A model of perceptual distance for group delays based on ellipsoidal mapping

Hideki Banno*, Kazuya Takeda and Fumitada Itakura

Graduate School of Engineering, Nagoya University/CIAIR, Furo-cho 1, Chikusa-ku, Nagoya, 464–8603 Japan

(Received 23 July 2001, Accepted for publication 7 August 2001)

Keywords: The human auditory system, Distance measure, Phase spectrum, Group delay, Ellipsoidal mapping

PACS number: 43.66.Nm

1. Introduction

Traditionally, the phase spectrum was believed to be insignificant in auditory perception. However, you can see that a white noise signal and a pulse train, made to differ only in their phase spectral characteristics, are perceptually very different in some cases. This implies that the human auditory system is not incapable of perceiving such phase differences. Many reports indeed support this assumption [1–3]. Nevertheless, no perceptual distance measures for phase spectrum fitting to the human auditory system is as yet available. Here we develop a computational method to estimate perceptual phase distance measures based on an ellipsoidal mapping.

The phase spectrum is sometimes transformed into the group delay spectrum which directly represents the timedelay value of the stimulus. Therefore, we used a group delay representation to design stimuli for perceptual experiments. Thus, we used a group delay representation to design stimuli.

2. Phase distance measure based on ellipsoidal mapping

Subjective similarity between stimuli which have a certain group delay value in a particular frequency band were obtained by conducting perceptual experiments [4]. These experimental results allowed us to develop a method of mapping to realize an optimum perceptual phase representation of a signal using a simple equation. Prior to formulating this equation, we discuss some of the limitations and constraints required.

Our experiments showed that when the group delay peak values of stimuli are between −1 ms and 2 ms, they are perceived to be equivalent to a pulse of zero phase regardless of their center frequencies and bandwidths. Moreover, when the peak values are less than −8 ms or more than 10 ms and the bandwidths are less than 1 ERB (equivalent rectangular bandwidth of the auditory filter), each stimulus pair is perceived to be similar. The similarity curves are almost symmetrical with respect to the vertical axis with a slight positive shift from the origin. If a stimulus is neither zero phase nor nonzero phase, the similarity scores are high where the group delay peak values are close to that of the stimulus, and low where the group delay peak values are far from that of the stimulus. However, the low similarity scores are improved when the group delay values are close to a negative value of the group delay value of the stimulus.

Based on these perceptual similarity results, we introduce an ellipsoidal function to explain such findings with a simple equation. We refer this ellipsoidal function as the similarity ellipse shown in Fig. 1. Stimuli are assigned to the ellipse boundary at uneven intervals according to the peak values of the stimuli. In this ellipse, zero-phase stimuli are assigned to around zero radian, and nonzero-phase stimuli are assigned to around $\pi$ radian. Stimuli with positive group delay and negative group delay are assigned to upper half area and lower half area, respectively. The similarity can be estimated by the Euclidean closeness between locations on the ellipse boundary. Limiting the perceptual similarity measures of the stimuli to an ellipsoidal boundary formulates the framework for our study.

A similarity score (in our experiments) between a stimulus pair can be estimated by

\[ s(t_1, t_2) = 5 - d(t_1, t_2), \quad 1 \leq s(t_1, t_2) \leq 5 \] (1)

where, $t_1$ and $t_2$ are peak values of two stimuli, respectively and $d(t_1, t_2)$ is a distance between the stimuli given by

\[ d(t_1, t_2) = |x(t_1) - x(t_2)|, \quad 0 \leq d(t_1, t_2) \leq 4. \] (2)

Here, $x(t)$ is a vector on an ellipse boundary as a function of a group delay value, described by

\[ x(t) = (a \cos \phi(t), b \sin \phi(t)), \quad a, b > 0 \] (3)

where, $a$ and $b$ are parameters defining the shape of the ellipse. The function $\phi(t)$ is a nonlinear function mapping a change in group delay varying between $-\pi$ and $\pi$, where $t$ is limited between −16 ms to 16 ms. This function is based on the sigmoidal function given by

\[ \phi(t) = \frac{2\pi}{1 + \exp(-(t - u)/v)} - \pi, \quad u, v > 0 \] (4)

where, $u$ describes the horizontal shift and $v$ the slope of the function. The parameters $a$, $b$, $u$ and $v$ were optimized based on the minimum-least-square-error criterion for $s(t_1, t_2)$ to fit with the similarity scores obtained by subjective experiments.

3. Evaluation

The parameters $a$, $b$, $u$ and $v$ were optimized by using the following cases independently, (1) $f_c = 1,000$ Hz and $c_x = 0.8$, (2) $f_c = 1,000$ Hz and $c_x = 2.0$, (3) $f_c = 4,000$ Hz and $c_x = 0.8$, and (4) $f_c = 4,000$ Hz and $c_x = 2.0$. $f_c$ is the
center frequency, $c_e$ is the bandwidth parameter that gives the bandwidth by multiplying $c_e$ and the equivalent rectangular bandwidth (ERB) together. $\tau_0$ is the group delay value at the center frequency.

Table 1 shows the optimum parameters and errors between the estimated similarity scores and the similarity scores obtained by subjective experiments. In this table, the errors obtained are about 0.3 in similarity score. These errors are quite small considering that the maximum similarity score is five and the subjective experiments have an inherent variance associated with them.

Figure 2 shows the resultant similarity ellipses. The ellipses become larger when $c_e$ is 2.0, that is, the bandwidth is 2 ERB. This is typical for the minor axes of the ellipses. Figure 3 shows the resultant similarity score representation by this optimization method and the similarity scores obtained by subjective experiments. It is found that the estimated similarity scores well approximate the similarity scores obtained by subjective experiments. These results imply that the human auditory system perceives the phase spectral differences based on a simple ellipsoidal function.

### Table 1 Optimum parameter values and errors.

<table>
<thead>
<tr>
<th>Stimuli type</th>
<th>$a$</th>
<th>$b$</th>
<th>$u$</th>
<th>$v$</th>
<th>Error</th>
</tr>
</thead>
<tbody>
<tr>
<td>(1)</td>
<td>1.5</td>
<td>0.4</td>
<td>0.4</td>
<td>4.2</td>
<td>0.31</td>
</tr>
<tr>
<td>(2)</td>
<td>1.9</td>
<td>0.8</td>
<td>0.3</td>
<td>6.2</td>
<td>0.30</td>
</tr>
<tr>
<td>(3)</td>
<td>1.8</td>
<td>0.4</td>
<td>1.0</td>
<td>5.8</td>
<td>0.27</td>
</tr>
<tr>
<td>(4)</td>
<td>2.2</td>
<td>0.7</td>
<td>0.8</td>
<td>6.8</td>
<td>0.31</td>
</tr>
</tbody>
</table>

---

**Fig. 1** The similarity ellipse for the estimation of experimental similarity scores.

**Fig. 2** The similarity ellipses. Upper figures and lower figures are the cases where $f_c = 1.000\,\text{Hz}$ and $f_c = 4.000\,\text{Hz}$, respectively. Left figures and right figures are the cases where $c_e = 0.6$ and $c_e = 2.0$, respectively.

**Fig. 3** Relationship between the group delay peak value $\tau_0$ and the similarity score when $f_c = 4.000\,\text{Hz}$ and $c_e = 0.6$. The solid line and the dashed line describe the estimated similarity score and the similarity score obtained by subjective experiments, respectively. Each panel shows the case where the group delay peak value $\tau_0$ is (a) $-16\,\text{ms}$, (b) $-8\,\text{ms}$, (c) $-4\,\text{ms}$, (d) $-2\,\text{ms}$, (e) $2\,\text{ms}$, (f) $4\,\text{ms}$, (g) $8\,\text{ms}$ and (h) $16\,\text{ms}$.

4. Summary

We developed a computational method to estimate perceptual phase distance measures based on an ellipsoidal mapping. Perceptual results obtained from our subjective experiments match this ellipsoidal-based computational method. Future work includes applying this method to speech.

References

