PRELIMINARIES TO ANALYSIS OF QUALITY IN SPEECH COMMUNICATION

Yoshiyuki OCHIAI and Teruo FUKUMURA

Department of Electrical Engineering

(Received October 31, 1957)

In this study we first give our reason for introducing confusion of timbre-signal in speechcommunication technology. We consider treatment of this point more important for our introduction than an ordinary quality treatment applied, for example, to articulation- and loudness-rating of communication systems. Then we proceed to analyse the mechanism of quality judgement in the confusion test carried out as our experimental study. In our judging we clarify two processes, those of identification and discrimination, the former associated directly with similarity and the latter with dissimilarity in timbre judgement. By introducing a concept of direction in confusion, we endeavor to derive from these two processes of judgement (identification and discrimination) the two confusion modes (incoming and outgoing) and to use them as the best available tool by which to dispose of confusion. Next, we stress the point that in speech-communication engineering we must advance our confusion study along a radical line — that of duality of quality; in other words, we must proceed not only with the confusion in phonemic aspect but also with the confusion in vocal aspect. The first is readily available for articulation transmission and the second for naturalness transmission either in telephonic blind communication or in broadcasting of speech programs by radio where, in the majority of cases, the aim of transmission is to transfer emotional information involved with speech and not to restrict the transference of intellectual information expressed by speech.

CONTENTS

Introduction
General consideration
System rating by articulation- and naturalness-quality
Confusion in transmission engineering
Characterization of transmission distortion
Inner processes in timbre judgement
Two proto-types: identification and discrimination
Confusion viewed through concept of

direction
Outgoing confusion
Incoming confusion
Confusion in dual quality aspect
Nature of intellectual and emotional information
Discussion
Conclusion
References

Introduction

In our previous reports, $^{1/2/3}$) where we described confusion of sustained vocalic sounds in naturalness and articulation aspects based on our original measurements, we stressed the importance of confusion study in the sense that confusion of timbre is no more than the problem of *timbre-discrimination* and that it has a positive relation to the problem of *timbre-sensitivity*. In those reports we did not completely

cover this point, thus leaving it open as an issue for future conceptional discussion, but we did introduce our own concepts of incoming- and outgoing-confusion as the most useful process for treatment of such a complicated and discursive subject as the phenomena of confusion, offering an illustration by employing and calculating the confusion data obtained in our experimental study. After reflecting upon the traditional method for treatment of confusion, we duly acknowledge that the confusion we have so far named vaguely is only an outgoing confusion by our terminology. Most significant in our quality theory is our introduction of the concept of incoming confusion itself and the fact that we are quickly led by this type of confusion to other important concepts, such as the true meaning of phonemic formant and vocal formant. In viewing confusion phenomena in this way, we were fully convinced that with the aid of these two modes of confusion we could reach our goal. We had not yet then fully explained why there must be such incoming confusion and why it is necessary for our quality theory and, further, why the outgoing confusion, which had always been a unitary concept of confusion, was no longer sufficient for our purpose. If we consider the fundamentals of judgement failure, we are obliged to give a clear and positive explanation of the mechanism of our failure to confirm sounds and to explain why our incoming and outgoing failures in judging are necessary in essence to the study of confusion. This is the main point of our conceptional study, because in clarifying the mechanism of failure we will be expected to elucidate the mechanism of confirmation itself.

General Consideration

We must consider at this time in our study of speech-communication engineering the most radical meaning of confusion because we are now forced to face this confusion problem after having finished many introductory studies on speech quality. Until we reached this confusion problem, our concern was only with the practical side of quality study. It is true that at that very time we did not think that the study of confusion had any other than a direct and practical meaning; but with the progress of our experimental study in quality aspect and by gradual maturing of our thinking, we were duly convinced that in addition to the practical meaning of confusion there is to be found another meaning, the scientific, which may be of paramount significance in the sense that through it we may come unerringly to the right interpretation of timbre perception. Without this latter meaning, any study of confusion would be but a tedious process promising no essential contribution to timbre theory. In this paper we discuss the most fundamental aspect of confusion before we step into the detailed treatment of data of actual measurements and observations in our confusion experiment.

Without touching upon the so-called conceptional study which is monotonous and uninteresting for those who are not familiar or concerned with the quality theory, it is better to commence our description by citing some actual illustrations in quality phenomena and to discuss some boundary conditions which regulate and orientate the quality phenomena in communication transmission.

System Rating by Articulation- and Naturalness-Quality

With due regard to an actual transmission-distortion in speech-communication system, the problem of system-rating is the most important of all the radical

problems in communication. There may be several attributes of quality available for this purpose; for the present we have three: Loudness, articulation and naturalness. That we most often base our system rating on the very ground of articulation is reasonable, because the aim of transmission of speech-communication system lies, in the majority of cases, in the intelligibility of speech through transmission. Such a rating, however, is merely a kind of quality estimation in communication system where the magnitude of quality to be rated is always expressed as the mean value of all the articulation values relative to all the speech-phones in the language Thus it is most natural that there are fluctuations and deviations in an articulation score which is measured for system rating. These values of articulation are, of course, applicable to system rating, and we must remember here the fact that they do reveal a one-sided estimation of quality based on the aspect of correct-or-incorrect judgement in speech-phone transmission. That the articulationquality rating based on the statistical judgement of correct-or-incorrect transmission is inadequate and not sufficient for merit judgement, will be proved and discussed later.

It is noted that any other kind of estimation of speech-communication system which utilizes a quality lower than articulation quality is unacceptable in principle in spite of the fact that the system rating in general has its origin in the field of practical application of transmission engineering.

The rating value based on articulation may perhaps be important to telephone designers, but for telephone users it would be less important. For telephone subscribers who actually use telephone equipment in their daily affairs, it is more important to know the cause and effect which governs the actual transmission phenomena than it is to be acquainted with average values of articulation data on the equipment they are now using. It is from this consideration that we now step into confusion as revealing the real phenomena in communication system in a sense that we are dealing essentially with not only the physical phenomena of transmission but also with the subjective phenomena related to both communicants, *i.e.*, speaker and hearer.

Present-day telephoning is but blind communication, *viz.*, communication between those who cannot see each other, because in ordinary telephone intercourse all visual help available in face-to-face communication is entirely cut off. Under such a condition we cannot overlook confusion in voice identification (in the sense of telephone-users) which has not yet been studied. The problem of the confusion between phonemes is, of course, important in our consideration. In short, when we try for an unerring approach to the confusion phenomena with the purpose in mind of a sharp rating of the speech-communication system, we must cover in our study these two sides of quality (phoneme- and voice-identification) from the standpoint of timbre quality.^{4) 5) 6)}

Confusion in Transmission Engineering

Let us focus our attention on what can occur during a face-to-face conversation between two persons when the hearer receives the direct voice of the speaker. Under such direct speaking-hearing conditions, there frequently arises a misunderstanding as to informational ideas notwithstanding the fact that the communicants have the advantage of watching the countenance and seeing the gestures of each

other, and also there may be some confusion of speech sounds in spite of the convenience afforded the hearer by his position favoring the watching of movements of the speaker's mouth.

Here we must employ confusion to help us detect any phoneme defect in pronunciation due to lack of phoneme training and incoherence because of local dialect, and also to aid us in evaluating hearing ability which is inadequate at the time either because of some physical defect or through excessive fatigue on the part of the hearer. As to the detection of a phoneme defect, the phonetician finds his specialized studies in the field of confusion, while the otologist can utilize the articulation test for the purpose of diagnosing the cause or causes of inadequate hearing. There is no doubt that confusion study applied to these fields of specialization is of great help because through it we can trace sharper differences and obtain more exact data than through the ordinary articulation test.

In natural sequence, we now discuss confusion as an aspect of speech communication engineering. The greatest significance comes to bear on the confusion problem when it is looked at in connection with conditions surrounding certain distortions with which we are confronted in actual telecommunication. Let us consider, for instance, the telephone conversation. More stress must be laid on telephone conversations than on face-to-face conversations because the *verbal exchange* over the telephone line must rely exclusively on hearing ability, having been refused all visual aids. Important from an engineering standpoint is the following feature. The nature and character of the distortions in communication systems reveal themselves most clearly in characteristics of confusion in phonemes and confusion in voices. That is to say, it is, in fact, the interrelation between confusion phenomena and transmission distortion that we must diligently pursue in our study.

The study of confusion from an engineering aspect must encompass as its reference the confusion study in a condition without distortion as well as the study of the confusion in a condition with distortion. What this comes down to is: The confusion study imposed on us from the standpoint of speech-communication must be treated in a general and broad sense. This includes confusion in conditions with and without distortion.

Characterization of Transmission Distortion

Our study on the interrelation between distortion and confusion is to be carried out for the following reason. Quality responses to distortion might vary according to type and magnitude of distortion. By making this relation clear, we are able to characterize several kinds of distortion, as we have shown in a previous paper covering a number of distortions of special kinds. We were not successful in clarifying these relations so long as we leaned only on quality indices which means that we were relying entirely upon the average value of quality response, such as ordinary estimation of articulation value. To illustrate, let us first take the case usually met with, having two types of distortion which, by estimation of articulation, are equal. Here we can neither characterize nor specify these two distortions by articulation quality. Next, there come those two types of distortion where quality characteristics are equal, a circumstance with which we seldom meet. In both cases we can, however, utilize the confusion characteristics for characterization

of distortion. In short, the mean articulation value can depict only the outline envelope of the quality phenomena and cannot bring out the most detailed figures of the innermost sphere to which the confusion characteristics alone can respond sharply and exactly.

Inner Processes in Timbre Judgement. Two Proto-Types: Identification and Discrimination

What do we mean when we say we can determine and localize unknown timbresignals impressed upon us under certain conditions through some particular distortions? When we consider a process through which we can properly judge timbre quality of given signals, we must assume that two processes, that of identification and that of discrimination, are carried out at the same time for the one judgement. By identifying a timbre-signal with a certain known timbre, the signal is at the same time distinguishable from all other timbres. That we can identify a given timbre-signal with a timbre already known and that we do not identify it with any other than that known timbre is surely the result of judgement of a certain type, a judgement in which we make use of some clues in timbre character. In other words, we correctly carry out the timbre-quality judgement, first, by marking the points in which the timbre given to us resembles some certain known timbre and second, by characterizing the points in which it differs from the others. As points to be marked for timbre judgement, we can list some factors in acoustical sense; but we must stress the importance of timbre-structure in such timbre judgement. When the degree of similarity between given timbre and known timbre is decreased for some reliable reason, for instance, by insertion of some distortion, then the degree of identification of the given timbre with the known timbre might diminish. In the same manner, when the degree of dissimilarity of the given timbre to the other timbres is decreased, then the degree of discrimination between the given timbre and the other timbres might also be decreased. This means that timbre confusion begins immediately at the time correct judgement is lost.

After reflecting upon one more element for our consideration, we will be ready to make an approach to our final conclusion. When we look back upon our daily experience in trying to recognize timbre as such, we are bound to feel that a completely established "map of timbre," * so to speak, might be based upon our already accumulated experiences with timbre. Our efforts are, in some cases, directed toward a gross recognition of an impressed unknown timbre as human speechtimbre or as an instrumental musical timbre, and, in other cases, toward making a finer distinction between a given timbre as the voice of some particular person or as the tone of some special musical instrument. Through the comparison possible with this very map, the impressed unknown timbre can be perceived and finally recognized. Although the mechanism of timbre perception is difficult to trace, that is to say, it is not easy to locate clues in the process of timbre-recognition, it is reasonable to assume that there is some level in the process of timbre perception where it is possible to detect an impressed timbre by coordinating it on the timbremap already established with the data accumulated within our memory through experiences with timbre. Locating a given signal on the mental timbre-map means

^{*} Here the term of "map of timbre" is used in a figurative sense only.

that the given signal is found in psychological timbre space where coordination is carried out with identification by the help of discrimination. The discriminating action has an associated and co-operative relationship with the identifying action. In other words, identification is facilitated by the co-operation of discrimination.

The timbre in referential condition, *i.e.*, under the condition without distortion, furnishes material for the imaginary timbre-map impressed on our consciousness covering data accumulated by our memory, and by referring to this mental map the similarity inherent in the timbres themselves is brought to the surface. When we introduce distortion into our study, we are led to some new conditions surrounding timbre locating in which correct correspondence is gradually lost resulting in some false correspondence which sensitively reflects the character of distortion. By carefully tracing the phenomena of such false correspondence, which we call "auditory timbre confusion," we can properly characterize and specify the transmission distortion.

As described in our previous report, (7) we could not bring out the distortion differences as represented by the ordinary quality characteristics where the quality response is expressed in relation to the magnitude of distortion. By using the representation of quality relationship between stationary sounds and transient sounds, we could, to some extent, bring to the surface the characterization of distortion, as shown in our experiments. For a more precise and comprehensive characterization of transmission distortion, the study of confusion characteristics offers the brightest prospect.

Confusion Viewed through Concept of Direction

The concept most important for an understanding of confusion in speech sounds under referential condition is one which is related to the concept of SIMILARITY Of speech sounds. This concept establishes and forms itself when the magnitude of confusion comes into balance between the two directions of confusion. As an illustration, we cite the example of timbre confusion. In such a balanced condition, confusion of "A" timbre toward "B" timbre equals the confusion of "B" timbre toward "A" timbre. This is true by actual measurement provided we have a sufficient number of observations in referential condition.²⁾ When some distortions are to be inserted the reciprocity of timbre-confusion no longer holds good in spite of a sufficient number of observations. The character of distortion reveals itself in a break-down of the trend of the confusion-reciprocity which is almost always established in the referential condition where there is no noticeable distortion. Interesting in a theoretical meaning and important even in a practical sense, it is essential that the concept of confusion-reciprocity can be given life by introducing the direction of confusion. What it comes down to is this-by this concept of confusion-reciprocity we are easily and quickly led to the two kinds of confusion, outgoing and incoming. Establishment of these two types of confusion enables us to treat with ease the analyses of confusion phenomena under distorted conditions.

Outgoing Confusion

For the purpose of illustrating the essential nature of confusion, we take an example of confusion by referring to the distortion of some kind. Assuming that a certain kind of timbre-signal "A" in some ensemble of timbres such as "A", "B",

"C", "D", etc., is impressed through certain distortion upon a listener to whom this ensemble of timbres is familiar, the timbre-signal deformed by distortion would be carried to his perception. Here the degree of deformity of the signal-pattern is correctly measured by comparing it with the original perfect pattern of "A". With the increase of distortion, and consequently with the increase of deformity of the pattern, the impression of similarity between the deformed signal given and its original perfect signal would, on one hand, decrease: and the impression of similarity between the deformed "A" and the other signals in the ensemble will, on the other hand, increase; in reversed expression, the dissimilarity in the latter, viz., the dissimilarity between the deformed and the other signals, also decreases. As a result of the simultaneous decrease in similarity and dissimilarity, we can assume a condition in which both the identifying function toward timbre "A" and the discriminating function against the other timbres begin to diminish in strength at the same moment, which results in some type of confusion outgoing from signal "A" to qualities other than "a", as expressed by using our technical terms in confusion theory. We show this implication schematically in Fig. 1.



FIG. 1. Schema of outgoing confusion from timbre-signal "A".

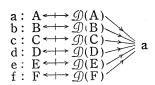


FIG. 2. Schema of incoming confusion toward quality "a".

Incoming Confusion

Under the condition of certain distortion, the outgoing confusion phenomenon described as relating to timbre "A" would occur in the same manner for each of the timbres other than "A", such as "B", "C", "D", . . . respectively. By reviewing and summarizing such outgoing confusions from "B", "C", "D", . . . , we can trace another type of confusion which also relates to timbre "A" and stands opposite to outgoing confusion—incoming confusion. From all possible individual outgoing confusions of each of the timbres "B", "C", "D", . . . wherein the confusion outgoes from the individual timbre-signals toward other qualities, we are able to single out some definite confusion types which are commonly directed toward quality "a", as schematically shown in Fig. 2.

It goes without saying that certain condition evokes this incoming confusion, as explained in the foregoing study on outgoing confusion, *i.e.*, incoming confusion is caused by the mechanism which decreases both identification and discrimination simultaneously with decrease of similarity and dissimilarity in timbre-signal configuration in distorted condition.

It is worthy of particular note that in referential condition where there is no noticeable distortion, the relation of similarity and dissimilarity is uniquely and inherently established, and that thereby, for localization of signal, the functions of identification and discrimination activate cooperatively and complementarily in

exact parallel to the relation of attractive and repulsive forces in the movement of particles. As these two functions, identification and discrimination, in referential condition, activate in such a confused way, we cannot separate one function from the other in a study of confusion phenomena where reciprocity always holds good. That we can segregate qualitatively and quantitatively one function from the other is only restricted to the condition of distortion where the reciprocity in confusion is broken down.

Confusion in Dual Quality Aspect

In the foregoing we consciously used timbre-signal as our signal in order to give to our description of confusion phenomena general formulation most universally available. It is, however, proper and natural that in the treatment of confusion we must cover all phases, dealing not only with phonemic but with vocal confusion. This attitude stems from the dual aspect of quality. As we have already shown in our interpretation of timbre, there is phonemic quality on one hand, and there is vocal quality on the other. It was only toward the phonemic quality that all developments in telephone engineering were directed in connection with quality-rating consideration. With reliable data, in our previous papers, we could point to the fact that in quality judgement vocal quality is quite different from phonemic quality. In other words, we verified that vocal quality can be uniquely oriented and located as standing in fine contrast to phonemic quality.

What can be immediately inferred and quickly understood is that this vocal quality must be neither *subordinated* nor *superordinated* to phonemic quality and *vice versa*. We might say, in short, that these two qualities represent, so to speak, two faces of the same shield. They do not exist on different layers; they are of the same status, and we cannot determine a higher or lower relationship between them. Insofar as timbre quality is concerned, we cannot but accept a coordination of this kind for quality ordering.

Nature of Intellectual and Emotional Information

We have finally come to that stage where we must consider the difference between the nature of *intellectual information* and that of *emotional information* in an aspect of timbre quality. In a previous paper, after we had differentiated between vocal quality and phonemic quality in an abstract sense, we were faced with the reality that it was difficult to separate one from the other when treating with the actual concrete speech sounds.

In exactly the same manner it might be difficult to decide whether a certain given message belongs to intellectual or to emotional information. As difficult as that may be, we dare to particularize General Information into two classifications, intellectual and emotional. We present two real examples. In the majority of cases, intellectual information is abundant in business telephone conversations, and it is for this reason that we must priorily stress the articulation viewpoint when designing telephone-communication system. It is rather on the contrary that in social telephone conversation and in speech in drama and singing by radio, we need stress more the emotional nature of information. Thus we need emotional information as well as intellectual information for an understanding of all types of speech in communication.

Of course, we do not believe that timbre is the only deciding factor in the

classification of information; but we do know that in this connection we must accept the priority claim of timbre. After many turns and twists in thinking, we are now inclined to believe that phonemic quality is the sole contributor to the establishment of the intellectual nature of a message and that it is the main problem of articulation. Of equal importance is vocal quality, which is the principal contributor to formation of the emotional nature of a message, must be the first consideration in a study of naturalness. We consider that the aspect of naturalness is the coloring constituent of emotional expression.

Discussion

It would seem to us that the true concept of quality in speech communication was long dormant in the field of communication technology, to be born in the very period when the highly advanced technique of transmission engineering, with its remarkable progress in colossal strides, came along to announce the arrival of the Modern Communication Age. It is indeed a significant fact that the brilliant achievements in the development of telecommunication engineering have been the immediate source which gave life to the innermost concept of communication. In other words, insofar as communication is concerned, the extremely rapid growth of the technical side of telecommunication was a strong motive for a series of psychological researches which were forced to acquire a serious approach to the very core of communication itself. We consider it natural from a historical standpoint that psychometrical studies were carried out through the medium of the estimation problem of the so-called transmission effect of speech, not being satisfied with the physical study of transmission phenomena of speech. In adhering to the radical thought on quality of speech and voice, we followed historical necessity. It is needless to say that the theme of psychological estimation of communication quality has something to do with the so-called philosophy of communication. Philosophy, or at least a philosophical consideration, is not to be found or needed in any other branch of technology.¹⁰⁾

Timbre in acoustics in general is difficult to deal with and at the same time, because of its delicacy, is hard to describe. Timbre perception is the most delicate problem of all human auditory sensations.

In the midst of these considerations, it can be easily understood that our researches based upon timbre quality should be planned neither by mere speculation without stubborn support of experimental data, nor by the simple accumulation of experimental data with no structure of radical and basic contemplation. For this reason we have usually carried out our researches in the following order: Before beginning an experiment, we think and think, conjuring up every possible angle. After making a decision, we proceed with our experiment. Following completion of the experiment, we again studiously inspect all experimental data in search for its interpretation.

Our research routine for speech and voice was, in fact, an off-spring of our study of duality of quality stemming from our original reflection on communication, and the methodology of representation of timbre-pattern had its origin in the same source. As to whether our idea of duality of quality really has merit, this must be proved by actual experiment. After having finished many lengthy experimental studies, we can say that, as a whole, our conceptional study is not wrong, although to bring it to perfection some refinement of theoretical development and some supplementary studies may be needed.

Conclusion

Timbre measurements, physical and psychological, are both difficult. Timbre measurement in subjective sense, *viz.*, timbre-quality measurement, must be sensitive and exact because in precision and exactness subjective measurements in general are far inferior to physical measurements. We are forced to adopt this view particularly when we consider the timbre quality subjectively obtained in connection with the timbre-pattern physically obtained. It is for this reason that the confusion study appears to be the most suitable for timbre-quality measurement. That the slightest variations in timbre construction have a considerable influence upon hearing is reflected in the phenomena of confusion. In other terms, the confusion as *timbre sensitivity* can, in some measure, be detected by such measurements, and this is quite reasonable because, through this confusion study, we deal with a sort of differential sensation of timbre which is most available to quality theory.

As an actual problem in the practical field of transmission engineering, this confusion is really important. It is thus with transmission of speech in telephone intercourse and also with transmission of speech programs in radio broadcasting.

As to the practical applicability of our quality-aspect study, it is important and necessary to treat this confusion problem not only from the side of phonemic quality but from the side of vocal quality. Telephone users in telephone conversation cannot see each other, and they must rely only on the impression of voice effect which is transmitted through the communication system. For radio dialogue in broadcasting, we may perhaps lay more stress on the voice quality than we can for telephone dialogue.

References

- 1) Y. Ochiai and T. Fukumura: Timbre Study of Vocalic Voices Viewed from Subjective Phonal Aspect. Part I—Preliminary Studies on Naturalness and Articulation Qualities Actually and Directly Measured with Respect to Band-Eliminating Distortion. M.F.E., Nagoya Univ., Vol. 8, No. 1, May, 1956.
- 2) Y. Ochiai, T. Fukumura and A. Hattori: Timbre Study of Vocalic Voices Viewed from Subjective Phonal Aspect. Part II (a)—Preliminary Studies on Timbre Confusion of Phoneme and Voice. M.F.E., Nagoya Univ., Vol. 8, No. 2, Nov., 1956. Part II (b)—Simple Description of Confusion Phenomena. M.F.E., Nagoya Univ., Vol. 8, No. 2, Nov., 1956.
- 3) Y. Ochiai and T. Fukumura: Timbre Study of Vocalic Voices Viewed from Subjective Phonal Aspect. Part III—Generalized Treatment of Timbre Confusion. Section A—Phonemic Confusion. M.F.E., Nagoya Univ., Vol. 8, No. 2, Nov., 1956. Section B—Vocal Confusion. M.F.E., Nagoya Univ., Vol. 8, No. 2, Nov., 1956.
- 4) Y. Ochiai: Mémoire sur les Sons des Voix Humaines. M.F.E., Nagoya Univ., Vol. 4, No. 1, July, 1952.
- 5) Y. Ochiai: Transmission of Quality. M.F.E., Nagoya Univ., Vol. 6, No. 2, Nov., 1954.
- 6) Y. Ochiai and T. Yamashita: On Timbre Quality. Part I—Introductory Consideration. M.F.E., Nagoya Univ., Vol. 7, No. 1, May, 1955.
- 7) Y. Ochiai and T. Yamashita: Characterization of Transmissional Distortions in Speech Communication. M.F.E., Nagoya Univ., Vol. 6, No. 2, Nov., 1954.
- 8) Y. Ochiai and T. Fukumura: Timbre Study of Vocalic Voices. M.F.E., Nagoya Univ., Vol. 5, No. 2, Sept., 1953.
- 9) Y. Ochiai and T. Fukumura: Study on Fundamental Qualities of Vocalic Timbre by Rotational Synchronous Distortion. M.F.E., Nagoya Univ., Vol. 8, No. 1, May., 1956.
- 10) Y. Ochiai: Reflection on the Basis of Transmission in Communication (in Japanese). Unpublished.