Towards the development of objective difficulty measure in Technical Ear Training tasks

Atsushi Marui* and Toru Kamekawa

Technical Ear Training is a method to improve the ability to focus on a specific sound attribute. It is also used to be able to communicate using the common language used in the industry such as Hz and dB. Although it is essential to gradually harden the task difficulty for successful technical ear training, the objective measure of the difficulty is still not known. Therefore, the tasks are decided by the teacher's own ears and experiences, leading to inefficiency when students want to train themselves in the teacher's absence. As the first step towards understanding this tacit ability of knowing the task difficulty, the authors investigated the correlation between the students' subjective ratings of the task difficulty and the physical measures calculated from the sound materials used in the training. A linear regression model (R-squared=0.629) which predicts the subjective task difficulty from residual of the linear fit through the spectra of the sound material was created. This model may provide a firm step towards the goal of developing objective difficulty measure in technical ear training.

Published by the Acoustical Society of America through the American Institute of Physics
INTRODUCTION

Technical Ear Training

Technical Ear Training (TET) is a method to improve the ability to focus on a specific sound attribute, and mainly used by students and professionals involved in audio and acoustics such as recording engineers, acousticians, and researchers/developers of audio equipment (Iwamiya et al., 2003). The training is also for improving the use of the common languages in the field such as Hz and dB. Only the knowledge of “6 dB is sound pressure level being twice as much” without understanding how it feels to the ears is not the ideal skill that can be used in the field. Furthermore, it is desirable to be able to use physical units such as Hz and dB instead of vague and arbitrary words such as “in the low end” or “a bit louder” when precise and prompt communication of the strength of a sensation to other people is needed. TET is a class of training methods to acquire the skill of relating physical parameters and perceptual attributes in the field of audio engineering and acoustics.

Setting Difficulty Levels in TET

It is essential to gradually raise the training difficulty for successful TET. Key factors that influence the difficulty include task procedure, response format, and sound materials.

Typical task procedures include pitch discrimination, sound level difference identification in dB, identification of center frequency/level/bandwidth for spectral changes, discrimination of different harmonic distortion amount, and identification/discrimination of codecs. In center frequency identification of a peaking filter (as shown in Figure 1), for instance, a gain of 12 dB at the center frequency is easier than the gain of 6 dB or 3 dB, and attenuation of 12 dB is more difficult than gain of 12 dB.

Response format is largely divided into active and passive training. In active training a student adjust parameters (Hz, dB, etc.) to reach the correct answer. In passive training a student compare and identify the difference between two sounds which are the reference and a modified sound. Absolute identification of the modification without listening to the reference is often done in the passive training as well. Response format is also influenced by the number of students to be trained at a time. Passive training is more suitable in a classroom situation where all students listen to the same sound and respond accordingly. There is a passive training for self-training situation (Moulton, 1993) as well. However, use of a computer system connected to a student’s past database giving personalized training program (Kaniwa et al., 2011), which is difficult in a group or passive training, can be realized and beneficial for a student in individual training.

Sound materials are sound source sampled from commercial recordings, synthesized, or recorded. The materials are used after being modified with effects processors in different ways suitable to the goal of the training. The materials can be any kind of music, speech, and other sounds. Puretones and pinknoise can be used depending on the purpose of the training, but the authors only use music since the course’s emphasis is on music recording. To the authors’ knowledge, common to all the training programs is that the choice of sound materials is done by the teacher or the developer offering the training program. The choice of sound materials should be dependent on the purpose and the goal of the training program and ability of the students. However, except for few parameters such as bandwidth of the sound material, the objective measure to choose the sound that is suitable for TET tasks is still not known. Therefore, the sound materials are chosen by the teacher’s own ears and experiences, leading to inefficiency and/or ineffectiveness in self-training by a student in the teacher’s absence.
Motivation

Objective measure for difficulty in TET is necessary for effective and efficient self-training of TET. As stated above, the difficulty is dependent on many aspects of the training. In the current study, the authors investigated the relation of the students’ subjective ratings of the task difficulty and the training scores against the physical measures calculated from the sound materials used in the training.

In this paper, a part of training method relevant to the study is introduced first. Then, the subjective difficulty evaluation is explained, followed by the objective difficulty counterpart. A trainer in TET is called a teacher and a trainee is called a student. Programs authors created that run on Mathworks Matlab R2012b were used for preparing and analyzing sound materials and for statistical analyses of the results.

TRAINING METHOD

Background

This section presents a part of the training method which is related to the evaluation of the subjective and objective difficulty discussed later in this paper.

The training is done within an university course with emphasis on sound recording. Because the training session is a part of the weekly class meeting, time and day of the training is fixed. About fifteen to twenty students participate in the training at once.

The training method developed for the current university course authors teach is inspired by the methods developed in Faculty of Design at Kyushu University, Japan (Kitamura and Sasaki, 1996, Chap. 3) and the method developed in Faculty of Music at McGill University, Canada (Quesnel and Woszczyk, 1994; Quesnel, 1996). McGill method can be seen as the computerized version of Timbre Solfeggio which originates in Chopin Academy of Music, Poland (Letowski, 1985; Miśkiewicz, 1992). An implementation inspired by the McGill method was created by Corey (2010), and currently a variant of it is used in McGill University. Our method uses hints of sound materials preparation and training methodology from McGill University and the format of group lesson from Kyushu University.

Task Procedure

The task under investigation is called the peak identification which a student listens to the original and the modified versions of a sound material sequentially and be asked to identify the center frequency of a spectral peak. The center frequency of the peak is one of the nine 1-octave spaced bands in between 63 Hz and 16 kHz inclusive, gain is +12 dB, and bandwidth is set in quality factor $Q = 2.0 \approx 0.714$ octaves wide. The shape of the filter used is shown in Figure 1. The student is informed with the possible frequencies, gain value, and $Q$ prior to the training and responds with a paper and a pencil.

The sound materials used are 10 to 15 seconds of music excerpts described next. When presenting to the students, intra-trial time is 0.5 seconds and inter-trial time is 3 seconds (Figure 2). One block of of ten consecutive trials is done for a particular task.

Sound Materials

Twenty-four music excerpts chosen widely from classical, jazz, rock, and pops genre are used as sound materials in the authors’ current TET course. The criteria of choosing the excerpts are bandwidth being sufficiently wide and having less variation in loudness level within the excerpt.
A few exceptions that do not follow the criteria such as solo performances of violin or piano are included in the set of sound materials.

Figure 3 shows 1/3-octave frequency spectra of two of the sound materials used in the training. The materials are taken from “Sunday Morning After” (in album Everybody’s Got a Story, 2001) by Amanda Marshal and “Leonore, No. 3, Op. 72b” (Georg Solti / Chicago Symphony Orchestra, 1975) by L. v. Beethoven, respectively. In each plot, levels are in relative dB, vertical dashed line shows spectral centroid (in Hz), horizontal whisker shows standard deviation about the spectral centroid (in Hz), and diagonal line shows linear regression line in 63 Hz–16 kHz range.

**EVALUATION OF SUBJECTIVE DIFFICULTY**

**Method**

In order to collect data for subjective difficulty of the training task mentioned in the preceding section, the students participated in the training were asked to listen to a sound material and rate “how difficult the sound is when presented in the peak identification task” in seven-point bipolar scale with end points being “very easy” and “very difficult,” for all 24 sound materials used in the training (Table 1).

A sound material was played back without any modification but with the same playback condition as the training session; that is, the original version of a sound material was played back twice with 0.5 seconds intra-trial interval and after 3.0 seconds of inter-trial interval the next sound material was played.
FIGURE 3: Frequency spectra of the sound material from “Sunday Morning After” (left) and “Leonore, No.3” (right). Levels are in relative dB, vertical dashed line shows spectral centroid (in Hz), horizontal whisker shows standard deviation about the spectral centroid (in Hz), and diagonal line shows linear regression line in 63 Hz–16 kHz range.

TABLE 1: Seven-point bipolar scale used for rating subjective difficulty of sound materials.

<table>
<thead>
<tr>
<th>Rating</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>+3</td>
<td>very difficult</td>
</tr>
<tr>
<td>+2</td>
<td>difficult</td>
</tr>
<tr>
<td>+1</td>
<td>somewhat difficult</td>
</tr>
<tr>
<td>0</td>
<td>moderate</td>
</tr>
<tr>
<td>−1</td>
<td>somewhat easy</td>
</tr>
<tr>
<td>−2</td>
<td>easy</td>
</tr>
<tr>
<td>−3</td>
<td>very easy</td>
</tr>
</tbody>
</table>

Participants to the rating were 19 students registered to the course with at least a half year (15 classes) of TET experience. Some students had experience of more than two years since the course can be registered multiple times. All students were informed that this rating is not related to the course grade in any ways.

Results

Figure 4 shows mean subjective ratings for 24 sound materials. Ordinate shows the means of subjective difficulty ratings. Abscissa of the graph shows the square root of residual of the linear regression on the spectra of a sound material in dB (called RMSres hereafter). RMSres is calculated as root mean square of the differences of power at each frequency band and the linear regression line shown in Figure 3, hence called “RMS of residuals.” Higher value of RMSres means higher variance in spectral power over frequencies, namely more bumps (peaks and notches) in the spectra. Symbols in the figure show 24 sound materials of different musical genre: circles (○) denote classical, diamonds (◇) denote rock/pops, squares (◻) denote jazz, triangles (△) denote solo instruments, and stars (☆) denote other types of sounds (pink-noise and applause). Classical music has higher RMSres values compared to rock/pops and jazz. Three of the four solo instrumental music have the highest RMSres values and also are rated higher in the subjective difficulty ratings.

From the result, a model of subjective difficulty ratings by independent variables including RMSres was created. Four independent variables (spectral centroid, spectral standard deviation about spectral centroid, slope of regression line on the spectra, and RMSres) and one dependent
variable (subjective difficulty ratings from all participants) were entered to a step-wise regression analysis \((p = .05 \text{ enter criterion and } p = .10 \text{ remove criterion})\). Following model having only \(\text{RMS}_{\text{res}}\) as an independent variable was obtained:

\[
\text{difficulty}_{\text{subjective}} = 0.148 \times \text{RMS}_{\text{res}} - 0.732
\]

The model's coefficient of determination is \(R^2 = .629\). ANOVA table of the regression is shown in Table 2.

**TABLE 2:** ANOVA table from step-wise regression for subjective difficulty ratings.

<table>
<thead>
<tr>
<th>Model</th>
<th>SS</th>
<th>df</th>
<th>MS</th>
<th>(F)</th>
<th>(p)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Regression</td>
<td>7.911</td>
<td>1</td>
<td>7.911</td>
<td>37.369</td>
<td>.000</td>
</tr>
<tr>
<td>Residual</td>
<td>4.657</td>
<td>22</td>
<td>.212</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>12.568</td>
<td>23</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**ESTIMATION OF OBJECTIVE DIFFICULTY**

**Method**

Subjective difficulty discussed in the earlier section is how difficult students feel. There are cases that a student feels a sound material is easy for the peak identification task but the score using the sound in the training is not as high as the student has expected. Hence, contrarily to the subjective difficulty, **objective difficulty** must be investigated. For the purpose of this study, objective difficulty is defined as the “number of correct answers when the sound material is used in the peak identification task.” For calculating the objective difficulty, 13 training results done in 12 weeks (April 10th to June 3rd, 2012) were used. However, because there were overlaps in sound materials used in the 13 training sessions, only eight sounds from 24 were used. Maximum number of correct answers is 10 since the sessions were done in a block of 10 trials.
Results

Pearson correlation coefficient between numbers of correct answers and subjective difficulty ratings was calculated. It was $r = -0.250$ ($p = 0.409$) showing no statistically significant correlation between them, suggesting that a sound material that seems easy to identify in the task is not necessarily the sound that can score high in the actual training.

Since it was hypothesized that students’ ability improve as the training progresses, partial correlation coefficient of number of correct answers and $\text{RMS}_{\text{res}}$ controlled for the number of weeks in the training was calculated:

$$\rho_{\text{#correct},\text{RMS}_{\text{res}}\cdot\text{#weeks}} = -0.518 \quad (p = 0.084).$$

This shows a moderate negative correlation between the number of correct answers and $\text{RMS}_{\text{res}}$, suggesting the possible inclusion of $\text{RMS}_{\text{res}}$ in the objective difficulty measure which is the goal of the study. However, when a step-wise regression analysis was done with spectral centroid, spectral standard deviation, slope of spectral linear regression, and $\text{RMS}_{\text{res}}$ as independent variables and mean numbers of correct answers as a dependent variable, the regression did not result in a statistically significant model.

CONCLUSION

Although gradually increasing difficulty levels is ideal for effective technical ear training, no objective measures for such training is available. Therefore, the tasks and sound materials used in the training are decided by the teacher’s own ears and experiences, leading to inefficiency when students want to train themselves in the teacher’s absence.

Towards the goal of developing objective difficulty measure in technical ear training, correspondence between students’ subjective difficulty ratings and physical indices calculated from the sound materials (such as spectral centroid and spectral standard deviation) were considered. As the result, a model which predicts the subjective task difficulty from residual of the linear fit through the spectra of the sound material ($\text{RMS}_{\text{res}}$) was created. Furthermore, $\text{RMS}_{\text{res}}$ may be a strong candidate for the variables in predicting objective difficulty measure in the peak identification task of technical ear training, since a somewhat meaningful partial correlation coefficient between the number of correct answers and $\text{RMS}_{\text{res}}$ was seen.

No strong correlation between subjective difficulty ratings and the number of correct answers, however, were seen. This means that a sound material that seems easy to be in the peak identification task is not necessarily the sound that can score high in the actual training session.

In order to provide the ability of critical listening with efficiency to students it is necessary to develop tools for measuring the effectiveness of the sound materials and the task procedures as well as a way to measure the students’ improvement. The authors plan to investigate on these aspects of technical ear training in the near future as well.

REFERENCES


