MIREX 2009: A THREE STAGES SCHEME FOR QUERY BY TAPPING

Shu-Jen Show Hsiao
Department of Computer science
National Chiao Tung University
Hsinchu, Taiwan ROC
show.cs95g@nctu.edu.tw

Tyne Liang
Department of Computer science
National Chiao Tung University
Hsinchu, Taiwan ROC
tliang@cis.nctu.edu.tw

ABSTRACT

Tapping is the intuitive way for people to represent music clip in mind, therefore Query by Tapping (QBT) system mechanism developed for years. We describe three kinds of rhythmic variations arise all over the query rhythm. This extended abstract presents three stages scheme (pre-processing, normalization and comparison) for onset sequence similarity measurement; each stage can efficiently decrease these variations respectively.

1 RHYTHMIC VARIATION

Onset sequence is the timestamp of each onset relative to the fist onset. Three kinds of rhythmic variation inaccuracies exist all over the query onset sequence. (1) *Inter-Onset-Interval Outlier Variation*: According to experiment people usually could not grasp well for long interval between onsets. The outlier possibly occurs in query rhythm when target rhythm cease onset for period of time. (2) *Tempo Variation*: Though the query and target express the same rhythm, the sequence may induce difference with nearly constant ratio because of playing in different tempo. (3) *Less or More Onset Variation*: Due to careless taping in query or noise note in target rhythm, the onset number of these two sequences may not be the same.

2 THREE STAGES SCHEMED FOR QBT

2.1 Pre-processing

Less research discussed this optional preprocessing step for onset sequence in QBT system; however the quality data can get quality retrieval result. This study applies data smoothing function to reduce *IOI Outlier Variation* effect. Nevertheless, this scheme must also consider the drawback of loosing the rhythmic distinction in collection by over smoothing.

In statistics, three-sigma rule, states that for a normal distribution, almost all values (99.7%) lie within 3 standard deviations of the mean. Given an IOIs sequence of length n: $IOIs\_sequence = \{ioi_k\}$, *k=1, 2…, n,*

It is still undetermined whether the rhythmic onset interval value is normal distribution or not, hence a Smoothing Threshold (ST) is defined:

$$ST = \mu + n * \sigma \quad (1)$$

where μ and σ is mean and standard deviation of IOIs sequence respectively. If inter onset interval ioi_k is greater than ST, suppose to be an outlier, the exceeding segment $(ioi_k - ST)$ will multiply by a reduction factor ε ($0 < \varepsilon \le 1$), larger the segment is, smaller ε is. The smoothing function is therefore computed as follows:

$$ioi_k^{'} = ST + \varepsilon * (ioi_k - ST) \quad \text{if } ioi_k \ge ST$$
$$= ioi_k \quad \text{if } ioi_k < ST \quad (2)$$

$$\varepsilon = \left(1 - \frac{ioi_k - ST}{ioi_k}\right) \qquad 0 < \varepsilon \le 1 \quad (3)$$

2.2 Normalization

Due to the nature of *tempo variation* in onset sequence, normalization techniques have to transform these onset values into common domain before similarity measurement. A good normalization scheme must be *robust* and *efficient*. Robustness refers to the insensitivity to the presence of outliers. Efficiency refers to the proximity of the obtained estimate to the optimal estimate when the distribution of the data is known [1]. This paper used SUM normalization to reduce tempo variation effect. The SUM normalization converts the summation of all ioi_k to 1; hence the normalization scores s_k are given by:

$$s_k = \frac{ioi_k}{\sum_{k=1}^{n} ioi_k} \quad (4)$$

2.3 Comparison

The *Less or More Onset Variation* effect occurs for several reasons. For examples, noise onset extracted from MIDI or careless tapping by query rhythm. Traditional linear one-to-one correspondence, Euclidean-base, similarity distance measure is inadequate for the unequal length sequence similarity measurement. Both Dynamic Time Warping (DTW) and Earth Mover's Distance (EMD) support variety sequence length similarity distance measurement and is extensively used in previous research. This submission adopt the assumption as Roger described in [2] that user is likely to lose onsets instead of gaining onset when generating query onset sequence. Assume length of query sequence is *n*; it compares with the first *n-1, n, n+1, n+2, and n+3* onset of reference sequence. Therefore every comparison takes five itera-

tions to compare with different reference sequence length, return the best result as similarity distance.

3 REFERENCES

[1] P.J. Huber: *Robust Statistics*, Wiley, New York, 1981

[2] JS Roger Jang, Hong-Ru Lee, and Chia-Hui Yeh: "Query by Tapping: A New Paradigm for Content-based Music Retrieval from Acoustic Input," *The Second IEEE Pacific-Rim Conference on Multimedia*, Beijing, China, 2001.