TOWARDS A PERSONALIZED TECHNICAL EAR TRAINING PROGRAM: AN INVESTIGATION OF THE EFFECT OF ADAPTIVE FEEDBACK

Teruaki Kaniwa[1], Sungyoung Kim[2], Hiroko Terasawa[1], Masahiro Ikeda[2], Takeshi Yamada[1], Shoji Makino[1]
[1]University of Tsukuba, [2]Yamaha Corporation
kaniwa@mmlab.cs.tsukuba.ac.jp, sungyoung@beat.yamaha.co.jp, terasawa@tara.tsukuba.ac.jp, masahiro_ikeda@gmx.yamaha.com, takeshi@cs.tsukuba.ac.jp, maki@tara.tsukuba.ac.jp

ABSTRACT

Technical ear training aims to improve the listening of sound engineers so that they can skillfully modify and edit the structure of sound. To provide non-professionals such as amateur sound engineers and students with this technical ear training, we have developed a simple yet personalized ear training program. The most distinct feature of this system is that it adaptively controls the training task based on the trainee's previous performance. In detail, this system estimates a trainee's weakness, and generates a training routine that provides drills focusing on the weakness, so that the trainee can effectively receive technical ear training without an instructor. We subsequently investigated the effect of the new training program with a one-month training experiment involving eight subjects. The result showed that the score of the group assigned to the proposed training system improved more than that of the group assigned to conventional training.

1. INTRODUCTION

Technical ear training aims to improve the ability to systematically discriminate and identify sonic differences. In the audio and music production/reproduction business, this ability has been regarded as essential for sound processing "in order to create the desired quality of sound" as Miskiewicz suggests [1]. Since Retowski's initial work [2], many institutions have developed systematic training programs for their junior employees or students [1-8].

Although the specific purposes of training programs vary according to the educational goal, the fundamental training method is to have a trainee compare a reference signal with its sonically modified version, comprehend the difference, and then repeat such comparisons until the trainee can reliably identify the sonic difference without a reference. In most cases, an instructor interactively guides the trainee; as in a music lesson, the trainee practices this method by observing the instructor's demonstration, accomplishing the given task, and receiving guidance or feedback to allow progress to the next training stage. In contrast, this method may be followed independently with existing static training materials such as the Ear-training audio CD [3].

Recent ear training programs [4-8] often utilize computer software that assists the instructor in the teaching task by, for example, administering the training schedule, and displaying the training scores to the trainee through the real-time visualization of electronically stored data. By incorporating these additional features from computer-assisted training programs, an instructor can effectively guide a trainee and offer appropriate advice to make the training faster than purely empirical acquisition.

The next step for the ear-training software that will take it beyond being a mere teaching aid is to guide the trainee in the same manner as a private lesson, so that a trainee can practice the training without an instructor. The related research field to this project is Intelligent Tutoring Systems(ITS), in which artificial intelligence systems provide customized instructions or discussions to students, i.e. without the intervention of human beings [9]. Topics of ITS includes to select problems at a level of difficulty appropriate to the student's overall performance, and such systems have been called "adaptive" and their sophistication lay in the task-selection algorithms. Moreover, many conferences of ITS have been held internationally, e.g. ITS2010 in Pittsburgh [10]. In recent e-learning studies, this adaptive and interactive feature is regarded as essential for maintaining a trainee's motivation even in a self-learning scenario [11].

In this study, we propose a new system for a computer-based ear training program that adaptively creates personal training routines based on an individual's training record. This system estimates the trainee's weakness, and generates a training routine, which provides drills focusing on the weakness, so that the trainee can effectively study technical listening without an instructor. We conducted a series of evaluation tests, and report the effect of this training system.

2. EAR TRAINING PROGRAM

As mentioned in Section 1, various types of ear training program have been proposed. The research group at McGill University in Canada proposed three types of training namely, “matching”, “removing”, and “absolute identification.”[5]. All three types adopt the training method introduced in the previous section and control the spectrum of a sound using the three parameters of a parametric equalizer [12] (center frequency, Q, and gain). Matching and removing training involves asking a trainee to modify the spectrum of a sound and make it equal to the given modification. On the other hand, the goal of an absolute identification task is to increase a trainee’s ability to identify a modified spectrum, describe it in terms of technical parameters (center frequency, Q, and gain), and eventually build a long-term memory of the internal reference on which a trainee will rely for future identification tasks.

Of these three types of training, we employed “absolute identification” for this study because it enables entry-level listeners to learn the identification quickly in a limited training period.

During our informal preliminary study, we found that the inherent ability to identify the modified center frequencies was not identical for all trainees. For example, a trainee may find it difficult to discriminate subtle differences between two adjacent low frequency bands such as 250 and 125 Hz, while another trainee shows similar confusion in the high frequency area. To the best of our understanding, the conventional training programs for self-study do not control such individual differences, which could make it more difficult for a trainee to overcome his/her weak points. Therefore, we hypothesized that a new system would assist a trainee more effectively if it could analyze a trainee’s previous training record and adaptively provide more training in the area in which he or she performed poorly. And we created a computer-based, personalized technical ear-training program, which manages individual training records with a database and provides personal training routines adaptively based on the record. Figure 1 shows a block chart of the new training system.

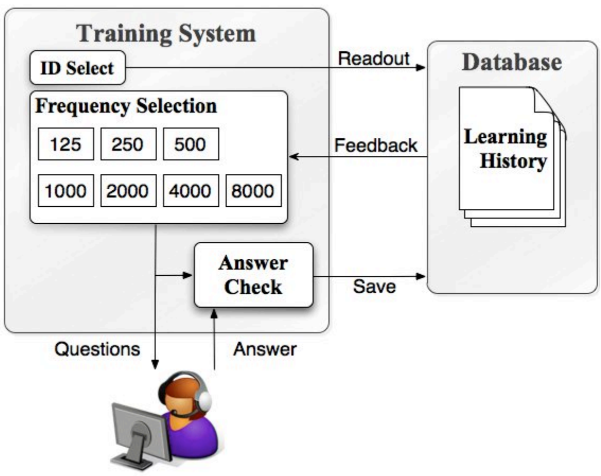

Figure 1. Block chart of new training system

3. GENERATION OF PERSONALIZED PROBLEM SET

As stated above, we proposed a new system that varies the composition of training questions depending on the individual study record. In our proposed system, the software reads the previous correct answer rate of the current trainee (i.e. how precisely the trainee identified the modified spectrum), and adjusts the probability of question appearance. The new probability of question appearance is calculated by dividing the relative weight of each frequency bandwidth by the entire weight as follows:

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

where n is the index number of the filter in this study as shown in Table 1, P_n is the probability of question appearance for bandwidth n, W_n is the relative weight of bandwidth n. W is the entire weight denoted as follows:

$$W = \sum_{n=1}^{7} W_n \tag{2}$$

W_N is calculated as follows:

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

where L is the minimum weight, and R_n is the correct answer rate for bandwidth 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 bandwidth by assigning a non-zero value to L. Currently L is set at 25 after several heuristic trials, resulting in a Wn value range of 25 to 100. With this weighting process, the system generates more questions to the low-score bandwidth, and fewer to the high-score bandwidth. In addition, this process is dynamic so that if the correct answer rate changes, the system automatically adjusts the question appearance accordingly. The consequent question was whether or not this “interactive and smart randomization” would provide self-trainees with more effective learning. To investigate this question, we conducted an experiment that compared the influence of the proposed system on learning the absolute identification of spectral modification.

N	Center frequency [Hz]
1	125
2	250
3	500
4	1000
5	2000
6	4000
7	8000

Table 1. Center frequency of the filter

4. EXPERIMENT TO EVALUATE PROPOSED SYSTEM

To evaluate the proposed system, we formed two groups of trainees: the Conventional group and the Proposal group. For the Conventional group , we set the software to generate questions using a non-weighted random function, while for the Proposal group, it generated questions using the weighted random function dynamically updated according to the trainee's previous training scores as described in Section 3.

In total, eight subjects participated in the experiment. Before the main experiment, we conducted a preliminary test of absolute identification (25 questions with +12 dB boosted pink noise) and divided the subjects into two groups so that the initial condition of each group was as close as possible. As shown in Table 2, the descriptive statistics of the preliminary test scores indicated that these two groups were similar to each other. Furthermore, we conducted a preliminary F-test to test the equal variance, followed by a two-tailed t-test (two independent samples with equal sample size and equal variance), which confirmed that there were no statistically significant differences between the two groups as shown in Tables 3 and 4.

Group	Conventional	Proposal
M [%]	64	63
SD	4.89	5.19

Table 2. Grouping of subjects

F (1, 6)	47.5
p	0.39

Table 3. F-test of initial test (Conventional group vs. Proposal group)

t (6)	2.45
p	0.63

Table 4. t-test of initial test (Conventional group vs. Proposal group)

The experiment took place in a recording studio at the University of Tsukuba using three types of sound files namely, "Pink Noise", "Orchestra (Symphony No. 3 by C. Saint-Saens)", and "Piano (Piano Sonata #2, Op. 36 by S. Rachmaninov)". We selected a portion of each sound file that contained the all target bands required for the training. Table 5 shows the equipment used for the experiment.

After an initial tutorial session on the operation of the ear training program, the subjects practiced with the training system by themselves twice a week for about 30 minutes each time, according to the training curriculum shown in Table 6. The training curriculum is designed so that the subjects find the task challenging and motivating with a level of difficulty that increases as the training proceeds. The total duration of the training was four weeks. In addition to actual training, the subjects could freely practice with the content shown in Table 7 on the first training day of each week to familiarize themselves with the kind of training task they would perform in that week.

Audio Interface	MOTU UltraLite-mk3
Headphone	SENNHEISER HD 650
Programming Language	Max MSP ver.5.1.5

Table 5. Equipment used for the experiment

Week	dB	Sound File
1	+12 dB, +6 dB	Pink Noise, Orchestra, Piano
2	+6 dB, +3dB	Pink Noise, Orchestra, Piano
3	−12 dB	Pink Noise, Orchestra
4	±12 dB, −12 dB	Pink Noise, Orchestra

Table 6. Training curriculum

Week	dB	Sound File
1	+12 dB	Pink Noise
2	+6 dB	Pink Noise
3	−12 dB	Pink Noise
4	±12 dB	Pink Noise

Table 7. Contents of practice

5. RESULTS

5.1 Analysis by Mean and Standard Deviation

We first analyzed the mean value and the standard deviation of the correct answer rate to investigate the effect of the proposed system. Figure 2 shows the overall average correct answer rate for each sound. From this, it was found that the average correct answer rate of the Proposal group was higher than that of the Conventional group. This result suggests that the proposed system raises the correct answer rate based on the fact that the initial average (before training) was about the same for the two groups.

Next, we analyzed the results for each sound. Figures 3, 4 and 5 show the mean value and the standard deviation of the correct answer rate for Pink Noise, Orchestra and Piano respectively. The analysis results show that the mean correct answer rate of the Proposal group is consistently higher than that of the Conventional group for all sounds. After verifying the equal variance with a preliminary F-test, we conducted a one-tailed t-test (independent two-sample, with equal sample size and equal variance) to evaluate the effectiveness of the Proposal group compared with that of the Conventional group. Table 8 shows the results of the t-test for each sound. We found a significant difference with Orchestra and Piano sounds but not with Pink Noise. This might be because the temporal variation of musical sounds (Orchestra and Piano) made it harder for a trainee to identify the spectral modification than Pink Noise. This result suggests that the proposed system is more efficient in terms of training for realistic and difficult tasks than the conventional training system.

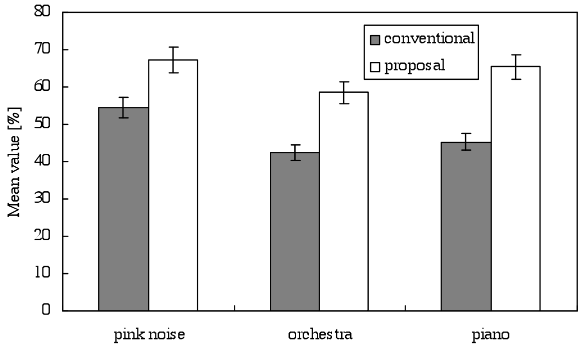

Figure 2. Average of each sound

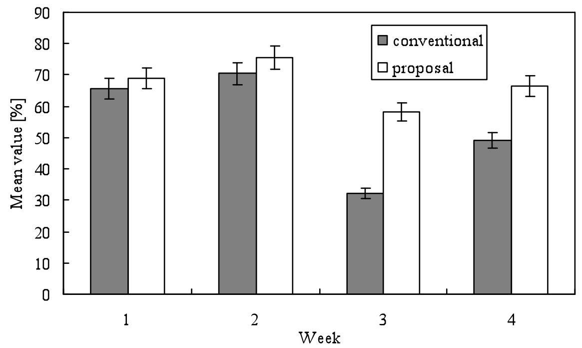

Figure 3. Result for Pink Noise

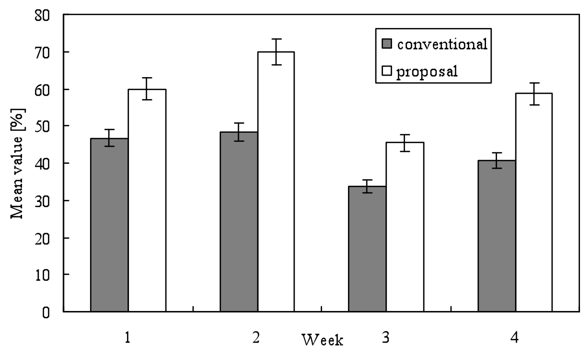

Figure 4. Result for Orchestra

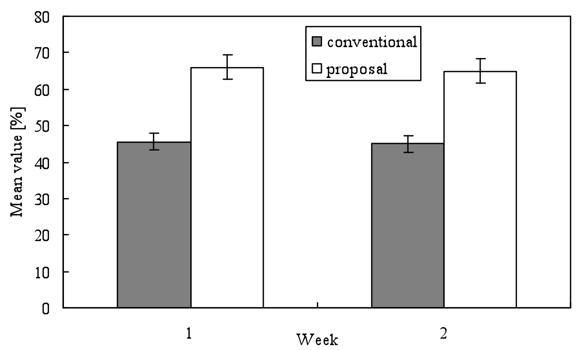

Figure 5. Result for Piano

Sound	Pink Noise	Orchestra	Piano
df	6	6	2
t (absolute)	1.36	2.67	36.2
p	0.11	0.018	0.00038

Table 8. t-test of training data

5.2 Analysis of Improvement for Low-Score Bands

In addition to the analysis in Section 5.1, we investigated the improvement for the low-score band, to confirm that use of the proposed system increased the correct answer rate in low-score bands as intended.

In this analysis, we counted the direction of change (i.e. positive, negative, and unchanged) in the correct answer rate in low-score bands, and compared the counts of the Proposal group and Conventional group. For each subject's data, we extracted three low-score bands that marked the lowest correct answer rates for every week, and analyzed the way in which the correct answer rates in those bands changed in the following week. The indices of change were "+ (improvement)", "－(decrease)", and "0 (no change)". Figure 6 shows the frequency distribution of the score change directions. The "－" element was smaller and the "+" element was larger for the Proposal group than for the Conventional group. To compare the frequencies of the score change direction quantitatively, we counted the sum of the score change by substituting each "－" element with "-1", each "+" element with "+1", and the no-change element with "0". The sum of the score change for the Proposal group was -3 whereas that for the Conventional group was -13, revealing greater improvements for the low-score band with the proposed system. (This sum naturally tends to be negative because the training was designed to become harder.)

Figure 7 shows the frequency distributions for the difference in the correct answer rates for the low-score bands. The Proposal group showed a higher frequency of positive score difference between 0% and 30% than the Conventional group. In contrast, the Proposal group showed a lower frequency of negative score difference between －20% and －10%. In other words, the Proposal group showed a strong tendency to improve their listening in low-score bands. This means that the Proposal group exhibited better progress in the low-score bands, which indicates that they conducted self-training more efficiently than the Conventional group.

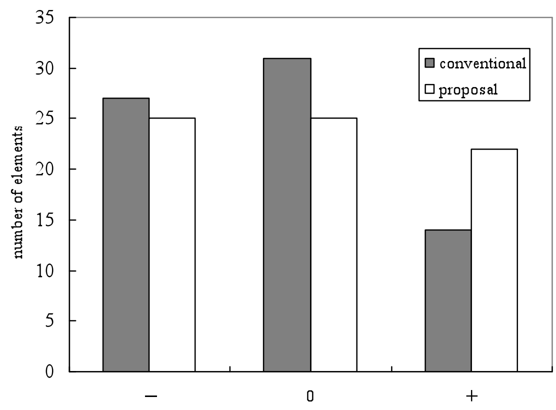

Figure 6. Result of analyzing with the index of the changed direction of the correct answer rate

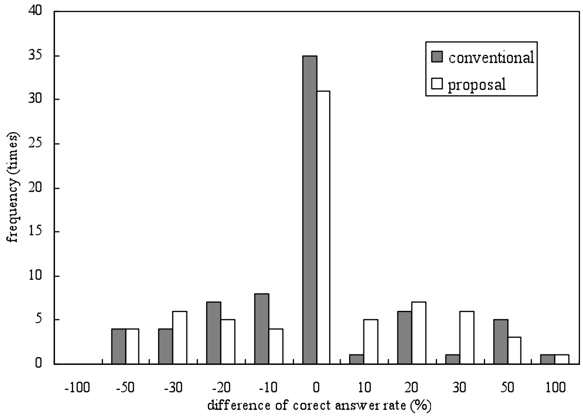

Figure 7. Frequency distribution

6. CONCLUSIONS

To support self-learners who are training their technical listening ability, we proposed a new training program that adapts to individual differences and provides more training with regard to a trainee's weak points. However, the proposed system exhibited different levels of effectiveness for each type of sounds, and that suggests interactions between the types of the test-sounds. In our future work, it would be desirable to reconsider the experimental design, so that analysis of variance can be conducted to test the effect of interactions. Even though this kind of software will not replace a human instructor, the results from the experimental test showed that our proposed system could assist the trainees to conduct more effective training by themselves, especially with realistic, therefore more difficult, identification tasks. To make this program more interactive and educative, we are investigating additional features that could offer a self-trainee fast and entertaining training, such as implementing the training program in a portable device.

7. REFERENCES

[1] A. Miskiewicz, "Timbre Solfege: A Course in Technical Listening for Sound Engineers," in *J. Audio Eng. Soc.*, 1922, 40(7/8): pp. 621 – 625.

[2] T. Letowski, "Development of technical listening skills: Timbre solfeggio," in *J. Audio Eng. Soc.*, 1985, 33(4): pp. 240 – 244.

[3] D. Moulton, *The Golden Ears Audio Eartraining Program*. Audio CD published by KIQ Productions.

[4] S. Iwamiya, Y. Nakajima, K. Ueda, K. Kawahara, and M. Takada, "Technical Listening Training: Improvement of sound sensitivity for acoustic engineers and sound designers," in *Acoustical Science and Technology,* January 2003, 24(1): pp.27 – 31.

[5] R. Quesnel, "Timbral Ear Trainer: Adaptive, interactive training of listening skills for evaluation of timbre," in *Proc. Audio Engineering Society 100th Int. Conv.*, Copenhagen, Denmark, 1996. AES. Preprint 4241.

[6] S. E. Olive, "A new listener training software application," in *Proc. Audio Engineering Society 110th Int. Conv.*, Amsterdam, Netherlands, May 2001, AES. Preprint 5384.

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

[8] N. Akira, "Auditory training system that uses TCP/IP networks and WWW browsers," in *Japanese Acoustical Society of Japan (J)*, 2006, 62(3): pp. 208 – 213.

[9] D. Sleeman and J. S. Brown, *Intelligent Tutoring Systems*. Academic Press, 1982.

[10] V. Aleven, J. Kay, and J. Mostow, *Intelligent Tutoring Systems 10th International Conference*, ITS 2010, Pittsburgh, PA, USA Springer-Verlag, 2010.

[11] R. Garris, R. Ahlers, and J. E. Driskell, "Games, motivation, and learning: A research and practice model," in *Simulation & Gaming*, 2002, 33(4): pp. 441 – 467.

[12] G. Massenburg, "PARAMETRIC EQUALIZATION," in *Proc. Audio Engineering Society 42nd Int. Conv.*, Los Angeles, USA, May 1972. AES.