IDENTIFYING EMOTION SEGMENTS IN MUSIC BY DISCOVERING MOTIFS IN PHYSIOLOGICAL DATA

Rafael Cabredo, Roberto Legaspi, Masayuki Numao
The Institute of Scientific and Industrial Research, Osaka University
{cabredo,roberto,numao}@ai.sanken.osaka-u.ac.jp

ABSTRACT

Music can induce different emotions in people. We propose a system that can identify music segments which induce specific emotions from the listener. The work involves building a knowledge base with mappings between affective states (happiness, sadness, etc.) and music features (rhythm, chord progression, etc.). Building this knowledge base requires background knowledge from music and emotions psychology. Psychophysiological responses of a user, particularly, the blood volume pulse, are taken while he listens to music. These signals are analyzed and mapped to various musical features of the songs he listened to. A motif discovery algorithm used in data mining is adapted to analyze signals of physiological data. Motif discovery finds patterns in the data that indicate points of interest in the music. The different motifs are stored in a library of patterns and used to identify other songs that have similar musical content. Results show that motifs selected have similar chord progressions. Some of which include frequently used chords in western pop music.

1. INTRODUCTION

Music has become a ubiquitous form of entertainment. People listen to music in various situations: while travelling, doing sports, studying, or relaxing. Music structure and features can be used to select music appropriate to the emotional interest of its listeners. This has been researched in various fields like music and emotion psychology, music information retrieval, and more recently affective computing.

Automatically detecting the emotion or mood content of music is still in its early stages. Some of the work involve manually annotating songs with emotion tags by individual human annotators [16], social tagging [13], and even using games to make the task more interesting for annotators [10].

Human assessment of music emotion or mood is based from what is heard. As such, a lot of work is devoted to understanding how various music features and music structure play a role in inducing emotion. A detailed review of these works can be found in [4, 8, 11].

The work of Livingstone, et al. [11] also demonstrates that by changing specific music elements, the emotion perceived by the listener also changes. A similar research is also done in [14] but instead of relying on verbal reports of feelings, emotion data is derived from analyzing change in activity in the autonomic nervous system.

Another approach to identifying emotion is using psychophysiological data. Researchers observed that changes in musical features lead to a change in physchophysiological response. For example, change in tempo lead to changes in respiration rate [3, 7]. Krumhansl [9] also noted increases in heart rate variability during sad, fearful and happy music. The use of physiological response also reflect an unbiased, objective emotional response to music listening as compared to self-reporting of emotions.

In this paper, we propose an approach for identifying music features that affect emotion. We identify patterns in psychophysiological data using a motif discovery algorithm and analyze the music elements used at the time the patterns were discovered.

We begin by defining some concepts and notations important for understanding the approach used. In section 3 and 4, we describe the framework used for the research. In section 5, we describe details of data collection and implementation of the algorithms presented. Next, results of our experiments are discussed together with observations made. Final section includes the conclusion and future work.

2. TIME SERIES MOTIFS

For clarity, first we define concepts and terminology needed to understand our work. These definitions are taken from [2]. The physiological signals are a continuous stream of real-valued data measured at a constant sampling rate. In data mining, this can be considered a *time series*. A time series T is defined as an ordered set of real-valued variables.

A *motif* is described as a pair of subsequences from the

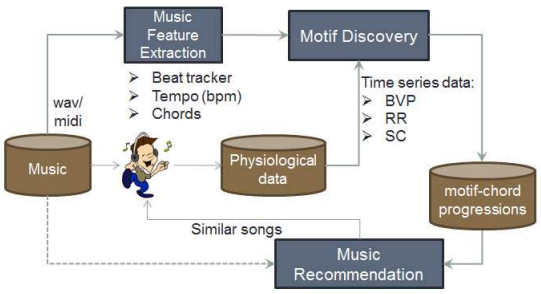

Figure 1. Architectural framework

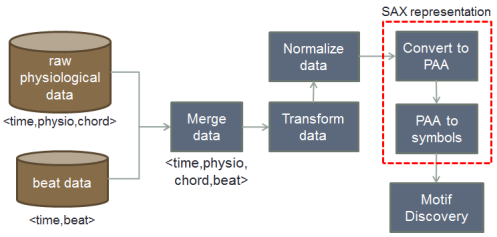

Figure 2. Data flow diagram for motif discovery

time series that are found to be similar. A *subsequence* C is a sampling of length n of contiguous positions in T.

Similarity between two subsequences is measured using a distance metric $D(C_i, C_k)$. It is possible to find many motifs in one time series, the most significant of these is referred to as *1-motif*. To ensure that the *1-motif* does not share elements with other motifs, a range R is defined such that $D(C_i, C_k) > 2R$, for all $1 \leq i < k$.

3. ARCHITECTURAL FRAMEWORK

The proposed framework of our system is shown in Figure 1. Our approach requires collecting psychophysiological data from a subject while he listens to music. We consider analyzing data from : blood volume pulse (BVP), respiration (RR), and skin conductance (SC). These are then passed on to a motif discovery module that attempts to discover patterns in the time series data. Details of this module are discussed in the next section.

A music feature extraction module is also included to determine various information from the music (i.e., beat occurrences, tempo, chords used, etc.). These are used by the motif discovery module to annotate discovered motifs.

Each motif is analyzed and annotated with music features that were present when the signal occurred. A library of different motifs is built and the data contained within is used by a music recommendation system that will generate a play list of songs that have similar music features. Intuitively, we expect that the subject will enjoy listening to music similar to that he has experienced.

This paper discusses the work done upto the motif discovery module using BVP data. The music recommendation system is currently being developed and will be described in future publications.

4. MOTIF DISCOVERY

The process of motif discovery is illustrated in Figure 2. This algorithm is adapted from the work in [2] where they used a projection algorithm by Buhler and Tompa [15]. The objective of the algorithm is to find signals that are very similar to each other. Physiological signals that keep on recurring would indicate that music passages heard at these points are interesting to the listener (i.e., it makes him relaxed, or he enjoys the music segment).

The motif discovery algorithm can be separated into 3 main parts: data preparation, conversion of the data to symbolic form using the Symbolic Aggregate ApproXimation (SAX) representation, and motif discovery using the projection algorithm. Each part is described in the following subsections.

4.1 Data preparation

Prior to motif discovery, the physiological data undergoes offset and amplitude scaling transformations using (1) and (2), respectively [1, 6, 17, 18].

$$Q_{offset} = Q - \frac{\sum_{i=1}^{n} q_i}{n}, \tag{1}$$

where Q is defined as a time series with n length and Q_{offset} is the time series after offset transformation.

$$Q_{scaled} = \frac{Q_{offset}}{\sigma}, \tag{2}$$

where σ is the standard deviation of the data and Q_{scaled} is the time series after amplitude scaling transformation.

In order to reduce further problems when comparing different subsequences, all data is normalized to the range [0,1] using (3).

$$Q = \frac{Q - min(Q)}{max(Q) - min(Q)} \tag{3}$$

4.2 SAX representation

The Symbolic Aggregate ApproXimation (SAX) representation is used to convert any time series into a string of symbols. By using SAX, powerful algorithms on string pattern analysis developed in other fields can be used. The first step is to convert the time series C of length n to a w-dimensional space by a vector $\bar{C} = \bar{c}_1, ..., \bar{c}_w$. The i^{th}

β_i \ a	3	4	5	6
β_1	−0.43	−0.67	−0.84	−0.97
β_2	0.43	0	−0.25	−0.43
β_3		0.67	0.25	0
β_4			0.84	0.43
β_5				0.97

Table 1. A lookup table containing breakpoints that divides a Gaussian distribution in an arbitrary number (from 3 to 6) of equiprobable regions

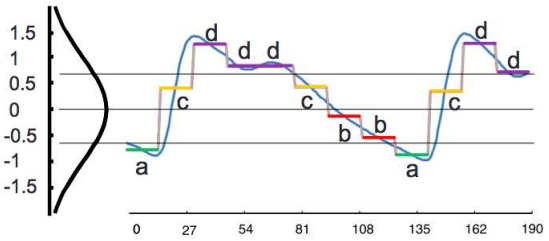

Figure 3. The physiological signal (thin smooth line) is discretized by first obtaining a PAA approximation and then using predetermined breakpoints to map the PAA coefficients into symbols (bold letters). In the example above, with $n = 190, w = 12$ and $a = 4$, the time series is mapped to the word **acddddcbbacdd**

element of $\bar{C}$ is calculated by the equation:

$$\bar{c}_i = \frac{w}{n} \sum_{j=\frac{n}{w}(i-1)+1}^{\frac{n}{w}i} c_j \quad (4)$$

Using this equation, the time series is divided into w equal sized frames. The average values of data in each frame is calculated and a dimensionality-reduced representation known as the Piecewise Aggregate Approximation (PAA) [5] is produced.

After transforming the time series into PAA representation, another transformation is applied to obtain the discrete representation. Assuming that the subsequences have a Gaussian distribution, we determine "breakpoints" that will produce equal-sized areas under the Gaussian curve. A *breakpoint* is a sorted list of numbers $B = \beta_1, ...\beta_{a-1}$ such that the area under a $N(0, 1)$ Gaussian curve from β_i to $\beta_{i+1} = 1/a$ (β_0 and β_a are defined as -∞ and ∞, respectively). a refers to the alphabet size used for SAX.

The breakpoints are stored in a look-up table similar to Table 1. Using the breakpoints, the time series can be discretized by going through each PAA coefficients. All coefficients below the smallest breakpoint are mapped to the symbol "**a**", all coefficients greater than or equal to the smallest breakpoint and less than the second smallest breakpoint are mapped to the symbol "**b**", etc. Figure 3 illustrates the idea.

The concatenation of symbols of the subsequence that is formed is defined as a *word*. Each PAA approximation is mapped to a symbol using Equation (5). a_i denotes the i^{th} element of the alphabet, i.e. a_1 = **a**, a_2 = **b**, etc.

	a	**b**	**c**	**d**
a	0	0	0.67	1.34
b	0	0	0	0.67
c	0.67	0	0	0
d	1.34	0.67	0	0

Table 2. A lookup table for MINDIST function. This table is for a SAX representation having $a = 4$. The distance can be obtained by matching the row and column. For example *dist*(**a**,**b**)= 0 and *dist*(**a**,**c**)= 0.67

$$\hat{c}_i = a_i \text{ iff } \beta_{j-1} \leq \bar{c}_i < \beta_j \quad (5)$$

The distance between two words can be measured by using a *MINDIST* function that returns the minimum distance between the original time series of the two words:

$$MINDIST(\hat{Q}, \hat{M}) \equiv \sqrt{\frac{n}{w}} \sqrt{\sum_{i=1}^{w} (dist(\hat{q}_i, \hat{m}_i))^2} \quad (6)$$

This function resembles the original Euclidean distance (7) used for comparing the distance between two time series Q and M. The function *MINDIST* uses a subfunction *dist*(), which can be implemented using a table lookup as illustrated in Table 2. The value in cell (r, c) for any lookup table can be calculated by the expression in (8).

$$D(Q, M) \equiv \sqrt{\sum_{i=1}^{n} (q_i - m_i)^2} \quad (7)$$

$$cell_{r,c} = \begin{cases} 0, \text{ if } |r - c| \leq 1 \\ \beta_{max(r,c)-1} - \beta_{min(r,c)}, \text{ otherwise} \end{cases} \quad (8)$$

4.3 Projection algorithm

The motif discovery algorithm proceeds by extracting subsequences from the SAX representation. Each subsequence of length w is placed into a matrix $\hat{S}$. Once the matrix has been constructed, we proceed to random projection. We randomly select $\frac{w}{2}$ columns of $\hat{S}$ to act as a mask. For example, given $w = 4$, columns {1,3} can be chosen to act as mask. Afterwards, all *words* in the $\hat{S}$ matrix are hashed into buckets based only on their values in the 1^{st} and 3^{rd} columns. If two words corresponding to subsequences i and j are hashed to the same bucket, we increase the count of cell (i, j) in a *collision matrix*.

This hashing process is repeated k times, with new, randomly chosen masks every iteration. Once completed, the highest value stored in the collision matrix correspond to

the candidate motif. For example, if the largest value in the collision matrix is at cell $(2, 43)$ then C_2 and C_{43} are the subsequences of the candidate motif. We confirm this by comparing the original time series data and using Euclidean distance to compute the distance.

At this point, it is possible to find other members of the motif. To find other members, we consider the other values of the collision matrix at $(i, 2)$ and $(i, 43)$. Once all the matching subsequences within R of C_2 and C_{43} have been found, results are reported to the user.

5. METHODOLOGY

5.1 Data Collection

For this research, we concentrate on analysing data from one subject (a 22-year male graduate student). The songs he listened to are part of the music dataset described in [12]. The collection includes 301 songs from various artists as well as annotations for song key, chords, beat and metric position, and segmentation (i.e. intro, verse, chorus, etc.). Songs for the experiments were selected based on three constraints. First, the song should not have any key and tempo changes. Second, the song should have complete chord and beat annotations. Last, the song is in a major key. Using this criteria, 83 songs were selected which include 77 songs from The Beatles, four Queen songs, and two Carole King songs.

Our subject listened to songs via audio-technica closed headphones (ATH-T400) connected to a computer in a controlled experiment room. Using the BioGraph Infinity System [1], the BVP was recorded. The sensor is attached to the subject as shown in the experiment setup in Figure 4.

Several sessions were needed for the subject to listen to all the songs without making him feel stressed. Each session took approximately 20 minutes, which allowed the subject to listen to seven to nine songs per session. One week was needed to complete the data collection. Sessions were held at the same time of the day throughout the week.

Before each session ended, the subject also self-reported the mood he had while listening to the songs. A scale of one to five was used to describe how happy and how exciting the song made him feel.

Although 83 songs were used for the data collection, only data from 64 songs are included for analysis for this experiment. Only songs that made the subject happy (i.e. songs rated three and above) and have a tempo between 76 – 168 beats per minute (bpm) are included. The tempo and key information of the music data set is shown in Table 3.

[1] About BioGraph Infinity System. Thought Technology Ltd. 14 May 2011. http://www.thoughttechnology.com

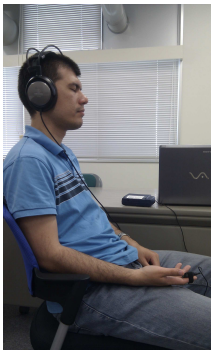

Figure 4. Data collection setup: BVP sensor worn on right index finger while listening to music via closed headphones

Key	Tempo			Total
	Andante	Moderato	Allegro	
C	1	1	3	5
D	1	1	7	9
E	3	3	8	14
F	2	1	2	5
F♯	0	0	1	1
G	5	2	3	10
A♭	1	0	0	1
A	5	4	5	14
B♭	1	0	1	2
B	1	1	1	3
Total	20	13	31	**64**

Andante : 76–108bpm Allegro : 120-168bpm
Moderato : 108–120bpm

Table 3. Summary of music included for motif discovery

5.2 Music feature extraction

Since the isophonics dataset already includes chord, beat, key and segment annotations for the different songs, only a simple text parser to read the different file annotations was needed. These annotations were manually done by music experts and students [12].

For the motif discovery, the physiological data is mapped to the chord information to determine what chord is being heard at that instance. The music features and the motif subsequences are stored in a file for cross-reference after motif discovery.

5.3 Motif discovery

All the 64 physiological readings were analyzed using three sets of parameters. Each set has varying sizes for motif length (n) and word size (w). However, all sets used an alphabet size of $a = 4$ and a range $R = 1.0$. The parameters used for each set are shown in Table 4. The motif length values were set as such to vary the chord progression length that was associated to a motif. The word size was adjusted to maintain a compression ratio of $\frac{n}{w} = 8$.

Set No.	*n*	*w*	Sequence length
1	1024	128	8 seconds
2	768	96	6 seconds
3	512	64	4 seconds

Table 4. Parameters used for the different sets

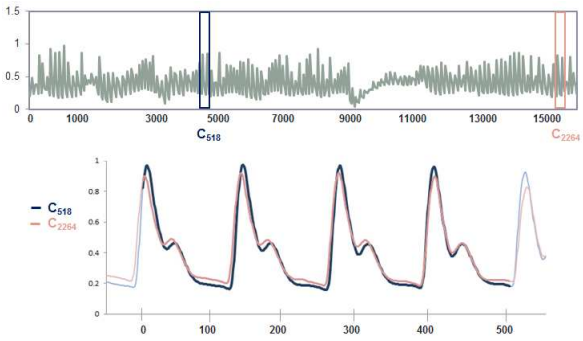

Figure 5. *(top)* The BVP signal of subject listening to *Please Mister Postman* from the Beatles has a motif of length 512 found as subsequence C_{518} and C_{2264}. *(bottom)* By overlaying the two motifs, we see the similarity of the two signals to each other.

6. RESULTS

Using the motif discovery algorithm, the most significant motif (*1-motif*) were obtained from the dataset. Figure 5 illustrates an example of a motif discovered.

We observe that the motif length is inversely proportional to the number of motifs found. Using set number 1 ($n = 1024$), for example, only the song *With A Little Help From My Friends* was identified to have a motif (see Figure 6). Analyzing the music features of the *1-motif* pair show that these have similar chord progressions : C_{764} has the chord progression F♯ - B - E - B - F♯m, and C_{1618} has B - F♯m - B - E - D - A. This suggests that using the chord progression will produce a similar physiological response. This phenomena can also be observed in most motif pairs taken from other physiological data. Table 5 shows the amount of motifs that were discovered to have similar chord progressions.

From the results of the motif discovery, on average, a motif length that will give four to six seconds of chord progression is desirable. The complexity of the chord progression will depend on the length of motif. Since the exact length of the motif is not known, an algorithm that does not use motif length as a parameter should be used instead.

Other chord progressions identified by motif discovery using parameter set 3 are found in Table 6. The chord progressions I-IV, I-IV-V and I-IV-V-I from the song *Please Please Me* are mapped to the *1-motif* for that song. These chords sound similar and possibly invoke the same emotional response for that song. Some motifs will have similar

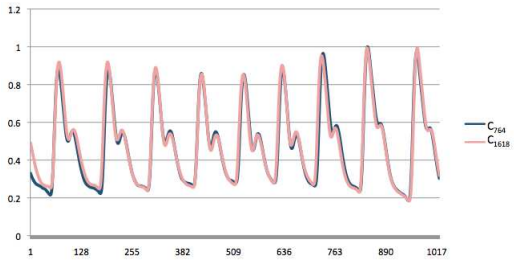

Figure 6. The motif discovered for the song *With A Little Help From My Friends* with $n = 1024$ occurring at subsequence C_{764} and C_{1618}.

Set No.	motif count		motifs with similar chord progressions	
1	1	$(1/64 = 1.5\%)$	1	$(1/1 = 100.0\%)$
2	25	$(25/64 = 39.0\%)$	17	$(17/25 = 68.0\%)$
3	61	$(61/64 = 95.3\%)$	39	$(39/61 = 63.9\%)$

Table 5. Number of motifs discovered for each parameter set and statistics for motifs with similar chord progressions

chord progressions but not in all cases. There are also motifs that have different chord progressions mapped to it, i.e. chords found in *Good Day Sunshine*.

Using motif discovery, we are able to discover chord progressions that are commonly used in western pop music. Given enough data, the library of motif could be used to identify the most frequently used chord progressions that invoke an emotional response by clustering similar psychophysiological motifs. This can be used in composing or recommending music with a desired emotion or mood.

7. CONCLUSION AND FUTURE WORK

In this work, psychophysiological readings from a subject listening to music was collected. A motif discovery algorithm was used to discover motifs from the BVP data. We observe that parts of music where the motif occur, have similar chord progressions and possibly other music features as well. By improving the algorithms used in this work, a library of different motifs can be built.

Future work includes additional analysis on the motifs to include other music features. Improving the motif discovery algorithm to dynamically identify motif length is also desired in order to have a more accurate account of the chord progressions that are important. Another round of data collection will also be done using a different set of participants. Analysis of other physiological data, (i.e. respiration rate and skin conductance) is also planned. A music recommendation system is also being designed that will use the information from motifs to generate a play list of songs that have similar emotion content.

Song	Key	Chord progression	
Act Naturally	G	G-D-G	I-V-I
		G-D	I-V
Dizzy Miss Lizzy	A	D-A	IV-I
		A-D	I-IV
		E-D-A	V-IV-I
For You Blue	D	D-A-D	I-V-I
		D-A	I-V
		D-A-G7	I-V-IV
Good Day Sunshine	A	B7-E7-A	ii-V-I
		F♯-B-F♯	vi-ii-vi
Please Please Me	E	E-A	I-IV
		E-A-B	I-IV-V
		E-A-B-E	I-IV-V-I
With A Little Help From My Friends	E	B-E-B	V-I-V
		F♯m-B-E	ii-V-I
Yesterday	F	B♭/7-Gm-C-F	IV-ii-V-I
		Gm-C-F-F7	ii-V-I-I

Table 6. Subset of results using parameter set 3

8. REFERENCES

[1] K. Chan and A.W. Fu. Efficient time series matching by wavelets. In *Proc. of the 15th IEEE Int'l Conf. on Data Eng.*, page 126–133. Sydney, Australia, Mar 23-26 1999.

[2] B. Chiu, E. Keogh, and S. Lonardi. Probabilistic discovery of time series motifs. In *Proc. of the 9th ACM SIGKDD Int'l Conf. on Knowledge Discovery and Data Mining*, page 493–498, 2003.

[3] J. A. Etzel, E. L. Johnsen, J. Dickerson, D. Tranel, and R. Adolphs. Cardiovascular and respiratory responses during musical mood induction. *International Journal of Psychophysiology*, 61(1):57–69, 2006. Psychophiosology and Cognitive Neuroscience.

[4] A. Gabrielsson and P.N. Juslin. Emotional expression in music. In R. J. Davidson, K. R. Scherer, and H. H. Goldsmith, editors, *Handbook of affective sciences*, page 503–534. New York: Oxford University Press, 2003.

[5] E. Keogh, K. Chakrabarti, M. Pazzani, and Mehrotra. Dimensionality reduction for fast similarity search in large time series databases. *Journal of Knowledge and Information Systems*, page 67–74, 2000.

[6] E. Keogh and S. Kasetty. On the need for time series data mining benchmarks: A survey and empirical demonstration. In *Proc. of the 8th ACM SIGKDD Int'l Conf. on Knowledge Discovery and Data Mining*, page 102–111. Edmonton, Alberta, Canada, July 2002.

[7] S. Khalfa, M. Roy, P. Rainville, S. Dalla Bella, and I. Peretz. Role of tempo entrainment in psychophysiological differentiation of happy and sad music? *International Journal of Psychophysiology*, 68(1):17–26, 2008.

[8] Y. E. Kim, E. M. Schmidt, R. Migneco, B. G. Morton, P. Richardson, J. Scott, J. A. Speck, and D. Turnbull. Music emotion recognition: A state of the art review. In *11th Int'l Society for MIR Conf.*, pages 255–266, August 2010.

[9] C. L Krumhansl. An exploratory study of musical emotions and psychophysiology. *Canadian Journal of Experimental Psychology/Revue canadienne de psychologie exprimentale*, 51(4):336, 1997.

[10] E. L. M. Law, L. von Ahn, R. B. Dannenberg, and M. Crawford. Tagatune: A game for music and sound annotation. In *Proc. of the Int'l Conf. on MIR*. Vienna, Austria, 2007.

[11] S. R Livingstone, R. Muhlberger, A. R Brown, and W. F Thompson. Changing musical emotion: A computational rule system for modifying score and performance. *Computer Music Journal*, 34(1):41–64, 2010.

[12] M. Mauch, C. Cannam, M. Davies, C. Harte, S. Kolozali, D. Tidhar, and M. Sandler. OMRAS2 metadata project 2009. In *10th Int'l Conf. on MIR Late-Breaking Session*. Kobe, Japan, 2009.

[13] F. Miller, M. Stiksel, and R. Jones. Last.fm in numbers. Last.fm press material, February 2008.

[14] M. Numao, T. Nishikawa, T. Sugimoto, S. Kurihara, and R. Legaspi. Constructive adaptive user interfaces based on brain waves. In *Proc. of the 13th Int'l Conf. on Human-Computer Interaction. Part II*, volume 5611 of *Lecture Notes in Computer Science*, pages 596–605, San Diego, CA, 2009. Springer-Verlag.

[15] M. Tompa and J. Buhler. Finding motifs using random projections. In *Proc. of the 5th Intl Conf. on Computational Molecular Biology*, page 67–74. Montreal, Canada, Apr 22-25 2001.

[16] D. Turnbull, L. Barrington, D. Torres, and G. Lanckriet. Semantic annotation and retrieval of music and sound effects. *IEEE Trans. on Audio, Speech and Language Processing*, 16(2), 2008.

[17] M. Vlachos, G. Kollios, and G. Gunopulos. Discovering similar multidimensional trajectories. In *Proc. of the 18th Int'l Conf. on Data Eng.*, page 673–684, 2002.

[18] B. K. Yi and C. Faloutsos. Fast time sequence indexing for arbitrary lp norms. In *Proc. of the 26th Int'l Conf. on Very Large Databases*, page 385–394, 2000.