Edinburgh, Scotland
EURONOISE 2009
October 26-28

Psychoacoustic criteria evaluates the quality of musical sounds

Masanobu MIURA[a]
Nozomiko YASUI [b]
Dept. of Media Informatics, Fac. of Sci.,&Tech., Ryukoku Univ.
Otsu, Shiga, 520-2194, Japan

ABSTRACT

This paper explains that psychoacoustic criteria evaluate the quality of musical sounds in terms of proficiency of performance. The tremolo performance played by mandolin is known as a performed sound produced by repetition of plucked strings, giving us the Fluctuation Strength. Authors confirms a relation that a musical sound evaluated as high from a psychoacoustic viewpoint will be evaluated as a proficient performance, which means the proficiency of music is highly correlated to a simple impression described by psychoacoustic criteria. Authors are also trying to construct an evaluation model for the aesthetics of four parts in the theory of harmony based on the Western classical music established in the seventeen century. The four parts are assumed here as a short musical signal, thought as appropriate for investigating the criteria on evaluating musical signals from a musical viewpoint. Here, the inhibition rules in the theory of harmony are investigated, by observing the characteristics of the inhibition rules. Therefore, four profiles of criteria are commonly found; a note, an interval, sequence of two notes and sequence of two intervals. These four profiles are thought as the basic criteria for evaluating the four parts in the theory of harmony. Finally, authors are trying to construct a conceptual and hierarchic model concerning the feeling the music subjectively, showing that all the impressions for musical signals are based on both simple (lower layer) and complex (upper layer) criteria. Relations between each layer of criteria are to be investigated in several subjective experiments, so that the model is expected to be clarified in the near future.

1. INTRODUCTION

Clarifying the recognition system of music has been thought as difficult from a scientific viewpoint, because of the difficulty of clarifying criteria for it, as well as mutual dependencies among them. On the other hand, several listening criteria for sounds, such as roughness, Fluctuation Strength(FS), and so forth, have been developed, for clarifying the system of listening to various kinds of sounds. Psychoacoustic criteria are thought as powerful, because they have almost no requirements for listening sounds, which means even if listening sound is music or noise, the criteria will be done in a certain manner. In this paper, we show scientific studies concerning the evaluation of quality of music, by using several criteria including conventional psychoacoustic criteria, as well as expecting that these criteria to will be applied to clarify the system of subjective evaluation of music performances.

2. PREVIOUS STUDIES OF FLUCTUATION STRENGTH AND ROUGHNESS

FS is one of evaluation indexes of psychological level, proposed by Fastl, in relation to hearing sensations concerning modulated sounds with a modulation frequency [1]. FS is elicited by modulated sounds with a modulation frequency of up to about 20 Hz or low frequency.

[a] Email address. miura@rins.ryukoku.ac.jp
[b] Email address. n-y@mail.ryukoku.ac.jp

"Roughness" which is elicited by modulated sounds with a modulation frequency of above about 20 Hz or high frequency, was suggested as the other evaluation index of hearing sensation. According to the finding of past studies, it has been clarified that FS for AM SIN with a modulation frequency within 4-8 Hz is larger than that in the range outside of 4-8 Hz, and the strongest sensation for a fluctuation subjectivity is obtained within 4-8 Hz[1]. Past studies concerning feeling of fluctuation in musical performances have been conducted using FS or roughness [1-4]. Andreas et al., extracted the physical characteristics of fluctuations that are critical to recognize rhythm patterns in various frequency bands for the current acoustical information using FS [3]. Kurakata et al., reported that the pattern of fluctuation of intensity in piano performances is related to FS [4]. Fujisawa et al., investigated fluctuation in frequency of the Japanese pressed voice, Dami-Goe, using roughness [5].

3. EVALUATING FLUCTUATED SOUND OF MUSIC PERFORMANCE

A. Calculating the amount of FS

Firstly, we investigated a fluctuated sound, known as tremolo, played by the mandolin. The mandolin is a kind of string instruments[6], played by plucking the strings alternatively using a pick. As sounds played on the mandolin is quickly attenuating, so players have to pluck strings repeatedly if they want a continuous sounds. Thus, tremolo is realized as a repetition of single sounds, therefore the fluctuation will be observed. In order to investigate the effect of the FS to the evaluation of proficiency, we need to obtain the amount of FS by calculation, so we defined the calculated FS as "physical FS", whereas we call the conventional FS proposed by Fastl as "psychological FS". Figure 1 shows the procedure for calculating physical FS. The levels of acoustic power of tremolos among players (S_1-S_7) were equalized so as to be approximately equal. Tremolos were divided into several waveforms (ID: 1 ~ N) respectively (f_1 (t) ~ f_N (t) shown in Fig.1). The wavelength and shift width of each waveform were 1 sec and 0.5 sec, respectively. FFT was performed for all acoustic data converted to absolute values for all sampled data. A band-pass filter for 4-8 Hz was used to cut signals of other regions of the converted data. IFFT was performed for them, and the averages of RMS values of the obtained waveforms were calculated. The physical FS of the tremolo of n Hz performed by player i ($F_{i,\, n\ \mathrm{Hz}}$) was obtained by summing up obtained RMS values over the divided waveforms.

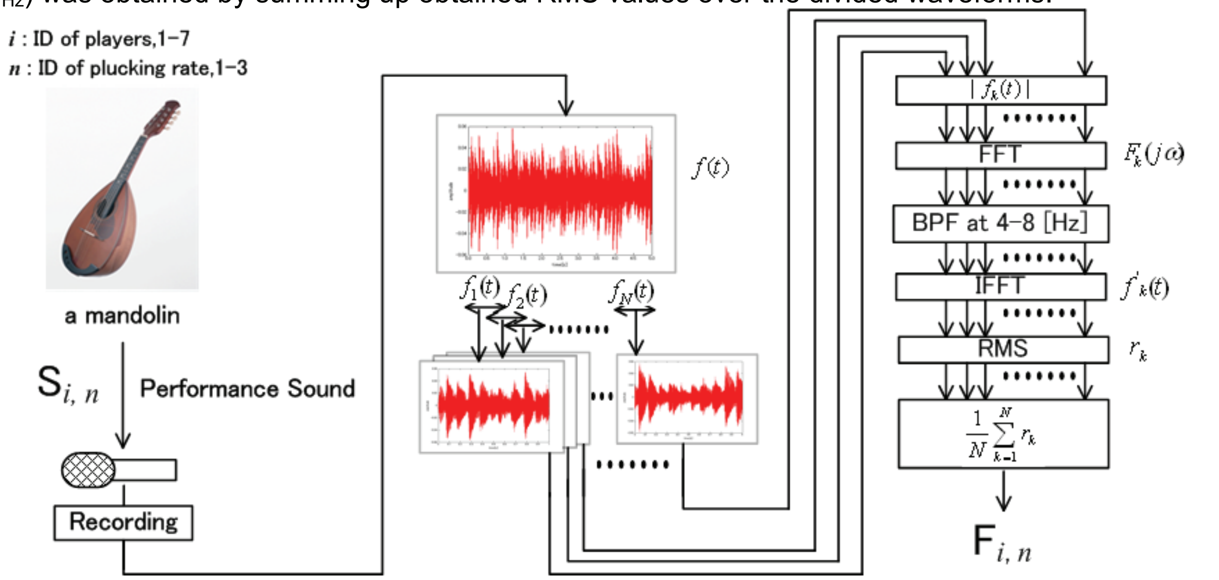

Figure 1. Physical FS calculation procedure

B. Investigation of actual performance

We investigated the maximum tremolo-plucking rate on a performance. Eight mandolin players with over a year of experience were asked to subjectively state their own maximum playable plucking rate on the basis of their tempi measured with a metronome. They were also instructed to state the tempo that they believed they were able to play for longer than thirty seconds. The results of the questionnaire are shown in Fig. 2. The upper horizontal axis represents the tempo, and the lower one represents the plucking rate. The circles in Fig. 2 represent the results of the questionnaire, which revealed that the maximum plucking rate was 11.7 Hz, standard deviation was 1.8. Thus, we found that mandolin players play tremolo with a plucking rate of up to approximately 12 Hz.

Next, tremolos played by seven mandolin experts were recorded by a microphone and a high-speed camera in a soundproof room. Seven mandolin players with over three years experience. They were asked to play tremolo at three plucking rates, 6, 8, and 9 Hz, for five seconds in a soundproof room. Then, they were instructed to play for longer than thirty-seconds at 90, 120, and 135 bpm (corresponds to 6, 8, and 9 Hz, respectively). Click sounds were presented with the tempo using a dynamic earphone from a metronome to only the left ear of the player because presenting the click sound in both ears would interfere with playing. Tone height of the tremolo was D4 (about 295 Hz), or the open tone for the third pair of strings. Intensity of sound was asked to be constant and to be "p (piano)", based on experts' comments that they can recognize easily the difference between an aesthetic tremolo and an unaesthetic tremolo when the intensity of sound is relatively small.

The results of onset deviations on actual performance are shown in Fig. 3. The horizontal axis represents the player, and the ordinate axis represents standard deviations on the IOI of plucking timing. The average of standard deviation on onset deviation was approximately 7 ms. As can be seen in Fig. 3, although the onset deviation has varieties that depend on the players, the range of standard deviation on onset deviation is approximately 3-14 ms.

Table 1 shows results of evaluating subjectively the tremolo performance proficiency for each player. As can be seen in Table 1, it is confirmed that tremolos performed at plucking rates of 8 or 9 Hz was evaluated as good. The chi-square test for goodness-of-fit was conducted determine to whether or not the results were significantly affected by plucking rates. It is found that the results were significantly affected by plucking rates of all players ($p < .05$).

Figure 4 shows the amount of physical FS for tremolos performed by each player. As can be seen in Fig. 4, faster rates correspond to smaller amounts of physical FS. However, it is confirmed in some cases that the amount of physical FS is small despite a slow plucking rate (for example, between P_7's 8 Hz and P_2's 6 Hz performance). This is thought to be because of the difference in timbre and/or irregularity of plucking.

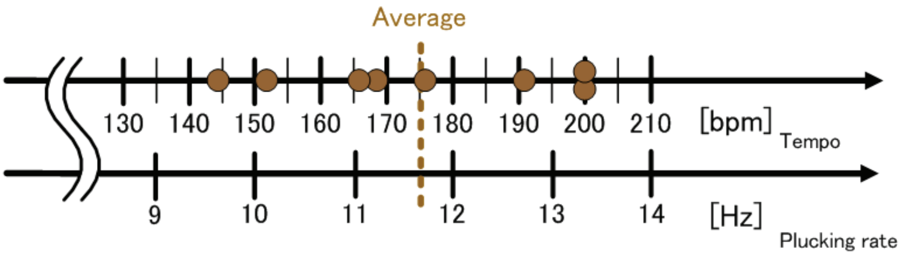

Figure 2. Results of questionnaire.

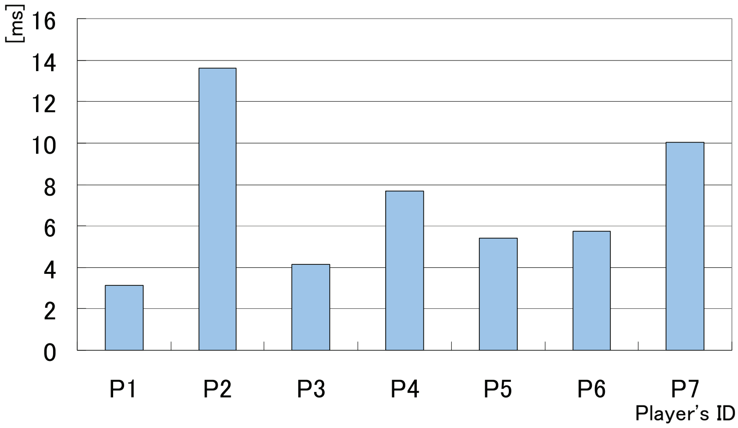

Figure 3 Onset deviations on tremolo

Table 1: Results of evaluating subjectively the tremolo performance proficiency performed by each player

	Plucking rate		
Player	6Hz	8Hz	9Hz
P_1	4* (11 %)	10* (28 %)	22* (61 %)
P_2	2* (6 %)	18* (50 %)	16* (44 %)
P_3	0* (0 %)	14* (39 %)	22* (61 %)
P_4	5* (14 %)	11* (30 %)	20* (56 %)
P_5	3* (7 %)	14* (39 %)	19* (54 %)
P_6	6* (16 %)	10* (28 %)	20* (56 %)
P_7	3* (7 %)	18* (50 %)	15* (43 %)

The number of cases evaluated as good (%) * : $p < .05$

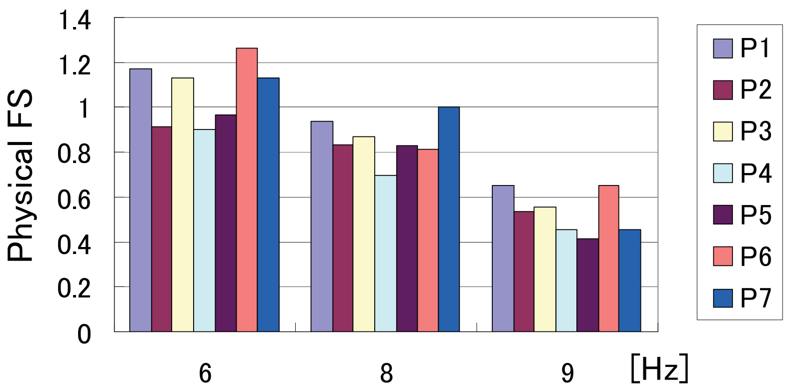

Figure 4 Physical FS for tremolo as specified rate of plucking

C. Investigation of simulated sounds

The investigation of actual performance is somewhat efficient in a sense. However, on the other side, investigations of actual situations are not enough for obtaining the criteria of recognizing the musical sounds. In order to investigate the criteria, we develop a method for simulating tremolo sounds, comprised of recorded sounds of playing. Two types of sounds, produced by plucking a specified pair of open strings, called here as "S_U" and "S_D", up motion or down motion respectively are recorded independently, The simulated tremolo sounds are the conjunction of the S_U and S_D alternatively, such as S_U, S_D, S_U, S_D,… . On generating the simulated tremolo, we can give two kinds of deviations, such as onset and dynamics. Both of the deviations are always observed in actual performance, so we need to give such deviations to simulated sounds to realize a natural tremolo. Onset deviations were investigated, and we found that the average of deviations of IOI in plucking is 7ms, and the amplitude deviations is observed by comparing the deviated sounds generated by our method, by determining the onset timings by the use of normal distribution, with recorded tremolo sounds. Preliminary results shows that actual performances have amplitude deviations, between 2dB and 6dB, 4dB is the average for recorded seven tremolos. So, we design three types of simulated sound, no deviation, 7ms and 2dB deviations, and 7ms and 4dB deviations.

Results of evaluating subjectively the tremolo performance proficiency are shown in Fig. 5. As can been seen in Fig. 5, it was also found that the tremolos with plucking rates of 8-11 Hz was evaluated as good. The chi-square test for goodness-of-fit was conducted to determine whether or not the results were significantly affected by plucking rates. We found that the results were significantly affected by plucking rates of all deviations ($p < .05$).

The amount of physical FS for tremolos synthesized by each deviation is shown in Fig. 6. As can be seen in Fig. 6, a plucking rate of 7 Hz corresponds to larger amounts of physical FS, and one after 7 Hz correspond to smaller amounts physical FS, exponentially.

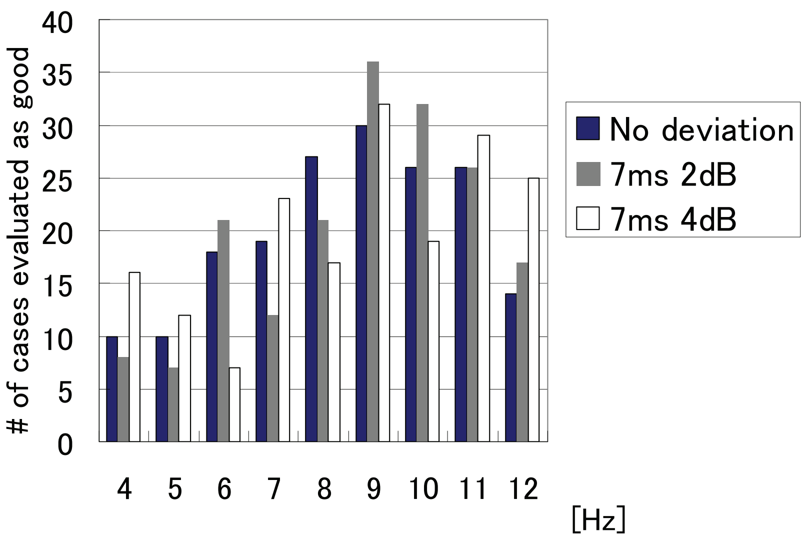

Figure 5. Results of subjective evaluation for proficiency of simulated tremolo

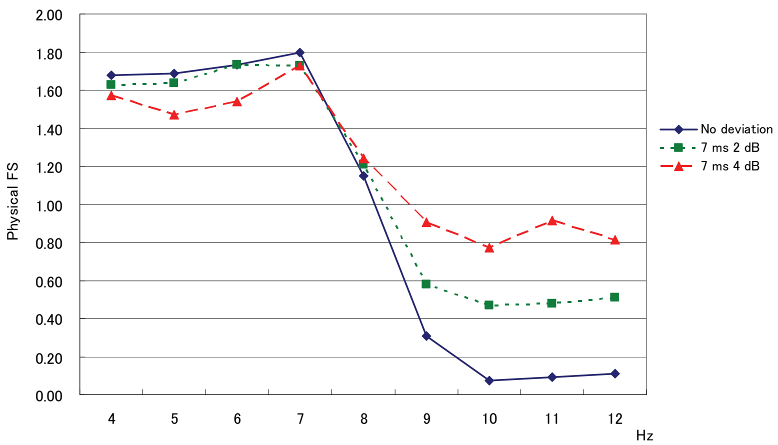

Figure 6. Physical FS for simulated tremolo sounds under three conditions.

D. Relation of criteria of psychoacoustic and musical criteria

As can be seen in Fig. 5 and Fig. 6, highly rated tremolos had a relatively low physical FS. Furthermore, the physical FS has a strong negative correlation with a high subjective evaluation (r = -.60(n=27)). Therefore, this shows that physical FS can be used as an evaluation index of the tremolo performance proficiency played on the mandolin.

The relation between results obtained in this study and psychological FS is shown in Fig.7. The "playable region" is obtained from the results of questionnaire. As can be seen in Fig. 7, two strong relations were confirmed among the calculated physical FS, the subjective evaluation results, and the psychological FS.

The findings show that "a tremolo played on the mandolin is best performed by avoiding higher levels of psychological FS under the playable region". In other words, a skillfully played tremolo is one that satisfies both a playing restriction and a psychoacoustical criterion.

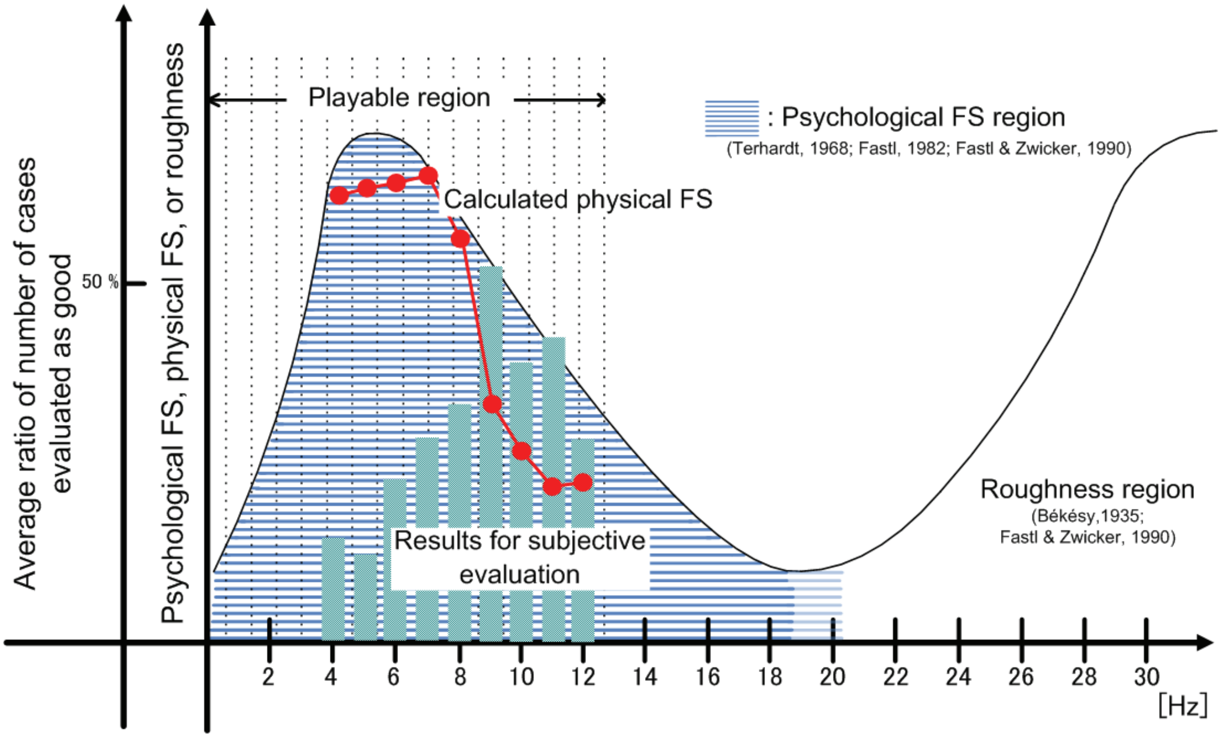

Figure 7. Relation between our results and psychological FS

4. NEW CRITERIA FOR THE EVALUATION OF FOUR PARTS IN THE THEORY OF HARMONY

A. Motivation

Fluctuation Strength is introduced as a basis of evaluation of quality of musical sound (tremolo , in this case). Here discusses about another layer of music perception, concerning the allocation of notes (i.e., musical composition). The basis of musical composition is given through learning the theory of harmony, which is based on the Western classical music, at musical schools. On learning the theory of harmony, students are required to solve given tasks, in which students have to allocate notes on musical score for given bass or soprano sequence. Students have to observe the inhibition rules, in which any allowable or inhibit profiles of notes are described. Although the rules in the theory of harmony are not global but specific criteria, the rules have a certain account for defining the appropriateness of music excerpts. Here, we discuss about the availability of evaluating four parts in the theory of harmony by just listening. Listening criteria for evaluating the four parts have not been clarified yet. In this study, we focus on the rules in the theory of harmony. Especially, a set of four parts, one of them is an allowable and the other is not allowable because of violation for rules. We try to investigate whether or not non-musician can judge the existence of violation of rules in the excerpts.

B. Analysis
At first, we need to extract the basic profile of the rules. Here 37 rules, extracted from a textbook of the theory of harmony[7], are chosen to be analyzed. In our previous study, the rules in the theory of harmony were implemented to a computer as a rule-base system[8], so we can analyze the profile of rules in terms of profiles of notes.

C. Four Criteria
All the rules are represented as vector containing 78 parameters, corresponding to "pitch of soprano", "interval between tenor and bass" and so forth. By holding together with same profiles, we obtain four profiles of notes, such as (i) height of single note, (ii) progression of single note, (iii) progression of an interval, and (iv) an interval. Figure 8 shows the concept of profiles. These four profiles are concerned with all the 37 rules except a rule of chord progression, so that the four profiles almost cover all the basic rules in the theory of harmony.

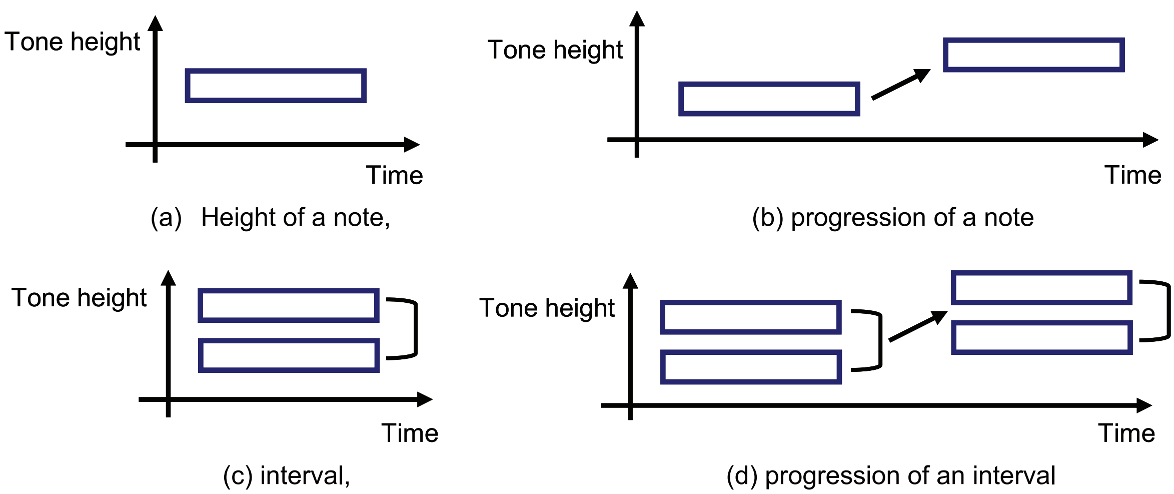

Figure 8. Four criteria for evaluating the four parts in the theory of harmony

D. Listening test
Four non-musicians participated were participated in this test as subjects. At first, one of the rules was taught by using just sounds (no musical score was used). Concretely, the difference between allowable and not allowable excerpts was explained before the subject answered as he/she could understand. After subject's answer, stimuli concerning the rules explained already were presented to the subjects, and the subject was asked to answer whether or not the stimulus is allowable or not.

Figure 10 shows the results of listening test, in which abscissa represents name of rules and ordinate represents average of correct answers. As is seen in Fig.10, the difficulty of evaluating the goodness of four parts depends on the profiles. Especially, the height of a note is quite easy, because in this case the mistake of allocating notes will give us the dissonance of the harmony, even if the consonance was introduced by previous study within one octave[9,10], it will give us a feeling something like dissonance. The progression of a note and progression of an interval was confirmed as be able to be heard for non-musicians. Criteria concerning such kinds of profiles have not been reported, so something criteria concerning this pattern of notes are expected to be developed. The interval of a single chord was not correctly evaluated. The profile of interval in a single chord is thought as difficult for non-musicians, but we need to investigate other patterns of examples.

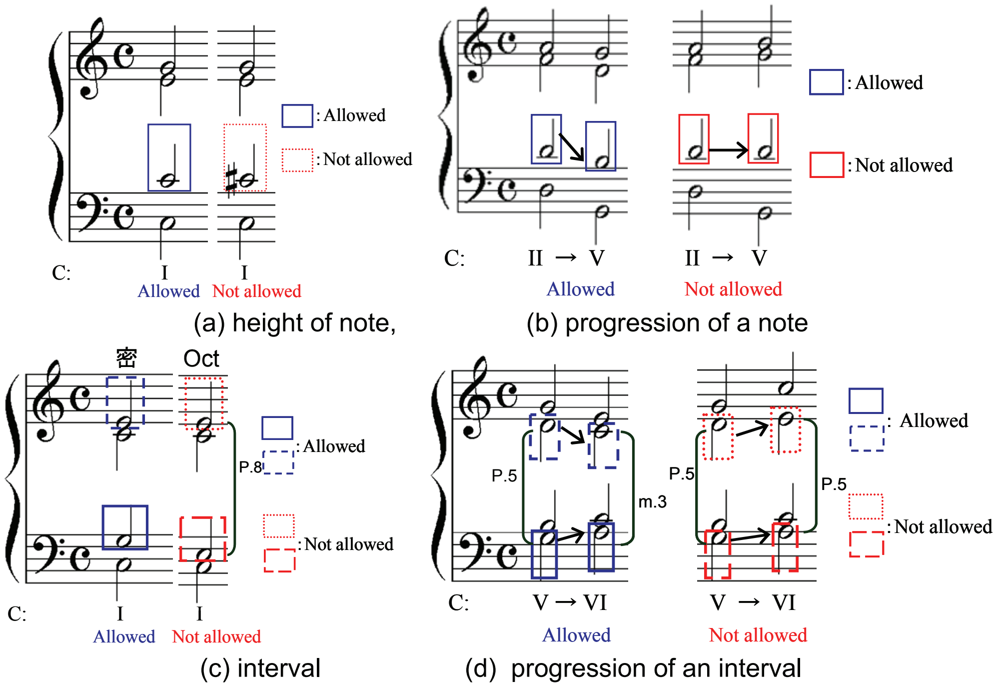

Figure 9. Examples of stimuli

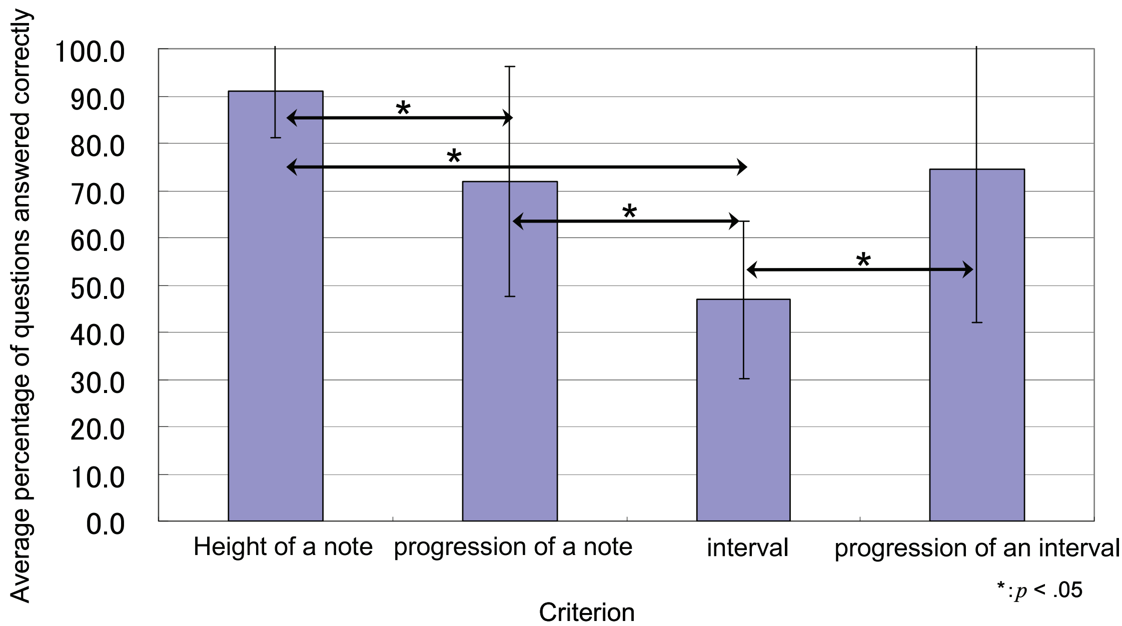

Figure 10. Results of listening test for evaluating the allowableness of given four parts.

5. A CONCEPTUAL MODEL FOR THE EVALUATION OF THE QUALITY OF MUSICAL SOUNDS

A. Motivation

Quality of music, i.e., appropriateness, or aesthetics of music, is evaluated in our daily life, however the system of feeling the music have not been clarified. Chapter 2 and 3 show examples of psychoacoustic criteria for evaluating the quality of music. Here, we try to discuss about a system of evaluating music by listening.

B. Hierarchy model for evaluating the musical sounds

In case of playing a solo on mandolin, any impression of a single note is described as the timbre, comprised of impression for powerful, rough, and metallic. But if the sound is played repeatedly, it will give us fluctuations, and finally, if an excerpt is played, we would observe lots of note sequences and interval sequences in the excerpts. According to this fact, there is a hierarchy model for evaluating the musical sounds, consisting of:

(1) Basic layer, comprised of the impression of a note, such as timbre, pitch, loudness, consonance, and so forth,

(2) Middle layer, comprised of the sequence of notes, such as fluctuations, roughness, part of rules in the theory of harmony, and so forth,

(3) Upper layer, comprised of the construction of notes, such as almost all of the rules in the theory of harmony, technique of music composition, and so forth.

In each layer, we have some kinds of subjective profiles, and by summing up all the impressions of each layer in the excerpt, we will evaluate the overall performance of the excerpt. This model is now quite ambiguous and has not been completed yet, we believe this representation will support to find the musical recognition system in near future.

6. CONCLUSIONS

This paper shows examples of psychoacoustic criteria for evaluating the quality of music. Although the relation between psychoacoustic criteria and musical criteria have been rarely reported, psychoacoustic criteria should be useful for recognition of music excerpts because of the universality of them. Here shows the relation of FS to musical sounds, and a possibility of listening criteria for listening a simple excerpt, four parts in the theory of harmony. Finally, we introduce a concept of recognition system as a hierarchy model. We will study about the other example of psychoacoustic criteria for recognizing music stimuli in near future.

ACKNOWLEDGMENTS

We are grateful to Prof. Hugo Fastl, with discussing for the application of Fluctuation Strength to actual musical contents, Prof. Yanagida and Prof. Yamada for their help in giving suggestions and ideas for this study. We also express our sincere thanks to OB/OG and the players belong to the Ryukoku University Mandolin Orchestra, players or listeners, for their help. This study is partly supported by “High-Tech Research Center” Project for Private Universities: matching fund subsidy from MEXT (Ministry of Education, Culture, Sports, Science and Technology), 2006-2011.

REFERENCES

1. H. Fastl and E. Zwicker, "Psychoacoustics Facts and Models", *pp.247-256,* Springer-Verlag (1990).

2. H. Fastl, "Fluctuation Strength and temporal masking patterns of amplitude-modulated broadband noise", Hearing Research, 8, pp.56-69 (1982).

3. Andreas Rauber, Elias Pampalk and Dieter Merkl, "Using Psycho-Acoustic Models and Self-Organizing Maps to Create a Hierarchical Structuring of Music by Sound Similarity", Proc. of ISMIR 2002, Paris, France, October 13-17 (2002).

4. K. Kurakata, S. Kuwano and S. Namba, "Factors determining the impression of the equality of intensity in piano performance", Journal of the Acoustical Society of Japan (E), 14, 6, pp.441-447 (1993).

5. N. Fujisawa, M. Yamada and I. Nakayama, "Fluctuation in frequency of Japanese pressed voice, Dami-Goe", Journal of the Acoustical Society of Japan (E), 21, 6, pp.365-367 (2000).

6. O. Fujita, "Illustrated the correct way to play Mandolin -for beginners to instructors with concert repertories-", Zen-On Music Company Ltd., (2005, in Japanese).

7. T. Ikenouchi et al., “Harmony, -theory and practice-”, Ongaku-no-tomosha, Tokyo, 1964 (in Japanese).

8. M. MIURA, and M. YANAGIDA "Constructing a validity Investigation system for answers to given exercises in the theory of harmony", Proc. of 18th International Congress on Acoustics, Mo2.C1.4,(2004).

9. R Plomp, and W.J.M. Levelt, “Tonal consonance and critical bandwidth”, Journal of Acoustic Society of America, 38, 4, pp.548-560 (1965).

10. A. Kameoka and M. Kuriyagawa, “Consonance theory part I: consonance of dyads”, Journal of Acoustic Society of America, 45, pp.1451-1459 (1969).