Perception of the Quality of Sound Amplitude-modulated with Triangular Waves

Kunihiro KUMAGAI¹, Kenji OZAWA², Yōiti SUZUKI³ and Toshio SONE⁴

¹Integrated System PF Engineering Department, ALPINE ELECTRONICS INC.,
20-1, Yoshima-Kogyo-Danchi, Iwaki 970-1192, Japan
²Faculty of Engineering, Yamanashi University, 4-3-11, Takeda, Kofu 400-8511, Japan
³Research Institute of Electrical Communication/Graduate School of Information Sciences, Tohoku University,
2-1-1, Aoba-ku, Sendai 980-8577, Japan
⁴Faculty of Systems Science and Technology, Akita Prefectural University,
84-4, Tuchiya-aza-Ebinokuchi, Honjoh 015-0055, Japan

Received December 28, 2000; Revised April 12, 2001

This paper discusses multidimensional perceptual space derived from the results of psychoacoustical experiments on similarities in the quality of sounds amplitude-modulated by nine triangular waves different in rise- and fall-times. For sinusoidal (1 kHz) and random-noise carriers, three experiments were conducted with different modulation frequencies, namely 8, 16, and 32 Hz. In each experiment, similarity in sound quality for stimuli with modulation depths of 0.4, 0.8 and 1.0, as well as the original carrier wave, were rated. The similarities were analyzed by a multidimensional scaling method (ALSCAL), and three-dimensional configurations were derived. A common configuration correlated with the modulation depth was seen irrespective of carrier wave and modulation frequency. There was a perceptual dimension related to shape of amplitude modulation for all conditions except for noise carrier with 32-Hz modulation. The difference in rise time caused larger differences in sound quality than similar differences in fall time up to a modulation frequency of 16 Hz, irrespective of the kind of carrier wave.

KEYWORDS: sound quality, amplitude-modulated sound, triangular wave, multidimensional scaling

1. Introduction

It is well known that the form of the amplitude envelope of a sound correlates with its sound quality [1]–[6]. For example, Miller and Carterette [1] carried out an experiment to evaluate similarities in sound quality for 27 complex tones with different amplitude envelopes, spectral envelopes, and fundamental frequencies. The similarities were analyzed with a Multi-Dimensional Scaling (MDS) technique and a three-dimensional configuration was derived. Though the effect of the amplitude envelopes was larger than that of the spectral envelopes, the effect was not systematically examined. Patterson [4], [5] observed differences in sound quality between sinusoidal waves with an exponential decay and their reverse in time. The results indicate the temporal asymmetry in the dynamic characteristics of the auditory system. Akeroyd and Patterson [6] carried out a similar experiment using wideband noise with the same amplitude envelope as the previous experiments [4], [5] and indicated the temporal asymmetry again. These studies compared the quality of a sound with a very steep rise and a gentle fall, and that of a sound with a gentle rise and a steep fall. It seems that these two sound stimuli were the extremes of that type of amplitude envelope. Thus, systematic changes in sound quality that would be caused by changes of amplitude envelope could not be observed by such a manipulation of the envelope.

In the previous research, the amplitude envelopes were so different that their results cannot be discussed comprehensively. To systematically investigate the relation between sound quality and amplitude envelopes, the envelopes should be varied systematically, not only in their shapes but also in their modulation depths. Furthermore, a systematic and detailed examination of the relation between sound quality and amplitude envelopes may result in a better understanding of the dynamic characteristics of the hearing system. With this in mind, psychoacoustical experiments were carried out in the present study to investigate the similarities in sound quality of stimuli using systematically varied amplitude envelopes and modulation depths for both sinusoidal and wideband-noise carriers. Then, perceptual spaces were derived for sound quality by use of an MDS technique.

2. Experimental Method

Nine different triangular waves, as shown in Table 1, were used as modulation waves, i.e., envelopes. These waves had the same repetition rate but different rise times given by \((n - 1)T/8\), where \(n\) is an integer from 1 to 9 and \(T\) is the repetition period. A triangular wave with rise time \(8 - (n - 1))T/8\) is the time-reversal of the wave with rise time \((n - 1)T/8\). Consequently, the Fourier power spectra of these two waves are identical so long as they have the same carrier wave, modulation frequency, and modulation depth.

A sinusoidal wave of 1 kHz and a broadband noise were used as the carrier signals. The carrier signals were am-
<table>
<thead>
<tr>
<th>n</th>
<th>Waveform</th>
<th>Modulation depth</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>0.4</td>
</tr>
<tr>
<td>1</td>
<td></td>
<td>41</td>
</tr>
<tr>
<td>2</td>
<td></td>
<td>42</td>
</tr>
<tr>
<td>3</td>
<td></td>
<td>43</td>
</tr>
<tr>
<td>4</td>
<td></td>
<td>44</td>
</tr>
<tr>
<td>5</td>
<td></td>
<td>45</td>
</tr>
<tr>
<td>6</td>
<td></td>
<td>46</td>
</tr>
<tr>
<td>7</td>
<td></td>
<td>47</td>
</tr>
<tr>
<td>8</td>
<td></td>
<td>48</td>
</tr>
<tr>
<td>9</td>
<td></td>
<td>49</td>
</tr>
</tbody>
</table>

![Fig. 1](image)

Time (s)

Amplitude-modulated by one of the nine triangular waves with one of three modulation depths, 0.4, 0.8 and 1.0, to give a set of sound stimuli. These modulation depths were selected for the following reasons. Firstly, a modulation depth of 1.0 was adopted as the most effective amplitude modulation condition. Secondly, a modulation depth of 0.4 was adopted as a small but adequately perceivable modulation depth [7]. Moreover, it is known that “roughness” of amplitude-modulated sounds is roughly proportional to the square of modulation depth [8]. Therefore, a depth of 0.8 was finally selected because its square (0.64) gives a value approximately intermediate between the squares of 1.0 (1.0) and 0.4 (0.16). Modulation depth is defined by \((|\text{MAX}| - |\text{MIN}|)/(|\text{MAX}| + |\text{MIN}|)\), where MAX and MIN are the maximum and the minimum values of an amplitude envelope, respectively. Stimuli without any modulation were also prepared. Therefore, the number of stimuli with the same carrier wave used in one experiment was 28 (\(=9\times3+1\)). For discussion, each stimulus is labeled with two- or three-digits, as shown in Table 1. The last digit represents the type of modulation wave, and the first one or two digits denote the modulation depth multiplied by ten.

Amplitude-modulated stimuli with a 1-kHz sinusoidal carrier were fed through a low-pass filter with a cut-off frequency of 5 kHz. The broadband noise was white noise low-passed at a cut-off frequency of 20 kHz. The broadband noise was independently generated at every presentation to eliminate the bias in spectra due to the limited duration of the stimuli. For both carriers, three independent experiments were conducted for modulation frequencies of 8, 16 and 32 Hz. All stimuli were generated on a computer and presented monaurally to the left ear of a subject via a D/A converter (Tucker-Davis Technologies DD1 with a sampling rate of 50-kHz) and a headphone (STAX SR-3). The RMS (Root Mean Square) values of the stimuli were equalized to eliminate the effect of loudness differences. They were presented at 70 dB SPL.

Similarity/dissimilarity in sound quality was rated with the rating scale method using a scale of seven categories (Category 1: indistinguishable–Category 7: entirely different) for all possible pairs of stimuli presented in random order with a temporal pattern as shown in Fig. 1. In the figure, stimuli [A] and [B] compose a pair to be rated their similarity/dissimilarity. As shown in the figure, this pair was presented repeatedly, where [A'] shows the first stimuli of the next pair. This repetition was performed to lighten the time order effect. In each independent experiment, the number of judgments was given by the permutation of all stimuli \((2\times6=756)\). This means
that all possible combinations ($\frac{1}{2} C_3^2 = 378$) were judged twice. The two judgments were averaged to obtain the similarity/dissimilarity for a combination. If the two judgments had differed by an average of more than one category for a subject, this subject should have been excluded due to lack of confidence. Fortunately, no subjects were excluded by this criterion.

Five males in their twenties and one male aged thirty-three participated as subjects in all experiments. Their hearing ability was ascertained to be normal ($<10 \text{ dB HL}$) from 100 to 10 kHz using a Békésy-type audiometer.

3. Analysis of the Experimental Result

3.1 Multidimensional Analysis of Similarities

The similarities obtained by the experiments were analyzed by a Multidimensional Scaling Method (Program: ALSCAL in SPSS; Model: EUCLID; Level: ORDINAL). The analysis was carried out using the EUCLID model [9] rather than the INDSCAL model [10] because the correlation among judgments for all subjects calculated as Kendall's rank correlation coefficient was statistically significant ($p < 0.01$). The high correlation coefficient indicates a small enough variance of the results among the subjects.

Three-dimensional configurations with a modulation frequency of 8 Hz are shown in Fig. 2 for two different carriers. The stress [11] for the sinusoidal carrier and the broadband noise carrier were 18.9% and 21.8%, respectively. Although the authors considered that the stresses were not sufficiently small, three-dimensional configurations were adopted because the authors were unable to find any meaningful relationship between the physi-

![Perceptual space](image)

**Carrier: Sinusoid**

![Perceptual space](image)

**Carrier: Random noise**

Fig. 2 Three-dimensional perceptual space for the sound quality derived by the multidimensional scaling method. The upper two panels are the results for a sinusoidal carrier. The lower two panels are those for a noise carrier. The modulation frequency is 8 Hz for both carriers.
cal parameters of the stimuli and the axes of the configurations for four- to six-dimensional solutions. In the figure, each open circle is identified by one of the stimulus numbers shown in Table 1. Stimuli with the same modulation depth are connected by lines; stimuli with modulation depths of 0.4, 0.8 and 1.0 are connected by solid lines, dotted lines and dashed lines, respectively. These configurations show very good agreement with those derived by the authors as preliminary experiments with twenty-one subjects [12], [13] and with subjects including two females [14]. This confirms the reliability of the present experiment carried out with six male subjects.

In Fig. 2, a distinct discrepancy is observed between the configurations for the sinusoidal carrier and that for the broadband noise carrier. For the sinusoidal carrier, stimuli with a last digit of either 1 or 9 in stimulus numbers, abbreviated respectively as *1 or *9 hereafter, protrude from the other stimuli along axis III. Meanwhile, there are no such protrusions in the configuration for the wideband-noise carrier. This discrepancy may be explained by differences in excitation patterns in the hearing system between both stimuli, which reflect the differences in their power spectra. In this paper, the excitation pattern of a sound is represented by the output of a gammachirp-filter bank [15].

Figure 3 shows the excitation patterns of the stimuli with the modulation frequency of 8 Hz and the modulation depth of 1.0. For the other modulation frequencies and modulation depths, the same tendencies were ob-

![Fig. 3 Excitation patterns of Stimuli 101 to 109 with an 8-Hz modulation frequency. A sinusoidal wave and noise were used as carrier signals for the upper and the lower panel, respectively. The horizontal axis represents the center frequency of an auditory filter.](image-url)
served in the excitation patterns. In the figure, the excitation patterns for Stimuli 101 and 109 are represented by solid lines and those for 102 to 108 by dotted lines. All excitation patterns for the noise carrier are almost identical (lower panel). For the sinusoidal carrier (upper panel), on the other hand, the excitation patterns for 101 and 109, which are almost identical, are clearly different from those for 102–108, which are almost identical, too. For the sinusoidal carrier, the steepest modulation waves of Stimuli 101 and 109 result in the excitation patterns being broadened to both sides of the carrier frequency. As a result, there was a big difference between the excitation patterns for stimulus numbers 101 and 109, and those for Stimuli 102 to 108. This difference may subjectively be perceived as a difference in sound quality, resulting in distinctively separate locations for 101 and 109 in the configurations (Fig. 2-(a)). Therefore, for the judgments of dissimilarity in sound quality between Stimuli *1 and

Fig. 4  Three-dimensional perceptual space derived by the multidimensional scaling method except for Stimuli *1 and *9. A sinusoidal wave was used as the carrier signal. The three columns correspond to the modulation frequencies (upper: 8 Hz; middle: 16 Hz; lower: 32 Hz).
*9 and others, it is considered that the subjects used a cue other than the temporal difference in amplitude envelope, i.e., the difference in the power spectra. Since the aim of this study was to investigate the effect of amplitude envelopes on the sound quality, further analysis was carried out after the results relating to Stimuli *1 and *9 were excluded. In advance of the analysis it had to be considered whether the subjects' judgments would differ from the present results if these stimuli had been excluded, because they might have acted as some anchor points for rating. However, it was concluded that this was not a problem because the finally subjects' judgments were based on the instruction that the category "entirely different" meant "the greatest difference among the stimuli of the set". Thus, the pairs judged as "entirely different" were not only those including Stimuli *1 or *9 but also those including Stimulus 0. That is, so long as Stimulus 0 is included in the analysis, we may assume that the pos-

Fig. 5  Three-dimensional perceptual space derived by the multidimensional scaling method except for Stimuli *1 and *9. Broadband noise was used as the carrier signal. The three columns correspond to the modulation frequencies (upper: 8 Hz; middle: 16 Hz; lower: 32 Hz).
sible deformation of the calculated perceptual space caused by the exclusion of Stimuli *1 and *9 is negligible.

Figures 4 and 5 show three-dimensional configurations derived from the dissimilarities other than those of Stimuli *1 and *9 for the sinusoidal carrier and the noise carrier. For the sinusoidal carrier, the stesses for 8-, 16- and 32-Hz modulation frequencies were 20.4%, 20.8% and 20.9%, respectively. For the noise carrier, the stesses for 8-, 16- and 32-Hz modulation frequencies were 21.6%, 20.2% and 15.3%, respectively. Although the stesses were not small enough, three-dimensional configurations were adopted. This was because there were only two parameters of the stimuli for the experiment (modulation depth and waveform of amplitude envelopes), and thus adoption of higher dimensions might have resulted in axes that do not have a meaningful relationship with physical parameters. In fact, the visual inspection of the results of four- to six-dimensional solutions brought about no meaningful relationships.

A configuration derived from the EUCLID model has freedom for rotation, parallel translation, expansion and contraction [9]. Thus, for further discussion, the configurations derived by ALSCAL were compared through a transformation given by the following operation:

$$Y = cXT + I_n \mathbf{a}' + \mathbf{E}$$

$$= \mathbf{X}_2 + \mathbf{E},$$

(1)

where \( \mathbf{X} \) and \( \mathbf{Y} \) are two configurations, constituted of \( m \)-dimensional \( r \) elements [16], to be compared among different experimental conditions. In the present examination, the number of dimensions was three, and the number of elements was equal to 22, which is the number of stimuli under consideration in an experiment \( 3^7(9 - 2) + 1 \). \( T \) is an \( r \times r \) orthogonal matrix representing rotation, \( c \) is a constant representing expansion and contraction, \( \mathbf{a}' = (a_1, a_2, \ldots, a_r) \) represents parallel translation, and \( \mathbf{I}_1 = (1, 1, \ldots, 1) \). Consequently, configuration \( \mathbf{X} \) can be transformed into the new configuration \( \mathbf{X}_2 \), which is as close as possible to \( \mathbf{Y} \), by minimizing error \( \mathbf{E} \). Here, the configuration for the sinusoidal carrier with the modulation frequency of 8 Hz was adopted as \( \mathbf{Y} \). For the other conditions the configurations transformed by Eq. (1) were calculated. Figures 4 and 5 show the resultant configurations.

The features shown in each configuration are enumerated as follows:

(a) On the I–III plane, horseshoe-shaped rows are seen. This structure seems to relate to the modulation depth.

(b) Except for the case of the noise carrier with the modulation frequency of 32 Hz (Fig. 5-(c)), the distribution along axis II corresponds to the systematic change in the waveform of amplitude envelopes. Furthermore, Stimuli *2 to *5 with modulation depths of 0.8 or 1.0 were more sparsely distributed than Stimuli *5 to *8. For the stimuli with the modulation depth of 0.4, a similar configuration is seen only for the sinusoidal carrier.

(c) In the configuration for stimuli with the noise carrier and the modulation frequency of 32 Hz (Fig. 5-(c)), the difference in the waveforms of amplitude envelopes is only slightly reflected in the distributions, while the difference in the modulation depths is clearly observed in the distribution.

### 3.2 Quantitative Evaluation on Commonality among Configurations

To quantitatively investigate the commonality of these features among the configurations, correlation coefficients between two sets of coordinates for every possible pair of configurations were calculated for each axis. Pearson’s correlation coefficients between every two axes obtained for every pair of configurations are shown in Table 2. In this table, experimental conditions are expressed as a combination of the kind of carrier and the modulation frequency, \( f_m \). Furthermore, “X” and “Y” shown in the table along with the experimental conditions correspond to \( \mathbf{X} \) and \( \mathbf{Y} \) in Eq. (1), respectively. For example, the appearance of three numerical values in a cell for the row of \( \mathbf{X} \): pure tone, \( f_m = 16 \) and the column of \( \mathbf{Y} \): pure tone, \( f_m = 8 \) indicates the correlations between coordinates on each axis of the configurations shown in Figs. 4-(a) and 4-(b). At the same time, Table 2 shows the results of the \( t \)-test for the correlation coefficient of each axis between both figures. The symbols ** and * indicate that the correlation is statistically significant beyond 1% and 5%, respectively.

Under the condition of 8- and 16-Hz modulation frequencies, it is seen that the commonality between the corresponding axes of two configurations is significant irrespective of the kind of carrier. This suggests that the effect of the amplitude envelope on its sound quality is independent of the kind of carrier wave. Furthermore, the configuration of the sinusoidal carrier with the modulation frequency of 32 Hz closely correlates with the configurations of modulation frequencies of 8 and 16 Hz for axes I and II. In the configurations for modulation frequencies of 8 and 16 Hz, axes I and III result in a horseshoe-shaped distribution of modulation depth, while the distribution of the modulation depth is nearly linear along axis I for the modulation frequency of 32 Hz. This results in a low correlation for axis III.

For the noise carrier, axis II in the configuration for the modulation frequency of 32 Hz does not correlate with axis II in the configurations of 8- and 16-Hz modulation frequencies. This is because the stimuli are systematically and linearly distributed along axis II for the modulation wave of 8 and 16 Hz, while they are clustered for 32 Hz. In the experimental results, axes I and III correspond to the modulation depth, and axis II cor-
Table 2. Correlation coefficients between corresponding axes in every two configurations by minimizing their differences by rotation, parallel translation, expansion and contraction. $f_m$ means modulation frequency [Hz].

<table>
<thead>
<tr>
<th></th>
<th>Y: pure tone</th>
<th></th>
<th></th>
<th>Y: noise</th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$f_m = 8$</td>
<td>$f_m = 16$</td>
<td>$f_m = 32$</td>
<td>$f_m = 8$</td>
<td>$f_m = 16$</td>
<td>$f_m = 32$</td>
<td></td>
</tr>
<tr>
<td>X: pure tone, $f_m = 8$</td>
<td>axis I</td>
<td>axis II</td>
<td>axis III</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.9413**</td>
<td>0.9365**</td>
<td>0.9118**</td>
<td></td>
<td>—</td>
<td>—</td>
<td></td>
</tr>
<tr>
<td>X: pure tone, $f_m = 16$</td>
<td>axis I</td>
<td>axis II</td>
<td>axis III</td>
<td>0.9316**</td>
<td>0.9288**</td>
<td>—</td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.9278**</td>
<td>0.9083**</td>
<td></td>
<td>0.2819</td>
<td>0.2674</td>
<td>—</td>
<td></td>
</tr>
<tr>
<td>X: pure tone, $f_m = 32$</td>
<td>axis I</td>
<td>axis II</td>
<td>axis III</td>
<td>0.8997**</td>
<td>0.8804**</td>
<td>0.8809**</td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.8786**</td>
<td>0.8793**</td>
<td>0.9112**</td>
<td>0.8690**</td>
<td>0.8527**</td>
<td>0.3273</td>
<td></td>
</tr>
<tr>
<td>X: noise, $f_m = 8$</td>
<td>axis I</td>
<td>axis II</td>
<td>axis III</td>
<td>0.9054**</td>
<td>0.8958**</td>
<td>0.8833**</td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.8710**</td>
<td>0.8810**</td>
<td>0.9253**</td>
<td>0.8969**</td>
<td>0.8764**</td>
<td>0.9089**</td>
<td></td>
</tr>
<tr>
<td>X: noise, $f_m = 16$</td>
<td>axis I</td>
<td>axis II</td>
<td>axis III</td>
<td>0.9058**</td>
<td>0.8855**</td>
<td>0.8507**</td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.9071</td>
<td>0.1255</td>
<td>0.9381**</td>
<td>0.8003**</td>
<td>0.8122**</td>
<td>0.8717**</td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.0971</td>
<td>0.1967</td>
<td>0.0617</td>
<td></td>
<td>0.0971</td>
<td>0.1255</td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.8003**</td>
<td>0.8122**</td>
<td>0.8717**</td>
<td></td>
<td>0.8855**</td>
<td>0.8122**</td>
<td></td>
</tr>
</tbody>
</table>

responds to the systematic change in the amplitude envelope. Consequently, for the noise carrier, it is considered that the effect of modulation depth on sound quality shows a similar tendency for different modulation frequencies, while the effect of the amplitude envelope is different.

With the modulation frequency of 32 Hz, the correlations have low values for axes II and III between the two carriers. This is due to both of the above reasons; the low correlation for axis II is attributable to the clustered distribution of the stimuli with the noise carrier at the modulation frequency of 32 Hz along the axis and that for axis III reflects the relatively linear configurations of the stimuli with the sinusoidal carrier at the modulation frequency of 32 Hz in the I–III plane.

4. Discussion

Feature (a) listed in Section 3.1, i.e., that the horseshoe-shaped distributions in the I–III plane are aligned with the modulation depth, shows that the magnitude of the fluctuation in the amplitude envelope influences the sound quality. The horseshoe shape implies that there is a non-linear relationship between the physical factor of the modulation depth and the subjectively perceived modulation depth.

For Feature (b), i.e., the non-linear distribution along axis II, the dissimilarity shown by the experiments is discussed in detail. Fisher’s rank test for the two dissimilarities between Stimuli *2 and *5 and between Stimuli *5 and *8 with the same modulation depth was conducted. The results are shown in Table 3. The values in this table show the probabilities beyond which the dissimilarities between Stimuli *2 and *5 are significantly larger than those between Stimuli *5 and *8. From this table, it can be seen that the dissimilarities between Stimuli *2 and *5 may be regarded as being larger than those between Stimuli *5 and *8 for the stimuli with modulation depths of 0.8 and 1.0, except only for the case of stimuli with noise carrier at the modulation frequency of 32 Hz. This difference appears in Figs. 4 and 5 as Feature (b), i.e. Stimuli *2 to *5 were sparsely distributed, while Stimuli *5 to *8 were closely distributed. As mentioned earlier, the difference in power spectra between Stimuli *2 and *5 is equal to the difference between Stimuli *5 and *8. This means that Feature (b) cannot be explained by the difference in the power spectra. The relationships between the form of the amplitude envelope and the sound quality was investigated by Patterson [5] and Akerooyd [6]. Stimuli *1 and *9 are similar to those used in their experiment. Their results showed that the stimulus corresponding to our Stimuli *9 was perceived to be more like a pure tone than the stimulus corresponding to Stimulus *1. The subjective distances between Stimulus 0 and Stimulus *1 and that between Stimulus 0 and Stimuli *9 which were derived in the present study, however, are nearly identical, as shown in Fig. 2. This result of ours, therefore, does not support their results. The reason for it may be the following: The amplitude envelope for their stimuli was an exponential function while those in the present study were a linear function. Therefore, our stimuli had milder peaks than their stimuli. Moreover, since the sound quality of Stimuli *1 and *9 were quite different from those of Stimuli *2 to *8, our subjects might have failed to express the difference between Stimuli *1 and *9 by use of the seven-category scale. Instead,
Table 3. Examination of the two dissimilarities between Stimuli *2 and *5 and between Stimuli *5 and *8 in the experimental results. The value represents the significant level when the distance between Stimuli *2 and *5 is longer than the distance between Stimuli *5 and *8.

<table>
<thead>
<tr>
<th>Modulation depth</th>
<th>0.4</th>
<th>0.8</th>
<th>1.0</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pure tone</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>( f_m = 8 )</td>
<td>0.002</td>
<td>0.001</td>
<td>0.000</td>
</tr>
<tr>
<td>( f_m = 16 )</td>
<td>0.001</td>
<td>0.004</td>
<td>0.000</td>
</tr>
<tr>
<td>( f_m = 32 )</td>
<td>0.000</td>
<td>0.087</td>
<td>0.019</td>
</tr>
<tr>
<td>Noise</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>( f_m = 8 )</td>
<td>0.852</td>
<td>0.023</td>
<td>0.000</td>
</tr>
<tr>
<td>( f_m = 16 )</td>
<td>0.159</td>
<td>0.005</td>
<td>0.050</td>
</tr>
<tr>
<td>( f_m = 32 )</td>
<td>0.159</td>
<td>0.070</td>
<td>0.121</td>
</tr>
</tbody>
</table>

Our results showed different types of asymmetric distribution between Stimuli *2 to *4 and *8 to *6. This means that the temporal asymmetry exists in the hearing system not only for sharp envelopes with abrupt changes in time but also for envelopes with milder changes in time. This temporal asymmetry seems to be one of the general characteristics of sound quality perception. The perceptual asymmetry of a signal with sharp rise/fall and gentle fall/rise as observed in this experiment could be accounted for by physiological mechanisms of the auditory system. In the auditory pathway, there are so-called On and Off type fibers, which respond to the steep rise- and fall-parts, respectively, of a burst. Thus, if the sensitivities of On-type and Off-type fibers for the same rate of temporal transients differ, the asymmetry might be attributable to the difference in the sensitivities.

Feature (c) shows that the difference in the shape of the amplitude envelope was not reflected in sound quality for the noise carrier at the modulation frequency of 32 Hz. For the sinusoidal carrier, even at the modulation frequency of 32 Hz, the change in the shape of the amplitude envelope resulted in a change in the sound quality. For the noise carrier, the hearing system seems to hardly detect the change in shape of the amplitude envelope at the modulation frequency of 32 Hz because of the inherent fluctuation of the carrier. In a sense, the intrinsic fluctuation in the wideband noise masked the fluctuation brought by the amplitude modulation at 32 Hz and the subjects could only evaluate the degree of modulation, but could not perceive the detailed pattern of the temporal fluctuation. This feature can also be accounted for by the physiological mechanisms of the auditory system. It is known that the primary auditory nerve fibers often exhibit an overshoot in their responses to a steep rise of a tone burst [17]. As the overshoot suddenly damps in 10 to 20 ms, the change in the responses evoked by the change in the amplitude envelopes might have been confused by the overshoot, because the reciprocal of 32-Hz, 31.25 ms, was comparable with the duration of the overshoot.

5. Conclusions

In this study, three-dimensional perceptual spaces for sound quality were obtained from dissimilarities in sound quality for a sinusoidal wave and broadband noise modulated by triangular waves by means of a multidimensional scaling method. The results show that the effect of modulation depth on sound quality occurred irrespective of the kind of carrier wave and modulation frequency. Moreover, it was shown that the systematic difference in the amplitude envelope did not always cause a corresponding difference in sound quality. It was also shown that the magnitude of the fluctuation in the amplitude envelope influences the sound quality. The change in the rise-time of the modulating signal resulted in a more prominent change in the sound quality than that of the fall-time, i.e., the difference in the rise-time was reflected as a significant change in the similarity in sound quality when the rise-time was short, while the smaller the change the longer the rise-time. This asymmetry of the amplitude envelopes appeared for both kinds of carrier waves at low modulation frequencies up to 16 Hz.

Acknowledgment

This study was partially supported by a Grant-in-Aid for Scientific Research (No. 09838003) by the Ministry of Education, Science, Sports and Culture of Japan.

REFERENCES