RESEARCH REPORTS

MEMOIR ON THE PROBLEM OF DESCRIPTION AND CLASSIFICATION OF TIMBRE IN GENERAL

Yoshiyuki OCHIAI, Kazuko TOKUDA and Koji MATSU-URA

Department of Electrical Engineering

(Received October 31, 1961)

Summary—Essential of speech sounds is made clear by considering *timbre* as a keystone. To study the speech sounds on the ground of timbre is of the utmost importance. Timbre aspect is, in this sense, the starting point of our study and at the same time the finishing one. In the classification of vowel-sound timbres for example, it is quite indispensable from the timbre aspect that the two elements, *phoneme* and *voice*, be made clear. In the timbre theory, phoneme classification is relatively easy, because a great many studies have already been made in this field. But the problem of voice classification would be much more difficult, because there is very little that has been done concerning this domain of research. It is essential that we distinguish most clearly these two elements, phoneme and voice, and that we do not confuse them. Distinction between phoneme and voice would lead us very far and bring us many new things of worth, but to have confusion between them promises us no advance. The consideration of dualism of quality is very serviceable for the general treatment of a problem such as "classification of timbres in broad and wide meaning". In order to describe the timbre-quality of musical instrument sounds for instance, the phoneme-aspect based upon human vowel is of some value. But for the quality estimation or quality rating of musical instruments based upon auditory preference or acoustical taste, the voice-aspect of human sounds is far more important and effective than the phoneme one. In this paper we tried to insist on this point of view by citing exclusively the example of violin sounds.

Introduction

In the long course of our studies on timbres of vowel sounds in order to classify and identify *phonemes* and *voices*, we have found or encountered with several views of importance. The first view is that Timbre Structure or Spectral Construction of Speech Sounds as an object of analytical study is apt to be considered from a too narrowly restricted viewpoint. It is rather necessary that this problem be investigated not only from the standpoint of speech sounds themselves as pronounced by human beings but also from that of general sounds in a more wide meaning. The second view is concerned with the interrelation between *timbre* and the two other attributes of sound quality, that is, *loudness* and *pitch*.

Even the very fact that timbre is the key-point in the acoustical study of

sounds must be based firmly upon a scientific background, and not upon a mere dogmatic belief. Speech sound is one of the most meaningful sounds that man can hear. And further man can say that without question human audition is inseparately bound up with human speech. In other words, human audition has developed in a way that makes the so-called speech domain the center of auditory sensation area. Simple hearing mechanism such as analytical and synthetical functions of audition, and furthermore advanced mood and feeling based upon audition, auditory taste and acoustical preference are all to be associated with, or more adequately to be interpreted on the central stage of speech domain in audition. In short, sounds of speech play a most important part in our acoustical and sensory life.

Preliminary Consideration on Timbre Study in Historical Aspect

The present current of the research of speech sounds finds its true orientation in the direction in which speech and audition are put together.[1)2)] As for the studies which have been executed up to the present time, it seems that the real activity in this line of research has been restricted to two kinds of study. One is the analysis of the most elemental and fundamental qualities (for example, loudness) based upon audition: and the other, the physical sound analysis in Ohm's aspect. But the main problem of interrelation between speech and audition, that is, the problem of differential mechanism and synthetical mechanism of timbre perception seem not to have been undertaken yet. Discrimination and identification of timbre are both practiced on the cerebral center by some way or other. There is not many which is made clear about them. However we can give one example. A. Tomatis has given a most clear verification of the close correlation between human audition and human speech[3)] (the so-called Tomatis' effect).

It should be noted that E. Thienhaus, H. Backhaus, M. Grützmacher, W. Lottermoser, F. Winckel (Germans), R. Husson, A. Moles (Frenchmen), W. S. Kazansky, S. N. Rschevkin (Russians) have dealt not only with acoustics of speech sounds but also with instrumental acoustics. Therefore, it was their characteristic that they attempted to interpret the results of analysis of speech sounds from the standpoint of general sounds. It is found from their excellent papers that the study of sounds of representative of European musical instruments, especially violin and organ are closely related with timbre of vowels. The typical European way to study sounds of speech which is a little different from the others will be considered later with concrete details.

Description of Timbre of Vowels

Over a long period of time of studies,[4)5)6)7)] and after logical reasoning[8)9)10)] and careful examination of experimental processes, the distinction between vocal quality and phonemic quality of vowels have been made possible experimentally in our Laboratory. The distinction and extraction of the vocal quality and the phonemic quality was a little difficult in our experiments because ordinary voices musically untrained were employed as our calling voices. The center of predominant regions of voices are named "vocal formants", and necessarily ordinary formants

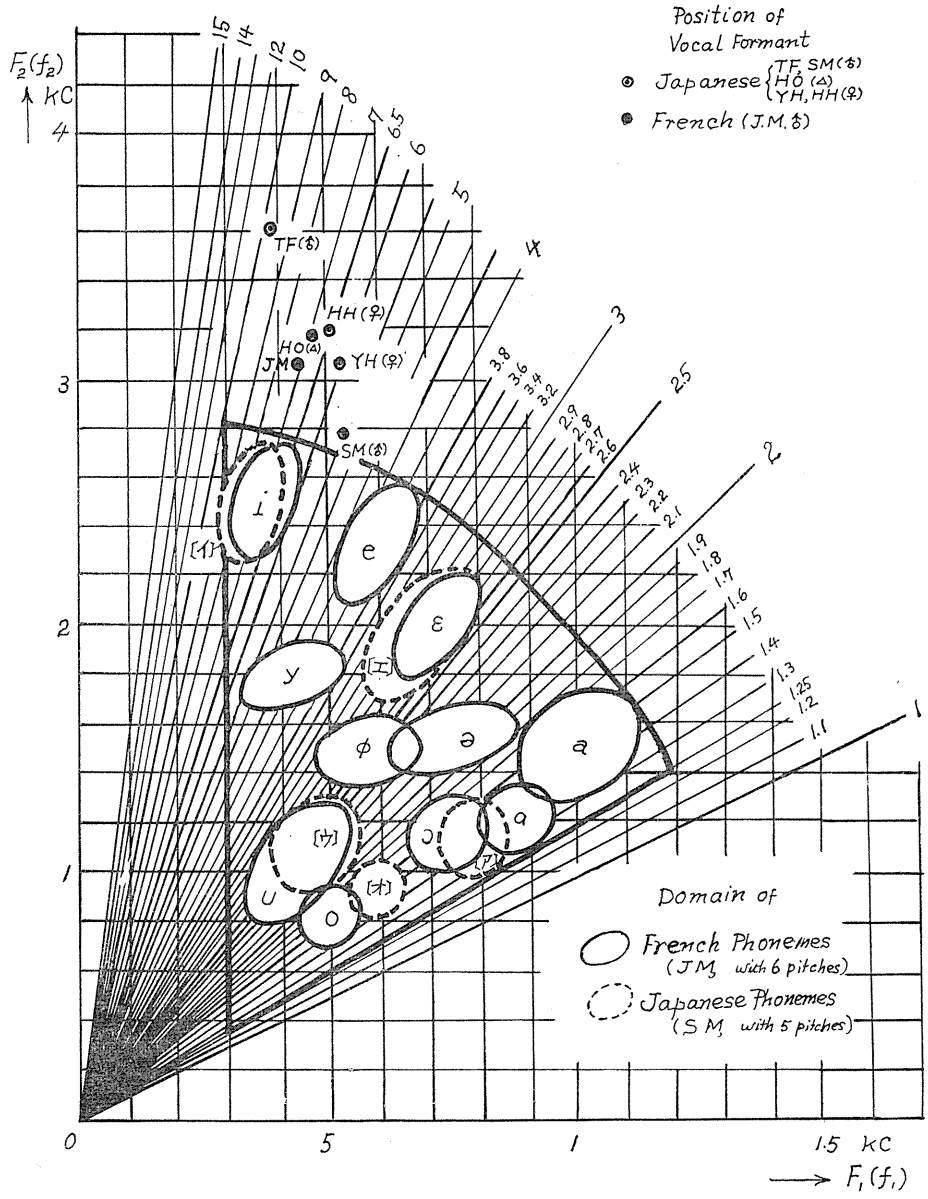

FIG. 1 (a). Phoneme chart of Japanese and French vowels with index of formant-quotient in two-dimensional plane.

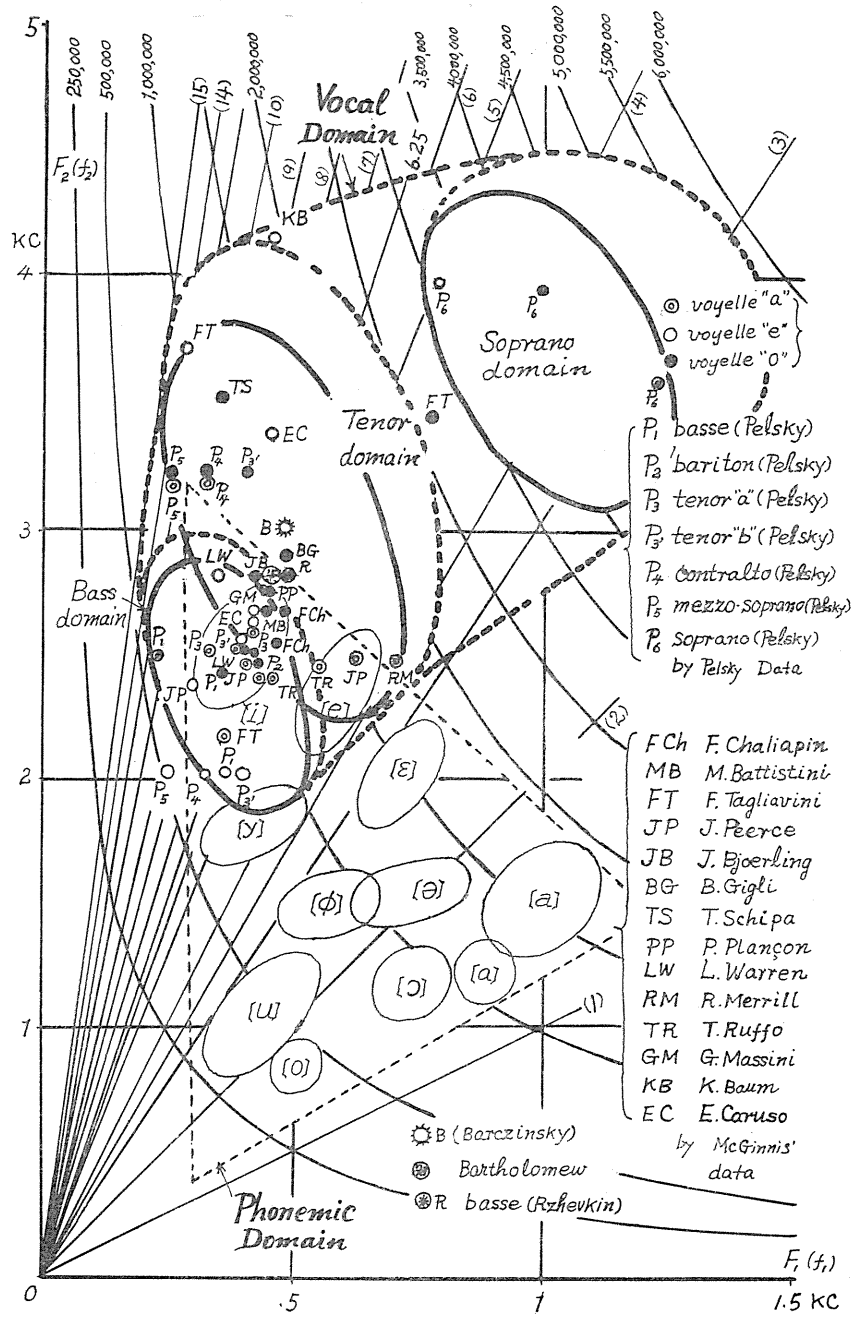

FIG. 1(b). Voice and phoneme chart by two indices of product and ratio in two-formants coordinate,

are termed "phonemic formants" in our Laboratory (1953). As to the vocal formants, we could generally find two centers, one in the higher frequency domain, and the other in the lower frequency domain (see Fig. 1 (a)). Meanwhile, by employing bass voices of trained musicians, Rzhevkin in Moscow University easily and clearly detected so-called "singing formants" (певческое форманты) with extremely sharp resonances[11] (1957). One singing formant is in the higher frequency domain, approximately 2800 cps, the other in the lower frequency domain, approximately 500 cps.* According to our experimental materials, that is, voices of male, female, and children, vocal formants were discovered on approximately 300–500 cps and approximately 3000–4000 cps, and these deviations in both formant regions are brought about by difference of ages and sexes of speakers. As to the sharpness of resonance of formants, differences between vocal formant and phonemic formant were scarcely found in our experiment. It was noticed by Rzhevkin that sharpness of resonance in vocal formant of a trained voice is different from that of an untrained voice, and that positions of resonance are also different in different speakers. In short, the most important fact is that vocal quality has the nature of double peak structures and they are distributed into the two positions, the higher and the lower. In our own studies,[13] it was noted that from the view-point of quality distribution in frequency domain, phonemic quality is characterized by concentrated distribution, and therefore only a limited part of pattern is contributed to the perception of phonemes. On the other hand, the distribution of vocal quality is relatively uniform and dispersive in frequency domain, and not only a part of formant (the focus region of vocal-quality distribution) but also details of envelop of pattern structure extending over the wide regions (both peak and glen structures) play important parts. In other words, for stationary or quasi-stationary construction such as spoken or sung "vowels", perception of vowel-phoneme is partially contributed by the predominant components in pattern which are distributed in comparatively narrow band regions. As to the vocal quality, voice perception is promoted by details of envelop of pattern extending in a comparatively wide band region, in addition to the contribution of vocal formants themselves.

In short, it seems that phonemic quality of normal oral vowel can be generally contributed only by the peak structure, but vocal quality of ordinal oral vowel can not be interpreted without refering to both peak structure (positive formant) and glen structure (negative formant). These are the summary of the final results which were obtained through our studies of characteristics of timbre confusion.

Bibliographical Study on Violin Sounds

Now let us look back upon the past studies on sounds of violin, for the reference of vowel study. It is true that from the view-point of mechanism of sound production, the mode of causing a string of the violin to creak is completely different from excitation of vocal cords on pronounciation and phonation, but timbre of violin can be compared with timbre of vowel sounds in respect to

* As to the two singing formants, 500 and 2800–2900 cps, before P. Rzhevkin, W. T. Bartholomew in U.S.A. has pointed out in his paper as one element of good voices of male singes.[12]

its delicacy and nuance.

According to the conclusion by Suominen,[14] violin sound is formed by several stages of transformation and transmission due to certain parts of instruments, for example, selective performance of resonance box, through which the primitive and unagreeable string vibration like saw-teeth is cultivated to finally the agreeable sounds, resembling a human voice. The secret of violin making and the art of violin music has resided and been rooted firmly upon these points. Judging from these points, studies on timbre of violin by German acousticians such as H. Backhaus, H. Meinel, W. Lottermoser and so on are very interesting. Their studies on violin were characterized by always comparing violin formant with vowel formant. These characteristics stand out in relief especially in the recent study by W. Lottermoser.[15] Fig. 2 indicates one of the examples. Timbre of violin is represented by level-distribution characteristics refered to formant position of vowel. Timbre-description or timbre-appreciation of violin is adequately performed with reference to the phonemic element of human vowel. In this figure, the nasal effect pointed out already by H. Meinel[16] and F. Winckel and high-formant effect of Sh-sound (so-called Zischlaut) are both taken in consideration well. Creaking instrument sounds are produced by special coupling (connection) of gut or steel string with so-called resonance-box of special woods having particular and complicated forms. It is very interesting that such sounds of creaking instrument are similar to those of vowels which can be produced by humans, so far as our concern is exclusively with the phoneme element of vowel sounds.

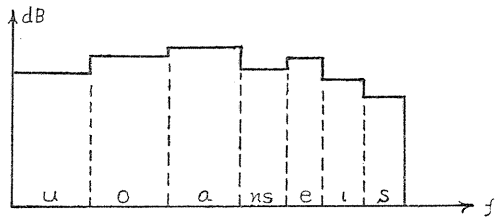

FIG. 2. Characteristics of resonance curve of violin, employing frequency scale corresponding to vowel formants (by W. Lottermoser)

Meanwhile, there is another way of description of violin timbre: H. Backhaus represented a mode of quality appreciation of violins[17] as illustrated in Fig. 3. This representation is very significant when we refer to voice element of vowel sounds. Attention must be focused on the following facts: Stradivarius (G. 1) and violin of old Italian type (G. 2) have plainly double peak structure and modern and good violin (G. 3) and inferior one made in a factory (G. 4) show a single peak structure instead of double peak structure. These relations in violin-quality recall to us another example, an organ instrument in which old Baroc-type organ gives regular construction of double peak structure, in opposition to the modern organ having irregular construction of single peak structure.[18] By employing Rzhevkin's data, the geometric mean of double peak formants in bass-voice is approximately $\sqrt{500\times2800}=\sqrt{1400000}=1183$; by using our data, it is approximately $\sqrt{400\times3600}=\sqrt{1440000}=1200$. This region of geometric means of vocal formants corresponds

exactly to positive formant region of nasals. By using Backhaus' data, the geometric means of excellent violin formants are $\sqrt{1800 \times 3600} = \sqrt{6480000} = 2546$ and $\sqrt{2100 \times 3100} = \sqrt{6510000} = 2551$. The geometric mean position of the violin formants is approximately twice greater than that of the human male voice.* As to the central position of timbre construction, a construction of single predominancy in this position seems to represent an inferior example of violin, and the predominancy of two formants situating to both sides of the central position seems to correspond to timbre of an excellent violin. Speaking more in detail, a violin of old Italian style has a somewhat narrow peak-interval and the Stradivarius has in some measure a wide peak-interval structure.

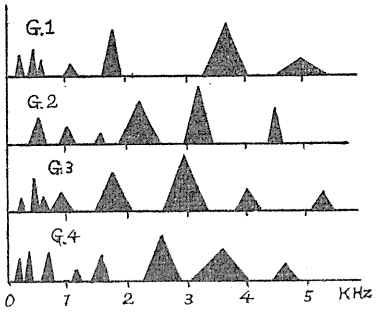

FIG. 3. Domain of resonances of excellent and minor-quality violins (by H. Backhaus).

There are a few points to be noticed. The first point is that a higher formant of Stradivarius type (about 4000 cps) is closely related with A. Tomatis' study,[19] in which he discovered that Caruso's voice has a very clear and higher formant of about 4000 cps in his best period. There must exist some particular tone or nuance which, with regard to its taste, is best adapted to a certain race at a certain period of history. As to the violin of Stradivarius, it seemed that the inborn passion of Italians was able to achieve the creation of this highly situated formant (4000 cps) which Russian acoustician A. A. Rozhdestvensky termed exquisitely "Italian Color" (итальянский тон) in his study on Stradivarius.[20] Here we give one example by which we insist on the fact that the mode of violin-formant can purposely be investigated by applying the conception of vocal formant. It comes into question where the criterion of comparison should be put, that is, whether timbre of violin is to be interpreted in the phonemic aspect or in the voice aspect. In short, the matter of criterion comes about by reason of the close resemblance of timbre of violin and human voice.

The second point is that when we describe timbre of violin by criteria of the vowel phoneme, a description of peak structure is enough, and therefore it is able to be described to some extent by rough characteristics of averaged level distribution within a particular band region. On the other hand, when we describe it by criteria of voice, just as in our studies where detailed structure of peak and glen

* In this connection, violin sounds seem to find their resemblance to the voices of Mezzo-Soprano or Soprano.

is necessary for expression of vocal element, it needs not only expression of position of peaks in main formant by Backhaus but also presentation of characteristics on glen structure as indicated in the study of E. Leipp and A. Moles[21] ("formants négatifs" by their terminology).

The third point is in the case of timbre representation by voice element criteria. It is the general principle on presentation of timbre that the detailed structure of timbre is to be considered on the ground of pitch. When sounds of speech by humans are strictly defined on the voice side, as idea of pitch is to be regarded as the basic parameter in consideration of timbre. On the contrary, when sounds of speech are defined on the phoneme basis, this limitation is almost unnecessary. It is well known that even in voiceless condition the phoneme is approximately established, but the voice can scarcely be formed in this condition. Without touching loudness, timbre structure is able to be described to some measure, (almost all the acousticians take this conventional standpoint), but it is absolutely meaningless to describe timbre without touching pitch.

To be more concrete, as far as vocal quality by humans is concerned, the relationship between vocal register and timbre structure is to be made thoroughly clear. Tenor, Alto, Bariton, and Bass show not only a classification of vocal register but also a classification of vocal timbre (see Fig. 1 (b))*. We must deal with vocal register as defining vocal timbre. Briefly, we must compare human voices in their own categories of registers. For example, in order to consider timbre of instruments of violin-family on the criteria of vocal element, subfamilies of violin, viola, violoncello, and contrabass should be considered in individual sections. This self-evident suggestion needs a little explanation. Pitch, intensity and timbre may be regarded as independent attributes when they are considered in the physical aspect, but when they are considered subjectively, these three attributes are no longer completely independent. Pitch is the first fundamental element of timbre and builds up a base of overall timbre in synthetic meaning. So it is almost impossible that pitch be treated as irrelevant to timbre and independent of timbre.**

If a Russian Bass is compared with a Japanese Tenor as an expedient, it does not mean a comparison of timbres in quality aspect, but merely a comparison in spectral structures. In the same manner, a guitar and a violin can be compared with each other only in their spectrum relation, but it does not mean a comparative study of timbre in a strict sense. Comparison of timbres is meaningless unless they are on the same pitch, or on the same register at least. This is one of the fundamental standpoints for the timbre study of vocal quality.

Descriptive Method of Timbre in General by using "Qualifier"

The above mentioned is concerned with the main boundary conditions on description or measurement of timbre. Now we reach the stage in which consideration of the contents of timbre signal (or the connotation of the expression of timbre pattern) is needed. It is necessary that timbre be described and appreciated on the scientific background, even though there is a question as to

* The representation of voice chart in this figure is made by employing the data of Persky,[23] McGinnis,[24], Barczinski and Thienhaus.[25]

** To compare timbres on the criteria of phoneme is naturally possible without assignment of pitch.

whether or not its quality of higher order and higher class is able to be measured directly.* But acousticians have used description by employing "qualifier" for timbre description up to the present time. It is well known that visual impression is adopted to describe acoustic impression, and it is also a matter of common knowledge for acousticians that the word "timbre" (*Klangfarbe* in German, *Ne-iro* in Japanese) comes from a close relation with color sensation. Technical terms for luminosity, for example, "dark" and "bright", are also often borrowed for the description of timbres. Similar impressions of the sounds of musical instrument are utilized for the description of human vocal color. For example, the terms, most frequently utilized by German acousticians, "offene Stimme" and "gedeckte Stimme" are based on the special effect of "offene Pfeife" and "gedackte Pfeife". As to the construction of pipe sounds, the open-pipe has all harmonic construction and the closed pipe has odd harmonic construction. A vocal impression with abundant partial construction is sometimes described as a metalic sound (metallischer Klang). The special impression based upon the difference between stringed instrunents with gut strings and those with steel strings has much influence upon the description of the special impression obtained. That is to say, impressions of familiar sounds of instruments are applied to the description of impressions of the human voice. It is very interesting to compare the above stated fact with the methods, in which phoneme and voice of human vowels are utilized as a criteria of analysis of timbre of violin. The special sound of stringed instruments is called "flageoletto", because its timbre impression brings to mind timbre of a pipe instrument. And "sul ponticello" or "sulla tastiera" is the name of techniques which give rise to particular sounds (indicating the specific position of string excitation), but they are a terminology founded on timbre impression which consists of rich or poor partial structures. Therefore, it can be said in conclusion that a consistent and systematic way for the description of timbre has never been found even in the traditional art of music. In other words, neither the science of acoustics nor the art of music has had a consistent system of description and presentation on timbre.

Appreciation of Timbre Quality

We can scarcely come to any consistent conclusion on the problem of timbre preference and timbre appreciation. Take, for example, the timbre appreciation of the violin. According to Hewlett[22], it is essential that an excellent violin must have a strong fundamental component. In the opinion of Backhaus[17], one of the main characteristics of Stradivarius is the strongness of [i] formant region. Meinel[27,28,29,30] and Winckel[31,32,33] called attention to the fact that the omission of nasal formant region is necessary for the timbre appreciation of violin. According to the Rimsky Korsakov,[26] the typical timbre of violin has comparable high and low symmetry which is measured from the central region of the strongest [a] formant. Suominen remarked that a characteristic of Stradivarius is in the establishment of partials in 3000-4000 cps region (Backenresonanzen), and then an additional vibrato-component appears in 2000-6000 cps region.[14]

* As for the quality of timbre, timbre-measurement of naturalness and articulation of vowels are nothing but indirect measurement.

Conclusion

When consideration is given to the spectral construction, it is necessary to point out clearly that the phonating condition requires two parameters, vocal intensity and vocal pitch. In addition to the physical construction study of timbre signals as objective description of the observer standing outside of the communication considered, the subjective appreciation of timbre impression as introspective description of the observer standing inside of the communication is most indispensable for classification of timbres in general on the background of "timbre" science, because it is most important to refer to the introspective description corresponding to the timbre construction as input stimulus to indicate the meaning and connotation of timbre signals in communication.

Acknowledgement—The authors take this opportunity to express their appreciation to the Nikko-Shoken Fund by the aid of which a main experimental part of this study could be carried out.

References

1) H. Fletcher: Speech and Hearing in Communication. Van Nostrand Co. Inc., 1953,
2) С. Н. Ржевкин: Слух и Речь в Свете соверменных физических Исследований. М-Л., 1936.
3) A. Tomaits: Relation entre l'audition et la phonation. Annales des Télécommunications, Tome **11**, Nos. 7–8, juillet-août, 1956.
4) Y. Ochiai and T. Fukumura: Timbre Study of Vocalic Voices. Mem. Fac. Engng., Nagoya, Vol. **5**, No. 2, 1953.
5) Y. Ochiai and T. Fukumura: Timbre Study of Vocalic Voices Viewed from Subjective Phonal Aspect. Parts I, II, III. Mem. Fac. Engn., Nagoya, Vol. **8**, No. 1 and 2, 1956.
6) Y. Ochiai: Memoir on Nasalics. Mem. Fac. Engng., Nagoya, Vol. **9**, No. 1, 1957.
7) Y. Ochiai and T. Fukumura: Timbre Study on Nasalics. Parts I, II, III. Mem. Fac. Engng., Nagoya, Vol. **9**, Nos. 1 and 2, 1957.
8) Y. Ochiai: Mémoire sur les sons des voix humaines. Mem. Fac. Engng., Nagoya, Vol. **4**, No. 1, 1952.
9) Y. Ochiai: Transmission of Quality. Mem. Fac. Engng., Nagoya, Vol. **6**, No. 2, 1954.
10) Y. Ochiai and T. Yamashita: On Timbre Quality. Mem. Fac. Engng., Nagoya, Vol. **7**, No. 1, 1955.
11) Р. Ржевкин: Некоторые Результаты Анализа Певческого Голоса. Акуст. Журн. Том II, Вып. 2, 1956.
12) W. T. Bartholomew: A Physical Definition of Good Voice-Quality in the Male Voice. J. Acoust. Soc. Amer. Vol. VI, No. 1, 1934.
13) Y. Ochiai and T. Fukumura: On the Fundamental Qualities of Speech in Communication. J. Acoust. Soc. Amer. Vol. XXIX, No, 1, 1957.
14) L. Suominen: Untersuchungen über die Schwingungseigenschaften des Geigenkörpers. Acustica, Vol. 8, No. 6, 1958.
15) W. Lottermoser: L'Examen acoustique des violons dans la Physikalisch-Technische Bundesanstalt. Acoustique Musicale (Colloques Internationaux du Centre National de la Recherche Scientifique. LXXXIV, Marseille, 27–28–29 mai 1958), Paris, 1959.
16) Г. Мейнель: О Научных Основаниях Конструирования Скрипок. Акустический Журнал. Том VI, Вып. 2, 1960.
17) H. Backhaus: Über Geigenklänge. Zeitschr. f. Techn. Physik, 8, Nr. 11, 1927.
18) M. Grützmacher: Über die Klänge von Glocken und Orgeln. Acustica, Vol. **4**, No. 1, 1954.
19) A. Tomatis: La correction de la voix chantée. Cours International de Phonologie et de Phoniatrie. Vendredi, 11 mars, 1960.

20) А. А. Рождественский: Смычковые Инструменты. "Музыкальная Акустика", Государственное Музыкальное Издательство, Москва, 1954.
21) E. Leipp et A. Moles: L'emploi du sonagraphe dans la détermination de la qualité des instruments à cordes. Annales des Télécommunications, Tome **14**, Nos. 5–6, mai-juin, 1959.
22) C. W. Hewlett: Analysis of Complex Sound Waves. Physical Review, Vol. **35**, 1912.
23) B. L. Pelsky: La structure de quelques voyelles chantées. Arch. Néerl, Phonét, Exp., **17**, 1941.
24) C. S. McGinnis, M. Elnick and M. Kratschman: A Study of the Vowel Formants of Well-Known Male Operatic Singers. J. Acoust. Soc. Amer., Vol. XXIII, No. 4, July, 1951.
25) L. Barczinski und E. Thienhaus: Klangspektren und Lautstärke deutscher Sprachlaute. Arch. Néerl. Phonét. Exp. Tome XI, 1953.
26) A. V. Rimski-Korsakov: Les recherches sur le timbre des violons et guitares et sur l'excitation des vibrations d'une anche d'harmonium. Acoustique Musicale (Colloques Internationaux du Centre National de la Recherche Scientifique, LXXXIV, Marseille, 27-28–29 mai 1958), Editions du Centre National de la Recherche Scientifique, Paris, 1959.
27) H. Meinel: Über Frequenzkurven von Geigen. Akustische Zeitschrift. II, 1, Jan. 1937.
28) H. Meinel: Akustische Eigenschaften klanglich hervorragender Geigen. Akust. Zeitschr. IV. 2, März, 1939.
29) H. Meinel: Akustische Eigenschaften von Geigen verschiedener Klangqualität. Akust. Zeitschr. V, 3, Mai., 1940.
30) H. Meinel: Über Frequenzkurven von Geigen. Akust. Zeitschr. V, 5, Sept., 1940.
31) F. Winckel: Die ästhetischen Komponenten der Stimmgebung. Acta Physiol. Pharmacol. Neerl. **5**, 1, July, 1956.
32) F. Winckel: Die psychoakustische Bewertung des Spektrums. Folia Phoniatrica, Vol. **10**, No. 2, 1960.
33) F. Winckel: Die Grenzanalyse des Komplexlautes in der Stimm- und Sprachforschung. Akt. Probl. der Phoniat. Logopäd., Vol. **I**, 1960.