ON THE MODEL BUILDING OF THE SOUND QUALITY RECOGNITION PROCESS

SUSUMU KONDO AND CHIKIO HAYASHI

(Received Feb. 2, 1968)

Summary

We have previously shown that listeners are divided into two groups, having different preferences, yet having nearly the same preference within each group [1]. In this paper we investigate the preference process and quality recognition process. It may be summarized as follows:

(1) The preference process can be expressed as a non-linear system, because preferences can be divided into at least two categories, excellent and poor. Therefore it is necessary to treat it in order to change it to a linear system.

(3) The preference process may be at the same level with that of the quality recognition process when the preference process has been changed to a linear system, so that if either response is obtained, the other would automatically be decided. It is generally reasonable and practical to construct a model of the recognition process in which the quality is decided by the preference.

(3) The process of changing a non-linear preference system to a linear one is described.

1. Introduction

The problems of evaluating listeners' preference or the quality of sound have been studied in regard to the development of broadcasting equipment and sound reproducing systems. There are many physical factors in regard to the quality of sound, for example, equipment circuit characteristics, the loud speaker, and the field of sound. The aim of this study is to construct a reasonable model of the preference process which is necessary in seeking the best physical conditions of equipments or transmission systems as shown by the preference of the listeners. It is therefore necessary to consider the recognition process for evaluation of the listeners who are the final receptors of audio information.

We have previously reported that the listeners were divided into two groups. Within each group their preference responses may be almost homogeneous, but between these two groups their preference characteristics were different [1]. In this paper we will describe the sound preference and quality recognition processes.

Similar problems with this consideration have been reported by T. Miura and T. Nakayama [2], but we assume a different method of approach to this problem.

2. Materials and methods

The listeners judge their preference and the quality of music reproduced from equipment with narrow-band-eliminated filters. The center frequencies of the filters were chosen to be 150, 300, 600, 1200, 2400, 3600, 4800, 7200 and 9600 cps, and the depths of the dips were chosen to be 10, 15 and 20 dB at the center frequency. And also their bandwidths at the 6 dB points was chosen to be 1 octave and 2 octaves. These

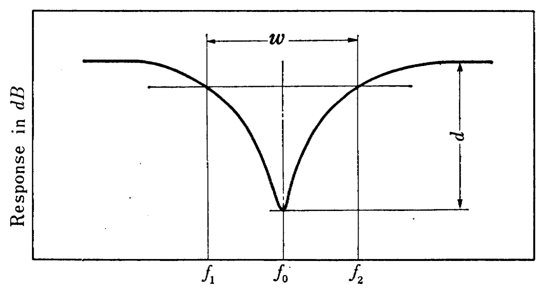

Fig. 1. Response curve of the narrow-band-eliminated Frequency filter 'w': the bandwidth, 'f_0': the center frequency, 'd': the depth of the dip

are shown in Fig. 1. Two sources of disc-recorded music were used: Beethoven's symphony No. 5 and the Coppélia suite. Reproduced sound pressure levels were 65 and 85 dB S.P.L. The number of filters was 228.

Judgement was done in two steps as follows. The first step was preference judgement. The listeners divided their preference into five categories, the criteria being excellent, good, fair (average), poor and very poor. Each category was weighted as +2, +1, 0, −1 and −2 respectively.

In the second step the listeners judged the quality of a series of 228 fragments of sound using twelve words* expressing conceptions of the

* These words were chosen by Osaka University and Tohoku University, and they were sixty or more. We selected twelve words from them excepting valueless words for this experiment.

quality of sound. These words were chosen freely from Table 1. The series of 228 fragments of sound were judged absolutely by the listeners.

Therefore we were able to make up three models with respect to the recognition process. Let "*P*" be the preference process and "*Q*"

Table 1. The words used to express the conceptions of the quality of sound (These words are symbolized by twelve letters)

a	clear, liveliness	*g*	not clear, not brisk
b	balancing	*h*	metalic
c	soft	*i*	sharp, stiff
d	smooth	*j*	rough
e	stereophonic	*k*	flat
f	splender, vivid	*l*	gentle, modest

the quality recognition process. Thus in Fig. 2 (1) "*P*" and "*Q*" are of the same level, while in (2) "*Q*" is decided after "*P*", that is "*Q*" is the function of "*P*". In (3) "*P*" is decided after "*Q*", or decided by the sum of a few essential factors of "*Q*". Miura and Nakayama's model stands on this consideration. In this figure each compartment shows a recognition process.

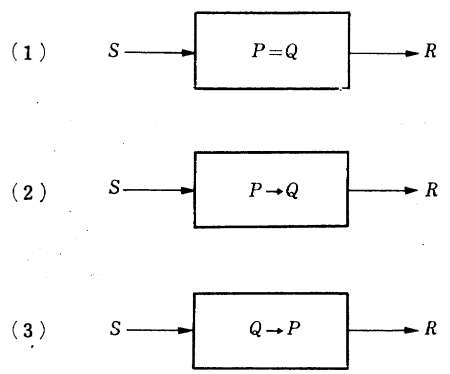

Fig. 2. These three compartments show the schematic models of the recognition process. '*P*' shows the preference process and '*Q*' shows the quality recognition process. '*S*' is the stimuli or input signal of the system and '*R*' is the response of the system.
(1) '*P*' and '*Q*' are equivalent on the recognition, that is if either '*P*' or '*Q*' is decided, the other is automatically estimated
(2) '*Q*' is decided after '*P*' has been decided
(3) '*P*' is decided after '*Q*' has been decided

The statistical treatment, of course, may determine according to which model is utilized. From the statistical standpoint, it is desirable that the system be linear. If the system is not linear, the relationship between the stimuli and the response can not be found simply, and it must be changed into a linear system by any method.

3. Relationship between preference and quality

The responses of the listeners who judged the series of fragments of sound are, for example, excellent—*a. b. c*, bad—*h. i. j*, fair—*a. c. l* and so on. (In this case the words may be freely chosen, and those felt to be unsuitable are of course not chosen.)

In Fig. 3 (a) the results of the recognition of two listeners who have had different judgements are shown. Each stimulus $S_1, S_2, \cdots, S_9$ was characterized by filters: $f_0=150$, 300, 600, 1200, 2400, 3600, 4800, 7200 and 9600 cps respectively, $d=15$ dB and $w=1$ octave. The ordinate shows the weights of their preference. From these results we clearly find that the same stimulus may give different responses, excellent or poor. In Fig. 3 (b) the relationship between preference and quality, except zero weight, is expressed in only two categories, excellent and poor. Fig. 3 (c) shows that the judgement of fair of the two listeners can again be divided into two categories as it follows the results of (b).

Figs. 4(a) and (b) also show the relationship between preference and quality. These are the cases of 228 stimuli and 10 listeners, and the ordinates show the frequency of each of the words selected by them. They are analogous to the results shown in Figs. 3(b) and (c). From these results we find that the words *a, b, d, e* and *f* belong to the excellent preference group, and that *g, h, i, j* and *k* belog to the poor preference group. We find also that *c* and *l* are used in both categories.

In Figs. 3(c) and 4 (b) we find that the judgement of fair itself can again be divided into two categories. Finally, these results show that the preference has basically only two categories, excellent or poor. This phenomenon shows that the preference process may be a non-linear system.

4. Relationship between preference and the combination of quality indicating words

In the previous section we have shown that the preference is basically alternative and that the quality words are divided into two groups: one group *a, b, d,* and *e* are used to indicate "excellent", the other group, *g, h, i, j* and *k* are used to indicate "poor". This relationship was obtained from standardizing the listeners' preference. In this section the kind of relationship between the groups of words and the listeners' preference will be shown.

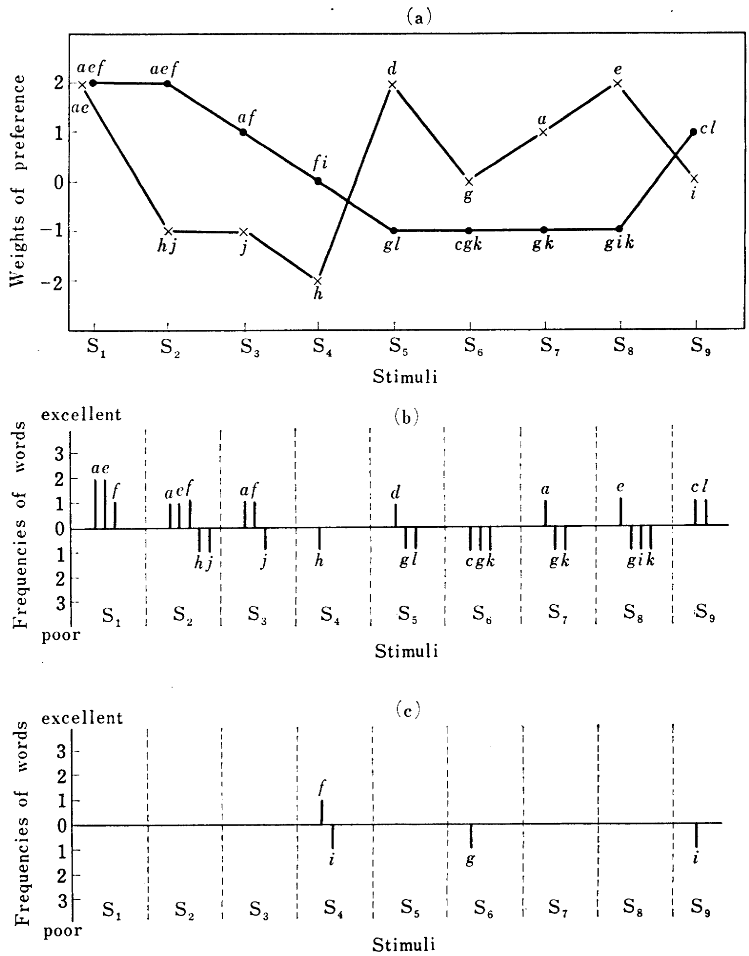

Fig. 3. An example of the relationship between preference and the quality of sound at each stimulus These show the typical results of two observers. (a) Relationship between the stimuli and the preference. Small letters on the curves show the conceptions of the quality (see Table 1). The ordinate shows the weight of preference, and abscissa shows the stimuli. (b) Relationship between the preference and the quality, except zero weight. These are shown by the histogram. (c) The histogram of the preference judged to zero only

Now we consider how to choose the twelve words $a, b, \cdots, l$ freely. There are, for example, a only, ab, abc, or cl, etc. In these combinations, of course, the order of choice is not the question. Then, the total number of combinations of these twelve words is 4095, but the actual number will probably be less than that because some words may not be selected.

Table 2 shows the correlation matrix of crosstabulated data concerning

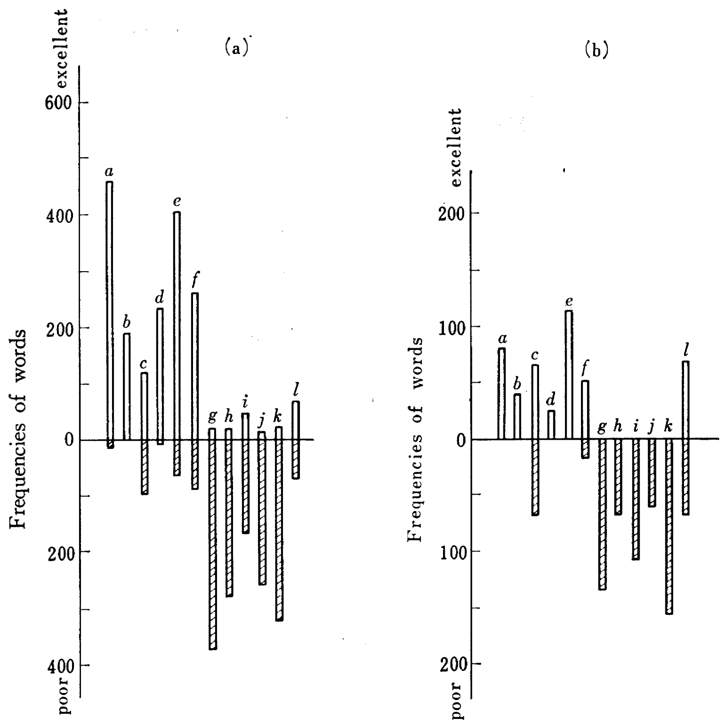

Fig. 4. Relationship between preference and the quality: ten observers and two hundred and twenty eight stimuli These are shown by the histogram. (a) Except the weight zero, and (b) the weight zero only

the combination of words. In this table, for example, the combinations of three words, a, b and c are shown as ab, ac, bc, so that each would be counted as one combination. We can now group the words from that correlation matrix by a method of "quantification"* [3]. That is, the words may be put on a two-dimensional psychological plane so that words with similar meanings are closely grouped but the distance between them increases with their dissimilarity. Fig. 5 shows the results.

In this figure $a, b, d, e, f,$ g, h, i, j, k and c, l are shown forming different groups on the plane. We now designate each small group as "U", "V" and "W". We can see that "U" corresponds to excellent,

* This is the method by which we seek the relationship of the resemblance of the individual in the group to others in the group by the measure expressed by Eq. (1). Namely, the individuals are located on the multi-dimensional space, as those who have a strong resemblance are in closer positions and those whose resemblance is weak are separated further.

$$D=-\sum_{i=1}^{N}\sum_{j=1}^{N} e_{ij}(x_i-x_j)^2 , \qquad (1)$$

where e_{ij} is the coefficient of resemblance of the individual "i" for the individual "j", larger e_{ij} shows larger resemblance, and x_i or x_j shows the number given to "i" or "j", $i\neq j$, $i, j=1,\cdots, N$ (N is the total sum of the individuals).

Table 2. Correlation matrix of cross tabulated data concerning the combination of the words
Each element shows the frequencies of the combination of letter '*i*' with letter '*j*'

i \ *j*	*a*	*b*	*c*	*d*	*e*	*f*	*g*	*h*	*i*	*j*	*k*	*l*
a		126	38	166	278	207	12	9	31	11	13	25
b			29	55	150	96	6	2	12	2	7	20
c				90	101	17	126	0	0	5	52	153
d					132	50	17	4	2	4	10	44
e						191	75	9	33	15	8	43
f							34	55	60	56	51	6
g								29	37	94	160	89
h									131	131	131	3
i										123	133	14
j											177	11
k												93
l												

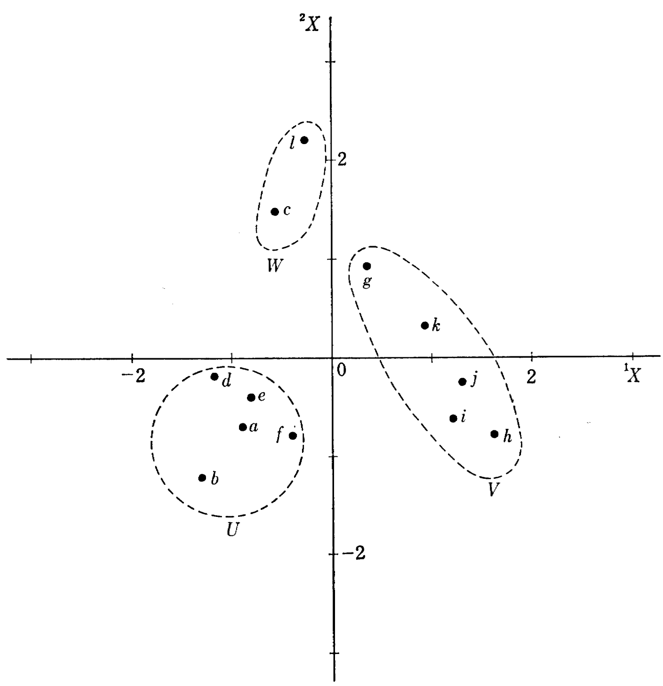

Fig. 5. Grouping of the words of the conceptions of the quality of sound
Small group '*U*' corresponds to excellent, '*V*' to poor and '*W*' correspond to both categories.

"*V*" to poor and "*W*" to both categories when they are compared with the results of Fig. 4. These groups of words belong to different quality concept categories, respectively.

This result is nothing but the reverse of the result shown in Fig. 4. It is, however, clearly found that "*Q*" is under the control of "*P*". That is to say, it may be natural to consider that "*Q*" is recognized after "*P*" is recognized. Namely, model 2 of section 2 applies.

5. The variances of the recognition process

In this section we will try to put in order the process involving "*P*" and "*Q*". That is, we shall estimate which process is more complex. For this purpose we state a hypothesis: Man's conception and recognition of sound may be related to his culture, education, mental activities and so on, and these things are not generally the same among men. Therefore, a more complex or in-depth response to sound stimuli will be indicated by a greater variance in the number of matched preferences among the listeners. If we would give numbers to these conceptions, we could calculate their variance. This variance may be utilized as a measure of the degree of the recognition process. In the high degree of recognition, the variance would be larger than in a low degree of recognition.

First we will consider the process of preference. The listeners who divide preference into two categories indicates a more simpler recognition than those who divide preference into five or seven categories*. Second, we will consider the process of quality recognition. In this process the listeners can choose from among 4095 possible combinations of the twelve words given in this example. This would correspond to recognizing 4095 possible conceptions, so that the expressions of the listeners concerning the quality would have more individuality than those concerning their preference and that the variance would be large. This shows that "*Q*" is a higher recognition process than "*P*".

For example, we compare two variances of equally preference-matched pairs. In the case of P_2, preference was divided into two categories and, in the case of P_5, into five categories being excellent, good, fair, poor, and very poor.

In Table 3, S/M, S being standard deviation, M being mean, and P_2 and P_5 are shown under four different conditions. In the left column of the table, *B* indicates a short period of Beethoven's symphony No. 5

* In general the number of categories is five or a maximum of seven, and the number of different conceptions of sound quality is generally larger than these categories of preference.

Table 3. Standard deviation/Mean, of 'P_2' and 'P_5'

	P_2	P_5
B-65	0.22418	0.23694
B-85	0.27570	0.33434
D-65	0.34341	0.35505
D-85	0.31989	0.33146

reproduced from a disc-record and D indicates a short period of the Copélia suite reproduced in the same manner. The figures of 65 and 85 indicate the mean sound pressure levels in dB at the listeners' positions. This table shows clearly that the psychological distances indicated by variances of the listeners' preference in the case of P_5 are larger than those in the case of P_2.

In the quality recognition process, the method of selection of quality indicating words would have more individuality. If the variance of equally-quality-matched-pairs of listeners' would be calculated, we could then see that the variance would be larger than that of "P". That is, S/M would be $P_2<P_5<Q$. Hence by the preceding hypothesis we can see that the quality recognition process is at a higher stage than that of preference process. This shows that the structure of Fig. 2 (2) would be supported.

6. The construction of a model of the recognition process and its considerations

In the preceding sections we arrived at the results that (1) the preference process has the character of being non-linear, and to treat it statistically, the process should be changed into a linear system. (2) It is convenient and reasonable to take the structure of, Fig. 2 (2) for the sound quality recognition and preference process model. From these two points, the quality recognition and preference process model may be considered as the structure shown in Fig. 6, which is constructed of two compartments "P" and "Q". When the stimulus or input signal "S" arrives at the ear organ "O", in compartment "P" the preference is decided. After that in the next compartment "Q" the sound quality is recognized and "R" is the response of this system. For example, the following structure is considered: $S \rightarrow \underset{\text{excellent}}{\langle P \rangle} \rightarrow \underset{a,c,d}{[Q]} \rightarrow R$. In this case, of course, it may be decided by the listener's free will that either "P" or "Q" be bypassed. If the P compartment were a linear system, then both "P" and "Q" would function at the same level. That is the structure shown in Fig. 2 (1). In Fig. 6, "fair (average)" is the neutral judgement not belonging to either the excellent or the poor category. In

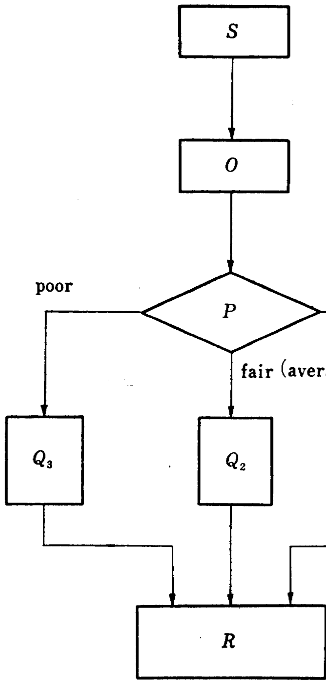

Fig. 6. A schematic model of the sound quality recognition and preference process
O: hearing organ
P: preference process
Q: quality recognition process
R: response of the system

the preceding section, however, it was shown that "fair" could be divided into two preference categories, but in practice this division may be difficult for the listeners. However, in this model it is permitted.

On the practical applications of this model, we must consider that the preference process is a non-linear system and should be converted to a linear one by some method, for example, the method of the "quantification", a division into groups of listeners with the same characteristics. In this case "P" is perfectly equivalent to "Q". For example, the relationship between "P" and "Q" is automatically decided as "excellent$=a$" or "poor$=h$". That is, the preference process may be at the same level with the quality recognition process, and the structure of Fig. 2(1) may be supported.

7. Conclusion

The relationship between the sound stimuli presented to man and his response to it is generally decided by the physical conditions or the quality of the sound reproducing equipments. When dealing with problems of sound quality and listener preference, how to reasonably understand the mental process of this relationship is an engineering requirement. For this purpose it is necessary to construct a sound quality recognition and preference process model.

In this paper we have obtained the results that the preference process was described as a non-linear function, and the judged quality of sound was related to this process. It would be suggested that the process described in this paper may be used as an approach to other problems with similar considerations.

HITACHI SYSTEM ENGINEERING CO., LTD.
THE INSTITUTE OF STATISTICAL MATHEMATICS

REFERENCES

[1] S. Kondo and C. Hayashi, "On the preference of sound quality," *Jour. Acoust. Soc. Japan*, 21 (1965), 216.
[2] T. Nakayama and T. Miura, "On the methodology of evaluating reproduced sound quality," *Jour. Acoust. Soc. Japan*, 22 (1966), 319.
[3] C. Hayashi, "On the prediction of phenomena from qualitative data and the quantification of qualitative data from the mathematico—Statistical point of view," *Ann. Inst. Statist. Math.*, 3 (1952), 93-98.