Acoust. Sci. & Tech. **34**, 6 (2013)

PAPER

Inter-subject differences in personalized technical ear training and the influence of an individually optimized training sequence

Sungyoung Kim[1,*], Teruaki Kaniwa[2,†], Hiroko Terasawa[3,4,‡], Takeshi Yamada[5,§] and Shoji Makino[2,¶]

[1]*Electrical, Computer and Telecommunication Engineering Technology, Rochester Institute of Technology, Rochester, NY, USA*
[2]*TARA Center, University of Tsukuba, Tsukuba, Japan*
[3]*Faculty of Library, Information and Media Science, University of Tsukuba, Tsukuba, Japan*
[4]*JST, PRESTO, Tokyo, Japan*
[5]*Faculty of Engineering, Information and Systems, University of Tsukuba, Tsukuba, Japan*

(*Received 13 January 2013, Accepted for publication 20 August 2013*)

Abstract: Technical ear training aims to improve the listening of sound engineers so they can skillfully modify and edit the structure of sound. Despite recent increasing interest in listening ability and subjective evaluation in the field of audio- and acoustic-related fields and the subsequent appearance of various technical ear-training methods, the subject of how to provide efficient training for a self-trainee has not yet been studied. This paper investigated trainees' performances and showed that an (inherent or learned) ability to correctly describe spectral differences using the terms of a parametric equalizer (center frequency, Q, and gain) was different for each person. To cope with such individual differences in spectral identification, the authors proposed a novel method that adaptively controls the training task based on a trainee's prior performances. In detail, the method estimates the weakness of the trainee, and generates a training routine that focuses on that weakness. Subsequently, we tried to determine whether the proposed method—adaptive feedback—helps self-learners improve their performance in technical listening that involves identifying spectral differences. The results showed that the proposed method could assist trainees in improving their ability to identify differences more effectively than the counterpart group. Together with other features required for effective self-training, this adaptive feedback would assist a trainee in acquisition of timbre-identification ability.

Keywords: Technical listening, Personalized training, Adaptive feedback, Spectral identification, Ear training

PACS number: 43.66.+y [doi:10.1250/ast.34.424]

1. INTRODUCTION

Technical listening refers to a cognitive process and its corresponding ability through which a listener can systematically discriminate and identify sonic differences. In the audio and music business, technical listening ability is considered essential for sound processing "in order to create the desired quality of sound," as asserted by Miśkiewicz [1]. It has been common that an assistant engineer learns this ability by observing how an experienced senior engineer fulfills requests from clients, and provides them with improved sound quality. While this kind of on-the-job training has been greatly respected and effective, it is not efficient, as it requires "a great deal of time [2]." A systematic training curriculum for technical listening could therefore reduce the time that is required for entry-level engineers to reach an expert level.

Since Retowski's initial work, *Timbre Solfeggio*, many institutions have developed systematic training programs of technical listening, often referred to as technical ear training, for their junior employees or students [1–8]. Whereas the specific purpose of the training program varies according to educational goals, the fundamental training method is to have trainees compare a reference signal with its sonically modified version, comprehend the difference,

*e-mail: sxkiee@rit.edu
†e-mail: s0913120@coins.tsukuba.ac.jp
‡e-mail: terasawa@slis.tsukuba.ac.jp
§e-mail: takeshi@cs.tsukuba.ac.jp
¶e-mail: maki@tara.tsukuba.ac.jp

and then repeat the process until they can reliably identify the sonic difference without a reference. This is similar to a general learning process. For example, Iwamiya *et al.* [2] compared this technical ear-training to "the simulation training which astronauts undergo." Just as astronauts can handle unexpected situations in space through numerous simulated missions and operational tasks, a technical ear-training program aims for a trainee to acquire fundamental principles of critical listening to "face the real job contexts" [2] at any time.

Traditionally, technical ear training was conducted in private or group sessions and required a dedicated instructor. In private training (such as [4]), an instructor provides the trainee with demonstrative work and the necessary feedback for efficient learning. While this private training is an ideal method, it is usually reserved for small numbers of trainees. Therefore, many institutions offer their students group training. The benefit of group training is that trainees can share the listening experience and develop common descriptors corresponding to physical variations of the auditory signals. Both students [2] and company employees [9] who have received group training have reported that the commonly developed descriptors helped them to communicate about acoustics- and audio-related topics in an effective and consistent way, thus reducing the chances of misunderstanding others. Another merit of group training is that the instructor can motivate trainees by stimulating soft competition among them. Marui [10] found that students, especially those who have just begun ear training, have a tendency to become more active when they notice the progress of their colleagues.

The benefits from the private and group training are not easily shared by individuals who are interested in the technical ear training yet have no instructor who can guide them. Individuals in this situation have to train themselves, using teaching materials available in various forms such as audio CDs [3], software [11,12], software with tutorial book [7], and an iPhone application [8]. This training can be categorized as self-training.

However, a self-trainee often faces difficulties in continuing the training. One problem is that self-training materials often do not evaluate the performance and inform the trainees about their progress. This could disincentivize the trainees, causing passive and unmotivated training. Another problem is that self-training material does not include appropriate feedback to guide the trainee to achieve effective learning. For example, if a trainee tends to repeat easy training that he or she seems to learn well enough, it would be necessary for the training material to judge whether the trainee should move on to new and more challenging material. Also, if one is struggling to learn a specific ability, a learning program should guide the trainee to repeatedly practice the related task to acquire the ability.

A potential solution to assist self-trainees is to develop a new training system that adaptively interacts with a trainee, which is common in e-learning and game-based learning. Garris *et al.* [13] reported that game-based learning can be effective because of "the intensity of involvement and engagement that computer games can evoke."

The new training system records a trainee's training history, analyzes his or her performance, and adaptively adjusts the presentation of training materials. In other words, the trainee can avoid redundant training and obtain individualized guidance for fast learning, with improvement of poor areas and effective self-study. We proposed that the new system can function as an instructor who acts upon "the available knowledge on its users and the subject matter at hand, to dynamically facilitate the learning process" [14].

With these goals, we conducted two experimental studies with two corresponding research questions: (1) Is there a listener-dependent idiosyncrasy in identifying spectral differences? and (2) Do adaptively regenerated training contents (for a specific trainee) assist in acquiring timbre identification ability more effectively than non-adaptive ones?

For the two experiments, we developed a personal ear-training program. The remainder of this paper includes a technical description of the program, experimental methods, results, and discussion.

2. PERSONAL TIMBRAL EAR TRAINER (PTET)

2.1. Operational Features of the Training Program

This section describes the operational features of a technical ear-training program, the Personal Timbral Ear Trainer (PTET), which was developed for self-trainees. The program was implemented using a graphics-based software development language, Max/MSP. As its name implies, PTET mainly focuses on technical-listening training associated with discrimination and identification of timbral difference. This program is based on the Timbral Ear Training developed at McGill University [4], but it is modified for self-training.

There are many tools [2,3,7,8] that provide a trainee with diverse and detailed categories of ear trainings that include detection and identification of timbral, dynamic, spatial, and temporal modification of a sound source. Such diversity is important in raising one's overall listening ability. We decided to focus mainly on timbral training because timbre-related attributes influence overall perceived sound quality of produced or reproduced sound field and because appropriate timbre balance is regarded as the most important factor in that perception [15,16].

According to Corey's book [7], there are two categories in timbral ear training: matching/removing and identification. In the matching/removing category, a trainee shapes the spectrum of the stimulus using parameters of a parametric equalizer [17] (center frequency, gain, and Q) until he or she makes the spectrum of stimulus equal to the reference. This kind training is similar to the practical work of mixing or sound-reinforcement engineers, who try to remove spectral peaks or dips of the given sound field and match it to their "internal" references. Without this internal reference, it would be hard to make the appropriate adjustment. This leads us to the second training category: identification. The goal of the identification training is to increase a trainee's ability to identify a small timbral difference, link it to the technical parameters, and eventually build a long-term memory using the technical parameter so that it can be used as an internal reference. While it is possible to configure PTET to satisfy both categories and it is recommended that an inexperienced trainee train in both tasks, we set PTET to provide only identification training. The decision was made because the identification training is for the acquisition of a long-term memory and is easier for inexperienced trainees to achieve.

PTET asks trainees to identify a given spectrum modified at one of seven center frequencies (125 Hz, 250 Hz, 500 Hz, 1 kHz, 2 kHz, 4 kHz, and 8 kHz) with one octave boost (peak) or cut (dip). For each question, a trainee can compare the non-processed sound at any moment during reproduction and report the identified center frequency. The program keeps reproducing the sound in a loop until the trainee provides the answer, and after the trainee confirms the frequency, the program displays the correct answer. Each training session is configured to have 25 identifications. This number was carefully chosen after heuristic trials conducted by the authors so that a single session can be effectively completed while maintaining a trainee's concentration. PTET lets a trainee conduct training with four sound sources: pink noise, orchestra, solo piano, and drums. We selected a portion of each sound file that contained all of the target bands required for the training, and the durations of the sound files were 1 min (pink noise), 1 min 37 s (orchestra), 1 min 39 s (solo piano), and 54 s (drums), respectively. The program also allowed trainees to select their own sound files (as a standard wave-file format) to be used for training purposes.

2.2. Adaptive Features of PTET

As explained in the Introduction section, the research objective of this paper was to experimentally investigate whether an adaptive training method would assist in effective learning by self-trainees. For this purpose, we manipulated PTET to operate in two modes: conventional and adaptive, as illustrated in Fig. 1. In a conventional mode, question-appearance probability is uniformly distributed, while in an adaptive mode a new probability is calculated according to an individual performance history. In other words, the adaptive mode generates a new probability of question appearance by calculating the weighting function of each frequency band for a specific trainee. It divides the relative weight of a specific frequency band by the entire weight as follows:

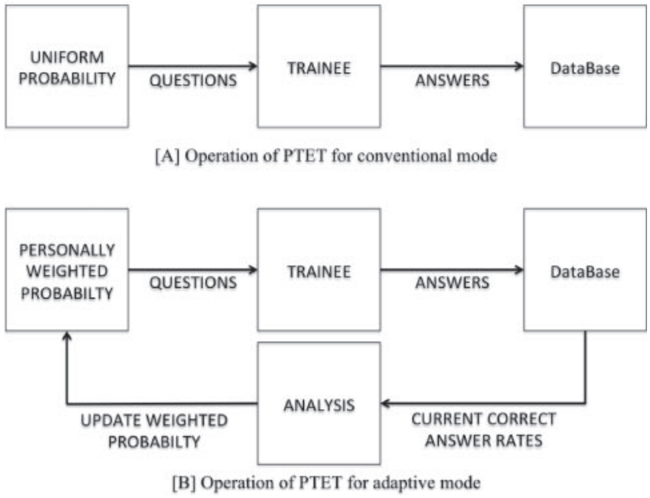

Fig. 1 The operational modes and associated block diagrams of PTET. The upper figure indicates that PTET uniformly generated questions regardless of personal performance, while the lower one reflects previous personal training history and the adaptively adjusted probability of the question appearance.

$$P_n = \frac{W_n}{W}, \tag{1}$$

where n is the index number of the seven frequency bands, P_n is the probability of the question appearance for the frequency band n, and W_n is the relative weight of the frequency band n.

W is the entire weight, denoted as

$$W = \sum_{n=1}^{k} W_n, \tag{2}$$

where k is the maximum number of bands (currently 7).

W_n is calculated as follows:

$$W_n = 100 - \frac{100 - L}{100} R_n, \tag{3}$$

where L is the minimum weight and R_n is the correct answer rate for band n. W_n decreases monotonously as R_n increases. To confirm whether a trainee has achieved a high correct rate by chance, the program needs to test even for a high-score band by assigning a non-zero value to L. Currently, L is set at 25 after several heuristic trials, resulting in a W_n value range of 25–100.

It should be noted that W_n in a conventional mode is always 100 for any value of n. Nevertheless, a single session (25 questions) would result in a different number of question appearances for frequency bands. We also confirmed that a conventional mode with multiple training sessions provided a uniformly distributed probability for all frequency bands.

3. EXPERIMENT I: INTER-SUBJECT DIFFERENCES IN IDENTIFYING SPECTRAL MODIFICATION

3.1. Methods

With the initial version of PTET, we evaluated the functional features and possible improvements of the program in order to make it better serve self-trainees. Four trainees participated in the first experiment. They were experienced listeners who had worked as audio signal-processing engineers and/or researchers for more than 10 years. Yet, the subjects were untrained listeners for this experiment in that they had no previous experience with technical ear training. The trainees regularly gathered once a week and practiced the week's task as a group under the supervision of the first author, who served as an instructor and explained the operational features of PTET and the mission of the week. The author did not provide feedback on any trainee's performance but rather observed whether there were specific tendencies in the trainees to identify spectral differences. All of the trainees had to complete each week's assignment before the following week's group practice. The training continued for eight weeks, and the training contents are listed in Table 1.

This curriculum was set to provide eight sessions. It started with a demonstrative session that was modified with a large spectral peak (+12 dB), then it generated questions of a smaller peak (+6 dB) and provided spectral dip (−12 dB) later. At the end of the training, the trainees were asked to identify random gain variations (either +12 dB, +6 dB, or −12 dB). During this training, the trainees were encouraged to evaluate the technical and operational features of the program and to share their evaluations with other group members. During this training, the trainees were encouraged to evaluate the technical and operational features of the program and to share their evaluations with other group members.

The participants conducted an hour-long group training session every week with an instructor (the first author) using a loudspeaker (Yamaha MSP-5), and they were required to conduct individual practice and assignments using a set of headphones (Beyerdynamic DT770).

After the eight weeks of training (group and individual), the trainees had to take a test that determined how consistently they remembered the spectral differences in terms of center frequency and gain. As stated in the Introduction section, the purpose of this test was to experimentally observe whether there is a listener-dependent idiosyncrasy in identifying spectral differences. The test was conducted in a recording studio using the same loudspeaker that was used for group training (Yamaha MSP-5). Since all trainees had to participate in the test simultaneously, they could not hear acoustically identical sounds. Yet acoustical differences associated with listener position were small enough that they did not generate characteristic idiosyncrasies that affected the testing results. We used four sound sources—pink noise, orchestra, drum, and piano—for the final test. For each sound source, we provided trainees with a set of 25 questions. Within a set, three gains (+9 dB, +3 dB, and −12 dB, representing a large peak, small peak, and dip) were randomly selected with equal probability. The +12-dB peak was excluded because it was too easy for the trainees, and the +3-dB peak was added as an anchor question to show whether the trainees could identify a smaller gain change.

Table 1 Curriculum of eight-week training program. The stimuli were pink noise and three musical sound sources (orchestra, drum, and piano).

Week	Training Content
1	12 dB peak (pink noise)
2	12 dB peak (musical sources)
3	6 dB peak (pink noise)
4	6 dB peak (musical sources)
5	12 dB dip (pink noise)
6	12 dB dip (musical sources)
7	12 dB or 6 dB peak, or 12 dB dip (pink noise)
8	12 dB or 6 dB peak, or 12 dB dip (musical sources)

3.2. Results

Figure 2 shows the results of the final tests. The ordinate represents the trainees' correct answer rates averaged over the four sound sources, while the abscissa represents the center frequencies of the seven bands. The test results highlight that the post-training ability to identify the given spectral modification was not the same for all trainees. This ability is a combination of inherent potential and learning through training. Even after the eight-week program, one trainee found it difficult to identify spectral differences in low-frequency bands, while another had similar confusion with high frequencies. In addition, the trainees commented that they felt that identifying the band that they already knew seemed ineffective and redundant, and they wanted to get training on the bands that were difficult for them.

The results indicate that the training would be more effective if it could provide trainees with more exercises in the areas where their identification performance is incon-

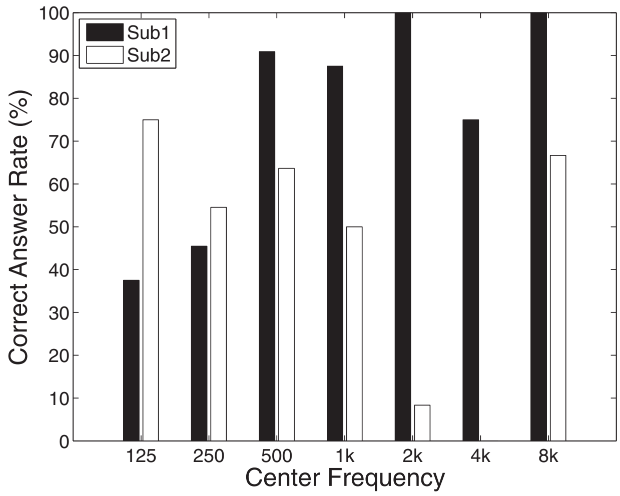

Fig. 2 Correct answer rates of the final test after the eight-week training program for two trainees, denoted as Sub1 and Sub2. The rates were averaged over the four sound sources. While the results of Sub1 showed lower correct rates in 125 Hz and 250 Hz (low frequencies), Sub2 showed lower rates in 2 kHz and 4 kHz (mid-high frequencies).

sistent. For example, trainee Sub2 would benefit from more training on the frequency band at 4 kHz to build a consistent internal reference to that spectral change, whereas for trainee Sub1, training with more weights on a low frequency would be helpful.

Previously, Quesnel [18] observed from his study that "all students' scores improved and the amount of improvement varied much between student." We also confirmed the individual dependency and further showed that such dependency could be associated with individual identification ability for each specific spectral band.

4. EXPERIMENT II: INFLUENCE OF ADAPTIVE TRAINING

4.1. Methods

After confirming that individual performance differed among trainees, we conducted an experiment that investigated whether such individual differences could be supported by an adaptive training sequence based on an analysis of each trainee's previous performance [19]. This study involved two trainee groups of four subjects that had similar descriptive statistics (average correct answer rates (64% and 63%) and variances (4.89 and 5.19)). Subsequently, we conducted a two-tailed t-test, which confirmed that there were no statistically significant differences between the two groups ($p = 0.63$, $df = 6$).

Two groups were labeled the Conventional Group, which was given training with a conventional method, and the Proposal Group, which was trained with the adaptive training method. The operational difference in the two groups was previously illustrated in Fig. 1, and the program generates a list of new probabilities of question appearances for a subject in the Proposal Group. For example, Table 2 shows a list of new probabilities for a subject (Sub2 in Experiment I), which was adaptively modified using the correct answer rates in Fig. 2. The process was also dynamic, so that if the correct answer rate changed in the middle of a training session the program automatically adjusted the question appearance accordingly.

This experiment was controlled so that subjects were required to attend weekly training sessions. The experiment took place in a recording studio at the University of Tsukuba using three types of sound sources—pink noise, orchestra, and piano—during a period of four weeks. We chose these three representative sound sources based on the results of Experiment I. While it would have been ideal to be able to conduct the same curriculum as Experiment I for the sake of comparison, the Experiment II was performed with students who were only able to participate for a limited time. Thus, the curriculum was reduced by half for the four weeks and, each week's training combined pink noise and two musical sources.

Subjects were asked to conducted two 30-minutes practice sessions and to complete the main assignment of the week. For this individual practice and training, subjects used a pair of headphones (Sennheiser HD 650).

4.2. Results

Figure 3 shows the results of Experiment II. The correct answer rates for the three stimuli (averaged over all subjects' responses to seven frequency bands and four-weeks of training) of the Proposal Group were higher than those of the Conventional Group. This trend was the same for the subsequent analysis of individual frequency bands. However, as Table 3 shows, the results of a one-sided t-test (significance level $\alpha = 0.05$, where the Bonferroni-adjusted significance level is 0.016) indicate that the Proposal

Table 2 Adjusted probability of question appearances based on the correct answer rates of subject Sub2 in Fig. 2 for seven octave bands. The initial probability was 14.3% (100/7) for all seven bands. Based on this adjusted-probability table, the program generated more questions for frequency bands where Sub2 performed poorly (for example, in the areas of 4 kHz).

Center Frequency	125 Hz	250 Hz	500 Hz	1 kHz	2 kHz	4 kHz	8 kHz
Correct Answer Rate	75%	54.5%	63.6%	50%	8.3%	0%	66.7%
Adjusted Probability	9.5%	12.8%	11.3%	13.5%	20.3%	21.7%	10.8%

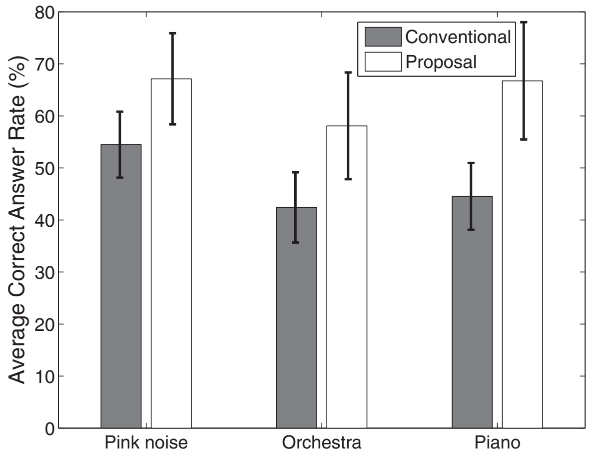

Fig. 3 The average correct answer rates and corresponding 95% confidence intervals of the two groups (averaged over seven center frequencies and four-weeks of trainings). There is a trend that the Proposal Group has consistently higher rates for all three sounds than the Conventional Group.

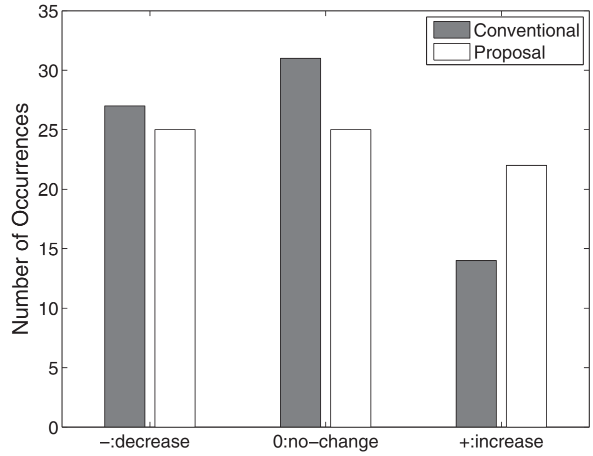

Fig. 4 The change of the correct answer rate of the three lowest-score bands. "+" indicates an increase in the correct answer rate; "−" represents a decreased correct answer rate; and "0" indicates no change. The Proposal Group tended to have a less decreased correct answer rate and a greater increased correct answer rate compared with the Conventional Group.

Table 3 The *t*-test results of evaluating the difference between the Proposal Group and the Conventional Group with the significance level $\alpha = 0.05$ (with the Bonferroni-adjusted significance level 0.016). Listeners' correct-answer rates are significantly different with the piano stimulus.

Sound	Pink Noise	Orchestra	Piano
p	0.11	0.018	0.00038*

Group and the Conventional Group became significantly different with the piano sound but not with the pink noise and the orchestra.

In addition, we investigated the improvement on the low-score band in order to verify whether use of the proposed method increased the correct answer rate in low-score bands as intended. For every week, we detected the three frequency bands where each subject scored lowest, counted the direction of change (i.e., increase, decrease, and no change) in the correct answer rate of those low-score bands in the following week, and compared the counts of the Proposal group and Conventional group. Figure 4 shows the occurrence number of the score change directions. The indices of change were "+" = increase, "−" = decrease, and "0" = no change, respectively.

These results contrast the change of decrease and increase of the two groups. The Conventional Group shows more counts on the decrease ("−") and no change ("0") cases than the increase case ("+"), whereas the Proposal Group shows more evenly distributed counts for the three cases. This observation implies the possibility that the Proposal Group did not decrease the score as much as the Conventional Group did, even though the training task became more difficult every week and thus naturally tended to generate fewer correct answers.

However, the results of the sign-test in the low-score bands provided no statistically significant evidence on this possibility—the null hypothesis for this sign test was "there was no score decrease," whereas the alternative hypothesis was "there was score decrease." The results of the sign test (small samples, single-sided, significance level $\alpha = 0.05$) were $p = 0.06$ for the Conventional Group, and $p = 0.77$ for the Proposal Group, with both not rejecting the null hypothesis, and therefore not showing a statistically significant contrast between the two groups. Although it is expected that the correct answer rate in the low-score bands would be improved by prolonging training time, further investigation on this effect remains for our future work.

5. DISCUSSION

Ear training in the audio and acoustic fields is still very young and growing, and thus the methodology has not been standardized among various programs and institutions. For example, the training program used in this paper presents a reference sound and a modified sound in parallel and allows a trainee to toggle between the two at any point. In contrast, other programs (including [2,3]) present two sequential sound sources. This sequential presentation of stimuli forces trainees to hold their impression of the first sound source and use it as a reference for the second sound source. While the latter requires more cognitive activity and thus seems to be more efficient in acquiring listening ability, the former is much more practical (for audio engineers in particular) as they usually compare a modified sound source with a non-modified (often referred to as a "bypass") version when applying a filter *in-situ*. Another example is the number of questions in a single training session. In our program, we have 25, which we arrived at during program development after considering training performance and listener concentration. This specific number will probably be different for other programs. Likewise, most other parameters have been optimized through heuristic trial-and-error, and therefore they are program dependent, which makes it difficult to establish standardized training for multiple institutions.

While this program was originally developed to replace a human instructor, it is still in its early stages, and a substantial amount of work will be required to transform the program into a smart and instructor-free ear trainer. Also, in order to fully validate the effectiveness of the proposed adaptive adjustment in the probability of question appearances, it will be necessary to conduct a validating experiment that compares three large groups of trainees: a self-training group without adaptive adjustment, a group with the proposed adaptive adjustment, and a group with a human instructor who would analyze and provide individual guidance for the trainees.

6. CONCLUSION

To support self-learning for identifying spectral modification, we developed a personal timbre-training program. The program is designed to provide trainees with multiple tasks of increasing difficulty along with adaptive feedback that produces more training with regard to each trainee's weak points. The results showed that listeners tend to obtain spectral identification for specific frequency bands first, which made it necessary to design a new training method that adaptively adjusts the test content for specific listeners considering their previous performance. The consequent study results showed that the personalized and adaptive feedback could assist trainees in conducting more effective training by themselves as the difficulty of training increases.

REFERENCES

[1] A. Miśkiewicz, "Timbre solfege: A course in technical listening for sound engineers," *J. Audio Eng. Soc.*, **40**, 621–625 (1992).

[2] S. Iwamiya, Y. Nakajima, K. Ueda, K. Kawahara and M. Takada, "Technical listening training: Improvement of sound sensitivity for acoustic engineers and sound designers," *Acoust. Sci. & Tech.*, **24**, 27–31 (2003).

[3] D. Moulton, The golden ears audio eartraining program, Audio CD published by KIQ Productions (1992).

[4] R. Quesnel, "Timbral ear trainer: Adaptive, interactive training of listening skills for evaluation of timbre," *Proc. Audio Eng. Soc. 100th Int. Conv.*, Copenhagen, Denmark (1996).

[5] S. E. Olive, "A new listener training software application," *Proc. Audio Eng. Soc. 110th Int. Conv.*, Amsterdam, Netherlands (2001).

[6] A. Nishimura, "Auditory training system that uses TCP/IP networks and WWW browsers," *J. Acoust. Soc. Jpn.*, **62**, 208–213 (2006).

[7] J. Corey, *Audio Production and Critical Listening* (Focal Press, Oxford, UK, 2010).

[8] M. C. Erickson, "Auricular — Mobile ear training apps for iOS devices," http://www.auriculaonline.com/ As of 2013. 4. 1.

[9] K. Kawahara, T. Ito, T. Kobayashi, S. Iwamiya and M. Takada, "Case study of curriculum development for technical listening training for employees of an acoustic related company," *Proc. 20th Int. Congr. on Acoustics*, Sydney, Australia, August (2010).

[10] A. Marui, Presented at the workshop "Listen Professionally or Train Your Ears!" during the *Audio Eng. Soc. 131st Int. Conv.*, New York, USA (2011).

[11] HARMAN International, "How to listen," http://harmanhowtolisten.blogspot.com/ As of 2013. 4. 1.

[12] L. Herranz, "Train your ears," http://www.trainyourears.com/ As of 2013. 4. 1.

[13] R. Garris, R. Ahlers and J. E. Driskell, "Games, motivation, and learning: A research and practice model," *Simul. Gaming*, **33**, 441–467 (2002).

[14] A. Paramythis and S. Loidl-Reisinger, "Adaptive learning environments and eLearning standards," *Proc. 2nd Eur. Conf. on e-Learning* (2003).

[15] A. Gabrielsson and H. Sjögren, "Perceived sound quality of sound-reproducing systems," *J. Acoust. Soc. Am.*, **65**, 1019–1033 (1979).

[16] F. Rumsey, S. Zieliński and R. Kassier, "On the relative importance of spatial and timbral fidelities in judgments of degraded multichannel audio quality," *J. Acoust. Soc. Am.*, **118**, 968–976 (2005).

[17] G. Massenburg, "Parametric equalization," *Proc. Audio Eng. Soc. 42nd Int. Conv.*, Los Angeles, USA (1972).

[18] R. Quesnel, *A Computer-Assisted Method for Training and Researching Timbre Memory and Evaluation Skills*. Ph.D. thesis, McGill University (2001).

[19] T. Kaniwa, S. Kim, H. Terasawa, M. Ikeda, T. Yamada and S. Makino, "Towards a personalized technical ear training program: An investigation of the effect of adaptive feedback," *Proc. 8th Sound and Music Computing Conf.*, Padova, Italy (2011).

Sungyoung Kim received a B.S. degree from Sogang University, Korea, in 1996, and a Master of Music and Ph.D. degree from McGill University, Canada, in 2006 and 2009 respectively. His professional work experiences include recording/balance engineer at Korea Broadcasting System (KBS), Seoul, Korea (1995–2001) and research associate at Yamaha Corporation (2007–2012), Hamamatsu, Japan. He is now an assistant professor at Electrical, Computer, and Telecommunication Engineering Department, Rochester Institute of Technology. His research interests are spatial audio and human perception, and efficient ear training method.

Teruaki Kaniwa is a graduate student of University of Tsukuba, Japan and he has been studying computer science and audio information processing with Shoji Makino at the Multimedia Laboratory. He researches the sonification of EEG steady state response now.

Hiroko Terasawa received B.E. and M.E. degrees in Electrical Engineering from the University of Electro-Communications, Japan,

and M.A. and Ph.D. degrees in Music from Center for Computer Research in Music and Acoustics (CCRMA), Stanford University, USA. Her research interests include timbre perception modeling and timbre-based data sonification. She currently serves as an assistant professor at University of Tsukuba.

Takeshi Yamada was born in Osaka, Japan. He received the B. Eng. degree from Osaka City University, Japan, in 1994, and the M. Eng. and Dr. Eng. degrees from Nara Institute of Science and Technology, Japan, in 1996 and 1999, respectively. He is presently an associate professor with Faculty of Engineering, Information and Systems, University of Tsukuba, Japan. His research interests include speech recognition, sound scene understanding, multi-channel signal processing, media quality assessment, and e-learning. He is a member of the IEEE, the IEICE, the IPSJ, and the ASJ.

Shoji Makino received B.E., M.E., and Ph.D. degrees from Tohoku University, Japan, in 1979, 1981, and 1993, respectively. He joined NTT in 1981. He is now a Professor at University of Tsukuba. His research interests include adaptive filtering technologies, the realization of acoustic echo cancellation, blind source separation of convolutive mixtures of speech, and acoustic signal processing for speech and audio applications.