ism: Improvisation Supporting System based on Melody Correction

Katsuhisa Ishida
Dept. of Information Sciences, Graduate School of Sciences and Technology, Tokyo University of Science
Noda City, Chiba 278-8510, Japan
kdi@mt.is.noda.tus.ac.jp

Tetsuro Kitahara
Dept. of Intelligence Science and Technology, Graduate School of Informatics, Kyoto University
Sakyo-ku, Kyoto 606-8501, Japan
kitahara@kuis.kyoto-u.ac.jp

Masayuki Takeda
Dept. of Information Sciences, Faculty of Sciences and Technology, Tokyo University of Science
Noda City, Chiba 278-8510, Japan
takeda@is.noda.tus.ac.jp

ABSTRACT

In this paper, we describe a novel improvisation supporting system based on correcting musically unnatural melodies. Since improvisation is the musical performance style that involves creating melodies while playing, it is not easy even for the people who can play musical instruments. However, previous studies have not dealt with improvisation support for the people who can play musical instruments but cannot improvise. In this study, to support such players' improvisation, we propose a novel improvisation supporting system called ***ism***, which corrects musically unnatural melodies automatically. The main issue in realizing this system is how to detect notes to be corrected (*i.e.*, musically unnatural or inappropriate). We propose a method for detecting notes to be corrected based on the N-gram model. This method first calculates N-gram probabilities of played notes, and then judges notes with low N-gram probabilities to be corrected. Experimental results show that the N-gram-based melody correction and the proposed system are useful for supporting improvisation.

Keywords

Improvisation support, jam session, melody correction, N-gram model, melody modeling, musical instrument

1. INTRODUCTION

Recent development of multimedia technology has enabled us to enjoy various styles of jam sessions. Jam session systems, for example, construct a virtual musician in a computer and provide us with an environment for jam sessions with the virtual musician. *Open RemoteGIG* [1] enables widely distributed musicians to join a worldwide jam session by using the Internet. Moreover, various novel electric musical instruments, such as a PDA-based portable type [2] and a wearable type [3], and new jam-session styles using these instruments have been proposed.

These studies aim at providing a new style jam session, not supporting a jam session for the people who cannot improvise. Since improvisation is the musical performance style that involves creating melodies while playing, to acquire the ability for improvisation, people need further training even if they can play an instrument with a score. There will therefore be many people who can play an instrument but cannot improvise, called *non-improvising players* in this paper, so that supporting such players' improvisation is an important task.

Improvisation can be separated into two techniques: playing an instrument and instantaneously creating melodies. Therefore, improvisation support should also be separated into support of the two techniques. Specifically, if an improvisation supporting system targets non-improvising players, the system should support the ability for only instantaneous melody creation because they already have skill for playing instruments.

In this paper, according to the above concept, we propose an improvisation supporting system, called ***ism***, for non-improvising players. This system automatically detects and corrects musically unnatural notes in the melodies of improvisation at real time. Because our approach is to add the melody correcting function to existing instruments, not to make a new instrument that makes improvisation easier, players can enjoy improvisation using their favorite instruments and their own playing techniques.

The main issue in achieving this system is how to detect notes to be corrected (*i.e.*, musically unnatural or inappropriate). We propose a method for detecting notes to be corrected based on the N-gram model. Our method uses N-gram probabilities calculated from a large-scale melody database to determine whether notes should be corrected or not. This N-gram-based determination makes it possible to solve the problem of correcting notes that should not be corrected.

2. ISM: AN IMPROVISATION SUPPORTING SYSTEM

The target of our study is the people who can play an instrument but cannot create melodies at real time. When we design an improvisation supporting system for such players, we should take into consideration that the players lack only the ability to create melodies instantaneously. Specifically, the system should satisfy the following requirements:

1. **Freedom of selecting an instrument**
   The supporting system should allow players to improvise using their favorite instruments. This is because players who can play instruments would want to use the instruments that they are the most accustomed to.

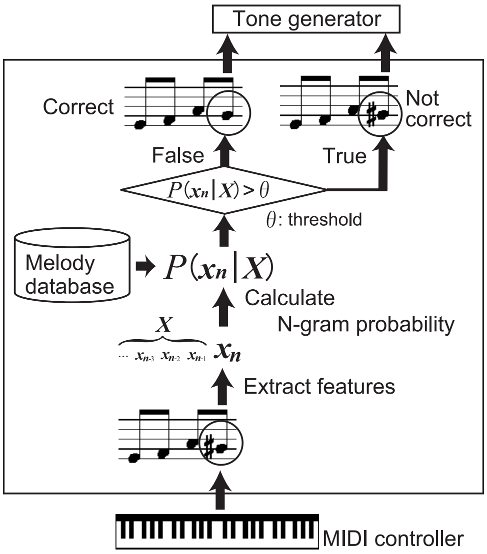

Figure 1: The overview of *ism*. The system first calculates N-gram probabilities of played notes, and then corrects only notes of which the N-gram probabilities are low.

2. **Avoidance of over-supporting**
   Even if players do not have the enough ability to create melodies, they do not always create musically unnatural/inappropriate melodies[1]. The system should not therefore do unexpected support while the player creates appropriate melodies.

Although some musical performance supporting systems have been proposed, these systems do not satisfy the above requirements. For example, *Coloring-in Piano* [4] is a musical instrument that corrects the player's wrong melodies using beforehand given score information. This method is usable for non-improvisation music, but it is not applicable to our purpose because scores of improvisation cannot be given beforehand. *RhyMe*, which is a subsystem of *MusiKalscope* [5], is an improvisation supporting system based on a fixed-function mapping. The fixed-function mapping is a new method for mapping between keys and notes according to the functions of the notes which depend on the context of the chord progression. This novel mapping can make it easier to choose keys that produce musically natural melodies. However, because the instrument using this mapping method is different from those that players are accustomed to, this improvisation support does not satisfy the first requirement. *INSPIRATION* [6] is an improvisation supporting system that corrects all of the notes out of the available note scale. Because these notes do not necessarily produce musically unnatural melodies, it is not desirable to correct all of them. It does not hence satisfy the second requirement.

In this study, we propose a novel performance supporting system called ***ism***, which detects unnatural notes in melodies based on the N-gram model and corrects them (**Figure 1**).

[1]Our investigation using 10 beginners and 15 intermediate players show that the rates of unnatural/inappropriate notes in the melodies of their improvisation are 12.03% and 8.22%, respectively.

Table 1: Elements of feature vector.

The kind of the note (chord tone, key tone, etc.)
The interval between the note and the last note (m2, M2, more than m3)
Whether the note is on eighth-note-level beats
Whether a rest exists between the note and the last note

Because this system can be added into existing MIDI systems, that is, it is set up between a MIDI controller and a MIDI tone generator, the player can use their favorite MIDI controllers such as a MIDI keyboard and a MIDI guitar. In addition, because this system determine whether notes should be corrected by comparing their N-gram probabilities with a threshold, the player can control the strength of melody correction (*i.e.*, how frequently melody correction occurs) by adjusting the threshold.

3. THE METHOD OF OUR SYSTEM

The main issue in achieving ***ism*** is how to detect notes to be corrected. One possible solution for this may be to correct all the notes (called *out notes*) that are out of the available note scale, which is a set of notes that can produce harmonic sounds. However, all of these notes do not necessarily produce disharmonious sounds, and they are frequently used in actual musical pieces. This method, therefore, causes unexpected correction and is not suitable.

In this paper, we propose a novel method for determining notes to be corrected based on the N-gram model. This method captures the tendency of note transitions by N-gram probabilities and judges that only notes with low N-gram probabilities should be corrected.

3.1 Feature Extraction

The 4-dimensional feature vector listed in **Table 1** is extracted from each note in a melody of improvisation. Let "note $\boldsymbol{x}$" be the note with feature vector $\boldsymbol{x}$.

3.2 Modeling melody by N-gram

The appropriateness of note transitions in a played melody is modeled by the N-gram model. This model gives the probability $P(\boldsymbol{x}_n|X)$ in which the note $\boldsymbol{x}_n$ exists behind the note sequence $X = \boldsymbol{x}_1 \cdots \boldsymbol{x}_{n-1}$. The N-gram model assumes that this probability is fixed by the $N-1$ notes $\boldsymbol{x}_{n-N+1} \cdots \boldsymbol{x}_{n-1}$ and calculates it by the following equation:

$$P(\boldsymbol{x}_n|X) = P(\boldsymbol{x}_n|\boldsymbol{x}_{n-N+1} \cdots \boldsymbol{x}_{n-1}) = \frac{P(\boldsymbol{x}_{n-N+1} \cdots \boldsymbol{x}_n)}{P(\boldsymbol{x}_{n-N+1} \cdots \boldsymbol{x}_{n-1})}.$$

3.3 Determining the notes to be corrected

When the out note $\boldsymbol{x}_n$ follows the note sequence X, its appropriateness is given by the N-gram probability $P(\boldsymbol{x}_n|X)$ calculated with a large melody database. In other words, if $P(\boldsymbol{x}_n|X)$ is high, $\boldsymbol{x}_n$ frequently follows X in melodies of actual musical pieces. Our method therefore judges that the out notes of which N-gram probability is lower than a threshold should be corrected.

3.4 Determination of after-correction pitch

The pitch maximizing N-gram probabilities is determined as an after-correction pitch.

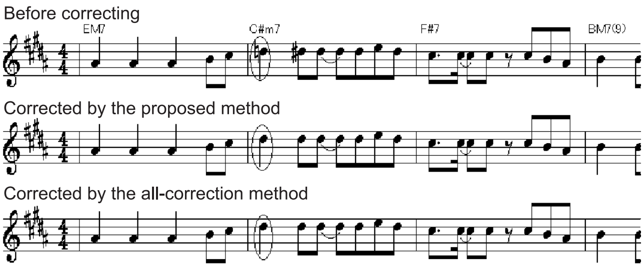

Figure 2: An example of melody correction. The marked note was corrected by both methods.

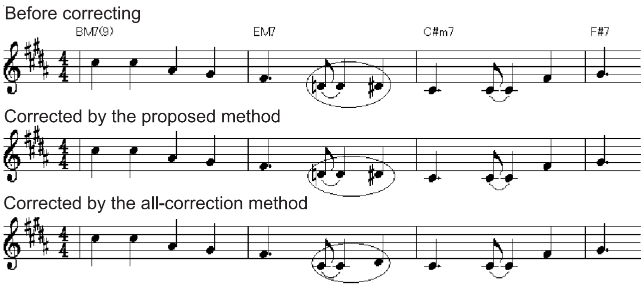

Figure 3: An example of unchanged melodies. The marked note is out of the available note scale, but was judged to be appropriate by the proposed method.

4. IMPLEMENTATION AND EVALUATION

4.1 Implementation

We built a prototype system of ***ism*** using the C language on Microsoft Windows. To construct a melody database, we used 208 songs' melodies of standard jazz. The total number of measures and notes of this database are 6,836 and 18,897, respectively. We adopted both the bigram model ($N = 2$) and the trigram model ($N = 3$) as the N-gram model. The threshold is 0.10.

This system has accompaniment data as standard MIDI files. While an accompaniment is played, a user plays improvisation to this accompaniment using the MIDI keyboard connected with ***ism***. Then, ***ism*** corrects the player's melody and the MIDI tone generator plays the corrected melody.

4.2 Examples of melody correction

Figures 2 and **3** show examples of correcting melodies. The top scores are melodies before correcting. The middle scores are melodies corrected by the proposed method. The bottom scores are melodies corrected by the method that corrects all the out notes (called *all-correction*). The marked note in **Figure 2** is an out note and was corrected by both methods. On the other hand, although the marked note in **Figure 3** is an out note, it was not changed by the proposed method. Actually, this note does not produce a disharmonious sound. These results mean that the proposed method captures harmoniousness of melodies better than the all-correction method.

Table 2: The details of subjects and labeled notes.

	# of players	Measure / player	Total notes	Correction-requiring
Beginners (under 1 yr.*)	10	64	3,108	12.03%
Intermediates (3–5 yrs.*)	15	64	3,177	8.22%
Experts (over 5 yrs.*)	12	64	2,660	3.38%
Total	37	64	8,945	8.11%

*Experience in playing instruments.

4.3 Evaluation of Determining Notes to be Corrected

4.3.1 Experimental conditions

We conducted experiments on determining whether notes in melodies should be corrected or not. The 37 non-improvising players listed in **Table 2** first played improvisation, and then the melodies of their improvisation were recorded. For each note in the melodies, we labeled whether it should be corrected by hand. The melodies were then corrected both by the proposed method and by the all-correction method, and finally the appropriateness of the correction was evaluated using recall rate R, precision rate P and F-measure F:

$$R = \frac{\text{Number of correction-requiring and actually corrected notes}}{\text{Total number of correction-requiring notes}},$$

$$P = \frac{\text{Number of correction-requiring and actually corrected notes}}{\text{Total number of actually corrected notes}},$$

$$F = \frac{2 \times R \times P}{R + P}.$$

4.3.2 Experimental results

Table 3 shows experimental results. Our method based on the bigram and trigram models improved the F-measure by 0.1093 and by 0.1080, respectively. Although the recall rates of the proposed method were 1–2% lower than that of the all-correction method, the precision rates were about 13% higher. These results mean that the proposed method achieved the improvement in over-correction, that is, correcting notes that should not be corrected.

The accuracies for the intermediate group with all the methods were high. Because many players in this group know that notes in the available note scale produce natural melodies, their out notes mainly appeared as a result of mistouching or a failure of challenging an advanced melody. The proposed method detected such clearly unnatural out notes with accuracy.

On the other hand, the accuracies for the expert group were not high enough. This insufficient accuracy was caused by the mismatch of players and the melody database; the melody database was constructed using jazz melodies whereas many players of this group have experience in classical music. It can be improved by constructing genre-dependent or player-dependent melody databases.

4.4 Questionnaire Evaluation

4.4.1 Method of evaluation

We conducted evaluation of users' feelings of our system by questionnaires. The subjects are three people, listed in

Table 3: Experimental results of determining notes to be corrected.

	Whole			Beginners (under 1 yr.*)			Intermediates (3–5 yrs.*)			Experts (over 5 yrs.*)		
	R	P	F	R	P	F	R	P	F	R	P	F
All-correction system	0.7822	0.3636	0.4964	0.7005	0.4242	0.5307	0.9123	0.5131	0.6568	0.7072	0.2012	0.3133
Ours (bigram)	0.7737	0.4977	0.6057	0.6628	0.5066	0.5743	0.9099	0.6622	0.7665	0.7072	0.2985	0.4198
Ours (trigram)	0.7682	0.4982	0.6044	0.6190	0.5078	0.5579	0.8969	0.6585	0.7594	0.7072	0.3032	0.4244

R: Recall rate, P: Precision rate, F: F-measure, *Experience in playing instruments.

Table 4: Musical Experiences of the Three Subjects.

	Playing	Composing	Improvising
Subject A	12 yrs. (Piano)	Yes	No
Subject B	11 yrs. (Electone)	No	No
Subject C	6 yrs. (Keyboard)	Yes	No

Table 5: Questionnaire results.

	Q1			Q2			Q3		
	all	bi	tri	all	bi	tri	all	bi	tri
Subject A	5	4	6	5	4	7	4	5	5
Subject B	5	7	6	1	4	6	6	6	7
Subject C	3	4	7	2	2	4	5	5	5
Average	4.33	5.00	6.33	2.67	3.33	5.67	5.00	5.33	5.67

all: The all-correction system
bi: The bigram system, tri: The trigram system

Table 4, who can play an instrument but have little experience in improvisation. They first played improvisation using our system and then answered the following questions:

Q1 Do you think the correction of your melodies is appropriate? (1: No, 7: Yes)

Q2 Did you feel strong sense of incongruity? (1: Yes, 7: No)

Q3 Did you enjoy your improvisation with this system? (1: No, 7: Yes)

4.4.2 Questionnaire results

Table 5 shows the results of the questionnaire. For any questions, the proposed method was superior to the all-correction method on the average of the three subjects. In particular, all the subjects answered that the trigram-based correcting method was better than the all-correction method. The proposed method is therefore effective in supporting improvisation of non-improvising players.

The results of the trigram-based system were better than the bigram one. This is because the trigram model captures the tendency of note transitions better than the bigram model since the former uses longer note sequences.

Subject A did not highly evaluate the bigram-based system in Q1. This is because chromatic phrases he frequently used were corrected by the system. However, some listeners say that the corrected melodies are more natural, so that this correction is not necessarily redundant.

When we focus on Q2, Subjects A and B highly evaluated our trigram-based system. They have long experience, more than 10 years, in playing instruments. It means that the melody correction by the proposed method does not give a strong incongruity feeling to players even if they have long experience in playing musical instruments.

We also obtained from subjects the opinion that it was good to hide their failure from listeners when they failed in improvisation. This opinion suggests that our system achieved mitigating their hesitation in improvisation.

5. CONCLUSION

In this paper, we proposed a new improvisation supporting system for non-improvising players using N-gram-based melody correction. This system is designed on the basis of the concept that the support of improvisation should be divided into those of playing instruments and instantaneously creating melodies. Supporting the ability of only instantaneous melody creation, according to this concept, made it possible to achieve improvisation support without making useless players' skill for playing instruments.

Moreover, we realized judgment of whether a melody is natural or not. It is considered difficult to automate this judgment due to the difficulty of extracting objective rules. We solved this problem by modeling melodies using the N-gram model, which is usually used for modeling natural languages. Our results suggest that such statistical techniques are usable for such musical judgments.

Future work will include construction of a large-scale database and evaluation of our system in more detail.

6. ACKNOWLEDGMENTS

The authors would like to thank Dr. Masataka Goto (National Institute of Advanced Industrial Science and Technology) for his great advice. The authors would also like to thank Mr. Takahiro Yanagawa, Mr. Yoshihiro Watanabe and many people for their cooperation.

7. REFERENCES

[1] M. Goto and R. Neyama: Open RemoteGIG: An Open-to-the-Public Distributed Session System Overcoming Network Latency, *IPSJ Journal*, **43**, 2, pp.299–309, 2002. (in Japanese)

[2] T. Terada *et al.*: A Portable Electric Bass Using Two PDAs, *Proc. IWEC*, pp. 286–293, 2002.

[3] K. Nishimoto *et al.*: Networked Wearable Musical Instruments Will Bring A New Musical Culture, *Proc. ISWC*, pp.55–62, 2001.

[4] K. Nishimoto *et al.*: A Musical Instrument for Facilitating Musical Expressions, *CHI2002 Extended Abstracts*, pp.722–723, 2002.

[5] S. Fels *et al.*: MusiKalscope: A Graphical Musical Instrument, *IEEE Multimedia*, **5**, 3, pp.26–35, 1998.

[6] A. Yatsui and H. Katayose: An Accommodating Piano which augments intention of inexperienced players, *Entertainment Computing Technology and Applications*, Kluwer Academic Publishers, pp.249–256, 2002.