QUALITY PROBLEMS IN PERCEPTION OF REPEATED DAMPED-SINUSOIDS

TERUO FUKUMURA

Department of Electrical Engineering

(Received May 30, 1959)

The conventional interpretation of the frequency spectra of vowels is that perceptional cues for vowel recognition are the correlates of frequency groups called the formants. Technically, it has been interested, on one hand, to synthesize the formant-like resonance modes having the perceptional impression of vocality, and, on other hand, to extract physically the parameters from speech waves, which are due representation of phonems. The short historical review concerning the synthesis and extraction technique brings us a large number of experiences, and almost the technical schemes adopted there are some kinds of analogue type. The dominating methodology, not the simulation by cut-and-try, cannot be found. In our belief, it was caused from the lack of quality studies which are fundamental in speech problems. How does the ear perceive the sounds having resonance modes, and discriminate them from each other so definitely that every sound becomes meaningful in human communication ? Upon this subject, we can not find, as far, many in the literature. Despite the importance of their nature, these problems are too delicate to be solved immediately because the events in which we are interested happen somewhere far from peripherals of sensory organs. Thus we are obliged to start our approach at the very beginning of the problems. The wave which has only single resonance mode takes the expressional form as

$$f(t) = Aexp(\delta + j\omega)t,$$

which is often called damped-sinusoid. In the daily life, the ear is impressed by sum of several such damped-sinusoids, repeating periodically or aperiodically. The wave forms are determined by parameters A, δ, and ω, and ears can discriminate the difference of wave forms caused by the variation of these parameters. Among the parameters. we select two; δ, damping coefficient, and ω, resonant frequency. A, amplitude, is excluded from our consideration because of its direct relation to loudness. Then the main task in the experiments below is to reveal the discriminatory behaviors of the ear against the change of δ and ω in damped-sinusoids. Since the discriminatory behavior strongly correlates to time condition under which the sounds be compared, experiments are divided into two parts according to the difference of experimental schemes, *i.e.*, the first; sheme of paired comparison where two sounds to be compared are continually presented. The second; scheme of identification where five or seven sounds of different wave forms are presented randomly in time at constant pause.

Experimental Procedure

The repeated damped-sinusoids are synthesized by introducing a narrow pulse-train with repetition rate of 200 p.p.s. through a selective amplifier tuned at a single frequency. Time width of each single pulse is about 150 micro-seconds. In the first experiment, where the scheme of paired comparison is adopted, two selective amplifiers, one of which is variable in selectivity and resonant frequency, are used. The equipment to obtain small variations of these parameters is as follows: To change the resonant frequency, a resistor of Twin-T network inserted into the feed-back loop of amplifier is cotrolled by hand as shown in Fig. 1. The change of selectivity is made possible by varying the amplification factor. By the inherent nature of such an active network, it is impossible to change both parameters independently. In our experiment, variations of parameters are limited not to exceed the extents within which the relative error of band width caused by the frequency change remains at several percentages, and that of frequency by the change of band width is smaller than 0.6%. This extent of variation in both parameters are also limited in the loudness aspects. For, changing parameters, especially in the case of frequency, happens to cause the great enhancement of amplitude of output wave, which necessarily leads to loudness change. Besides these limitation of variation range, exhaustive trainings for the subjects are executed to exclude loudness factor from their judgment of quality difference between two successive sounds each having duration of one second. In the second experiment, five or seven sounds having different parameters are recorded in the magnetic tape. The order of recording is randomized, and the recording levels of each sound are uniformly matched at a constant level. The time interval between adjacent sounds is three seconds, and each sound has duration of two-half seconds. The subjects are asked to identify the reproduced sounds with their memory learned in several preliminary tryings. In both the experiments, the sounds are impressed into subjects' ears through the dynamic receivers with equalizing networks at low frequency. Four young students having normal hearing acuity take part in listening throughout the experiments.

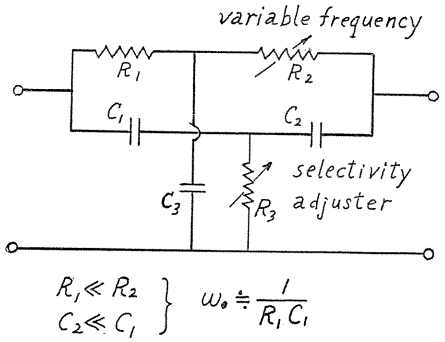

FIG. 1. Circuitry device to obtain the small variation of resonant frequency almost independently of selectivity change.

Results

First experiment

There are several of data processing methods in the experiment of this kind. A method we adopted here is only for convenience's sake and is illustrated in Fig. 2, where the abscissa represents shifts of parameter and the ordinate represents the percentages of equal judgment. Then the data plotted on this plane constitute a histogram. From this, we can calculate the standard deviation σ, and

call it, "insensitivity" with respect to the change of the concerned parameter.

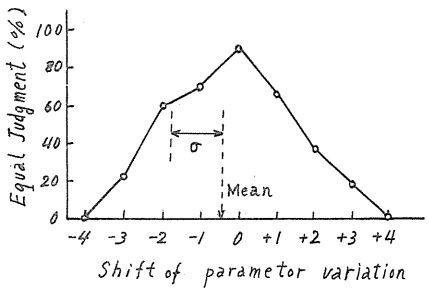

FIG. 2. Curve of "equal" judgment distribution showing the probability of judging the variable signal as bearing reference quality. σ: the standard deviation is the measure of dispersion of quality judgment and the "insensitivity" is defined by this quantity.

Insensitivity to resonant frequency

The insensitivities σ_f with respect to the change of resonant frequency are obtained at several different values of selectivity and resonant frequency. The data are plotted in Figs. 3 (A) and (B). The abscissa represents the resonant frequency in both figures. The ordinate stands for absolute value of σ_f in (A) and relative value σ_f/B, where B denote the half-power band width of selective amplifier. In (A), the parameter attached to each smoothed curve, which is connected through the observed points represented by different signs for different parameters, is the value of selectivity, Q. The legend of sign of observed point in (B) is the same as in (A). From (A), we can easily see that the discrimination of the frequency change becomes more insensitive as the resonance system reduces its selectivity, and further, even at the same condition of selectivity, the insensitivity, takes larger value at higher resonant frequency. To display these discriminatory characteristics more simply, it is convenient to take the relative value σ_f/B in the ordinate representation as in (B), where every observed point, although scattered in (A), is approximately in a row parallel to the abscissa. This means that the discriminatory ability depends upon the band width in a relatively simple manner[1].

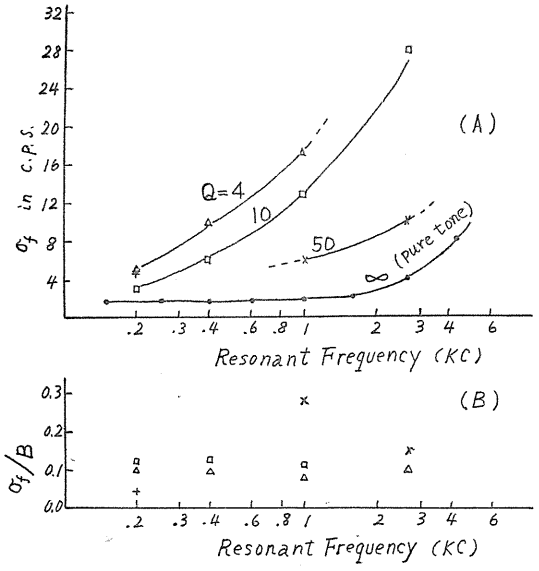

FIG. 3. Insensitivity to the change in resonant frequency; (A) Absolute values and (B) Relative values in comparison to the half-power band width B.

Insensitivity to damping coefficient

Measurement of the discriminatory insensitivity to the change in damping coefficients, σ_δ, are carried out at several different resonant frequencies. In every case, the selectivity of the amplifier for the comparing sounds in AB testing method is kept constant at $Q=10$. The data are plotted in Fig. 4, where the

abscissa represents the resonant frequency of each test. Half-power band widths of amplifier, corresponding to the selectivity at each resonant frequecy, are also entered under the abscissa. The ordinate represents the value of insensitivity σ_δ. As immediately seen in the figure, each point is plotted on a almost straight line, gradient of which is nearly 45°. This means that the discriminatory ability of ear to the change in damping coefficient of sounds relates almost linearly to its absolute values of δ, or B.

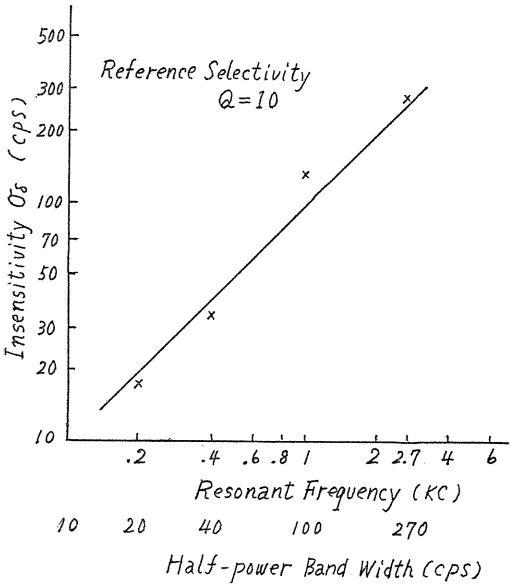

Fig. 4. Curve shows the nearly linear relation between damping coefficient (band width) and insensitivity to its change.

Second experiment

In the second experiment under the scheme of identification method, the subjects are expected to give judgments basing only on their memory. Then the data obtained can be all tabulated in the confusion matrices. The constituents of signal ensembles used are illustrated in Table 1. From each matrix for each ensemble, the mean percentages of correct judgment are calculated, and some of them are shown in Fig. 5 for the sake of comparison of identifiabilities between pure tones and complexes. The most interesting facts shown in this figure are that the pure tone, having the smallest insensitivity to the frequency change, does not possess the highest

TABLE 1. List of Signal Ensemble

Ensemble		Freq (kc)	0.23	0.37	0.52	0.60	0.72	1.0	1.4	1.6	2.0	2.7	4.3
Pure	g_1			○		○		○		○		○	
	g_2	○		○		○		○		○		○	○
	g_3			○	○		○	○	○		○	○	
Complex	G_1	$Q=2$		○		○		○		○		○	
	G_2	4		○		○		○		○		○	
	G_3	10		○		○		○		○		○	
	G_4	50		○		○		○		○		○	

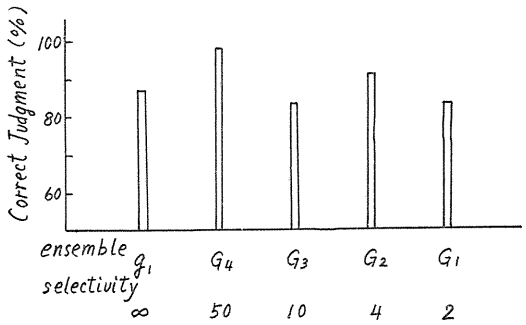

Fig. 5. Percentages of correct judgments for various signal ensemble in identification test.

identifiability. The highest happens in the case of complexes having the selectivity of $Q=50$. But the complexes do not show the monotonous relation between identifiability and selectivity. It will be noticed that, in the ensemble G_2, which takes the higher percentages of second order, the band widths of ensemble members are most close to that of vowel formants.

To reveal the further information about the subjects' judgment, we manipulate the confusion matrices as described below. Denote the frequencies of input sounds by X_K ($K=I, II, \ldots, N$), and the received frequencies by X_i ($i=1, 2, \ldots, n$). The larger integer of indices monotonically corresponds to higher frequency. Then we define judgment error by

$$K_{Kj} = X_K - x_i.$$

Restricting our consideration only to the ensemble member on the center position of frequency continuum, we calculate, in each direction of confusion,[2] the standard deviation of judgment error with respect to this member on the center position; that is, in incoming direction, σ_m, and in outgoing direction, σ_M. The use of these quantities is practically illustrated in the data shown in Figs. 6 (A) for pure tone and (B) for complex sound. In both the figures, the abscissa stands for the kinds of ensemble and the ordinate represents the standard deviation. (The standard

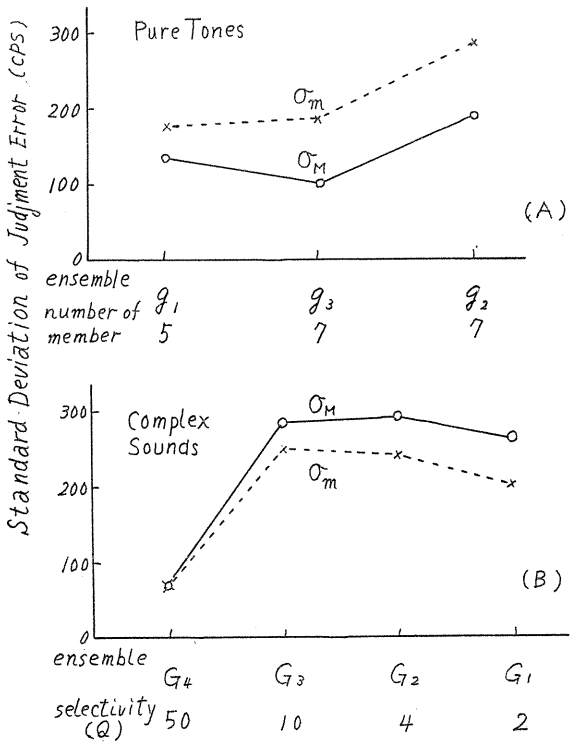

FIG. 6. Standard deviations of judgment error in two confusion directions, σ_m: in incoming direction, σ_M: in outgoing one. (A) is for signal ensembles of pure tones and (B) for that of complex sounds.

deviation thus calculated may be considered as insensitivity in the test of identification.) In Fig. A, the inequality $\sigma_m > \sigma_M$ is remarkable. This is resulted from the fact that the subjects tend to judge the different sounds as center member rather than confuse it with other sounds. In other words, the center frequency is most probable to present itself in the subjects' responses. This definite confusion trend may be inherently caused by auditory nature of pure tone, which can be described by the simple position on one dimension. In the case of complexes, as shown in (B), above inequality is not held and general trend is $\sigma_M > \sigma_m$. From the above illustration for pure tone, it can be easily seen that in these cases, the frequencies most probably appeared in the subjects' responses are not already at the center position. The most probable frequency shifts onto the other position having particular band width which corresponds to the value of Q given to the ensemble. In short, in the case of complexes, the frequency position of each sound cannot be the only cue for the identification. Of course, the most probable response depends upon the definitness of quality possessed by each sound. Then it may be safely said from our data that, in the quality formation of repeated damped-sinusoids, frequency and band width do not perform such a co-operated function that the optimum Q is constant throughout every resonant frequency.

Discussion

If the experimental results obtained above have any relation to the human hearing mechanism, one can suppose a hypothetical model of the auditory system which does not contradict these results. Customery assumption of auditory system is that the ear can be considered as the assembly of many parallel-connected band-pass filters each having different band positions[3]. These filters also have been assumed to have finite band width called critical band width, which was early suggested by Fletcher[3] from his experimental data of masking by white noise. Recently, Zwicker *et al.*[4][5] found the facts that the sounds having line spectra also display the critical band phenomena, and their band widths are several times wider than that obtained from white noise data. To answer the question why the ear, in spite of its comparatively wide band width, can discriminate the frequency changes of pure tones so sharply as indicated, for example, in our results, Huggins[6] suggested the phase principle in frequency discremination by band pass filter. Considering further the case where the input signal was the damped sinusoids, he concluded that the parallel connection of two or more filters facilitate the discrimination of frequency changes in both pure tone and damped sinusoid. These propositions may be considered as seemingly related to our experimental data which reveals the simple relation between quality discrimination and damping coefficient. If the assumption of the ear being a linear band-pass filter, is accepted, its transfer function can be written in the form:

$$g(t) = \sum_i A_i \exp(\delta_i + j\omega_i)t.$$

The input wave of damped sinusoid takes also the same form, and the output of the filter is determined by calculating the convolution integral of these two exponential functions. Now we assume $A_1 = A_2 = \ldots = A_0$ and $\delta_1 = \delta_2 = \ldots = \delta_0$,—this

does not restrict the generality of our consideration—and that the input consists of only one reasonance mode as is in our case. At time when the natural vibration modes of every component filter vanish, the output will be

$$F(t) = AA_0 exp(\delta + j\omega)t \sum_i 1/(\delta - \delta_0) + j(\omega - \omega_i),$$

where δ is damping coefficient and ω is natural frequency of input wave respectively. From this analytical expression, we see that the ouput wave, the vector sum of each output from component filter, changes its amplitude and phase by the changes of δ and ω in input wave.

Now we further assume that the subjects succeeded to exclude the loudness factor from his quality judgment, then there remain only phase parameter which probably relate to subjects' discrimination. In regard to the sensory mechanism which organize such a number of outputs from component filters, we can also venture to propose several hypothetical organizing schemes suggested from human social behavior. One of them is as follows: Denote the phase angle of any one component output by θ_i with reference to input, then we assume that the ear organizes every component so that

$$(d\Theta)_{w \text{ or } \delta} = \max(\sum_i w_i (d\theta_i)_{w \text{ or } \delta}),$$

where w_i is the weighting factor for component i. One of such a maximizing operation is to select each w_i so that all $w_i d\theta_i$ have positive value. By this positiveness of each term in the avove formula, every variation of individual $d\theta$ by the change of δ is uniformly added to get total variation of $d\Theta$. If this phase principle and maximizing operation can be taken as a working hypothesis to illustrate our experimental data, increment of δ must result in decrement of $d\Theta$. But from the nature of function,

$$\theta_i = \tan^{-1}\omega - \omega_i / \delta - \delta_0,$$

we can not find any tendency that the second derivatives $(d(d\theta_i)_\omega)_\delta$ and $(d(d\theta_i)_\delta)_\delta$ remain constantly negative in the variation of difference values between two damping coefficients, $\delta - \delta_0$, the former being used in the experiment and the latter being inferred from critical band width. This contradiction between hypothesis and experimental result does not necessarily confute the whole hypothesis. For example, we can say that the phase principle is held, but the organizing operation is not optimum such as;

$$d\Theta = \sum_i u_i d\theta_i - \sum_j v_j d\theta_j$$

where u_i, v_j are both weighting factors making every phase positive. In other words, the ear can not utilize the all phase informations conveyed by every output wave of component filters. It is probable that the components passed through the superfluous filters perturb the significant components rather than contribute to overall phase information of output.

In this connection, we must not overlook the spectral construction of periodically repeated damped-sinusoid. The frequency spectrum of such a sound, having the shape of something like a resonance curve, consists of many descrete fre-

quency components placed at a constant interval corresponding to repetition rate. How the ear perceives the ensemble of many frequency components ? As for the loudness, it is said that the ear summs up the loudness of each component. This sensory operation is very simple. The sensory phenomena, which can be illustrated by such a simple principle, usually show the larger DL in the larger physical dimension. The spectral informations, which seems most directly correlating to the quality discrimination of the subjects in our experiment, are the position and dispersion of spectra. The data show that the DL's for the spectral configuration become larger as the components distribute wider. With reference to this, it can be inferred that the spectral dispersion on the frequency continum may be transformed psychologically into the extent in Euclidean space. Physiological facts, which are accompanied with these psychological discrimination of DL, seems relatively peripheral. As the data of our experiments by identification judgment indicate, pure tones, which behave as having smallest DL, do not convey the maximal information for the best identifiability. The discrimination, which is executed, for example, with the continually successive two sounds, will not necessitate the quality judgment of higher level. The speech qualities which are fundamental in communication, must be identified isolately in time, then presumably necessitate the sensory activity of higher level. The quality which is recognized by subjects in the first experiment is not the same as that in the second experiment. However, it can not immediately be determined if these two qualities must be treated in the manner of, for example, Dualism. This important problem is left for further study.

Summary

We made here some approach to the problems of synthesized sounds, and revealed several basic properties of quality judgment for the repeated damped-sinusoids. The main conclusion deduced from our experimental results are as follows. Such a quality of complex sounds, as detected under the experimental condition of paired comparison with small time interval, is somewhat trivial in quality problems of speech. The quality, which is interested and fundamental in the quality theory, must not be affected by time condition in comparison test. The quality of repeated damped-sinusoid, which can be identified isolately in time, depends upon the band width of frequecy spectra of the sounds. Although the optimum relation between reasonant frequency and band width was not yet strictly determined, it became clear that the selectivity, Q, of resonant circuit or that calculated from frequency spectra has not so great significance in the interpretation of vowel patterns.

Acknowledgment

I am especially indebted to Dr. Y. Ochiai for his helpful comment and criticism. Also, I wish I could adequately thank all those participated in this laboursome tasks.

References

1) J. L. Flanagan: J. Acoust. Soc. Am., 27 (1955); Our data agree fairly well with the

Flanagan's data obtained by the speech synthesizer MIT POVO. We regret that we have not yet opportunity to have the literature of K. N. Stevens' exhaustive work about the frequency DL of damped-sinusoids.

2) Y. Ochiai, T. Fukumura: Memo. Fac. Eng., Nagoya Univ., 8 (1956) 293.
3) H. Fletcher: Rev. Mod. Phys., 12 (1940) 47.
4) E. Zwicker, *et al.*: J. Acoust. Soc. Am., 29 (1957).
5) B. Scarfer: J. Acoust. Soc. Am. 31 (1957).
6) W. H. Huggins: J. Acous. Soc. Am., 24 (1952) 582.