International Journal of Advanced Engineering and Management Research
Vol. 2 Issue 5, 2017

http://ijaemr.com/ ISSN: 2456-3676

PITCH ESTIMATION BASED ON WAVELET TRANSFORM FOR MUSIC THERAPY

Tien-Hui Luo[1] and Chih-Hsu Hsu[2*]

[1]Department of Nursing, Mackay Junior College of Medicine, Nursing, and Management,

No.92, Shengjing Rd., Taipei City 112, Taiwan, R.O.C.

[2]Department of Applied Information and Multimedia, Ching Kuo Institute of Management and Health,

No.336, Fu Hsin Rd., Keelung 203, Taiwan, R.O.C.

ABSTRACT

Music is widely used for various purposes, it contributes to the treatment of patients since people appreciate it and use it for music therapy. Music therapy is a clinical practice that engages both client and therapist in dynamic musical interaction. In music pitch estimation experiments, the database is adopting western classical music and the original songs. We proposed a multi-resolution feature extraction technique to deal with music therapy signals. We utilize adaptive resonance theory to search their clusters of features. The effectiveness of the clusters is encouraging.

Key Words: music therapy, pitch estimation, wavelet transform, neural networks, adaptive resonance theory.

1. INTRODUCTION

In this paper, a musical pitch estimation system based on wavelet transform is proposed. Music is widely used for various purposes, it contributes to the treatment of patients since people appreciate it and use it for music therapy. Music therapy is a clinical practice that engages both client and therapist in dynamic musical interaction. Found in all cultures around the world, music has the power to touch the human spirit at a deep level, often without the use of words. The disciplines of music therapy and music technology are both relatively recent, having emerged late in the 21th century, so combining the therapy with the technology is still considered novel. In response to the social changes, nursing profession requires multiple abilities of care. Music therapy is considered to be one of the multi-mode therapies [1].

Music signal processing researchers are expected to propose more efficient and effective solution for real world. Using real world data is one of the good ways to think such solutions. In addition, we are required to grow the real fields by ourselves where the technologies potentially can be used [2]. In music signal, we utilize the technology of signal processing to analyze the features of each music melody. Music signal could be extracted by a set of features of each frame from wavelet transform and Fouier transform.

Pitch is a common parameter in many types of music signal processing. It is defined as the perceived fundamental frequency of a signal and is used in many applications of music signals. Pitch estimation algorithms can be classified in two separate categories, spectral-domain based and time-domain based detection. Spectral pitch detectors estimate the pitch period directly using windowed segments of music signal. However, time based pitch detectors estimate the pitch period by determining the glottal closure instant and measuring the time period between each "event" [3].

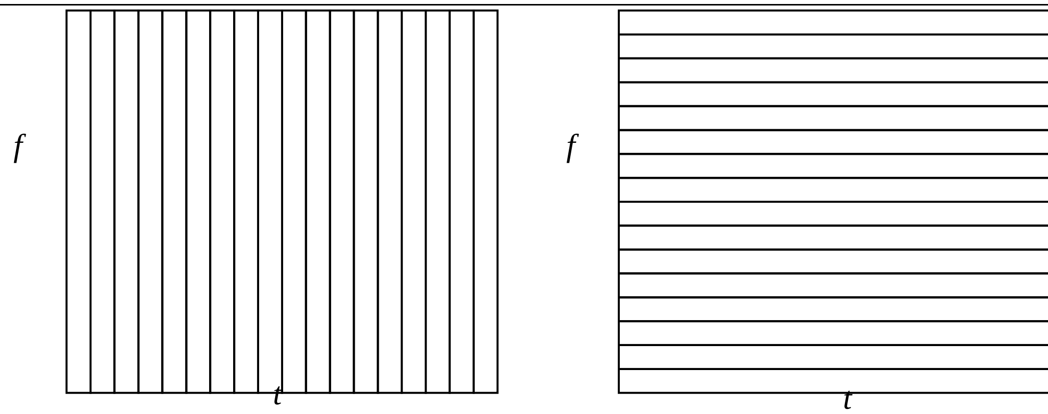
This paper is organized as follows. In section 2 we discuss the pitch estimation based on wavelet transform. Section 3 briefly describes the adaptive resonance theory. The experimental results are given in Section 4. Finally, some concluding remarks are presented in Section 5.

2. PITCH ESTIMATION BASED ON WAVELET TRANSFORM

2.1 Wavelet Transform

The multi-resolution formulation of wavelet transform is obviously designed to represent signals where a single event is decomposed into finer and finer detail, but it turns out also to be valuable in representing signals where a time-frequency or time-scale description is desired even if no concept of resolution is needed. In many applications, one studies the decomposition of a signal in terms of basis function. For example, stationary signals are decomposed into the Fourier basis using Forier transform. For non-stationary signals (i.e. signals whose frequency characteristics are time-varying like music, speech, image, etc.) the Fourier basis is ill-suited because of the poor time-localization. The classical solution to this problem is to use the short-time (or windowed) Fourier transform. However, the short-time Fourier transform has several problems, the most severe being the fixed time-frequency resolution of the basis functions. Wavelet techniques give a new class of bases that have desired time-frequency resolution properties. The "optimal" decomposition depends on the signal studied.

Each function in a basis can be considered schematically as a tile in the time-frequency plane, where most of its energy is concentrated. Non-overlapping tiles can schematically capture orthonormality of the basis functions. With this assumption, the time-frequency tiles for the standard basis and the Fourier basis are shown in Fig. 1.



(a) (b)

Fig. 1 (a) Standard time domain basis, (b) Standard frequency domain basis.

The discrete wavelet transform is another signal independent tiling of the time-frequency plane suited for signals where high frequency signal components have shorter duration than low frequency signal components. The discrete wavelet transform coefficients are a measure of the energy of the signal components in the time-frequency plane, giving another tiling of time-frequency plane. Fig. 2 shows the corresponding tiling description, which illustrates time-frequency resolution properties of a discrete wavelet transform basis.

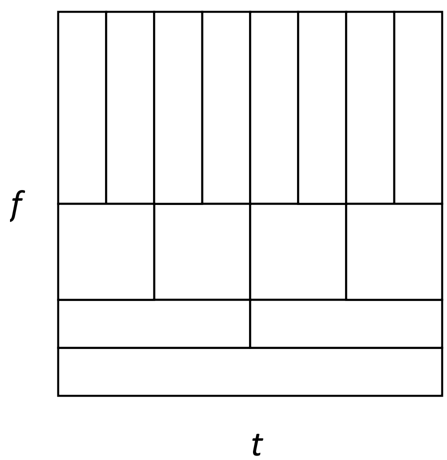


Fig. 2 Three-scale of wavelet basis.

The definition of the scaling function $\phi_{j,k}(t)$ and wavelet function $\psi_{j,k}(t)$ is given by [4].

$$\phi_{j,k}(t) = 2^{j/2}\phi(2^j t - k) \qquad j,k \in Z \qquad (1)$$

$$\psi_{j,k}(t) = 2^{j/2}\psi(2^j t - k) \qquad j,k \in Z \qquad (2)$$

This two-variable set of basis function is used in a way similar to the short time Fourier transforms. A signal space of multi-resolution approximation is decomposed by wavelet transform in an approximation (lower resolution) space and a detail (higher resolution) space. In order to generate a basis system that would allow higher resolution decomposition at higher frequencies, we will iterate the wavelet transform recursively to divide the approximation space, giving a left binary tree structure. The wavelet packet was allowed a finer and adjustable to particular signals or signal classes. The wavelet packet decomposes the detail spaces as well as approximation ones.

2.2 Pitch Estimation

Pitch estimation is necessary for a variety of applications. In musical field, pitch extraction from audio files could be used for music transcription. Pitch estimation is commonly treated as a problem of classification. Several approaches have been proposed for classification problems. Some works focus on conventional probabilistic and deterministic classifiers [5]. Another approach uses neural networks to classify patterns [6]. In music signal, we utilize the technology of signal processing to analyze the features of music pitch. Music signal could be extracted by a set of features of each frame from wavelet transform and Fouier transform. The windowed Fouier transform has uniform resolution over the time frequency plane. It is difficult to detect sudden burst in a slowly varying signal by Fouier transform. Wavelet transform overcomes the problem of fixed resolution, using adaptive window sizes, which allocate more time to the lower frequency and less time for the higher frequency [7]-[9].

3. ADAPTIVE RESONANCE THEORY

Adaptive resonance theory describes a family of self-organizing neural networks, capable of clustering arbitrary sequences of input patterns into stable recognition codes. Grossberg attempted to address the stability-plasticity dilemma: how can a learning system remain plastic (adaptive) in response to new, unseen information, yet remain stable in response to irrelevant information? How can a system preserve its already acquired knowledge and at the same time be flexible enough to accommodate new information to be store? How can the system decide when to alternate from the stable to the plastic state and vice verse? Grossberg’s answer to the stability-plasticity dilemma was the adaptive resonance theory. In an adaptive resonance theory network, information reverberates between the network’s layers. Learning is possible in the network, when resonance of the neural activity occurs. Resonance occurs when an already learned pattern is presented and the network recalls / recognizes it and when a novel input pattern is presented, the network realizes that the pattern constitutes new information and then enter resonant state to memorize it. Fig. 3 shows the block diagram of adaptive resonance theory [10].

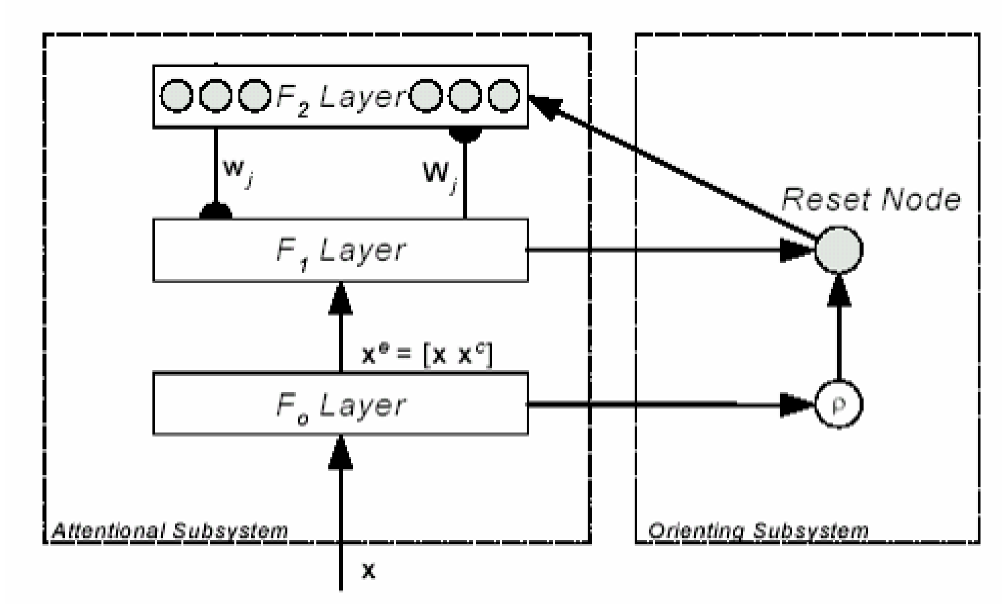


Fig. 3 Block diagram of adaptive resonance theory.

Adaptive resonance theory is comprised by two major subsystems. The subsystem consists of three layers of neurons. If the dimensionality of input patterns is M, the module's F0 layer has M nodes and is a pre-processing stage that complete encodes the input patterns. It requires that its input vectors have their entire feature values normalized between 0 and 1. In other words, the input domain of adaptive resonance theory is M-dimension. F0 transforms an input vector to 2M-dimension, which serves as an input vector to the F1 layer. Although complement coding doubles the size of input patterns, it turns out to be essential for adaptive resonance theory to perform clustering. Layer F1 has 2M nodes, while F2 has a large enough number of nodes that will allow the adaptive resonance theory module to perform its learning task. All nodes in F1 are interconnected with all nodes in F2 via bottom-up and top-down weights.

Fuzzy adaptive resonance theory employs localized rather than distributed learning. The later one applies to multi-layer perceptrons, in order to learn a single pattern; many weights have to be updated. However, learning a particular pattern in fuzzy adaptive resonance theory only involves the template modification of a single node, whether a category updates or a category creation takes place. Thus, updating weights in fuzzy adaptive resonance theory learning lies primarily in the search (the combination of repetitive node competition and performing the vigilance test) for a suitable category.

In general, fuzzy adaptive resonance theory training may produce overlapping categories, that is, the intersection of some hyper-rectangles corresponding to fuzzy adaptive resonance theory categories might be non-empty. In has been shown in [11], that a pattern located inside the intersection of some categories will choose the categories of smallest size.

4. Experimental Results

In music pitch estimation experiments, the database is adopting western classical music and the original songs. We utilize the technology of signal processing to analyze the features of music signals. Music signals could be extracted by a set of features of each frame from wavelet transform and Fourier transform. The windowed Fourier transform has uniform resolution over the time frequency plane. It is difficult to detect sudden burst in a slowly varying signal by Fourier transform. Wavelet transform overcomes the problem of fixed resolution, using adaptive window sizes, which allocate more time to the lower frequency and less time for the higher frequency.

The melody and rhythm play important roles in music care. Then, we utilize adaptive resonance theory to search their clusters of features. Fig. 4 shows the results of pitch estimation. Thus, the effectiveness of the clusters is encouraging. The results of this study can be applied as the reference of music intervention.

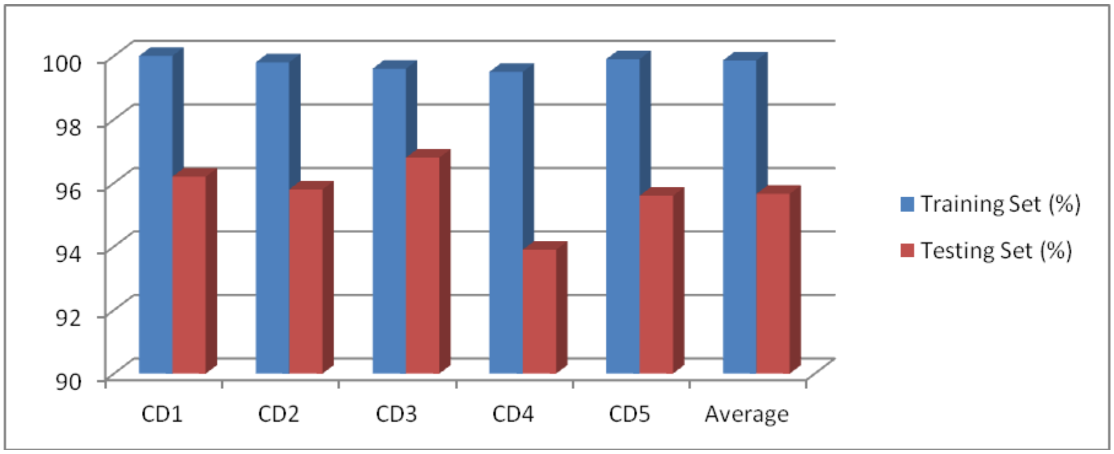


Fig. 4 The results of pitch estimation.

5. Concluding Remarks

In this paper, the pitch estimation for music therapy signals was presented. We proposed a multi-resolution feature extraction technique to deal with music therapy signals. We utilize adaptive resonance theory to search their clusters of features. The wavelet transform has many applications on music signal processing. The correlation of wavelet coefficients will be smaller than the wavelet coefficient of the corresponding temporal process. Thus, the wavelet transform can transfer complex signals to simple signals. We extracted the features of pitch from music signal, but they need more precise frequency. Therefore, we should combine the complexity of wavelet transform and the precision of the time frequency analysis in the future.

References

[1] A.W. Lipe, E. York, E Jensen, "Construct validation of two music-based assessments for people with dementia," Journal of music therapy, 44(4), pp. 369-387, 2007.

[2] N. Kosugi, M. Kondo, "Community site for music therapists based on the session records for music therapy," In Proceedings of the 2012 15th International Conference on Network-Based Information Systems, Melbourne, Australia, pp.319-325, Sep. 26-28, 2012,.

[3] C. Wendt and A.P. Petropulu, "Pitch determination and speech segmentation using the discrete wavelet transform," In Proceedings of IEEE International Symposium on Circuits and Systems, pp.45-48, 1996.

[4] C.S. Burrus, R.A. Gopinath, and H. Guo, Introduction to Wavelets and Wavelet Transforms, Prentice-Hall, 1998.

[5] X. Huang, A. Acero, and H.W. Hon, Spoken language processing: A guide to theory, algorithm, and system development, Prentice-Hall, 2001.

[6] J. Jose de Jesus Guerrero-Turrubiates, S.E. Gonzalez-Reyna, S.E. Ledesma-Orozco, and J.G. Avina-Cervantes, "Pitch Estimation For Musical Note Recognition Using Artificial Neural Networks," In Proceedings of International Conference on Electronics, Communications and Computers, Cholula, Mexico, pp.53-58, Feb. 26-28, 2014.

[7] W.C. Chen, C.T. Hsieh, and C.H. Hsu, "Two-stage vector quantization based multi-band models for speaker identification," In Proceedings of International Conference on Convergence Information Technology, Gyeongju, Korea, pp.2336-2341, Nov. 21-23, 2007.

[8] T.H. Luo and C.H. Hsu, "Pitch Estimation Based on Neural Network for Music Therapy," In Proceedings of International Scientific Conference on Engineering and Applied Sciences, Okinawa, Japan, pp. 732-736, July, 2015.

[9] T.H. Luo and C.H. Hsu, "Signal Analysis for "Kagaya Miyamoto Shiki" Music Therapy," In Proceedings of International Scientific Conference on Engineering and Applied Sciences, Singapore, pp. 283-284, Aug. 15-17, 2014.

[10] G.C. Anagnostopoulos, Novel approaches in adaptive resonance theory for machine learning, Ph. D. thesis, University of Central Florida, 2001.

[11] G. A. Carpenter, S. Grossberg, and D. B Rosen, "Fuzzy ART: an adaptive resonance algorithm for rapid, stable classification of analog patterns," International Joint Conference on Neural Networks, Vol. 2, pp. 411 – 416, July 8-12, 1991.