-d),$$
$$\Gamma_c(+)=F_c(+d),\qquad \Gamma_c(-)=F_c(-d).$$

In the extent of the distortion that allows the validity of quality-product principle, the following relation can be established

$$\Gamma_s(J) = F_v(d) \cdot F_c(d).$$

Now, if we plot the values of Γ_s on the plane determined by the two coordinates of F_v and F_c, we shall have a locus of quality function, which is explicitly independent of the variable d, of which the representation is called F_v-F_c(or V-C, for the sake of simplicity) expression. This V-C representation is more convenient than the usual representation of $F(d)$.

Quality Characteristics of Actual Distortion of Simple Nature

Basing upon the fundamental materials of articulation responses caused by several types of distortion, we can show real examples of V-C diagram in Fig. 1 where the distortons are considered relatively simple and pure, and are arranged in such an order: attenuation, band elimination, masking (by noise of pure tone of 1,000 c.p.s.), over-amplification, revolutional synchronous- and carrier synchronous-distortion. The types of distortion we meet in actual transmission systems are not at all restricted to these: these distortions given here are only what we have set to work in studying their quality responses. By observing the characteristics in Fig. 1 we notice: nearly all characteristics seem to find themselves only in the domain of upper triangular area of V-C diagram (refer Fig. 2) that means: this group of distortions can be characterised by the fact that it causes a greater loss of consonant quality than that of vowel quality. As far as our experience with distortion is concerned, we cannot find any distortion that clearly falls in the lower triangle domain (refer Fig. 3) which says: a type of distortion which predominantly influences the vowel quality is not found actually. Only the synchronous distortion of negative shift type is nearly in the neighbourhood of the bisector of V-C area, starting, however, from the upper region, crosses the bisection-line in the mid-course, and ends by invading a little the lower region. This distortion is worthwhile to be studied: when the distortion is controlled restrictedly to some certain extent, its quality response comes to pass in such a way that it yields a greater loss in consonant quality than in vowel; meanwhile, as long as the distortion is increased beyond certain extent, its quality response begins to reverse, causing this time comparatively a greater loss in vowel quality than in consonant. Because, the bisector of VC rectangle justly corresponds to the balanced type of distortion as to the consonant- and the vowel-quality.

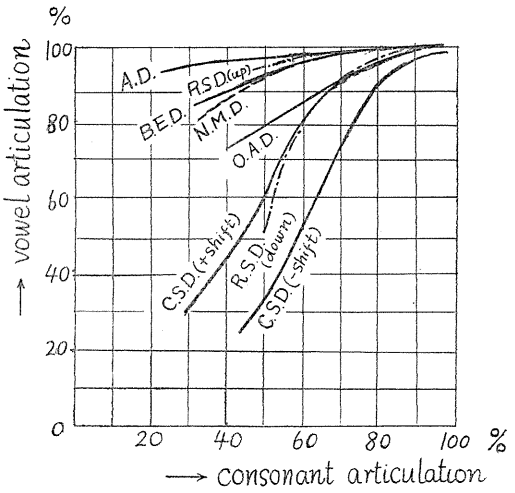

FIG. 1. Characteristics of several types of distortion represented by some performances of quality response. In the diagram the distortions are investigated: A.D.; Attenuation distortion. B.E.D.; Band elimination distortion. N.M.D.; Noise masking distortion. O.A.D.; Over-amplification distortion. C.S.D.; Carrier synchronous distortion. R.S.D.; Revolutional synchronous distortion.

To catch hold of the distortion characteristics by *V-C* representation means to make use of the phenomena in which the working responses of distortion reveal themselves very differently according as the phone is stationary or transient. As a variety and diversity of speech sounds in phonal system in a stated language brings influences upon quality response of distortion, so the difference of pronounciation or vocalization of the uttering subject may also cause some change of quality response which we will describe below.

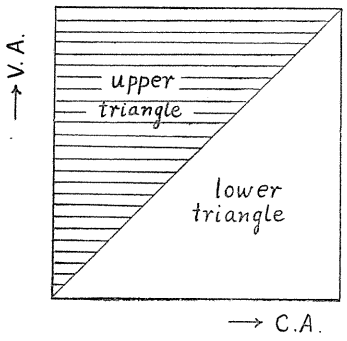

FIG. 2. Explanatory diagram of *V-C* representation.

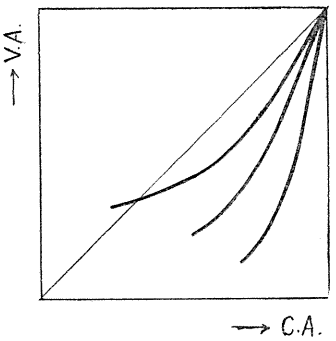

FIG. 3. Explanatory diagram of characteristics of some sorts of distortion that were never found hitherto.

Effect of Mode of Utterance

To detect the effect of utterance, we will take first two cases: one, usual voiced utterance; the other, whispered or voiceless utterance. And, as distortions by which the difference of uttering mode is detected, we choose here only two: one, carrier synchronous distortion (CSD); the other, revolutional synchronous distortion (RSD). This time, we will study the effect of distortion more closely in all two directions.

Effects of Voicing and Whispering

The result obtaind for the distortion CSD is given in Fig. 4, and that for RSD in Fig. 5. The change of characteristics due to the differnce between voicing and whispering can be considered as being in relation of parallel displacement: the characteristics due to whispering are in general placed in the left side (that is, fairly in the upper triangle domain) to those of voicing. In either mode of utterance the positions of characteristics of the distortion of positive shift type is left to those of negative shift: the most conspicuous influence takes place in the case of negative shift distortion under the uttering condition of voicing; on the contrary, less grave influences are obtained for the positive shift under whispering condition. With the distortion RSD the affair is quite different; less serious influence happens in the step-up distortion under the voicing condition and the most serious in the step-down distortion under the same condition of voicing: the two characteristics due to whispered utterance come to lie between those two characteristics in voicing. Speaking more closely, the up-distortion characteristics are of

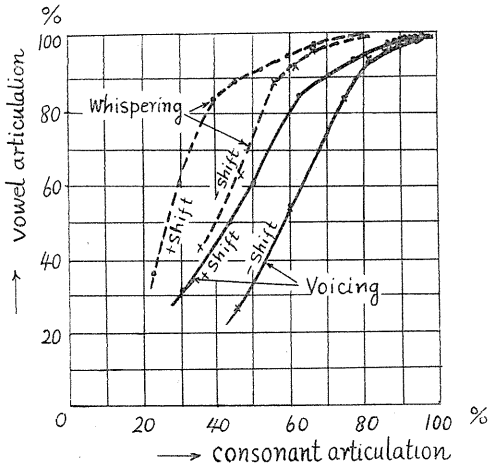

FIG. 4. Characteristics of distortion CSD observed in its two directions both of positive and of negative shift and under the two conditions of utterance of whispering and voicing.

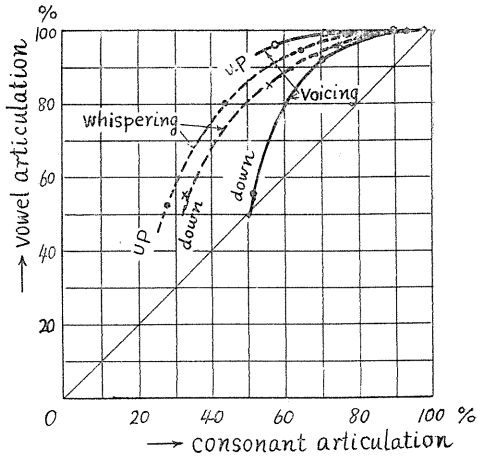

FIG. 5. Characteristics of distortion RSD in its two directions (up- and down-disortion) and under the two modes of utterance of voicing and whispering.

predominant consonant-loss type something like that of attenuation distortion; and the down-distortion characteristics are not far off from the bisector like the CSD of positive shift type in voiced utterance. It is to be noticed that the up- and down-distortion of the same distortion RSD appear as quite different in the speech phone in voiced utterance; in the speech phone by whispering they appear as rather similar.

Further Studies in RSD

In order to account for the intervening characteristics of whispering phenomena between the two characteristics in utterance, we will try still more studies: the study

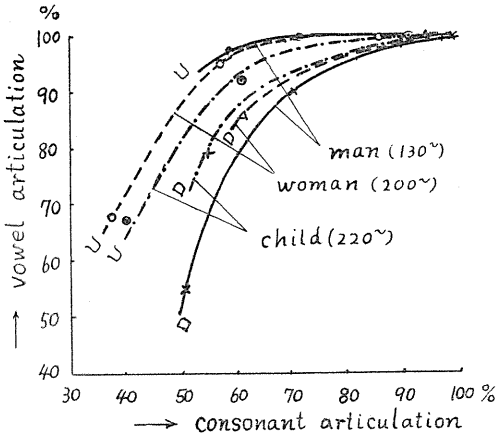

FIG. 6. Characteristics of RSD due to the change of uttering subjects whose pitches are variable.

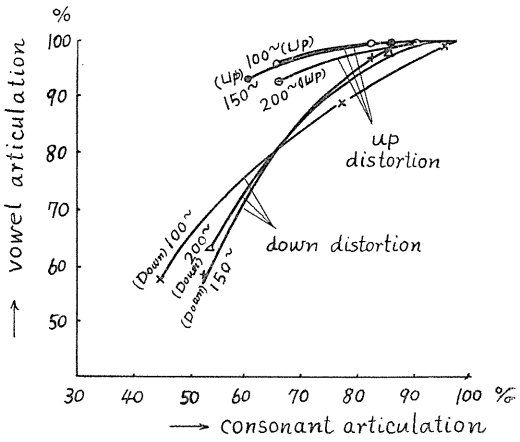

FIG. 7. Characteristics of RSD due to the change of uttering pitch in the same subject.

of the effect of pitch difference, and the study of the effect of difference of uttering subject. In the former test, three pitches of the same male subject, having respectively 50~ difference each from other, such as 100~, 150~, 200~ are chosen, and the characteristics obtained are shown in Fig. 6. And for the latter test, three subjects, a young male, a young female and a boy are selected, whose pitches are not well matched, being left utterly to their free selection: as a result, the pitch employed by the male is 130~, by the female 200~, by the boy 220~, the voice of the female is something like contralto, and the boy is before the change of voice (mutation). The characteristics of three subjects are shown in Fig. 7. We do not think this test can detect the pure effect of vocal quality. For, the pitches employed were not well matched; therefore we must think that the characteristics in Fig. 7 show probably most powerfully also the effect of pitch difference. Only the characteristics due to down-distortion in this figure are a little out of order. But we can assert generally a considerable good ordering in the arrangement of characteristics. As the result of these experiment, we may arrange possibly the characteristics shown in Fig. 8 which is nothing but an explanatory diagram: The distortion RSD probably in such a way that the diversity between characteristics due to up- and down-distortion is maximum in the case of low pitched voice and is diminished as the pitch is increased in height, and moreover it becomes minimum when a whispering is used, coming to lie always between the two characteristics of extreme diversity in lower pitched voicing.

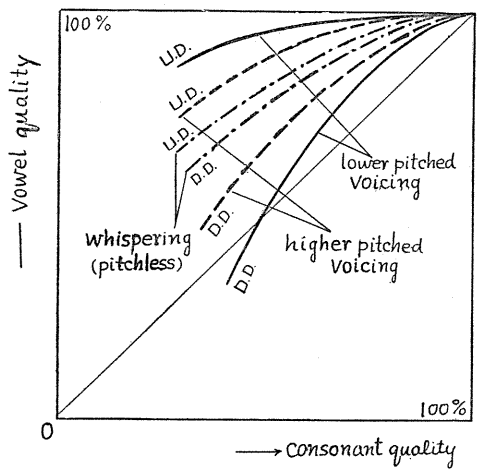

FIG. 8. Explanatory diagram showing the trend of characteristics of RSD (in two directions of up and down) influenced by the change of pitch in voicing and compared with the influence of whispering.

Conclusion

The problem of characterization of distortion, that is, a problem of classification of distortion in transmission system has been discussed from a viewpoint of phoneme quality. We have tried to express the type of distortion in the proportional representation between the consonant- and vowel-quality which we called the *V-C* representation. The Japanese language is of the type of balanced distribution as to the consonants and vowels, as shown in the previous paper: and its syllable-units are composed of *C* and *V*. *V-C* representation that is, therefore, most interesting and important in the phone system of our language. But the idea which underlies this *V-C* representation is essentially full of meaning in all other languages. Because it is based upon the fact that the type of distortion can be characterized by the difference in its working performance for speech phones of transient and stationary nature. Several distortions have been examined and their

characteristics have been obtaind by such a representation. However, considering the problem as one of timbre appreciation, we have primarily studied the distortions CSD and RSD which are considered as typical timbre distortions. We have detected furthermore the effect of vocalization, though it is limited only to the effect of pitch change. By this investigation we are inclined to conclude that the curves of characteristics of distortion given here must be generally admitted with certain range of margins which come from several factors due to the changes of uttering pitch and uttering subject. It seems to us that some feeble effect of the vocal quality upon phoneme quality (that must be the subtle subject in quality study) is more or less detectable. But the very problem of the performance of distortion in an aspect of vocal quality will be studied in another paper. Also the problem of compound distortion is put off to the future study.

References

1) Y. Ochiai and T. Fukumura: Studies on qualities of speech and voice by timbre distortion. Memoirs of the Faculty of Engineering (MFE), Nagoya Univ., Vol. 4, No. 2, Nov. 1952.
2) Y. Ochiai et S. Saito: Sur la représentation de la structure phonétique des langages. MEF., Nagoya Univ., Vol. 4, No. 2, Nov. 1952.
3) Y. Ochiai and T. Fukumura: Timbre study of vocalic voices. MEF., Nagoya Univ., Vol. 5, No. 2, 1953.