CONTOUR STRING MATCHING FOR QUERY BY TAPPING: A MIREX 2014 SUBMISSION

Jorge Herrera **Hyung-Suk Kim** **Blair Kaneshiro**
CCRMA
Stanford University
{jorgeh,hskim08,blairbo}@ccrma.stanford.edu

ABSTRACT

Query-by-tapping systems aim to retrieve a song based upon a rhythmic query tapped by the user. Here we apply Rhythmic Contour String encoding to rhythm and pitch dimensions of tapped queries and ground-truth excerpts to match queries with candidate ground truths.

1. INTRODUCTION

Query by tapping (QBT) is the process by which a tapped representation of a musical excerpt, usually the rhythm, is matched with ground-truth candidates. While conceptually similar to query by humming (QBH), QBT systems have historically been more difficult to implement successfully, perhaps due to the inability for rhythm alone to uniquely characterize an excerpt. Recently, a new QBT dataset incorporating both rhythm and pitch information into the tapped queries has been published [1].

2. ALGORITHM

The current algorithm uses Rhythmic Contour String encoding [2], as described in our paper accompanying the release of the QBT-Extended dataset [1].

2.1 String Encoding

Each ground-truth MIDI file and tapped query is encoded separately in the following manner:

1. Derive vector of inter-onset intervals (IOIs), X_{IOI}, by computing successive pairwise differences of elements in X, the vector of onset times;
2. Compute μ_{IOI}, mean IOI, by dividing time of last tap or note onset by total number of onsets;
3. Compute $X_\mu = X_{IOI}/\mu_{IOI}$;
4. Derive normalized IOI transitions, X_T by taking successive pairwise differences of elements in X_μ.

These entries were then mapped to “down”, “same”, or “up” encodings using a threshold of 0.2, such that element i of X_E is defined as follows:

$$X_E^i = \begin{cases} \text{D}, & \text{if } X_T^i < -threshold \\ \text{S}, & \text{if } -threshold \leq X_T^i \leq threshold \\ \text{U}, & \text{if } X_T^i > threshold \end{cases}$$

An excerpt containing N tap or note onsets thus produces an IOI vector of length $N-1$, and an IOI transition vector of length $N-2$.

2.2 String Matching

We used Levenshtein distance [3] between strings, computed by the approximate matching procedure [2], to match queries with ground truths. For a given query, distances were computed over the full set of candidate ground truths; results are returned in decreasing order of closeness (increasing order of distance).

2.3 Two-Dimensional Queries

Encoding of the pitch dimension (for QBT-Extended queries) was performed in a manner similar to the rhythm dimension:

1. Derive inter-onset positions Y_{IOI} by computing successive pairwise differences of elements in Y, the vector of pitch positions;
2. Compute the normalized difference vector $Y_N = Y_{IOI}/\max(\text{abs}(Y_{IOI}))$, so that difference values fall between -1 and 1.

Following this, we encoded the length-$(N-1)$ pitch contour to a string as described in §2.2, using a threshold of 0.1. The string-matching precedure was implemented by calculating the Levenshtein distance for both the rhythmic contour string and the pitch contour string, and then taking the L2 norm of the two values. Our current algorithm gives equal weighting to rhythm and pitch dimensions for two-dimensional queries.

3. DISCUSSION

We have used Rhythmic Contour String encoding as a first attempt to implement a working QBT system. Future implementations will explore other thresholding techniques within the current procedures, as well as alternate procedures. As the QBT-Extended dataset also includes annotations of user familiarity and ability, we may also refine relative pitch and vector weightings for queries from this dataset according to these annotations.

4. REFERENCES

[1] Blair Kaneshiro, Hyung-Suk Kim, Jorge Herrera, Jieun Oh, Jonathan Berger, and Malcolm Slaney. QBT-Extended: An annotated dataset of melodically contoured tapped queries. In *Proceedings of the 14th International Society for Music Information Retrieval Conference*, pages 329–334, 2013.

[2] Geoffrey Peters, Diana Cukierman, Caroline Anthony, and Michael Schwartz. Online music search by tapping. In *Ambient Intelligence in Everyday Life*, pages 178–197. Springer, 2006.

[3] Vladimir I Levenshtein. Binary codes capable of correcting deletions, insertions, and reversals. *Soviet Physics Doklady*, 10(8):707–710, 1966.