Audio Engineering Society

Convention Paper 9945

Presented at the 144th Convention
2018 May 23 – 26, Milan, Italy

Does Spectral Flatness Affect the Difficulty of the Peak Frequency Identification Task in Technical Ear Training?

MARUI, Atsushi[1] and KAMEKAWA, Toru[1]

[1] *Faculty of Music, Tokyo University of the Arts*

Correspondence should be addressed to MARUI, Atsushi (marui@ms.geidai.ac.jp)

ABSTRACT

Technical ear training is a method to improve the ability to focus on a specific sound attribute and to communicate using the common language and units used in the industry. In designing the successful course in a sound engineers' educational institution, it is essential to have the gradual increase of the task difficulty. The authors had investigated the correlation between the students' subjective ratings on the task difficulty and the physical measures calculated from the sound materials used in the training. However, the objective measure of the difficulty is still not known. Authors created the training materials with different spectral envelope but having the same music content and tested them in the ear training sessions.

1 Introduction

Technical Ear Training is a method to improve the ability to focus on specific sound attribute. It is also used to acquire the ability to communicate using the common language and units used in the industry such as Hz and dB [1, 2, 3]. Technical ear training is widely accepted as one of the building blocks of a sound engineers' education curriculum. In designing the successful course in an educational institution, it is essential to have the gradual increase of the task difficulty. However, the objective measure of the difficulty is still not known, and thus the tasks are decided by the instructor's own ears and experiences, which leads to inefficiency when students want to train themselves in the instructor's absence.

In the past, the authors have investigated the correlation between the participants' subjective ratings on the task difficulty and the physical measures calculated from the sound materials used in the training [4]. For the identification task for frequency spectrum peaks, a regression model which predicts the subjective task difficulty from residual of the linear fit through the spectral envelope of the sound was created. For example, Fig. 1 shows power spectra of two different music excerpts. The subjective difficulty rated by the trainee was higher for the excerpt on the lower panel which has higher sum of residuals of the linear fit.

The regression model shows that increase of the residual (*i.e.*, the spectrum being less "flat") makes the subjective impression of peak identification task more difficult. In this paper, the term spectrally "flat" means having less residual from the linear fit to the spectral envelope of the original signal. Fig. 2 shows the scatter

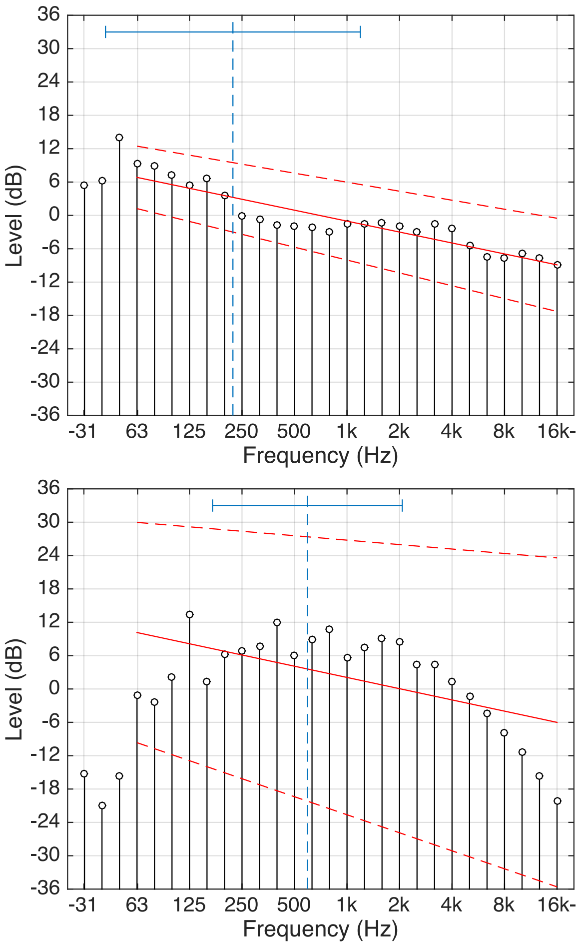

Fig. 1: Power spectra of two music excerpts. Circles show the power at each 1/3-octave band. Solid line is the linear fit through the spectrum between 63 Hz and 16 kHz (with dashed lines above and below denoting 1 standard deviation). Vertical dashed line shows spectral centroid with a horizontal whisker showing 1 standard deviation around the centroid. Upper panel has less residual from the linear fit than the one on the lower panel.

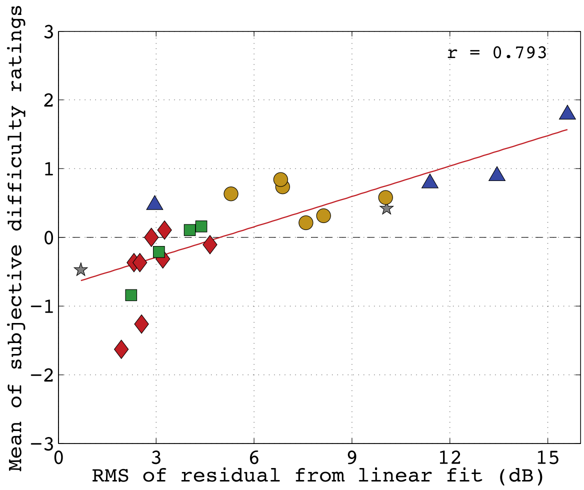

Fig. 2: Subjective difficulty plotted against the residual from the linear fit to the spectral envelope. Each symbol denote a music excerpt used in the training. Symbol shapes represent different musical styles. The figure was cited from [4].

plot of the residual and the subjective difficulty ratings for 24 music excerpts used in the training tasks. Each symbol denote a music excerpt used in the training. It can be seen that the subjective difficulty rating increases as the residual increases. Please refer to [4] for more detail.

While subjective impression of task difficulty is an important part of the training, it is not identical to the objective measure of difficulty. In order to see the relation between subjective and objective measures for the task difficulty with different spectral flatness, authors created a pair of training materials with different spectrum but having the same music content and tested them in the actual ear training sessions.

2 Methods

2.1 Stimuli

A 12 second music excerpt from Ozric Tentacles' "Knurl" (in the album *Paper Monkeys*, 2011) was processed to have two versions of spectral envelopes. The first is unmodified version and the second was equalized to have little residual from the linear fit through the spectral envelope of the original sound hence having "flat" frequency spectrum. Two versions were labeled as "original" and "flat," and the spectra are shown in Fig. 3. Note that their spectral centroid and overall bandwidth (measured using standard deviation of the spectrum) were not affected by this process. The song and its excerpt was chosen because it has wide bandwidth and somewhat constant spectra and beats over time, as well as the participants had no prior information about the band and the piece.

2.2 Task

Participants' task was to answer the center frequency of an octave-band peak raised by 12 dB with quality factor of 2.0. In the training session, a sound without equalization was presented followed by the sound with equalization (Fig.4). Participants had about three seconds to respond which frequency was raised in written form. The participants were allowed to write down the response during the playback of the processed sounds. Ten such tasks were given in random center frequencies (among 63, 125, 250, 500, 1000, 2000, 4000, 8000, and 16000 Hz, which the participants were instructed beforehand). The equalization filters were biquad peaking filters programmed in Matlab using formulae in [5]. Filter shapes are shown in Fig. 5.

2.3 Participants

The tests were done in Technical Ear Training course offered in Department of Musical Creativity and the Environment, Faculty of Music, Tokyo University of the Arts. Students ranged from first year to third year in the university. The course consisted of 30 weeks in an academic year, and the above-mentioned sounds were presented in two different weeks at around 20th week. Two versions (original and flat) were given in a random order in two different weeks. The stimuli were used without revealing to the students that there are two versions. The test was repeated for four times for each version during three consecutive years from 2015 to 2017[1].

3 Results

Total of 469 responses for original sounds and 390 responses for flat sounds were collected from total of 32 different participants in eight ear training sessions (four original and four flat sessions) executed over three

[1] June 6, 2015 (flat), June 22, 2015 (original), May 24, 2016 (flat), May 31, 2016 (original), June 27, 2016 (flat), May 23, 2017 (original), May 30, 2017 (flat), and June 6, 2017 (original).

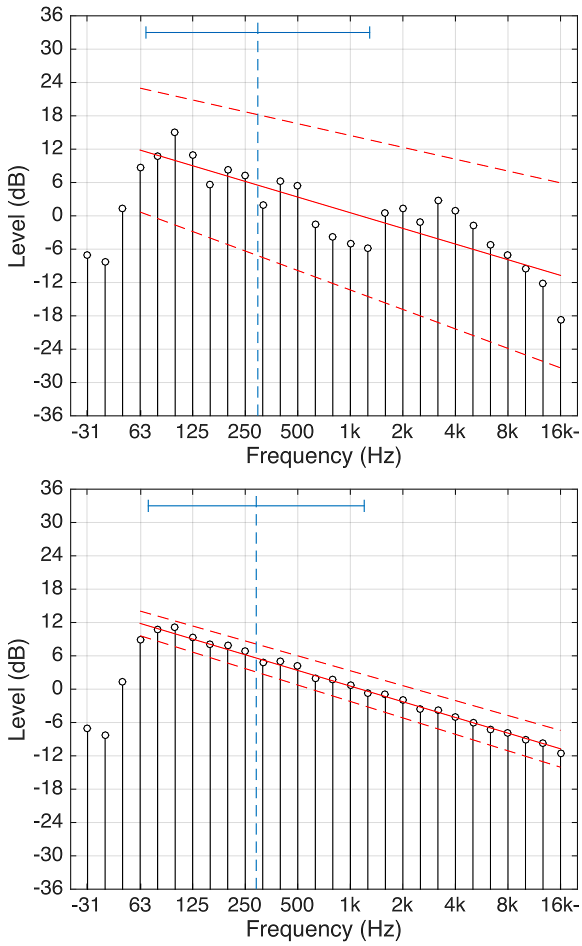

Fig. 3: Power spectra of the music excerpts used in the test. Upper panel is unprocessed "original" and lower panel is "flat" processed version. Circles show the power at each 1/3-octave band. Solid line is the linear fit through the spectrum between 63 Hz and 16 kHz (with dashed line above and below it denoting 1 standard deviation). Vertical dashed line shows spectral centroid with a horizontal whisker showing 1 standard deviation around the centroid.

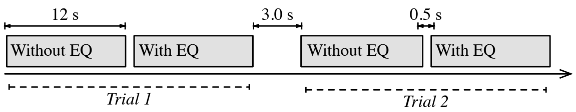

Fig. 4: Stimuli presentation time-line. "Without EQ" is presented followed by "With EQ" version. There were total of 10 trials in one test session (only 2 are shown in the figure).

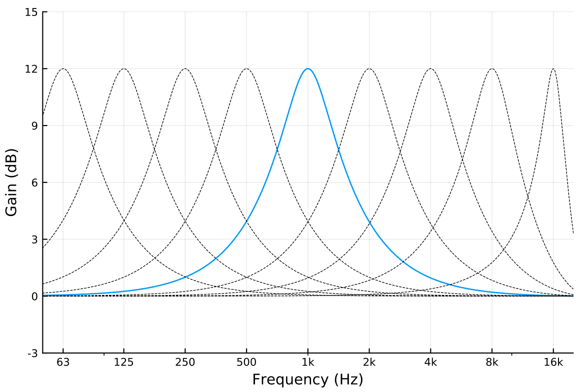

Fig. 5: Biquad peaking EQ used for creating the stimuli (gain = 12 dB, Q = 2.0) [5]. Solid curve shows the filter centered at 1 kHz and dashed curves show the filters of the other center frequencies. (Line style differences are for presentation only and there are no difference in the filter algorithm)

Table 1: Correct ratio (average score, higher is better and shown in **bold**) for center frequencies of the peaks.

Frequency	Original		Flat
63 Hz	**0.864**	>	0.857
125 Hz	0.759	<	**0.800**
250 Hz	0.745	<	**0.897**
500 Hz	0.780	<	**0.821**
1 kHz	**0.783**	>	0.686
2 kHz	**0.708**	>	0.583
4 kHz	**0.783**	>	0.690
8 kHz	**0.776**	>	0.692
16 kHz	0.771	<	**0.880**
overall	**0.780**	>	0.767

years. In each session, about 11 students participated in the tasks.

The responses were graded as either correct (= 1) or wrong (= 0), and the average score for each of the spectral envelope versions were calculated by taking the ratio of the correct response to the total number of responses. The correct ratio of the original version was 0.780 and the flat was 0.767 with no statistically significant differences. Detailed look at correct ratios for each frequency did not reveal systematic patterns (Table 1).

Then, the score was graded with how many octaves the response was off from the correct answer by (*e.g.*, 2 octaves off if the correct answer was 2000 Hz and the response was 500 Hz). A slight decrease in wrong response was seen for "flat" (avg. 0.292 octaves) in comparison to the "original" (avg. 0.320 octaves), although with little statistical significance (Table 2).

In the first grading method of correct/wrong decision, original version was answered correctly than the flat version. In the second grading method of octaves answered wrong by, flat version was answered correctly than the original version. In frequency bands from 1 kHz to 8 kHz, original version was answered with higher correct rate in both grading methods. However, the effects are small and no statistically significant differences were observed in both grading methods.

4 Discussion

The results that there were no significant difference between original and flat versions is counter intuitive

Table 2: Mean number of octaves answered wrong (lower is better and shown in **bold**) for peak frequencies.

Frequency	Original		Flat
63 Hz	0.153	>	**0.143**
125 Hz	0.241	>	**0.240**
250 Hz	0.255	>	**0.103**
500 Hz	0.373	>	**0.308**
1 kHz	**0.233**	<	0.314
2 kHz	**0.417**	<	0.533
4 kHz	**0.275**	<	0.310
8 kHz	**0.293**	<	0.333
16 kHz	0.943	>	**0.280**
overall	0.320	>	**0.292**

to what authors first thought, that is, sounds that have flat spectrum (such as pink-noise) are thought to be easier to identify the spectral changes. However, there were no significant differences in correct ratio between them.

The previous research finding [4] was that altering the spectral flatness affects the participants' subjective difficulty for the task. The current finding is that the flatness does not affect the participants' ability to respond correctly to the task, namely the objective difficulty. The results from the current study combined with the previous findings suggest that altering the spectral flatness affects the participants' perception of how difficult the task feels but does not affect their ability to respond correctly to the task.

However, having no statistically significance between the two versions may be due to the characteristics of the task that "without EQ" version was always presented before "with EQ" version, so the participants could build their reference during the comparison task and hence there might be little effect from the original or flat sources used in the tests. Another possibility is that the spectral differences between the original and the flat version were not large enough for the effect to have statistical significance. Further study is necessary.

5 Summary

In order to investigate whether spectral flatness affect the objective difficulty of peak identification task in Technical Ear Training, experimental test sessions were carried out. Authors created the training materials with different spectra ("original" and "flat" versions) but having the same music content and tested them in the actual ear training sessions.

From the test results, no statistically significant differences from the spectral flatness were seen in the number and amount of correct responses. Thus, combined with the authors' previous research, it was found that different spectral flatness affects the participants' perception of how difficult the task feels but does not affect their ability to respond correctly to the task.

A possibility was seen for controlling subjective difficulty independent of objective difficulty in Technical Ear Training from the results. However, the findings are limited only to the sounds used in the mentioned training sessions, and further study is necessary to both confirm it and to find the objective measure for the difficulty of the Technical Ear Training tasks.

References

[1] Letowski, T., "Development of Technical Listening Skills: Timbre Solfeggio," *Journal of Audio Engineering Society*, 33(4), pp. 240–244, 1985.

[2] Miśkiewicz, A., "Timbre Solfege: A Course in Technical Listening for Sound Engineers," *Journal of Audio Engineering Society*, 40(7/8), pp. 621–625, 1992.

[3] Iwamiya, S., Nakajima, Y., Ueda, K., Kawahara, K., and Takada, M., "Technical Listening Training: Improvement of Sound Sensitivity for Acoustic Engineers and Designers," *Acoustical Science and Technology*, 24(1), pp. 27–31, 2003.

[4] Marui, A. and Kamekawa, T., "Towards the Development of Objective Difficulty Measure in Technical Ear Training Tasks," in *Proceedings of Meetings on Acoustics*, volume 19, 21st International Congress of Acoustics (ICA2013), Montreal, Canada, 2013.

[5] Zölzer, U., *Digital Audio Signal Processing*, Wiley, 1997, ISBN 0-471-97226-6.