PRESENTATION OF CHARACTERISTIC DATA ON PHONEMIC QUALITY DISTRIBUTION OF JAPANESE VOWEL SOUNDS, FROM THE STANDPOINT OF SPEECH COMMUNICATION

YOSHIYUKI OCHIAI and KAZUKO TOKUDA

Department of Electrical Engineering

(Received May 31, 1961)

Summary—We present here the most reliable characteristic data on phonemic quality distribution of five Japanese oral vowels in frequency dimension. These representative characteristics of five Japanese vowels are obtained as an average of seven results in seven experiments carried out during about five years in our Laboratory of Audiology. As data on frequency distribution of phonemic quality of vowels, we use so-called articulation scores in the two directions of distortion of frequency-cutting, from low to high and high to low. From the inclination of characteristics on one hand and from the intersection point of characteristics on the other hand, we can study several phases of quality distribution of vowel-phonemes in frequency dimension. We discuss here the existence of *negative importance* regions in addition to *positive importance* regions, from the standpoint of phonemic quality distribution of vowels.

In our previous papers,[1)2)3)] according to the dual aspect of speech quality, we have defined the phonemic pattern of vowels as well as the vocal pattern of speakers. These patterns are of course defined as frequency-amplitude construction in frequency dimension or overtone-amplitude structure in consideration of harmonics. However, from the general consideration of timbre pattern, it is very important to examine: whether, or not, the phonemic quality in subjective sense is uniquely and completely determined by the representation of phonemic pattern thus defined as frequency construction; and, whether, or not, the vocal quality in subjective sense is absolutely and solely determined by the representation of the vocal pattern in harmonic structure. When the problem comes down to these essential points, we must prepare another kind of experiment, covering the insufficient side of nature of speech, in order to explain these points with certitude and clearness. This means that we are forced to execute an experiment of psychological nature.

To examine definitively and precisely the subjective meaning of timbre pattern in frequency dimension, it is necessary to utilize a system of frequency distortions by which band-cutting conditions are easily and smoothly available for our speech studies. Over a period of five years, by changing the combination of speaker-group and listener-crew, as shown in Table 1, we have performed this type of experiment seven different times with the view of measuring phonemic quality as well as vocal quality.

It must be borne in mind that even though it is not difficult to define the vocal quality in general, it is very difficult and troublesome to obtain the concrete and

TABLE 1. List of Experiments Executed in Laboratory of Audiology, 1953–57

Number	Year	Uttering Pitch	Voice-Group	Uttering Level	Vowel-Group	Signal-Pause	Listener-Group	Transducer	Number of Observations per Condition of Distortion
1	1953	280~	5 F°(♂) S°(♂) C°(△) Y°(♀) H°(♀)	*mezzo-forte*	5 I. E. A. O. U.	4 sec–4 sec	4 Fu(♂) Fa(♂) Na(♂) Ha(♂)	Loud-Speaker	2,000
2	1954	140~	4 K(♂) Y(♂) N(♂) F(♂)	*mezzo-forte*	5 I. E. A. O. U.	4 sec–4 sec	4 K(♂) Y(♂) N(♂) S(♀)	Loud-Speaker	1,200
3	1954	240~	5 K(♂) Y(♂) S(♀) N(♂) F(♂)	*mezzo-forte*	5 I. E. A. O. U.	4 sec–4 sec	4 K(♂) Y(♂) N(♂) S(♀)	Loud-Speaker	1,500
4	1955	240~	5 K(♂) Y(♂) F°(♂) C(♀) S(♀)	*piano*	5 I. E. A. O. U.	4 sec–4 sec	4 K(♂) Y(♂) C(♀) S(♀)	Loud-Speaker	1,500
5	1955	240~	5 K(♂) Y(♂) F°(♂) C(♀) S(♀)	*forte*	5 I. E. A. O. U.	4 sec–4 sec	4 K(♂) Y(♂) C(♀) S(♀)	Loud-Speaker	1,500
6	1956	220~	4 K′(♂) M(♂) C′(♀) S(♀)	*mezzo-forte*	5 I. E. A. O. U.	3 sec–3 sec	4 K′(♂) M(♂) C′(♀) S(♀)	Loud-Speaker	1,200
7	1957	150~	4 T(♂) S′(♂) M′(♂) H(♂)	*mezzo-forte*	5 I. E. A. O. U.	2 sec–3 sec	4 T(♂) S′(♂) H(♂) D(♀)	Dynamic Receiver	2,000

(♂): male voice, (♀): female voice, (△): child voice actual vocal nature of each individual speaking subject. But for phonemic quality, it is possible to determine all conceivable phonemic natures of all phonemes in the Japanese vowel system. Our objective of study is the phonemic quality found in the finite and limited group of the five Japanese vowel sounds. It is therefore our present aim to define and discuss the phonemic nature of individual vowels in the Japanese language system from the standpoint of subjective quality.

In our study throughout the seven experiments, the five oral vowels in sustained and isolated state were pronounced by a total of 24 speaking subjects (men and women, sometimes children) at different six pitches, almost always on *mezzo-forte* but sometimes on *piano* and *forte*.

In Fig. 1, we give the final overall characteristic of Japanese vowel articulation quality, based upon the whole data of seven experiments and averaged ultimately and separately for five vowels. The average vowel sound for the five actual vowels can be called "white-vowel" for the sake of convenience.* It must be noted that in the Japanese vowel system where there is characteristically no nasal and no neutral, the region of about 1,200~1,300 cps has either no significance or only a small one in a phonemic sense. At this region two characteristics in High-Cut and Low-Cut Distortions intersect, meaning thereby that this region corresponds to the apparent center of phonemic quality distribution of Japanese vowels; and corresponds really to the almost empty gap of quality distribution, that is, to the major vocal glen in pattern.[6] The slopes of characteristics are extremely smooth and gradual, meaning thereby that the distribution of quality of phonemes in an average sense changes smoothly and continuously. It is not unnecessary to add that this pair of quality characteristics does not come in perfect symmetry as regards the center of characteristics-intersection when the frequency is scaled in logarithmic or even in *mel* measure.

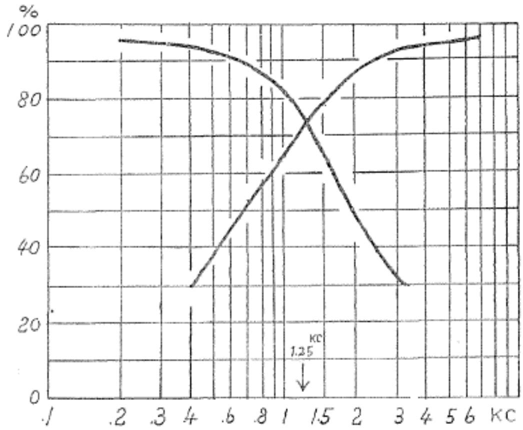

FIG. 1. Phonemic quality characteristics of Japanese oral vowels in final mean representation. Intersection point of characteristics is about 1,250 cps, as center of quality distribution in frequency dimension.

In Fig. 2, we give individual vowel-articulation characteristics corresponding to seven experiments. There is some deviation between these characteristics. After most attentive examination, we have come to the following conclusion. Inspite of deviation of characteristics, the intersection point of each characteristic does not differ considerably according to the difference in experiments, showing that differences in pitch or level bring no influence or only a little upon the mode

* By white vowel, we designate an ideal vowel in an average sense for all possible vowels. We have used "white", because the sum of all possible colors (timbres) becomes white. This does not mean *voce bianca* in vocal music.

of distribution of phonemic quality.*

In Fig. 3, in five columns, we give five representations of phonemic quality for five vowels, each of which is given by seven characteristics in seven experiments. From these representations, we can see that there is considerable deviation among characteristics and further that the degree of deviation is different, in some measure, according as the vowel is different. For instance, single-formant vowels such as "A" and "O" show considerable deviation even from the beginning condition of distortion. On the contrary, at the starting condition of characteristics, double-formant vowels such as "I" and "E" show very little deviation and it is only after the augmentation of the distortion that the deviation is quite considerable. The Japanese vowel "U" seems to belong to the latter type of vowels, judging from the mode of quality characteristics. From the standpoint of symmetry in a pair of characteristics, we can point out a sole example of vowel "A" in which Low-Cut characteristic is quite symmetric to High-Cut characteristic with regard to the intersection point. Even for another single-formant vowel "O", we find some degree of unsymmetry with regard to the inclinations of Low- and High-Cut characteristics. This means that the vowel "A" is situated in the true center in phonemic aspect amidst the five vowels in a Japanese vowel system and further that the very center of the distribution of phonemic quality of this single-formant vowel is really in the region of about 1,000 cps.

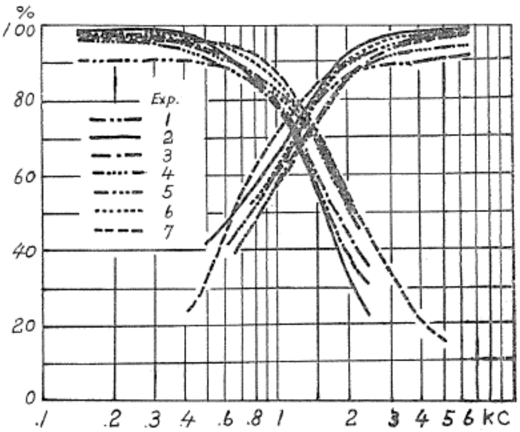

FIG. 2. Group of mean representation of phonemic quality characteristics of Japanese oral vowels obtained in seven experiments.

As the most remarkable point of quality changing phenomena in such distortions, we must attach importance to the particular tendency of quality recovering such as found most clearly in vowels "I", "E", "U". In the mean-value representation of phonemic characteristics for individual phonemes, as shown in Fig. 4, we can clearly trace the quality-recovering trend for these vowels. This trend is by no means caused by any experimental error. We found out clearly in our very ealier studies[4)5)] that for some particular vowels there exist certain irregular points in the phonemic characteristics for distortions of band-elimination type. That is, in these distortions, the characteristics are not necessarily of a monotonously decreasing type, and there is very often such cases in which we experienced the characteristics not only of simple hesitating type but also quality-recovering type.

As reasonable interpretation for this irregular point of quality characteristics, we must propose the most important hypothesis concerning the basis of quality theory. We consider, as mode of quality distribution in frequency dimension, two

* This point stands in big contrast with the distribution of vocal quality. Vocal quality distribution is considerably influenced by pitch change and by level variation.

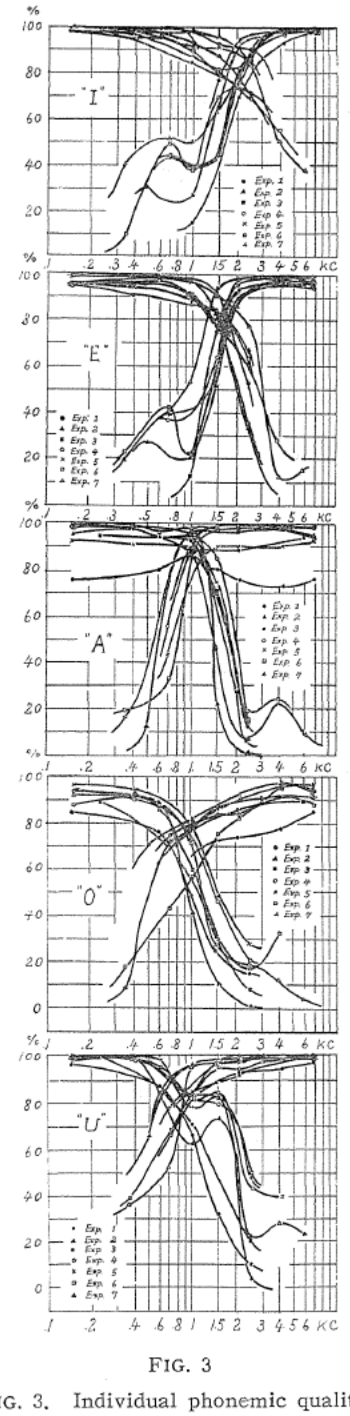

FIG. 3

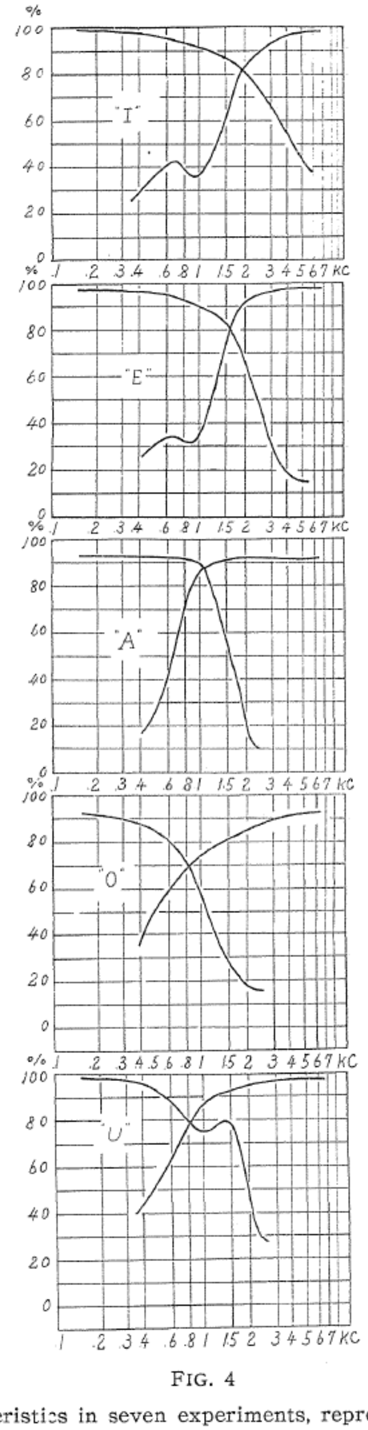

FIG. 4

FIG. 3. Individual phonemic quality characteristics in seven experiments, represented for five vowels.

FIG. 4. Individual phonemic quality characteristics of five vowels, each characteristic being averaged for seven results in seven experiments.

kinds of regions: region of *positive importance* and region of *negative importance.* For an establishment of phonemic quality, any phonemes have their own proper regions of both positive importance and negative importance. Positive importance region is positively necessary and useful for the establishment of phonemic quality of a phoneme in question, and negative importance region is not necessary but harmful for the formation of its quality. By suppressing the positive importance region, quality is deteriorated. By eliminating the negative importance region, quality is formed. When the distortion comes to the band of negative importance after passing through the band of positive importance, we experience the phase of characteristics in which the quality has appearance to be recovered to some measure. We can probably explain the quality recovering characteristics which occurred in the course of distortions for vowels "I", "E", and "U". Because, for these vowels of double-formant type, the region of negative importance is actually situated in the mid-way between two regions of positive importance.

Further we must explain the reason why the quality-recovering trend appears only in High-Cut distortion (not in Low-Cut distortion) for vowels "I" and "E", and for vowel "U" it appears only in Low-Cut distortion (not in High-Cut distortion). For this point, we must consider the influencing power of each "importance" region. For vowels "I" and "E", for example, the higher region of positive importance is more powerful than the lower, with regard to the formation of phonemic quality. On the contrary, for vowel "U", the lower region of positive importance is more significant than the higher. By suppressing first the less significant region of (positive) importance, there is either no deterioration or only a little, because there remains more significant region of importance in the deformed pattern. On the contrary, when the pattern is suppressed first from the region of more significant importance, the quality begins to drop considerably, because there remains only, in the pattern, a part of far smaller quality distribution. It is only after heavy drop of quality that we experience most clearly the quality-recovering phase of characteristic.

It becomes very interesting problem whether, or not, the vowels of single-formant type show the so-called quality-recovering trend in their characteristic courses. In other words, there exists, or not, the so-called negative importance region for these vowels. From the results of our experiments discussed here, we can not give a definite answer to this. But probably, in the outskirts of the region of heavy concentration of single positive importance, there may exist a relatively weak negative importance region which gives rise to a tendency of slight recovering of quality. For the verification of this point, we are obliged to make further experiments of much precision, based upon, for instance, a special consideration and observation of phoneme confusion.[6)7)]

In Fig. 5, we show the five individual quality characteristics in ensemble. By this representation, we can easily see how one vowel is different from another in the quality distribution mode in frequency domain. For example, the center of quality distribution is different in position according to the difference of vowels, as shown by the marks on abscissa. From the forms of quality characteristics, we can imagine approximately the mode of quality distribution of vowels. For vowel "A", the phonemic quality is distributed symmetrically upward and downward, having its center (that is, the point of maximum concentration), in an average sense, in about 1,100 cps. For vowel "O", the mode of distribution is quite unsymmetric: viewed from the center of about 850 cps, the distribution in lower side is relatively dull and the distribution in upper side is rather sharp.

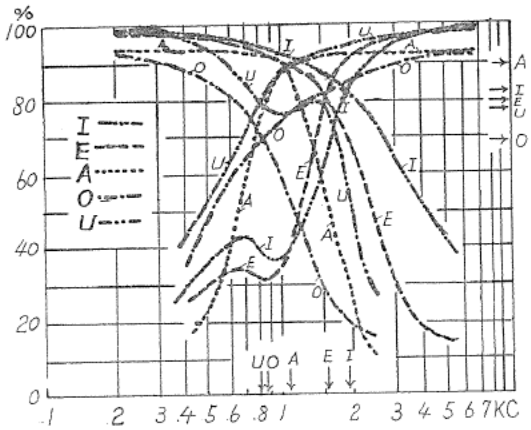

FIG. 5. Individual representation of phonemic quality characteristics for five Japanese oral vowels, each of which is obtained by averaging the results of a total of seven experiments.

References

1) Y. Ochiai et T. Fukumura: Sur les qualités essentielles des vocales. Cahiers d'Acoustique des Annales des Télécommunications, Tome **15**, No. 11–12, Novembre-Décembre, 1960, pp. 277–291.

2) Y. Ochiai: Phoneme and Voice Identification Studies Using Japanese Vowels. Language and Speech, Vol. **2**, Part 3, July-September 1959, pp. 132–136.

3) Y. Ochiai: Fondamentales des qualités phonémique et vocalique des paroles par rapport au timbre, obtenues en employant des voyelles japonaises vocalisées par des sujets japonais. MFE of Nagoya University, Vol. **10**, No. 2, November, 1958, pp. 197–210.

4) Y. Ochiai: Nihon-Onin no Denwa-Hizumi-Gensho to sono Shukan-teki Sokutei. Denshin-Denwa-Gakkai Zashi, Showa 10 Nen, 2 and 5 Gatsu (February and May, 1935).

5) Y. Ochiai: Nihon-Onin no Hizumi-Gensho to Denwa-Meiryodo ni kansuru Kenkyu. Dissertation at Tohoku University, 1943.

6) Y. Ochiai and T. Fukumura: On Fundamental Qualities of Vowel Sounds in Speech Communication. Review of the Eelctrical Communication Laboratory, Nippon Telegraph and Telephone Public Corporation, Vol. 8, No. 9–10, 11–12, 1960.

7) Y. Ochiai and T. Fukumura: Preliminaries to Analysis of Quality in Speech Communication. MFE of Nagoya University, Vol. **9**, No. 2, November, 1957, pp. 306–315.