

# TIMBRE STUDY OF VOCALIC VOICES VIEWED FROM SUBJECTIVE PHONAL ASPECT

## PART II (a)—PRELIMINARY STUDIES ON TIMBRE CONFUSION OF PHONEME AND VOICE

YOSHIYUKI OCHIAI, TERUO FUKUMURA  
and AKIRA HATTORI

*Department of Electrical Engineering*

(Received October 10, 1956)

In the present report we deal with the confusion problem of vocalic timbre. Guided by the dual aspect of quality in timbre, we developed our study along an original line of inquiry treating not only *phonemic confusion* but also *vocal confusion*. The subject of vocal-confusion is novel and, because it stands in close relationship to transmission of the voice-quality in speech-communication engineering, is of inexhaustible interest. Complicated and divergent as the phenomena of confusion are, they are none-the-less quite available for, first, the interpretation of quality-loss, such as articulation- and naturalness-loss, and second, they give precious clues for perfect understanding of vocalic timbres themselves for the simple reason that the problem of confusion is but a problem of *timbre-discrimination*, and for the further fact that an experimental study of confusion also supplies us with an effective system of methods of inquiry on *timbre sensitivity* which is of vital importance. On the other hand, we must sharpen our observations on confusion phenomena by introducing *incoming confusion* as well as employing *outgoing confusion* by the help of which our description of the phenomena is animated and our treatment systematized.

### CONTENTS

Introduction	—Confusion-modes of phonemic and vocal confusions
Research aim	Phonemic-confusion characteristics
Procedure of experiment	Vocal-confusion characteristics
Commentary on timbre pattern	Special observation on confusion
Some remarks on the description of experimental results	—Judgement of sex-voice-quality
Simple description of confusion phenomena	Discussion
Timbre confusion in normal condition	Summary
Graphical representation of timbre confusion as a whole	Postscript
	References

### Introduction

In the first part of our timbre study viewed from subjective phonal aspect,<sup>1)</sup> we dealt primarily with quality characteristics in dual aspect, utilizing thereby only one distortion of Band-Eliminating type (BED). Then we stepped into the next study, treating all possible individual quality characteristics, *i.e.*, articulation and naturalness characteristics for individual phonemes, and doing the same for individual voices. There we further attempted to interpret those quality characteristics

which are functions of distortion, making allowances for the various representations of timbre patterns.

In this second part of the timbre study, we must try a more sensitive approach to the quality problem by adopting a compendious and pertinent method to grasp an understanding of more minute and concrete phases in quality phenomena. Thus we come to a problem of immediate importance—that of timbre confusion. By help of this understanding, we can substantially explain the behavior and movement of quality characteristics and can elucidate the real content of quality-loss caused by the distortion under consideration. This study is based upon data of the original experiment carried out at our Laboratory throughout the winter of 1953.

### Research Aim

Our experiment on auditory confusion is, in fact, an offspring of the development in our Laboratory of the study on quality, which we touched upon in our Introduction. Through this experiment and by careful consideration afterwards we obtained results from which we could bring into focus the three following important points, either direct or indirect, as the aim of our present research.

(1) To gain the best interpretation of characteristic in quality-loss phenomena toward the distortion BED. There is no phase other than confusion appearing so prominently in the quality-loss phenomena.

(2) To verify the theory of timbre quality itself in its most fundamental meaning and to elucidate the naturalness-loss characteristic on the same basis as we explain articulation-loss characteristic, we must be in a position to deal with vocal confusion which stands uniquely responsible for naturalness quality phenomena just as we deal with phonemic confusion which possesses unique responsibility for articulation quality phenomena. In short, to get a real understanding of both naturalness and of articulation on the common basis of timbre viewpoint, vocal and phonemic confusion must be equally accessible and as easily recognizable under the consideration that, in the nature of things, each of these confusions is but a sort of *timbre discrimination*. Thus the problem of timbre confusion becomes something which is no more than a problem of some kind of *timbre sensitivity* when viewed from timbre aspect.

(3) To re-examine our method of timbre representation in a more practical sense, we have already devised and developed two sets of representation of timbre: *Phonemic pattern* and *vocal pattern*. We must now re-examine whether such representations are adequate or not by actually following a close comparison of subjective quality-discrimination results with configurations of timbre patterns physically obtained. In short, this comparatively short treatise on auditory confusion plays only a preliminary and prelusive role toward a more comprehensive study promised for the future.

### Procedure of Experiment

We can omit here a description of our experimental procedure as it is the same as given in the preceding report—Part I. However, we must again note that we employed five calling voices, two adult males, two adult females and one adolescent boy, all with orthoepical ability. These subjects were requested to utter five Japanese oral vowels at the voice pitch of 280 cycles and on the voice level

of about *mezzo forte*. Further, for timbre judgement, four listening subjects were employed, all young male students with normal hearing acuity. For timbre signals to be tested, twenty-five kinds of timbre in total, five vowels by five voices, were prepared. Instead of listing them all, for convenience we show in Fig. 1 only the mean phoneme pattern for individual vowels averaged for all five voices and the

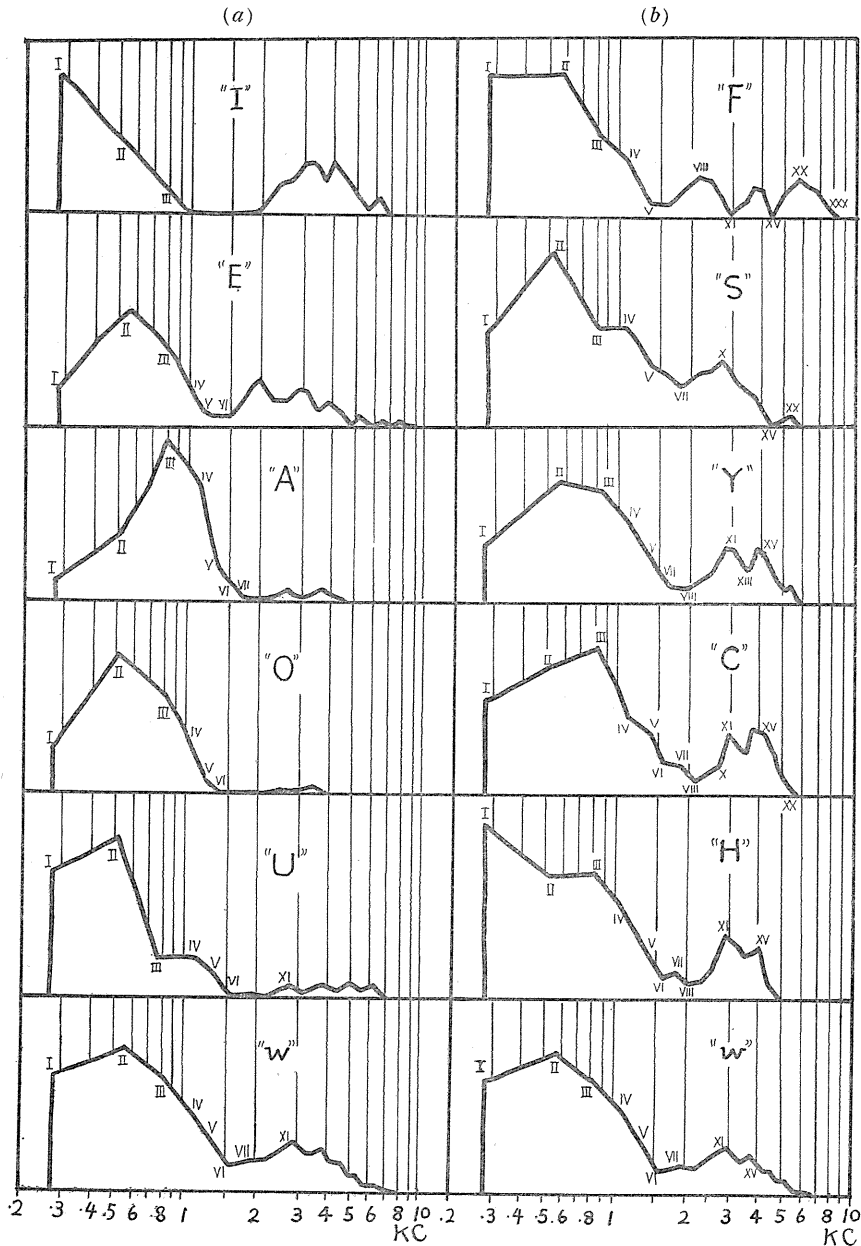


FIG. 1. Representation of timbre patterns. (a): phonemic patterns, (b): vocal patterns with the addition of white-voice pattern in the last row. "F", "S" are male voices; "Y", "H" female voices; "C" boy's voice.

mean vocal pattern for individual voices averaged for all five vowels. For distortion, we used only the Band-Eliminating type (BED), employing three conditions,  $f_c = 1.5, 2.1, 4.2$  kc, of distortion in high-cut direction and four conditions,  $f_c = 0.6, 1.0, 1.5, 2.5$  kc, in low-cut direction, adding one referential condition without distortion. In the neighborhood of their frequency cut-off, each of the seven conditions of distortion has the transmission characteristic of about 60 db per octave. Finally, upon cutting the band frequency, the power level of the transmitted voice-signals lost through distortion was never recovered to the original level of the uncut signal.\* The phenomena of confusion are observed and measured in this way and under these conditions.

### Commentary on Timbre Pattern

It is suitable, even indispensable, to first consider the details of timbre pattern because later in discussing the results of confusion, we must lay much stress on the pattern figure. Of the five phonemes, vowels "I", "E" and "U" have patterns more widely distributed in frequency domain; Japanese "U" (u) is characterized by a pattern of wider extension in contrast to the English and German vowel "U" (u) which is essentially of narrower extension. The speaking vowels "A" and "O" have a much narrower pattern than the singing vowels "A" and "O" and since this study is only on the spoken vowel, it is reasonable that we ignore the singing vowels and concentrate on their spoken aspects. On only the speaking condition of single-formant vowels is there no appearance of pre-eminent upper pattern, *i.e.*, that part distributed beyond vocal glen. It is to be noted that for an examination of phoneme pattern we found it sufficient to study only the *frequency structure* with the precaution that we inscribed the harmonics number in the pattern figure. There was, of course, more or less deviation in each of the five phoneme patterns due to the variance in callers' voices, a deviation which will be found in the detailed studies of References (2) and (3). By consulting the patterns in Fig. 1 (a) we can conclude that there exist some gradual yet conspicuous differences in phoneme patterns, a fact which is helpful in a subjective quality judgement where, in reality, timbre discrimination is carried out on the basis of phoneme value, particularly in the sense of observation of the whole. Next we come to an examination of the vocal pattern where there seem to exist more delicate differences between patterns, as is well illustrated in Fig. 1 (b). In other words, the patterns of voices of the adult female subjects (H, Y) and the adolescent boy before mutation (C) are generally of simple construction, remaining in a region of narrower range notwithstanding the fact that the 280~ pitch used is natural and proper for female conversation; their patterns are evidently composed of two parts, the lower, that part distributed below vocal glen, and the upper pattern of regular fixed form, contrary to the vocal patterns of the male adult subjects (F, S), as shown in the two upper rows of Fig. 1 (b). Of these two male patterns, the voice of F, the tallest of all in stature, has the widest pattern, extending up to 8-9 kc in frequency region with the highest harmonic number—about the 30th, and involving many conspicuous glens. So far as vocal pattern is concerned, we are bound to discuss the further fine point of *harmonics structure*, in addition to the consideration of

\* Recovery was improved in succeeding studies.

wideness in frequency structure to both of which we have already alluded in a previous paper.<sup>4)</sup> On harmonics structure, we find that voices "S" and "F" are characterized by a dominancy of even-number harmonics. This is especially true for voice "F" where it can be particularly seen that the components corresponding to glens are occupied almost entirely by odd-number harmonics. On the contrary, voices "C" and "H" are clearly of a dominant odd-number harmonics type where all peaks are of odd-number harmonics and nearly all glens correspond to even-number harmonics.\* In the white-voice pattern, *i.e.*, the mean pattern of all five vocal patterns, these two types of dominancy cancel out each other, thus resulting in the characterless pattern seen in the last row, Fig. 1.

### Some Remarks on the Description of Experimental Results

In the nature of things, timbre-confusion characteristics are most essential in the sense that they reveal the innermost sphere of quality phenomena. Our approach to this confusion problem promises fruitful results if preliminary processes in our experiments are suitable for timbre treatment in a pure sense and, further, if our methods of observation and our descriptions are pertinent to these ends.

As our present object is to treat the two types of confusion, phonemic and vocal, simultaneously, we begin with confusion in ensembles of suitable size. The fact that the system of Japanese oral vowels is formed of no more than five, is fortunate and most convenient for our purpose. In pace with this phoneme system limited to a total of five vowels, two front, two back and one intermediate, we can easily select our voice system which is formed by two male and two female voices and one indeterminate nature.

We are now in a position to describe our experimental results on confusion phenomenon, a phenomenon so vitally important but far too complicated and divergent for all points to be covered here. As a matter of fact, it has been our long-cherished desire to establish a definite methodology for treatment of this phenomenon, but we are not able to say that our goal has yet been fully attained. However, it is natural for us to describe confusion phenomena in the order of the importance of our viewpoint and the processes so far devised and developed. Here in this Report, Part II, we therefore give a quite naïve description under the title "Simple Descriptipn of Confusion Phenomena," considering the subject from the standpoint of simple practicability where the purpose is only to interpret the quality-loss characteristics as treated in our previous study, Part I. As Part III of the series we will present under the title "Generalized Treatment of Timbre Confusion" another comprehensive description. This last will serve to bring us into a more intimate relationship with the remotest question of confusion phenomena by surpassing the limitation of ordinary treatment, and to carry us into an original approach through giving a fresh meaning to this most important phenomenon.

Viewed from the process of development of our confusion-theory, these two modes of description have their own inherent meaning and therefore have their own distinguishing traits. The former is more suitable to describe the detailed phases of confusion and the latter is better fitted to grasp the general trends of confusion.

---

\* Two of the most conspicuous differences between "C" and "H" voices consist in the difference in intensity of their fundamentals, and the third harmonic of "C" is superior to its fundamental while in "H" the relation is reversed.