PROXIMITY ANALYSIS OF PITCH PERCEPTION OF COMPLEX TONES IN ENDLESS SCALE

Meiko Sugiyama* and Kengo Ohgushi**

Experiments were carried out on the pitch of two groups of computer-generated complex tones. The sequence of the ten complex tones in group A are perceived as an endless scale when they are heard repeatedly (Shepard, 1964). Complex tones in group B were synthesized by slight modification of frequency structure of complex tones in group A. The experimental results were analyzed by eij-type quantification theory. The configuration of ten complex tones in group A is considerably circular. On the other hand, the configuration of ten complex tones in group B shows a considerably one-dimensional character. This difference in configuration was explained by the place cue and the time cue originated from neural activities of auditory neurons.

1. Introduction

The pitch of a tone is generally considered to be the one-dimensional quantity. The pitch unit of ‘mel’ is adopted on the assumption of transitive law regarding three tone pitches ‘a’, ‘b’ and ‘c’ that ‘a<c’ holds if ‘a<b’ and ‘b<c’. However, Shepard (1964) synthesized a specially contrived set of complex tones on a digital computer. Each complex tone consists of 10 frequency components spaced at octave intervals shown in Fig. 1(a). When we hear this special set of complex tones successively and repeatedly, each tone is always perceived as higher in pitch than the preceding. This effect can be called as an endless scale. Regarding these complex tones, transitive law does not hold for pitch. It is impossible to arrange these complex tones on one-dimensional scale with respect to pitch. Then, the following problems are raised. How should we understand the interrelation among these complex tones with respect to pitch? What are the factors governing the pitch of complex tones? To find a solution to these problems, we apply Hayashi’s quantification theory (1952) to our experimental results concerning the pitch of these complex tones and explore the factors governing the pitch of complex tones.

2. Psychological Experiment

2.1 Generation of Complex Tone

We used two groups of complex tones: groups A and B. Each group was composed of ten complex tones and all complex tones were synthesized on an electronic computer. The spectrum of each complex tone in group A and in group B has the structure illustrated in Fig. 1(a) and 1(b). The spectral envelope is constant in each

Key Words and Phrases: pitch of complex tones, endless scale, auditory nervous system, quantification theory of eij-type, paired comparison
* NHK Public Opinion Research Institute, 2-1-1, Atago, Minato-ku, Tokyo, 105
** NHK Broadcasting Science Research Laboratories, 1-10-11, Kinuta, Setagaya-ku, Tokyo 157
Fig. 1 (a) Frequency spectra of complex tones in group A

Fig. 1 (b) Frequency spectra of complex tones in group B
complex tone, but each frequency of the corresponding components is shifted to a position $2^{10}$ times higher.

We recorded a series of tones from No. 1 to No. 10, in that order repeatedly on magnetic recording tape. When we hear the No. 10 tone after the No. 1 tone, the pitch was perceived to become higher. Therefore, if this magnetic tape is played back, an endless scale was formed where the pitch continued to rise.

Since spectral information is a principal factor to govern the pitch of complex tones, each complex tone in group B was composed by slightly changing the spectrum of each complex tone in group A. The spectral envelope of each complex tone in group B is the same shape as that in group A. However, the peak of the envelope in group B always corresponds to the sixth component and shifts to the higher frequency as the number of the complex tone increases. When the series of complex tones in group B were heard repeatedly, they could be perceived as a somewhat imperfect endless scale, since the timbre difference between the No. 10 and No. 1 complex tones was clearly perceived.

2.2 Experimental Procedure

The phenomenon of endless scale is probably subjected to contextual influence. To evaluate the pitch of each complex tone without contextual influence, a paired comparison was made of the pitch for all two sets of each complex tone (ninety sets available by replacing the sequence). Each tone pair as shown in Fig. 2 was presented to the right ear of the subject through a headphone in a sound-proof room every three seconds. Nine subjects with normal hearing were required to compare the pitches of two complex tones and to decide which of the two tones was higher in pitch. Sound pressure level was set at about 30 dB (Sensation Level). The experiment was made twenty times for each subject.

Fig. 2 Schematic representation of the stimulus timing

2.3 Results

Tables 1(a) and (b) show the number of times where the complex tone $(i)$ was judged to be higher in pitch than the complex tone $(j)$ for all nine subjects. Comparing the pitches between complex tones $(i)$ and $(j)$ the following interesting phenomenon is observed.

Assuming that the "$\alpha$" is a positive integer from 1 to 9 and

$$j = i + \alpha$$ (1)

where, if $j > 10$, "$j - 10$" is defined as "$j$".

Then, when $\alpha = 1$, the pitch of the complex tone $(i)$ was judged to be lower than that of the complex tone $(j)$ in most cases. On the other hand, when $\alpha = 9$, the pitch of the
complex tone \((i)\) was judged to be higher than that of the complex tone \((j)\) in most cases. Generally, with the increase of the value of \(\alpha\), the frequency increases that the complex tone \((i)\) is judged to be higher in pitch than the complex tone \((j)\). This trend can also be observed in Table 1(b), though it is considerably decreased. If transitive law holds as to the pitch of the tones perfectly, the half column of the right top in Tables 1(a) and (b) should show zero.

The results in Tables 1(a) and (b) show that each group of complex tones cannot be arranged on the one-dimensional scale. The interrelation among these complex tones will be explored in the following section.

3. Analysis of the experimental result

3.1 Procedure of Analysis (Quantification Theory of \("e_{ij}\" Type)

The following attempts to describe the special arrangement of ten complex tones, based on the judgements of complex tone pitch obtained from the paired comparison method. As described in the previous section (2.3), as far as Table 1(a) is concerned,
almost all the judgements come to the same as to the pitch of the two complex tones, when tone stimulus number are close to each other. When tone stimulus number are separated 6 or 7 from each other, a divergency is observed in the judgements. With the further separation of the numbers, a higher percentage of agreements is observed in the judgements.

From the relationship represented in Table 1(a), the following relationship can be set up between the proximity of two complex tones and the judgements regarding the pitch of two complex tones:

\[
\text{Easy to judge} \quad \longleftrightarrow \quad \text{greater proximity} \quad \quad (2)
\]

\[
\text{Difficult to judge} \quad \longleftrightarrow \quad \text{smaller proximity}
\]

Since twenty experiments were performed, the value "20" or "0" is described, when perfect judgements were made as to the pitch. On the other hand, in cases where judgements diverge fifty-fifty as to the pitch, the value "10" is described. Thus, the experimental data is converted into data showing the proximity, according to the following conversion table (Table 2).

<table>
<thead>
<tr>
<th>Original data</th>
<th>20</th>
<th>19</th>
<th>18</th>
<th>17</th>
<th>16</th>
<th>15</th>
<th>14</th>
<th>13</th>
<th>12</th>
<th>11</th>
<th>10</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>6</td>
<td>7</td>
<td>8</td>
<td>9</td>
<td>10</td>
</tr>
</tbody>
</table>

| Conversion data | 11 | 10 | 9  | 8  | 7  | 6  | 5  | 4  | 3  | 2  | 1  |

Based on the conversion data, the complex tones with greater proximity are placed close to each other, while those with smaller proximity are placed distant from each other. We will employ the quantification theory of "e_{ij}" type developed by Chikio Hayashi. Here e_{ij} is a numerical value of proximity between i and j. The larger e_{ij} means the more proximity. The value of x_i (where i = 1, \ldots, 10) is determined for the complex tone in such a way that "Q" will become maximal, where

\[
Q = -\sum_{i=1}^{10} \sum_{j=1}^{10} e_{ij}(x_i-x_j)^2
\]

However, as observed in expression (3), the size of "Q" can be discussed only when the variance of "x_i" is a fixed value. Thus, the value of "x_i" (where i = 1, \ldots, 10) is determined in such a way that the value of "Q" will be maximal on the condition that a variance of "x_i" is a fixed value and that the mean value of "x_i" can be taken as zero for convenience of calculation, that is, in such a way that the following value will be maximal:

\[
G = \frac{Q}{\frac{1}{10} \sum_{i=1}^{10} x_i^2}
\]

This is, the following expression (5) should be solved with respect to "x_i":
where \( l = 1, \ldots, 10 \).

This solution is the characteristic vector corresponding to the maximum characteristic root in the characteristic equation obtained by (5).

### 3.2 Theoretical Case

If the relationship between the judgements of the complex tone pitches and proximity of the complex tones is as illustrated in expression (2), the case shown in Table 3 is an ideal case. This shows that there is no error in the judgements of the tone pitches between the close complex tones. On the other hand, a divergency is observed in the judgements of complex tone pitches which are separated by five or six levels.

<table>
<thead>
<tr>
<th>Table 3</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Ideal case</strong></td>
</tr>
<tr>
<td>( j )</td>
</tr>
<tr>
<td>( i )</td>
</tr>
<tr>
<td>1</td>
</tr>
<tr>
<td>2</td>
</tr>
<tr>
<td>3</td>
</tr>
<tr>
<td>4</td>
</tr>
<tr>
<td>5</td>
</tr>
<tr>
<td>6</td>
</tr>
<tr>
<td>7</td>
</tr>
<tr>
<td>8</td>
</tr>
<tr>
<td>9</td>
</tr>
<tr>
<td>10</td>
</tr>
</tbody>
</table>

Fig. 3 Ideal configuration by \( e_{ij} \),

The table 3 is subjected to the conversion shown in Table 2, and is solved by the "\( e_{ij} \)" method. The result is shown in Fig. 3. In Fig. 3, and almost perfect circle is obtained by connecting the complex tones in the order from No. 1, No. 2, No. 3, etc. This is an ideal arrangement in the endless scale.

### 3.3 Groups A and B

The result of analysis of nine subjects by the "\( e_{ij} \)" method differs according to each subject. To remove this spreading we will do the analysis after totaling nine subjects. To summarize the results of nine subjects, we do not totalize the exact experiment results of each subject, as shown in Tables 1(a) and (b). Instead, we have totalized the results of nine subjects after preparing the conversion data from Table 2.
The result of applying the "$\epsilon_{ij}$" method is shown in Figs. 4(a) and (b). By plotting the tones in the order of numbers, we obtain an almost circular form in group A (Fig. 4(a)). It (Fig. 4(a)) looks like a theoretical case (Fig. 3). On the other hand, in group B by plotting the tones in the order of numbers, we can not obtain a circular form. (Fig. 4(b)).

4. Discussion

The tones, such as sinusoidal tones and musical instrument tones can be perfectly arranged on the one-dimensional scale in order of the pitch, namely the fundamental frequency.

Two complex tones in an ideal endless scale whose frequency ratio is near 1 or 2 would be arranged at near points and two complex tones whose frequency ratio is around 1.5 would be arranged in the distance. In other words, the configuration of the complex tones in the ideal endless scale would form a circle on a two-dimensional plane.

Let us consider the reason why the configuration of complex tones in group A is almost circular and that in group B shows rather one-dimensional character as shown in Figs. 4(a) and 4(b).

Auditory neurons are known to be arranged in order of their characteristic frequency (CF or the frequency to which the neuron is most sensitive) on an auditory neuron layer (Tsuchitani and Boudreau, 1966). In other words, frequency of tone is converted into place in the auditory nervous system.

Pitch can be divided into two attributes; tone chroma and tone height. On the other hand, it is known that two types of cues for pitch perception are transmitted in the auditory nervous system. The place pitch cue is transmitted as a spatial response pattern on a neuron layer. Furthermore, auditory nerve fibers discharge impulses corresponding to peaks of the stimulating waveform for frequencies up to about 5kHZ (Rose et al 1967). The time pitch cue is transmitted as interspike intervals (time intervals between two successive impulses) of auditory nerve fibers.

Tone chroma changes regularly with one octave periodicity for frequencies below about 5 kHz. However, the regularity of the relationship between frequency and
tone chroma disappears over that frequency (Bachem, 1948). This suggests that tone chroma is based on the time cue, or the temporal distribution of neural impulses. On the other hand, tone height is perceived for all audible frequencies and the place cue is available for those frequencies. This suggests that tone height is based on the place cue, or the spatial distribution of firing neurons.

Since auditory nerve fibers emit impulses corresponding to peaks of tone stimulus waveform, interspike intervals of the auditory fibers are nearly equal to peak intervals of the stimulus waveform. Then, time intervals between prominent peaks (exceeding about 75 percent of the maximum peak) in the temporal fine structure of the waveform of complex tones were calculated and shown in Figs. 5(a) and 5(b). The abscissa represents time intervals between prominent peaks of each complex tone in group A and the ordinate shows tone stimulus number.

![Fig. 5 (a) Time intervals between prominent peaks in the waveform of complex tones in group A](image1)

We can observe in these figures that the corresponding peak intervals become shorter with increasing tone number. This tendency can be seen in the case that the No. 1 tone follows the No. 10 tone. This suggests that tone chroma changes so that tones are perceived higher in pitch when the series of tones from No. 1 to No. 10 is heard repeatedly.

As mentioned previously, tone height and tone chroma are considered as clues for comparing pitches of two complex tones. Tone height of each complex tone in group A is almost the same by equalizing the spectral envelopes. Therefore, tone height little serve as a clue in comparing pitches of complex tones in group A. Thus, tone chroma provides a main clue for this comparison. On the other hand, complex tones in group B show much more a one-dimensional configuration than those in group A. This can be explained by the difference in tone height originated from the difference in place cue among complex tones in group B. The difference in place cue is based on the frequency structure in which the peak frequency increases with the number of complex tones. However, the difference in tone chroma among complex tones in group B expected from Fig. 5(b) is considered to prevent the configuration from showing a perfect one-dimensional form.
We can, then, explain the configurations of complex tones in group A and in group B by the place cue and the time cue originated from neural activities of auditory neurons.

References


(Received June, 1979)