

A Music Search Engine for Therapeutic Gait Training

Zhonghua Li¹, Qiaoliang Xiang¹, Jason Hockman², Jianqing Yang¹, Yu Yi¹
Ichiro Fujinaga², Ye Wang¹

¹School of Computing, National University of Singapore, Singapore

²Schulich School of Music, McGill University, Canada

{lizhongh, xiangqiaoliang, yiyu09, wangye}@comp.nus.edu.sg, jianqing@nus.edu.sg
jason.hockman@mail.mcgill.ca, ich@music.mcgill.ca

ABSTRACT

A music retrieval system is introduced that incorporates tempo, cultural, and beat strength features to help music therapists provide appropriate music for gait training for Parkinson's patients. Unlike current methods available to music therapists (e.g., personal CD/MP3 library search), we propose a domain-specific search engine that utilizes a database of music found on YouTube. We independently evaluate the efficacy of our tempo, cultural, and beat strength features on a music database extracted from YouTube. Results from our user study demonstrate the effectiveness and usefulness of our search engine for this application.

Categories and Subject Descriptors

H.3.3 [Information Search and Retrieval]: Information filtering, Query formulation; H.5.5 [Sound and Music Computing]: Signal analysis, synthesis, and processing; J.3 [Life and Medical Science]: Health

General Terms

Algorithms, Design, Experimentation, Human Factors

Keywords

Music Therapy, Gait Training, Parkinson's Disease

1. INTRODUCTION

Parkinson's disease is a neurodegenerative disorder that afflicts millions of people worldwide. The primary symptoms of Parkinson's disease are abnormal motor performance, irregular gait, and the progressive loss of mobility. At present there is no cure for this condition, and most patients rely on physiotherapy to improve their quality of life. Music therapy offers an alternative treatment method for Parkinson's patients, known as *rhythmic auditory stimulation* (RAS) [7, 5], which seeks to improve control over the motor system by utilizing the effects of heard rhythms during gait training.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, to republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

MM'10, October 25–29, 2010, Firenze, Italy.

Copyright 2010 ACM 978-1-60558-933-6/10/10 ...\$10.00.

In order to conduct RAS, music therapists select music according to tempi appropriate for individualized training levels for each patient. Despite its clear benefits, widespread adoption of RAS has been limited by the time-intensive task of producing an individualized music selection for each patient, as no one type of music is suitable for all patients.

While commercial sites (e.g., iTunes¹) allow users to find and purchase desired songs, search functionality on these sites is limited to querying by keywords such as an artist's name only. Last.fm² offers streaming music retrieved by query with user-contributed tags. Discrete musical tempi are not reliably available as search terms, and as audio is not available for analysis, such a database is not ideal.

YouTube is currently the largest video sharing website on the Internet, with over 14 million music videos contributed by people across the globe. While YouTube's freely available database is an excellent potential resource for music therapists to find desired songs for patients, searching for relevant videos is difficult due to the lack of a retrieval architecture specifically suited for music therapists.

In order to help music therapists find music suitable for patients performing the RAS technique, we propose a domain-specific search engine for YouTube content. This search engine enables music therapists and patients to find user-preferred music using a simple interface offering three fields for desired *tempo*, *culture style*, and *beat strength*. The tempo value in a query can be derived from a gait sensor attached to the patient. Cultural style covers the four major ethnic groups in Singapore: Chinese, Malay, Indian, and Western. Beat strength, a feature that defines the perceived emphasis of the musical beat, may be one of three levels: Weak, Medium, and Strong. We have also integrated a tempo stability measure, as music therapists have emphasized the need for stable tempi during training.

The remainder of the paper is organized as follows. In Section 2, we present a system overview and details of our implementation. Section 3 presents both the system evaluation and results. Conclusions and future work are provided in Section 4.

2. PROPOSED SYSTEM

We now present an overview of our search engine for music therapists and Parkinson's patients. Music features widely used in music therapy are directly integrated into our system. These features include tempo calculation, tempo sta-

¹<http://www.apple.com/itunes/>

²<http://www.last.fm/>

bility, cultural style, and beat strength. A ranked music list is then generated by fusing results from each feature aspect.

2.1 Tempo Estimation

Many tempo estimation algorithms have been developed and evaluated [1, 6], however, as the cultural style of music in the presented work differs from that of previous evaluation datasets, we could not be certain that the top performing algorithm in these evaluations would perform best in our system. We therefore conducted an evaluation of four state-of-the-art algorithms. The algorithm of Klapuri et al. [2] achieved the best results in our evaluation, and has therefore been incorporated into our system. Each song is divided into 30-second segments, and the overall song tempo is calculated as the median value of the tempi estimated from all segments using this algorithm.

2.2 Tempo Stability Feature

The intended usage of our system requires that retrieved songs elicit consistent tempi throughout the duration of an audio file. We therefore calculate the tempo stability within each song by analyzing the variation of tempi across all segments of the audio.

As there are no established metrics for tempo stability, we chose to use the standard deviation to evaluate this variation. However, it will be severely and disproportionately affected by having tempi at different metrical levels. Therefore, before calculating standard deviation, tempi across all segments of a song are first adjusted to the same metrical level using the dominant tempo calculated in Algorithm 1. As shown in Algorithm 1, a precision window Γ is defined to identify two similar tempi with this difference. Γ is set as 4% in view of the tempo tracking evaluation in [1].

Algorithm 1 Dominant Tempo

Input: T : Tempo List of Song i ;

Output: Dominant Tempo;

Description:

```

1: initialize 2 arrays:  $S$  for # of related tempo values on the
   same metrical level and  $D$  for # of related tempo values on
   different metrical levels;
2: for all term  $t \in T$  do
3:   for all term  $k \in T$  do
4:     if  $t$  and  $k$  are within  $\Gamma$  difference of each other then
5:        $S\{t\} = S\{t\} + 1$ ;
6:     else if  $t$  and  $k$  are within  $\Gamma$  difference of  $\frac{1}{2}$ ,  $2$ ,  $\frac{1}{3}$  or  $3$ 
       times of each other then
7:        $D\{t\} = D\{t\} + 1$ ;
8:     end if
9:   end for
10: end for
11: find tempo  $t$  from maximum of array  $S$ 
12: if unique  $t$  values  $> 1$  then
13:   return first tempo  $t$  from maximum array  $D$ ;
14: else
15:   return tempo  $t$ ;
16: end if

```

2.3 Culture Classification

A typical music culture classification involves audio feature extraction and classification using machine learning algorithms. Timbral, rhythmic, wavelet-based, and musicology-based features are extracted from each song as in [4]. We tested the performance of a variety of feature combinations with two classifiers: k -Nearest Neighbor (k NN) and Support

Vector Machines (SVM), to achieve the following optimal parameter settings for our system.

Each song is divided into 30-second non-overlapping segments, which are represented by 140-dimensional feature vectors comprised of 56 timbral, 10 rhythmic, 24 wavelet-based, and 50 musicology-based feature dimensions. Classification is then performed on each segment by SVMs using a radial bias kernel, and cultural style is determined by majority voting.

2.4 Beat Strength

Beat strength is a rhythmic feature that attempts to describe the prominence of a beat within a piece of music. It is useful in determining how well listeners can perceive the induced rhythm and discriminate songs with the same tempo. For our measure of beat strength, we follow the implementation of pulse clarity explained in [3], which relies on statistical analysis of autocorrelation coefficients.

Beat strength is presented as real values between 0 and 1, which proves difficult for users to indicate in their queries. To facilitate users to customize this option, we define two thresholds T_1 and T_2 , and values between T_1 and T_2 are defined as the medium level, while values in $[0, T_1)$ and $(T_2, 1]$ are considered as weak and strong levels, respectively.

2.5 Composite Ranking

Based on a user’s customized query, a ranked music list is returned by integrating all ranked lists from each dimension. Culture information simply functions as a binarization filter by assigning score 1 to all songs with matched culture and 0 to all mismatched songs. The beat strength ranking is applied linearly.

2.5.1 Tempo Stability Ranking

Tempo stability is represented using both the number of metrical levels m and the standard deviation σ of the adjusted tempi in each music piece. We propose a stability ranking according to the following rule: songs with lower standard deviations and fewer metrical levels are ranked higher. Standard deviation is of greater importance as metrical level determination of tempo-tracking algorithms is known to be not very discriminative. Specifically, the number of metrical levels will be analyzed only when songs have the same σ value. We first rank song i according σ_i to get an initial score

$$S_{si} = \frac{\sigma_{max} - \sigma_i}{\sigma_{max} - \sigma_{min}}. \quad (1)$$

Next, we sort songs with the same initial score in increasing order of their respective number of metrical levels m . The final score S_s is derived by equally distributing the score among different metrical levels as in Equation 2.

$$S_{si+1} = S_{si} - (n_{i+1} - 1) * \frac{S_{si} - S_{sj}}{n_{j-1}}, \quad (2)$$

in which j is the song with a nearest and lower initial score to song i , and n_j represents the number of unique metrical levels for songs between i and j .

2.5.2 Tempo Ranking

While we could retrieve all songs with the exact queried tempo t_q , such a method fails to consider tempi close to t_q or on different but related metrical levels. We therefore utilize a Linear distance score S_{tL} (Equation 3) and a Gaussian

distance score S_{tG} (Equation 4) to measure the similarity between a given tempo t and t_q .

$$S_{tL}(t_q, t) = \begin{cases} \frac{1}{t_w}t - \frac{t_q - t_w}{t_w} & \text{if } t_q - t_w < t < t_q, \\ -\frac{1}{t_w}t + \frac{t_q + t_w}{t_w} & \text{if } t_q < t < t_q + t_w, \\ 0 & \text{otherwise} \end{cases} \quad (3)$$

t_w is the tolerance window size, and has been set to be $t_w = 0.1t_q$ through consultation of music therapists. Similarly, this tolerance corresponds to 3σ in the Gaussian distance score.

$$S_{tG}(t_q, t) = \begin{cases} \frac{1}{\sigma\sqrt{2\pi}} \exp -\frac{(t-t_q)^2}{2\sigma^2} & \text{if } t_q - 3\sigma < t < t_q + 3\sigma, \\ 0 & \text{otherwise} \end{cases} \quad (4)$$

A final tempo score $S_{tG}(t_q, t)$ is provided for each song by integrating the weighted scores of the same, half, and double the query tempo. For example, in Equation 5, we use the Gaussian score:

$$S_{tG}(t_q, t) = w_0 S_{tG}'(0.5t_q, t) + w_1 S_{tG}'(t_q, t) + w_2 S_{tG}'(2t_q, t), \quad (5)$$

where w are weights given to each metrical level and may be adjusted according to their influence in gait training.

2.5.3 Fusion of Multiple Ranking Lists

A linear combination method has been adopted for our present system to rank each music item according to its final score. To measure the importance of different aspects, we communicated with two music therapists. However, since music therapy aims at optimizing the performance for each patient, the therapeutic procedure is highly individualized according to the patient’s music preference. Many factors, including cultural background and musical perception, are involved and may affect the music selection. As it is difficult to distinguish the priority of one aspect over the others, we treat all these music features with equal weight.

3. SYSTEM EVALUATION

We now present a subjective evaluation of our system using results from a survey designed to gauge user satisfaction for both the various features used, and the overall system. The evaluation was performed using a subset of our database originally extracted from YouTube, and we first explain the creation of this dataset below.

3.1 Dataset Extraction Methodology

We have collected 787 full songs from YouTube. As most Parkinson’s disease patients are elderly, this database should reflect a variety of music with which the elderly are likely to be familiar.

An approximately equal number of songs from the four main cultural groups present in Singapore (i.e., Western, Indian, Chinese, and Malay) were selected. The eras of origin for the songs range from the 1930s to 1990s. The cultural style of each song was verified manually during collection. Ground truth tempo values were achieved by dividing each song into several 30-second segments, which were annotated by two amateur musicians with over ten years of musical experience. The ground truth annotation integrates both results by checking whether they agree with each other on the same or different metrical levels within precision window Γ , as described in Section 2.5.1. Annotations that agree on two metrical levels were both maintained.

Songs which have zero or only one manual annotation, or two annotations exceeding Γ were not included for testing against the tempo estimation algorithms. The reduced dataset was comprised of 560 songs (148 Western, 132 Chinese, 128 Indian, and 152 Malay). As beat strength is possibly more subjective than tempo estimation, we chose to annotate beat strength with the algorithm described in Section 2.4. In our system, thresholds T_1 and T_2 were set to be 0.1 and 0.4.

3.2 System Evaluation

A user survey was conducted to evaluate both the satisfaction of the retrieval accuracy and the usability of our system. 21 subjects (17 male, 4 female) participated in our survey. 4 subjects were amateur musicians, and the remainder were music hobbyists; none had physical difficulties tapping along with the music. The subjects were from China, India, Canada, Romania, Singapore—five countries that comprise the four cultural styles that we studied. In addition, 19 of the subjects were students, and 2 were professional music therapists.

3.2.1 Query Design

The two music therapists were also involved in the design of the search queries. Seven types of queries were generated using the following criteria: tempo only (TP), culture only (CT), beat strength only (BS), tempo and culture, tempo and beat strength, culture and beat strength, and all features.

As suggested by music therapists, tempi from 60 beats per minute (BPM) to 120 BPM are the most frequently used queries in gait training, and we thus tested values in the set of $\{60, 70, \dots, 110, 120\}$ in our experiment. Four cultural styles were used (Western, Chinese, Indian, and Malay), and three degrees of beat strength were offered (Weak, Medium, and Strong). Tempo stability is considered as a default feature, and is integrated into our final algorithm, as stable songs are required in gait training, and also could be potentially difficult to evaluate.

3.2.2 Methodology

For each subject, the same procedure was conducted. A brief introduction was first conducted to ensure that they understood the experiment procedure and how to use our system. In order to evaluate the tempo dimension, users were asked to perform a simple tapping task to guarantee that they could provide reliable tempo judgements. A simple web-based tempo algorithm was provided along with the survey to verify that the BPM at which the subject was tapping, was in fact correct. Every user was asked to test all seven query cases as discussed in Section 3.2.1, following the order provided: from queries of lower complexity (single feature dimension) to higher complexity (more feature dimensions). For each query, subjects could choose arbitrary values for each dimension (from the lists presented in Section 3.2.1), and then assess whether the top ten ranked results matched the dimensions selected in the initial query. The number of dimensions that matched the desired dimensions in their queries was also indicated.

3.2.3 Results and Analysis

To evaluate the retrieved music, we utilized precision at each rank position for a single query and Mean Average Pre-

cision (MAP) for multiple queries. The precision for the top ten songs ($P(10)$) and MAP results for all seven query types are presented in Table 1. For queries by a single dimension, searching by culture has produced the best results, followed by searching by beat strength and tempo, respectively. As can be expected, queries involving more dimensions achieve worse results than those with a single dimension due to the strict requirement of matching all query dimensions. However, culture-related combinations still achieved very high results.

Table 1: Precision ($P(10)$) and Mean Average Precision (MAP) for all seven query cases

Query	$P(10)$	MAP	Query	$P(10)$	MAP
TP	0.711	0.818	TP+BS	0.505	0.696
CT	0.946	0.978	CT+BS	0.735	0.830
BS	0.789	0.878	TP+CT+BS	0.373	0.546
TP+CT	0.689	0.754	Overall	0.635	0.757

Figure 1 shows the precision for each evaluation result by rank. It shows a rough downhill trend from the 1st to the 10th result, which indicates that our system is for the most part retrieving songs according to their true ranks.

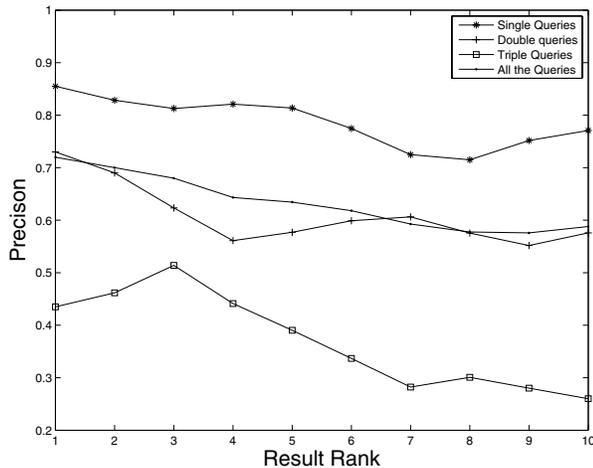


Figure 1: Average precision for different query types

We have also compared the evaluation results of the two music therapists and 19 non-music therapists. Their evaluations for precision ($P(10)$) and MAP are respectively 0.583 and 0.709, which are not far off from the overall results shown in Table 1. At the very least, it appears that the evaluations by non-music therapists should be fairly reliable.

To measure the usability of our system, subjects’ responses were collected and mapped to numeric values between 1 and 5, representing “Very Unsatisfied” and “Very Satisfied”, respectively. Table 2 shows the mean and median results across all the responses. It can be seen that subjects’ results are roughly consistent with the quality of the search results for each dimension, with culture being rated the highest. The music therapists were both “Satisfied” with the tempo and culture classification of the search results. They were “Neutral” with respect to the beat strength feature as they found the perception of beat strength to be too subjective as a feature. When asked about the potential usage of our system in the music therapy domain, they responded with “Useful”, the highest of the 5 response choices. Since our

Table 2: Mean/median of user satisfaction results by query type

	Tempo	Culture	Beat Strength
Mean	3.95	4.32	3.89
Median	4	4	4

system integrates specific search criteria and presents the estimated results, it saves them a great amount of time from listening and evaluating each song using traditional search engines or CDs.

4. CONCLUSIONS AND FUTURE WORK

We have presented a domain-specific music search engine to help music therapists find appropriate music for Parkinson’s disease patients. Songs of a specific tempo, cultural style, and beat strength are proven to be easily retrieved based on our experiments.

At present, we have only analyzed a few selected features in this system. There are certainly more relevant features that could potentially be explored within our method, such as era of musical origin (e.g., 1930s, 1960s) and genre.

There are additional therapeutic uses for our search engine, such as a method for finding music for progressive relaxation exercises for patients with Parkinson’s disease. In our future work, we plan to explore the applicability of the presented system for such scenarios. We also intend to examine the inclusion of additional search fields, as well as to perform further system evaluations to assess the scalability of our present system.

5. ACKNOWLEDGMENTS

The work was supported by Singaporean MOE grant R-252-000-421-112. We appreciate music therapists Dr. Patsy Tan and Ms. Ng Wang Feng for their instructive suggestions and evaluations. We also thank Dr. Anssi Klapuri of Queen Mary, University of London for the kind permission to use his source code in this project.

6. REFERENCES

- [1] F. Gouyon, A. Klapuri, S. Dixon, M. Alonso, G. Tzanetakis, C. Uhle, and P. Cano. An experimental comparison of audio tempo induction algorithms. *IEEE Trans. on Audio, Speech, and Language Processing*, 14(5):1832–1844, 2006.
- [2] A. P. Klapuri, A. J. Eronen, and J. T. Astola. Analysis of the meter of acoustic musical signals. *IEEE Trans. on Audio, Speech and Language Processing*, 14(1):342–355, 2006.
- [3] O. Lartillot and P. Toiviainen. Mir in matlab (ii): A toolbox for musical feature extraction from audio. In *International Conf. on Music Information Retrieval*, 2007.
- [4] Y. Liu, Q. Xiang, Y. Wang, and L. Cai. Cultural style based music classification of audio signals. *IEEE International Conf. on Acoustics, Speech, and Signal Processing*, pages 57–60, 2009.
- [5] G. C. McIntosh, S. H. Brown, R. R. Rice, and M. H. Thaut. Rhythmic auditory-motor facilitation of gait patterns in patients with parkinson’s disease. *Neurology, Neurosurgery, and Psychiatry*, 62(1):22–26, 1997.
- [6] M. McKinney, D. Moelants, M. Davies, and A. Klapuri. Evaluation of audio beat tracking and music tempo algorithms. *New Music Research*, 36(1):1–16, 2007.
- [7] M. H. Thaut, G. C. McIntosh, R. R. Rice, R. A. Miller, J. Rathbun, and J. Brault. Rhythmic auditory stimulation in gait training for parkinson’s disease patients. *Movement Disorders*, 11(2):193–200, 1996.