Three Dimensional Continuous DP Algorithm for Multiple Pitch Candidates in Music Information Retrieval System

Sung-Phil HEO† **Motoyuki SUZUKI**‡ **Akinori ITO**‡ **Shozo MAKINO**‡
†Graduate School of Information Sciences Tohoku University
‡Graduate School of Engineering Tohoku University
05 Aza-Aoba, Aramaki, Aoba-ku Sendai, 980-8579 Japan
{hsphil, moto, aito, makino}@makino.ecei.tohoku.ac.jp

Abstract

This paper treats theoretical and practical issues that implement a music information retrieval system based on query by humming. In order to extract accuracy features from the user's humming, we propose a new retrieval method based on multiple pitch candidates. Extracted multiple pitches have shown to be very important parameters in determining melodic similarity, but it is also clear that the confidence measures feature which are obtained from the power are important as well. Furthermore, we propose extending the traditional DP algorithm to three dimensions so that multiple pitch candidates can be treated. Simultaneously, at the melody representation technique, we propose the DP paths are changed dynamically to be able to take relative values so that they can respond to insert or omit notes.

1 Introduction

In researching MIRS (Music Information Retrieval System), retrieval keys have been made to not only using text information, such as a singer's name, the composer, the title of the piece of music, or the lyrics of a song, but also the use of music information has been accomplished.

There are several important issues in building MIRS. One is that an input query may have various errors due to uncertainty of the user's memory or the user's singing ability. Next problem, it is still difficult to implement a 100% accurate system for transcribing the hummed signals into musical symbols, which are needed for melody matching in the next step (Kosugi, N. *et al.* 2000, Pauws, S. 2002).To tolerate these errors, we need an effective representation of the hummed melody and a musically reasonable approximate matching method. Therefore, we can consider the following problems: the feature extraction problem such as harmonic frequencies was extracted, the melody representation problem when insertion or omission of notes occurred and the melody matching problem connected with such problems. We believe the above problems are the key points for complete, robust and efficient MIRS.

2 Extraction of multiple pitch candidates

Pitch is a perception that is defined as the characteristic of a sound that gives the sensation of being high or being low. A pitch extraction algorithm is applied to the humming input, but it often captures incorrect frequency.

The accuracy of the pitch extraction greatly affects the system's performance. Therefore, we consider multiple pitch candidates to enhance the performance of retrieval.

The pitch extraction accuracy was calculated by comparing the true pitch value with the extraction result. The references were labeled by human. The pitch extraction accuracies when 1st rank, within the 2nd rank and within the 3rd rank obtained 88.4%, 96.3% and 99.7%, respectively. Harmonic frequencies (double or half pitch of the true pitch) were extracted at the most frames with incorrect results. This result shows that three pitch candidates are enough for the following processing.

The pitch extraction is based on cepstral analysis. Multiple pitch candidates are passed to the query engine without choosing one candidate in the feature extraction part. The confidence measure is calculated as the cepstral value of the peak divided by that of the top candidate.

3 Melody representation

When a user inputs humming, its key and tempo may differ from that of the corresponding music in the database. Therefore humming data needs to be normalized. The traditional melody representation method normalizes the humming data by calculating relative pitch ratio and relative span ratio against the successive note (Sonoda, T. *et al.* 1998). However there are several problems in the traditional melody representation method. When insertion or omission of notes occurs, relative values are changed.

Therefore we propose a new melody representation strategy to treat this kind of mismatch in Figure 1. When a note is matched in an assumption that the previous note in humming is separation, the relative pitch of the current note in humming should be calculated against the second note before the current note.

This means deciding dynamically the relative value to match. Besides in omission, the determination method of the relative pitches is dynamically decided by the same manner.

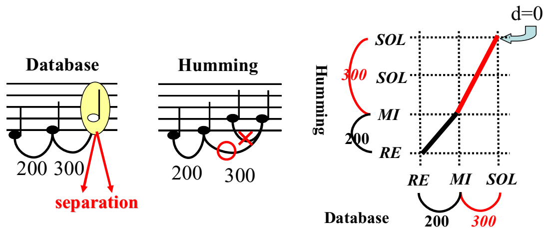

Figure 1: Relative pitch values conversion method when separation is occurred. (Unit is cent)

4 Three dimensional continuous DP algorithm

The features obtained from humming and the features of the musical piece in the database are matched using continuous DP (Paulus, J. *et al.* 2002). However, the query engine must be extended so that multiple pitch candidates along with confidence measure can be utilized. Furthermore, the DP equation must be changed. The relative notes should be dynamically calculated considering the insert or omit of notes. Therefore, the DP algorithm is extended into three dimensions shown as equation (1). Here, $g(i,j[k])$ is the accumulation distance of the *k-th* pitch candidate value in the *j-th* humming note and the *i-th* musical piece note.

$$g(i,j[k]) = \min\begin{cases} \min_l\{g(i-2,j-1[l]) + d_1(i,j[k],j[l])\} \\ \min_l\{g(i-1,j-1[l]) + d_2(i,j[k],j[l])\} \\ \min_l\{g(i-1,j-2[l]) + 2d_3(i,j[k],j[l])\}. \end{cases} \quad (1)$$

$$d_e(x,y,z) = \beta\{\alpha p_e(x,y,z) + (1-\alpha)c_e^{-1}(y,z)\} + (1-\beta)t_e(x,y). \quad (2)$$

This score is a weighted sum of pitches with confidence measure and span scores, $p_e(x, y, z)$, $c_e(y, z)$ and $t_e(x, y)$, the former being the weight assigned for a particular distance in pitch, the next being the confidence measure obtained from the power value, and the last being the weight assigned for the distance in duration. Moreover, $d_e (e=1,2,3)$ equation (2) corresponds to local path constraint, respectively (Heo, Sung-Phil *et al.* 2003). The factors α, β can be varied to reflect the relative contribution of pitch, confidence measure and span.

The humming is matched with the database at the DP plane extended to three dimensions. Therefore the mechanism of the matching algorithm extended to three dimensions. When considering matching the humming to the database, the proposed algorithm calculates the combination in all candidate points. Finally, the algorithm will determine the optimal candidate points and paths.

5 Experimental conditions and results

The music database consists of children's songs, and 155 pieces of music from Japan and foreign countries. We also used the query corpus which is 5 subjects hummed, totally 320 queries. All subjects were inexperienced singer. They were allowed to use a headset microphone and start humming any part of a song with free key and free tempo.

Evaluation measure is the most important factor to evaluate the performance of a retrieval method. The performances of the systems are compared using '*retrieval accuracy*'.

Table 1: Comparisons of the accuracy by various features (%)

Features	1st rank	Ranked in the top 10	Weight values
Span + Pitch	73.9	91.1	α=1.0, β=0.6
Coarse-to-Fine	78.4	89.6	none
Category 27	81.6	91.5	none
Span + Pitch(3) + CM	86.5	94.1	α=0.5, β=0.7

The experimental results for music retrieval are shown in Table 1. To a performance evaluation here, the Coarse-to-Fine and Category 27 method are used (Sonoda, T. *et al.* 1998).

Pitch *(m)* refers to the use of *m* multiple pitch candidates. *CM* stands for confidence measure. By using multiple pitch candidates, especially the accuracy of the 1st rank was improved, and thus the validity could be conformed. Retrieval accuracies are improved by using multiple pitch candidates. When we used three pitch candidates the first rank was 86.5%.

Even if we compared Category 27 and the proposed method (Span + Pitch (3) + CM), the retrieval accuracy is improved from 81.6% to 86.5% at the 1st rank.

6 Conclusion

In this paper, we implemented a practical query-by-humming system, which can find a piece of music in the database based on a few hummed notes. Moreover, the retrieval engine was extended by the three dimensional continuous dynamic programming so that multiple pitch candidates could be treated. Actually, the influence by individual character, such as a difference in key and a difference in tempo, an error like insertion or omission of notes, etc. are contained in a user's humming. Considering these problems, we represented the notes by taking relative ratio against the optimum neighbor note dynamically determined according to the DP paths.

Furthermore, even if the hummed queries are perfect, it is still difficult to retrieve the pitch perfectly from the hummed queries. To consider the pitch extraction errors, we proposed to use multiple pitch candidates. Using the proposed method, 99.7% of pitch extraction accuracy was shown within the third rank. Moreover the similarity measuring algorithm extended the search space of DP plane into three dimensions for robust matching against pitch extraction errors in the query processing.

References

Heo, Sung-Phil *et al.* (2003). Multiple Pitch Candidate based Music Information Retrieval Method for Query-by-Humming. *In International Workshop on AMR2003*.

Kosugi, N. *et al.* (2000). A Practical Query-By-Humming System for a Large Music Database. *ACM Multimedia 2000*, (pp. 333-342).

Pauws. S. (2002). CubyHum: A Fully Operational Query by Humming System. *In Proc. 3rd ISMIR2002*.

Paulus, J. *et al.* (2002). Measuring the Similarity of Rhythmic Patterns. *In Proc. 3rd ISMIR2002*.

Sonoda, T. *et al.*(1998). WWW-based Music Retrieval System. *In Proc. ICMC'98* (pp. 343-352).