Subjective Evaluation of Correlation of Melody Length and Melodic Similarity Based on a Music Theory

Sakurako Yazawa
Graduate School of Systems and Information Engineering, University of Tsukuba, Tsukuba, Japan
Email: sakurko_@_music.iit.tsukuba.ac.jp

Masatoshi Hamanaka
Department of Clinical System Onco-Informatics, Kyoto University, Kyoto, Japan
Email: masatosh_@_kuhp.kyoto-u.ac.jp

Takehito Utsuro
Faculty of Engineering, Information and Systems, University of Tsukuba, Tsukuba, Japan
Email: utsuro_@_iit.tsukuba.ac.jp

***Abstract*—This paper studies to measure similarity of melodies based on Implication-Realization Model (IRM), a music theory that abstracts music and then expresses music through symbol sequences based on information constituting the music such as pitch, rhythm, and rests and so on. The model employed in this paper is an extended one that is much more appropriate to measuring similarity of melodies. Compared with the symbols of the original IRM, the extended model employs finer grained symbols by simply distinguishing up and down of interval directions and by dividing each most symbols of the original IRM into two extended symbols. Based on such a fundamental framework of measuring similarity of melodies by an extended IRM, the major focus of this paper is to claim that the similarity measure studied in this paper has a certain correlation with subjective human judgments on melodic similarities. Furthermore, this paper examines the correlation between melody length and subjective human judgments on melodic similarities. It is quite remarkable to note the result of this analysis: i.e., the smaller the melody length is, the more similar the pair of melodies is judged to be by human subjects, even though their similarities based on the extended IRM are measured to be almost in the same range. Thus, we conclude that it is necessary to invent a much finer-grained similarity measure which is designed to be more sensitive to melody length and to be coincide with the results of subjective human judgments.**

***Index Terms*—music theory, melodic similarity, implication-Realization Model (IRM)**

I. Introduction

This paper studies to measure similarity of melodies based on Implication-Realization Model (IRM) [1], [2], a music theory that abstracts music and then expresses music through symbol sequences based on information constituting the music such as pitch, rhythm, and rests and so on. The model employed in this paper is an extended one that is much more appropriate to measuring similarity of melodies. Compared with the symbols of the original IRM, the extended model employs finer grained symbols by simply distinguishing up and down of interval directions and by dividing each most symbols of the original IRM into two extended symbols. In this framework of the extended IRM, a parser is also developed, where it transforms tone sequence of an input melody into a sequence of the extended IRM symbols. When implementing this extended IRM parser, parameters such as thresholds of intervals are examined and an optimal set of parameters is obtained through empirical evaluation.

Based on such a fundamental framework of measuring similarity of melodies by an extended IRM, the major focus of this paper is to claim that the similarity measure studied in this paper has a certain correlation with subjective human judgments on melodic similarities. Furthermore, this paper examines the correlation between melody length and subjective human judgments on melodic similarities. It is quite remark- able to note the result of this analysis: i.e., the smaller the melody length is, the more similar the pair of melodies is judged to be by human subjects, even though their similarities based on the extended IRM are measured to be almost in the same range. Thus, we conclude that it is necessary to introduce a much finer-grained similarity measure which is designed to be more sensitive to melody length and to be coincide with the results of subjective human judgments.

Previous approaches to measuring similarity of melodies include those based on acoustic features of melodies such as based on spectral analysis [3], [4] and discrete Fourier transform [5]. Our approach is different from those previous works in that we represent melodies

Manuscript received May 10, 2015; revised September 3, 2015.

doi: 10.18178/ijsps.4.4.269-275

through symbols of a music theory rather through acoustic features. One of the most important advantages of our approach is that it is quite easy for our approach based on a music theory to abstract melodies in terms of symbols of the music theory. With this advantage, it becomes possible to realize flexible similarity measure that is quite suitable to human's sense of melody similarity.

Among other approaches to measuring similarity of melodies, Grachten *et al.* [6] proposed to measure similarity of melodies based on the symbols of the original IRM. However, it has been already shown that the original symbols of IRM perform worse than the extended IRM symbols [7].

II. Related Work

There exist several approaches to measuring similarities of melodies proposed so far. One approach is based on technique of considering users' preferences that can be collected from users' records of selecting favorite melodies. Hoashi *et al.* [8] proposed a method for selecting the user's favorite melodies by generating vectors representing the user is preference from the melodies selected by the user and by collecting melodies that are similar to the user's favorites. Vignoli and Pauws [9] also proposed to represent vectors of melodies according to the weighted sum of the features such as tones, genre, tempo, year of the release, and the atmosphere of the music, and then to incorporate users' preference into the vector representation by asking users' to tune the weights of the features. Lampropoulos *et al.* [10] also proposed to select similar melodies using neural networks trained with acoustic features, aiming at realizing personalized music similarity.

Another approach to representing similarities of melodies is based on representing similarity of melodies through objective acoustic features. Previous work includes a technique of employing spectral analysis of melodies and then introducing a distance metric of the analyzed results [3], [4], that of applying discrete Fourier transform to the melodies and detecting patterns of phrases [5], that of introducing MFCC features [11]), and that of introducing features based on spectrum shape [12].

One of other approaches is an attempt to representing melodies through symbols rather than through acoustic features. Doraisamy and Ruger [13] proposed to transform a tone sequence into a symbol sequence and then to measure the similarity of the symbol sequences. One of the disadvantages of this approach is that it is not capable of considering structure of the melodies when measuring the similarity of melodies. One other approach is to employ the results of analyzing melodies through a music theory in the task of measuring similarity of melodies. GTTM (Generative Theory of Tonal Music) [14], [15] is an example of implementing a music theory. Hirata *et al.* [16] proposed to measure similarity of melodies through GTTM, where the similarity is measured according to the way the hierarchical time span tree branches, as well as tones under the branches in the tree. One of the major drawbacks of this approach is that it can measure the similarity of melodies only when the two melodies are almost the same.

Compared with those previous approaches, the method employed in this paper is based on an extension of another music theory, IRM, and we realize to abstract melodies through the framework of the extended IRM symbols. One of the major advantages of this approach is that the similarity measure studied in this paper is capable of measuring similarity not only of the almost the same melodies, but also of rather less similar pairs of melodies appropriately.

III. Annotating Symbols of Extended Implication-Realization Model

A. Implication-Realization Model (IRM)

The IRM is a music theory, proposed by Eugene Narmour. The IRM abstracts music and then expresses music according to symbol sequences based on information constituting the music such as pitch, rhythm, and rests and so on (Fig. 1).

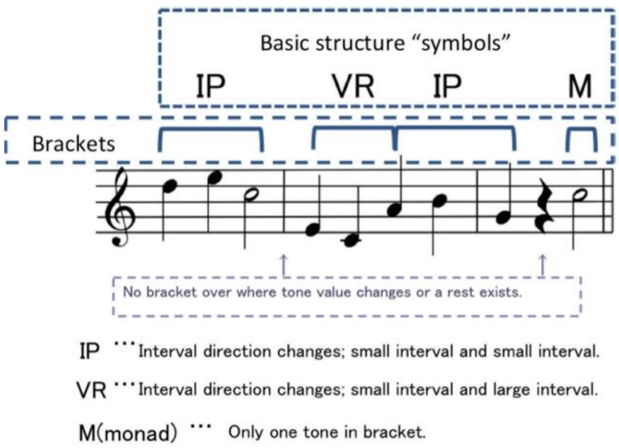

Figure 1. An example of analyzing melodies according to IRM.

When analyzing melody using the IRM, we have the following two steps. The first step is to enclose the tones successively with a bracket. The bracket is an important structure when abstracting melodies. In the procedure of bracket abstraction, first, a large tone column group is created in order to detect the location where the bracket is interrupted. A bracket containing three successive tones is then formed from the beginning to the end of the group. A set of three tones cannot form a bracket, if there are only one or two tones. In such a case, we re-structure the tones sequence and then form a bracket. The second step of analyzing a melody using the IRM is to assign a symbol to each bracket. Tones enclosed in brackets are assigned a symbol and are called basic structures.

There are two important points in assigning symbols. The first point is the pitch of the current two to three consecutive tones. The second point is the interval direction.

There are ten basic structures in the IRM; eight types of symbols include three tones in a bracket (Fig. 2), one "dyad" includes two tones in a bracket, and one "monad" includes one tone in a bracket (Fig. 3).

For example, “IP” in Fig. 1 includes three tones in a bracket and is assigned the feature of “down sound up a narrow pitch”. The tonal row with tones enclosed in various brackets is analyzed based on the basic structures.

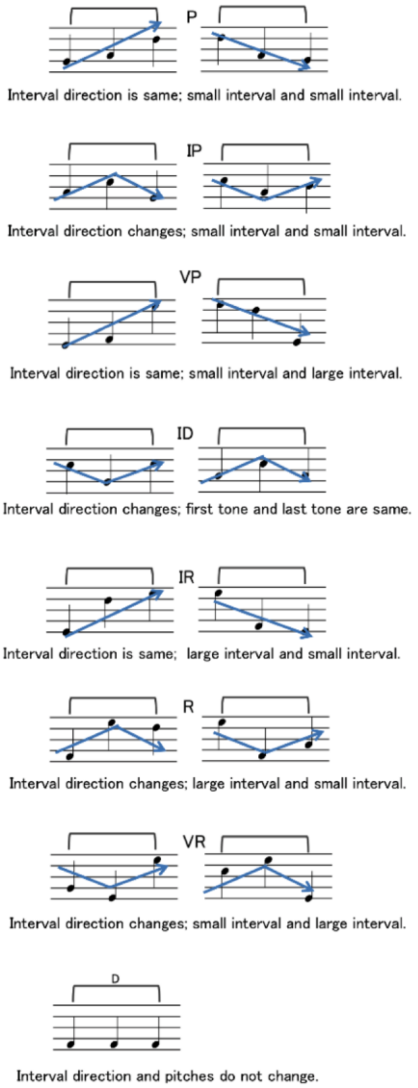

Figure 2. Basic structure symbols.

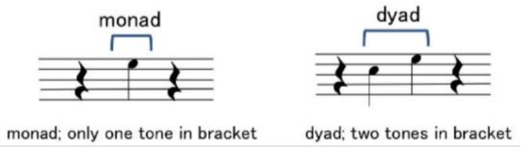

Figure 3. “dyad” and “monad”.

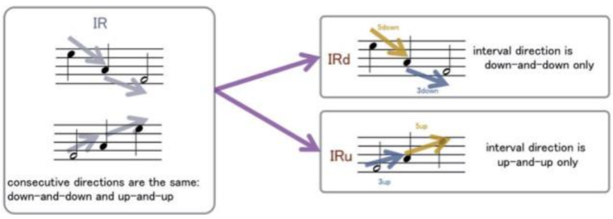

Figure 4. An example of dividing a symbol into two symbols.

B. Extended Symbols of Implication-Realization Model

Original IRM has eight symbols, where they are too coarse- grained in terms of abstracting melodies into symbols and measuring similarities of melodies represented by symbols. This section describes how to extend seven of original symbols into 14 symbols. Fig. 4 shows an example of introducing two extended symbols by dividing one symbol of the original IRM into two symbols of the extended IRM. In this example, as shown in the figure, the original symbol “IR” represents that, for the three tones in a bracket, directions of the changes between two consecutive tones are the same, i.e., down and down, or up and up. Then, these two cases “down and down” and “up and up” are divided into two extended symbols as “IRd” (representing “down and down”) and “IRu” (representing “up and up”).

Fig. 5 illustrates all the cases of dividing seven symbols of the original IRM into 14 extended symbols. More specifically, “P”, “IP”, “VP”, “ID”, “IR”, “R”, and “VR” are divided into “Pu” and “Pd”, “IPu” and “IPd”, “VPu” and “VPd”, “IDu” and “IDd”, “IRu” and “IRd”, “Ru” and “Rd”, and “VPu” and “VPd”, respectively.

C. Implication-Realization Model Parser for Annotating Extended Symbols

This section describes how to implement IRM parser for annotating extended symbols of IRM.

Suppose that we are given a sequence t_1, …, t_m of tones. First, the parser detects the changes of tone values or rests, and segments the sequence into sub-sequences so that each sub- sequence includes neither tone value change nor rest. Then, the parser parses each sub-sequence of tones into a sequence of brackets, by segmenting 3 consecutive tones within a sub- sequence into a bracket. Here, the parser segments 3 tones in a way that two consecutive brackets share a tone which locates at the third tone position of the preceding bracket as well as the first tone position of the subsequent bracket. The way the parser segments a sub-sequence of tones into a sequence of brackets is categorized into four cases according to the number k of extended symbols in the sub-sequence:

1) $k = 1$,
2) $k = 2$,
3) $k = 2n + 1$ $(n \geq 1)$
4) $k = 2n + 2$ $(n \geq 1)$.

When the sub-sequence consists of only one or two tones, then the parser segments the tones into a single bracket, where the symbols “monad” or “dyad” are to be annotated. Otherwise, the sub-sequence is segmented into a sequence of 3 consecutive tones (case 3) or a sequence of 3 consecutive tones plus 2 consecutive tones (case 4).

Next, to each bracket consisting of three consecutive tones, the parser annotates one of the 15 extended symbols listed in Fig. 5 according to the definition of those extended symbols in Fig. 5. Here, we follow the examples in the original IRM literature [1], [2] and define “large interval” as where the interval is larger than or equal to 6 degree, while “small interval” as where it is smaller than or equal to 5 degree. In the evaluation of this paper, we obtained the best performance with this definition among other definitions of “large”/“small” intervals.

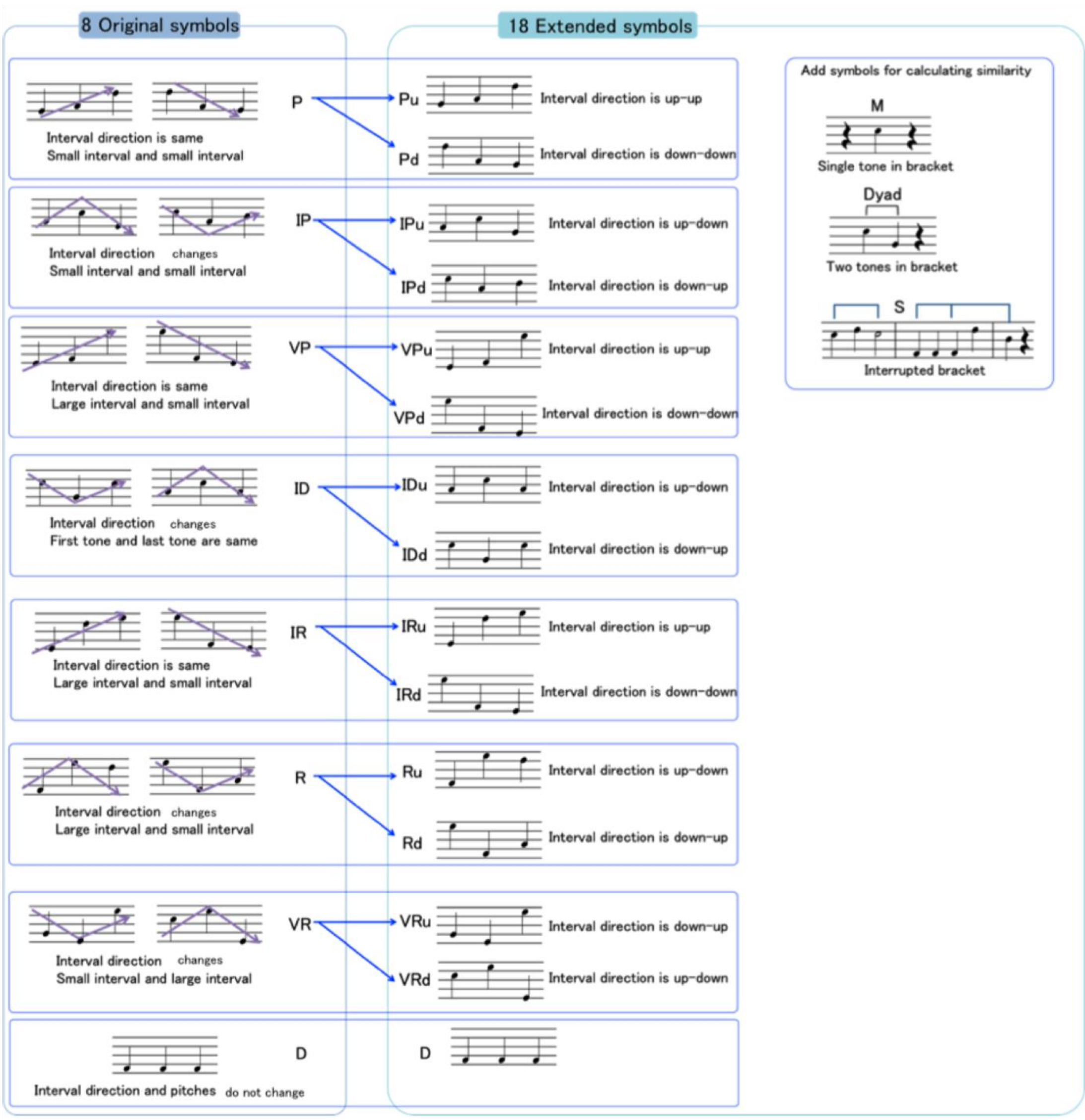

Figure 5. 18 symbols of the extended IRM (including seven symbols of the original IRM divided into 14 symbols).

IV. SIMILARITY OF EXTENDED IRM SYMBOL SEQUENCES

This section describes how we calculate the similarity between sequences of symbols of extended IRM as well as those of the original IRM. First, let Seq_1 and Seq_2 be the two sequences of symbols of the extended IRM or those of the original IRM. We measure the similarity of Seq_1 and Seq_2 by a python library for calculating n-gram similarity of two strings, which is available at https://github.com/gpoulter/python-ngram. Let N be the length of the fragmental sub-sequence considered in this procedure of calculating the similarity. Then, this library measures the similarity of Seq_1 and of Seq_1 and Seq_2 by counting the number of fragmental sub-sequence of symbols which has the length N and is shared by the two sequences. Let M be the number of fragmental sub-sequences of symbols that are shared by the two sequences. The following formula gives the definition of the similarity Sim of two sequences Seq_1 and Seq_2 of symbols:

$$Sim(Seq_1, Seq_2) = M / A$$
$$A = (|Seq_1|+2(N-1)) + (|Seq_2|+2(N-1)) - 2N - M + 2$$

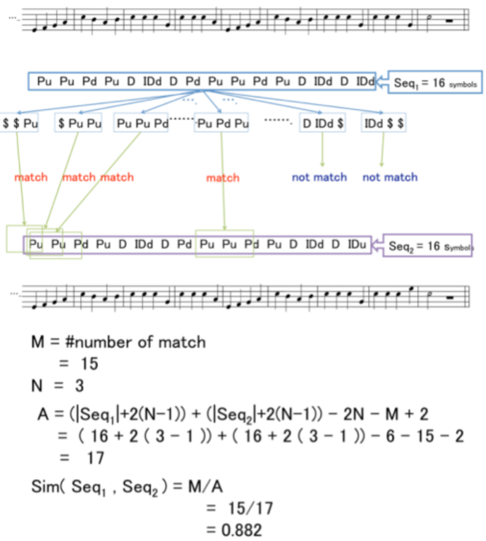

Figure 6. Procedure of calculating similarity between sequences of symbols of extended IRM (1).

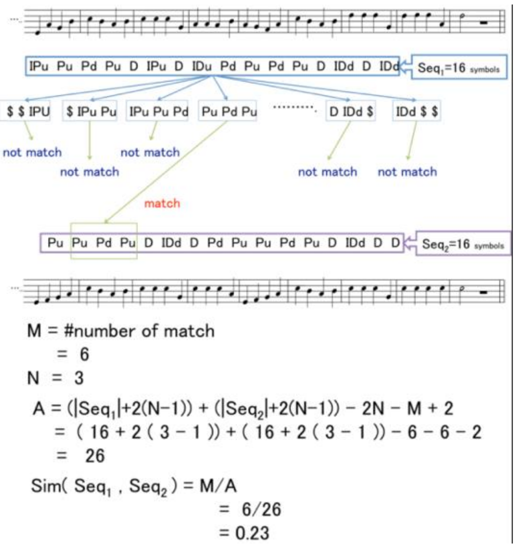

Figure 7. Procedure of calculating similarity between sequences of symbols of extended IRM (2).

TABLE I. EXAMPLES OF CALCULATING SIMILARITIES BETWEEN SEQUENCES OF SYMBOLS OF EXTENDED IRM (RED SYMBOLS ARE DIFFERENCES OF *SEQ*$_1$ AND *SEQ*$_2$)

Sim(Seq_1,Seq_2)		Extended Symbols
0.8951	Seq_1	Rd IDd IDd dyad Ru Pd monad ... Ru Pd IDd Pd Pu ... Pd Pu IDu IPu Pu Vpu Ru IPd
	Seq_2	Rd IDd IDd dyad Ru Pd monad ... Ru Pd Pd IDd Pd Pu ... Pd Pu IDu IPu Pu Vpu Ru IPd
0.8224	Seq_1	Ru Pd IDd Pd Pu IPu dyad .... Pu Pd IDd Pd .... IPu Pu dyad IPd Pd
	Seq_2	Pd IDd Pd Pu IPu dyad VRu .... Pd Pd IDd Pd .... IPu Pu dyad IPd Pd
0.6941	Seq_1	D IPd Pd monad IPd Pd ... IDd Pd IPd Pd IDd Pd ... IDd Pd Pu IPu Pu IPu Pd Pu IDu
	Seq_2	D IPd Pd monad monad Pd ...IDd Pd IPd Pd IDd Pd ...IDd Pd Pu Pd IPu IPu Pd Pu IDu
0.6058	Seq_1	...Pu IPu dyad VPu ...Pd IDu Pd Pu ... dyad Pd D D IPu Pd IDd Pd Pu monad IPd ...
	Seq_2	...Pu IPu dyad VPu ...Pd dyad Pu Pu ... dyad Pu D D IPu Pd IDd Pd Pu monad VPd ...
0.5174	Seq_1	D D monad IRd .... D D IRd Ru .... Pu D D IDd Pd ...
	Seq_2	D D monad IRd .... D dyad D D .... Pu Pd Pu monad D ...

Fig. 6 and Fig. 7 illustrate the procedure of calculating the similarity between two sequences of symbols.

For each of the similarity values around 0.5, 0.6, 0.7, 0.8, and 0.9, Table I also lists examples of calculating similarities between sequences of symbols. Those are actual examples of calculating similarities between melodies used in the evaluation of this paper.

V. EVALUATION

A. Evaluation Procedure

We evaluate the similarity measure using the 5,000 songs extracted from the Essen Folksong database[1]. For each of the 5,000 songs, the IRM parser for the extended symbols parses the sequence of tones into the sequence of the extended IRM symbols.

Since one of the major objectives of this paper is to examine the correlation between melody length and subjective human judgments on melodic similarities, we first restrict the number of constituent symbols of the songs for evaluation within the range from 4 to 90. Then, we have 3,159 songs out of the total 5,000, where their distribution of the number of symbols constituting each song is as shown in Fig. 8. Next, we divide the set of those 3,159 songs into the following nine subsets according to the number of symbols constituting each song:

(a) songs constituting 4∼10 symbols
(b) songs constituting 11∼20 symbols
(c) songs constituting 21∼30 symbols
(d) songs constituting 31∼40 symbols
(e) songs constituting 41∼50 symbols
(f) songs constituting 51∼60 symbols
(g) songs constituting 61∼70 symbols
(h) songs constituting 71∼80 symbols
(i) songs constituting 81∼90 symbols

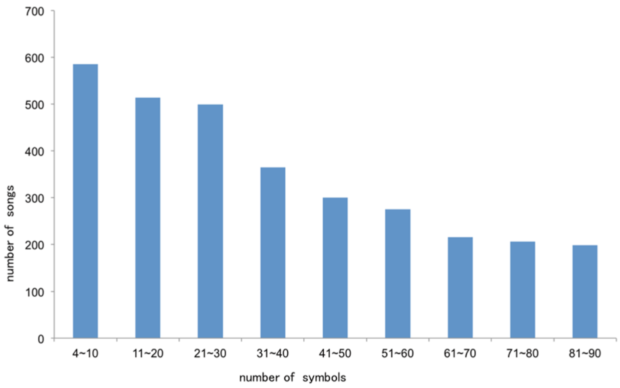

Figure 8. Distribution of the Number of Symbols Constituting Each Song.

Then, within each subset, we calculate similarities of all the pair of two songs. Note here that we calculate similarities of the pair of two songs whose lengths are only within the same range. And, for each subset, we consider the following (at most) nine similarity ranges:[2]

$$0.87 + 0.02i < \text{similarity} \leq 0.87 + 0.02\,(i + 1)$$
$$(i = 0, \ldots, 3)$$

$$0.95 + 0.01i < \text{similarity} \leq 0.95 + 0.01\,(i + 1)$$
$$(i = 0, \ldots, 4)$$

Then, for each subset and for each of the total nine similarity ranges, we randomly pickup five pair of two songs and (at most) 45 pairs in total for subjective human judgment 3. Fourteen subjects participated in subjective human judgments evaluation. Each subject listened to all the pairs of two songs in random order without duplication. Every time he/she listened to a pair of two songs, he/she was asked as how similar were the two songs and was requested to rank with a 5-point scale; 5-point: very similar, 4-point: similar, 3-point: neutral, 2-point: different, 1-point: very different.

B. Evaluation Results

For the total nine subsets (a) to (i), each of Fig. 9 to Fig. 13 compares the results of subjective human judgments among the (at most) nine similarity ranges. Within each subset, by comparing the results of subjective human judgments among (at most) nine similarity ranges, it is clear that the similarity measure has a certain correlation with subjective human judgments: the higher the similarity value is, the higher the score of subjective human judgment is. Furthermore, by comparing the results of subjective human judgments among the nine subsets (i.e., Fig. 9 to Fig. 13), it is also clear that the smaller the number of constituent symbols is, the higher the score of subjective human judgment is. In other words, the smaller the melody length is, the more similar the pair of melodies is judged to be by human subjects, even though their similarities based on the extended IRM are measured to be almost in the same range. From this subjective evaluation result, we conclude that, as future work, it is necessary to invent a much finer-grained similarity measure, which is designed to be more sensitive to melody length and to be coinciding with the results of subjective human judgments.

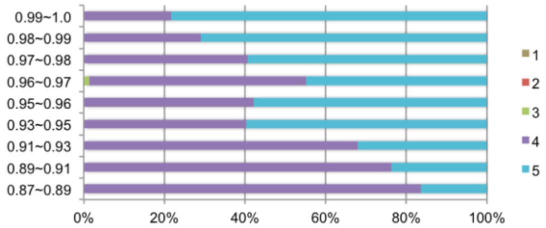

Figure 9. Results of subjective human judgments (number of constituent symbols: 4~10).

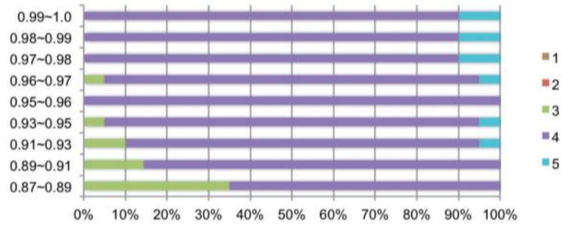

Figure 10. Results of subjective human judgments (number of symbols: 21~30).

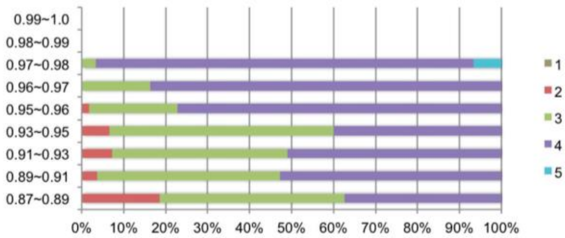

Figure 11. Results of subjective human judgments (number of symbols: 41~50).

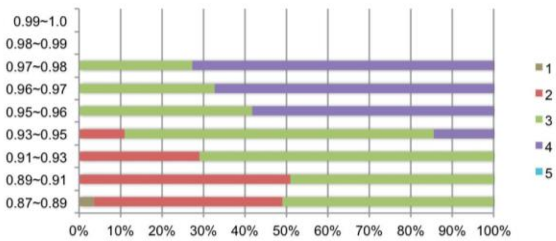

Figure 12. Results of subjective human judgments (number of constituent symbols: 61~70).

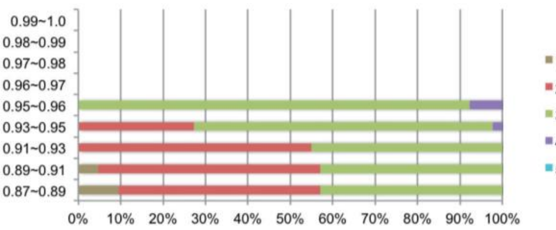

Figure 13. Results of subjective human judgments (number of constituent symbols: 81~90).

VI. Conclusion

This paper studied to measure similarity of melodies based on Implication-Realization Model (IRM), a music theory that abstracts music and then expresses music through symbol sequences based on information constituting the music such as pitch, rhythm, and rests and so on. The model employed in this paper is an extended one that is much more appropriate to measuring similarity of melodies. Then, this paper successfully claimed that the similarity measure studied in this paper has a certain correlation with subjective human judgments on melodic similarities. Furthermore, this paper examined the correlation between melody length and subjective human judgments on melodic similarities. It is quite remarkable to note the result of this analysis: i.e., the smaller the melody length is, the more similar the pair of melodies is judged to be by human subjects, even though their similarities based on the extended IRM are measured to be almost in the same range. We finally concluded that it is necessary to invent a much finer-grained similarity measure which is designed to be more sensitive to melody length and to be coincide with the results of subjective human judgments.

Future work includes applying the similarity measure to melodies of genres other than folk songs. And, based on those experimental evaluation results, it is definitely necessary to design a much-finer grained similarity measure that is sensitive to melody length. Another future work includes further extending the extended IRM symbols studied in this paper by incorporating long distance structures such as typical musical structures. We are also planning to invent a machine learning based IRM parser, which overcomes limitation of the rule-based IRM parser, employed in this paper.

References

[1] E. Narmour, *The Analysis and Cognition of Basic Melodic Structures*, The University of Chicago Press, 1990.

[2] E. Narmour, *The Analysis and Cognition of Melodic Complexity*, The University of Chicago Press, 1992.

[3] E. Pampalk, "A Matlab toolbox to compute music similarity from audio," in *Proc. 10th International Conference on Digital Audio Effects*, 2001.

[4] B. Logan and A. Salomon, "A music similarity function based on signal analysis," in *Proc. IEEE International Conference on Multimedia and Expo*, 2001.

[5] G. Velardea, T. Weydeb, and D. Mereditha, "An approach to melodic segmentation and classification based on filtering with the Haar- Wavelet," *Journal of New Music Research*, vol. 42, pp. 325-345, 2013.

[6] M. Grachten, J. L. Arcos, and R. L. D. Mantaras, "Melody retrieval using the implication/realization model," in *Proc. MIREX 2005 Symbolic Melodic Similarity Results*, 2005.

[7] S. Yazawa, M. Hamanaka, and T. Utsuro, "Evaluating melodic similarity based on an extended implication-realization model," in *Proc. 29th Annual Conference of JSAI*, 2015.

[8] K. Hoashi, K. Matsumoto, and N. Inoue, "Personalization of user content-based music retrieval on relevance feed back," in *Proc. ACM Multimedia*, 2003, pp. 110-119.

[9] F. Vignoli and S. Pauws, "A music retrieval system based on user-driven similarity and its evaluation," in *Proc. International Society of Music Information Retrieval*, 2005.

[10] S. Lampropoulos, D. N. Sotiropoulos, and G. A. Tsihrintzis, "Individualization of music similarity perception via feature subset selection," in *Proc. IEEE International Conference on System Man and Cybernetics*, 2004, pp. 552-556.

[11] M. F. McKinney and J. Breebaart, “Features for audio and music classification,” in *Proc. International Society of Music Information Retrieval*, 2003.
[12] E. Pampalk, A. Flexer, and G. Widmer, “Improvements of audio-based music similarity and genre classification,” in *Proc. International Society of Music Information Retrieval*, 2005, pp. 11-15.
[13] S. Doraisamy and S. Ruger, “Polyphonic music retrieval system using n-grams,” in *Proc. International Society of Music Information Retrieval*, 2004, pp. 204-230.
[14] F. Lerdahl and R. Jackendoff, *A Generative Theory of Tonal Music*, MIT Press, 1983.
[15] K. Hirata, T. Satoshi, and M. Hamanaka, “Fatta: Full automatic time-span tree analyzer,” in *Proc. International Computer Music Conference*, 2007, pp. 153-156.
[16] K. Hirata, T. Satoshi, and M. Hamanaka, “Cognitive similarity grounded by tree distance from the analysis of k. 265/300e,” in *Proc. CMMR*, 2013, pp. 415-430.

Sakurako Yazawa received her B.E. and M.E. degrees in engineering from University of Tsukuba in 2011 and in 2014. She is now a student of doctor course in Department of Intelligent Interaction Technologies, Graduate School of Systems and Information Engineering, University of Tsukuba. She is a member of the Society for Music Perception and Cognition and is a student member of the Information Processing Society of Japan (IPSJ) and the Japan Society for Artificial Intelligence (JSAI).

Masatoshi Hamanaka received his D.Eng. degree from University of Tsukuba in 2003. He joined Graduate School of Systems and Information Engineering, University of Tsukuba as an assistant professor in 2007, and clinical research center in Kyoto University Hospital since 2014. His research interest is in medical science, drug discovery and music information technology. He received the Journal of New Music Research Distinguished Paper Award in 2005.

Takehito Utsuro received his B.E., M.E., and D.Eng. degrees in electrical engineering from Kyoto University in 1989, 1991, and 1994. He has been a professor at the Division of Intelligent Interaction Technologies, Faculty of Engineering, Information and Systems, University of Tsukuba, since 2012. His professional interests are in natural language processing, Web intelligence, information retrieval, machine learning, spoken language processing, and artificial intelligence. He is a member of the Association for Computational Linguistics (ACL) the Institute of Electronics, Information and Communication Engineers (IEICE), Information Processing Society of Japan (IPSJ), the Japan Society for Artificial Intelligence (JSAI), and the Acoustical Society of Japan.