International Conference on Applied and Theoretical Information Systems Research,
February 10-12, 2012, Taipei, Taiwan

ON SYMBOLIC MUSIC CLASSIFICATION USING WAVELET TRANSFORM

Gissel Velarde,
gisselvelarde@yahoo.com

Tillman Weyde,
City University, London
t.e.weyde@city.ac.uk

ABSTRACT

This work investigates a new approach for symbolic music classification via continuous Haar wavelet transform, used as alternative representation and segmentation technique. The method was experimentally tested on the J.S. Bach's 15 two-part inventions for keyboard. In music theory, the inventions are often described as developed coherently from a short melodic pattern that dominates the whole piece. Based on this premise, we take fragments of the inventions and try to identify the invention they belong to using a nearest neighbour classifier, which is based on the initial sections of all inventions. We compare combinations of wavelet and pitch vector representation and segmentations, as well as different lengths for the initial sections, and different ways of segmenting the sections and representing the segments, including melodic variations. The classification rates achieved by the combination of wavelet representation and segmentation are far better than those of pitch vectors and constant length segmentation, while including melodic variations yields only minimal rate changes.

Keyword: symbolic music classification, continuous wavelet transform, melodic segmentation

1. INTRODUCTION

Music classification and retrieval are important tasks in today's multimedia research community. Typical retrieval tasks involve identifying pieces of music from a melodic fragment entered by singing/humming or using a musical or computer keyboard. The aim of this study is to test whether using wavelet analysis of symbolic music data can capture musically relevant properties that help music classification.

In this paper we use Johann Sebastian Bach's two-part inventions as a small dataset for an initial test of wavelet representation. The baroque invention form is a short piece of keyboard music, with the inventions by Bach as the most prominent examples. Bach himself described the inventions in his foreword as a demonstration of how to develop musical ideas ("inventiones")[8]. Music theorists broadly agree that the inventions are developed around a short pattern – the musical idea or, in musicological terms, the thematic subject – from half a measure to four measures in length and in the development the theme or pattern is changed in various ways, but remains recognisably similar for the listener [13].

The classification task we address is the recognition of the similarity of the initial theme to fragments of later parts of the invention, where the theme is changed and transformed. With this task, we aim to investigate the suitability of wavelet representation and segmentation for detecting this similarity.

We do not explicitly identify the themes or model the complexity of individual relations between melodic segments, as would be required for a paradigmatic analysis [3]. The complexity of this task and the need to evaluate the result on the basis of subjective human judgements would have made it harder to evaluate the effect of wavelet representation. Therefore we preferred a simpler setup with objective ground truth that allows a more straightforward interpretation of the results.

The paper is organized as follows: after this introduction we discuss previous work related to this technique. Section 3 is dedicated to describing the method and the experimental setup. The results are presented in section 4, followed by a discussion in section 5, and finally in section 6 conclusions are presented, including future work.

2. RELATED WORK

Classification of melodies has been studied extensively in the last 15 years in the music information retrieval community, while the specific subjects of Bach's 2-part inventions and of wavelet representation of music have been covered less often. In the following we give a brief overview of related work in these areas.

2.1 Classification of melodies

Classification of symbolic melodies has so far mostly been applied to determine composer, genre or origin of melodies. The methods used vary widely, but they normally include some form of feature-extraction or other pre-processing followed by the application of a computational classification model. For example, Ponce and Iñesta [14] compare Bayesian classification, k-nearest neighbour (kNN), self-organising maps for the distinction between jazz and classical styles on the basis of various features relating to pitch, pitch intervals, note durations, syncopations and diatonic notes. Hilleware et al. [15] compare a wide selection of global statistics of melodies with n-gram models based on combinations of viewpoints (intervals, pitch, degree, direction of movement, etc.) as the bases for classification with naïve Bayes, decision tree, support vector machine, and kNN classification. They find that the n-gram models outperform the global models. Conklin [4] uses segment class representations to classify melodies by style (Bach chorales vs. Nova Scotia folk songs), achieving high accuracy for pitch class patterns on beat-length segments. Li and Sleep [11] use 1-nearest neighbour classification with a similarity metric based on Kolmogorov complexity for classifying melodic styles with good success.

2.2 Bach's two-part inventions

Bach's two-part inventions have been the centre of attention of previous studies. Adiloglu et al. [10] defined a significance measure for melodies via the number of repetitions of a given melody and its close variations in the piece, and extracted the melodies, which appear more often than a threshold value, and clustering them. The evaluation is based on comparing the results to the interpretation of a human analyst. Pinto [1] developed motif discovery based on eigenvectors, highlighting the centrality of bars and considering the relevance of motifs included within ranked bars. This method is also evaluated on a given musicological analysis.

2.3 Wavelet representation of music

Wavelets have become increasingly popular as a technique for audio and image processing. Although they have to our knowledge not been used for symbolic melody classification, there have been other applications to music analysis. Smith and Honing applied the continuous wavelet transformation, which we use in this paper, to audio signals in order to decompose an interval representation of a musical rhythm into a hierarchy of short-term frequencies, trying to make the structure of a rhythm signal explicit [12]. In this paper, we apply the continuous wavelet transform to symbolic music data instead of audio. Symbolic music data represent the notes as they are written in a score with time and pitch information. To our knowledge, Pinto was the first to use wavelets on symbolic music material [2], employing multi-resolution analysis based on Haar transform. The purpose of that study was the indexing of long melodic sequences without losing information. Velarde [6] described the use of wavelet coefficients to visualise symbolic music to support visual identification of musical patterns.

3. METHOD

Our aim is to investigate the effectiveness of wavelet coefficients to represent musically relevant properties of symbolic melodies. Therefore, in contrast to Adiloglu [10] and Pinto [1], we chose to use a method that does not involve subjective human judgement as ground truth (i.e. the classification accepted as correct). We choose a task with a ground truth that leaves no room for interpretation as a music fragment belongs to a specific invention or not. Therefore we can focus on comparing different representation in relatively simple setting, without the need to statistically interpret the ground truth.

We used the 15 two-part inventions by Johann Sebastian Bach (BWV 772-786) in the MIDI encodings provided by the MuseData database [7]. The inventions were divided into sections, with the initial section, the exposition, being used for defining the classifier. The sections were segmented into short melodic segments, which we used for classification of the sections with the 1-nearest-neighbour method.

We used two approaches for segmentation and two approaches for pattern representation. The patterns were represented as normalized pitch vectors or as wavelet coefficients. We segmented into constant length segments or by wavelet coefficients. The details of these steps are described in the following sections.

3.1 Pitch vector representation

We use MIDI files as input, which we convert into pitch vectors *v* with length *l*, and resolution *r* in number of samples per quarter note. *v(t)* is defined as a continuous time signal with a pitch value for every time point. Rests are assigned the value zero.
We normalize vectors, when not represented as wavelet coefficients, by subtracting the average pitch, excluding rests, in order to make the representation invariant to transposition. After the normalization, rests are assigned the value zero.

3.2 Wavelet coefficients representation

We use the wavelet coefficients in comparison to pitch vectors, because we assume that this representation might be well suited for detecting melodic similarity. In the rest of this section we give a short explanation of wavelets roughly following Mallat's presentation [16].

Any finite energy signal *v* can be represented by its wavelet coefficients. We selected the Haar wavelet, which is a piecewise constant function

$$\psi(t) = \begin{cases} 1, & if \quad 0 \le t < 1/2 \\ -1, & if \quad 1/2 \le t < 1 \\ 0, & otherwise \end{cases} \tag{1}$$

from which an orthonormal basis is generated by dilatations (scaling) and translations (shifting)

$$\left\{ \psi(t) = \frac{1}{\sqrt{2^s}} \psi\left(\frac{t - 2^s u}{2^s} \right) \right\}_{(u,s) \in Z^2} \tag{2}$$

The wavelet coefficients of the pitch vector *v* for scale *s* and shift *u* are defined as the inner product

$$\langle v, \psi_{u,s} \rangle = \int_{-\infty}^{+\infty} v(t) \psi_{u,s}(t) dt \tag{3}$$

The transformation is performed over all *u,* from *1* to *l,* at a particular scale *s*.

3.3 Segmentation

We use two segmentation approaches: constant segmentation and wavelet segmentation. In constant segmentation, we divide the melodies into segments of equal length, as a simple baseline for segmentation.

Wavelet coefficients obtained by expression (3) are used not only to represent melodies, but also to segment them into smaller musical units. The continuous wavelet transform at a particular scale produces a 1-dimensional function. We conjectured that coherent musical units might be found between sign changes of the coefficients. Our wavelet segmentation points are therefore set at zero crossing points of the wavelet coefficients at one given scale. Usually, smaller scales will produce more segmentation points than larger scales.

Constant segmentation produces vectors of equal length, which is necessary for calculating distances. Wavelet segmentation, however, returns segments of different lengths. We therefore use the maximal length of the segments as the length for all vectors, and padded shorter segments with zeros at the end. This does create a bias for shorter patterns, as they produce smaller distances, which we are not compensating for in any way.

Figure 1 shows an example of the first 16 quarter notes of the upper voice of invention 1 in the four different combinations used for the experiments: wavelet segmentation – vector representation (ws-vr), wavelet segmentation – wavelet representation (ws-wr), constant segmentation – vector representation (cs-vr) and constant segmentation – wavelet representation (cs-wr). The segmentation points are shown as vertical dotted lines.

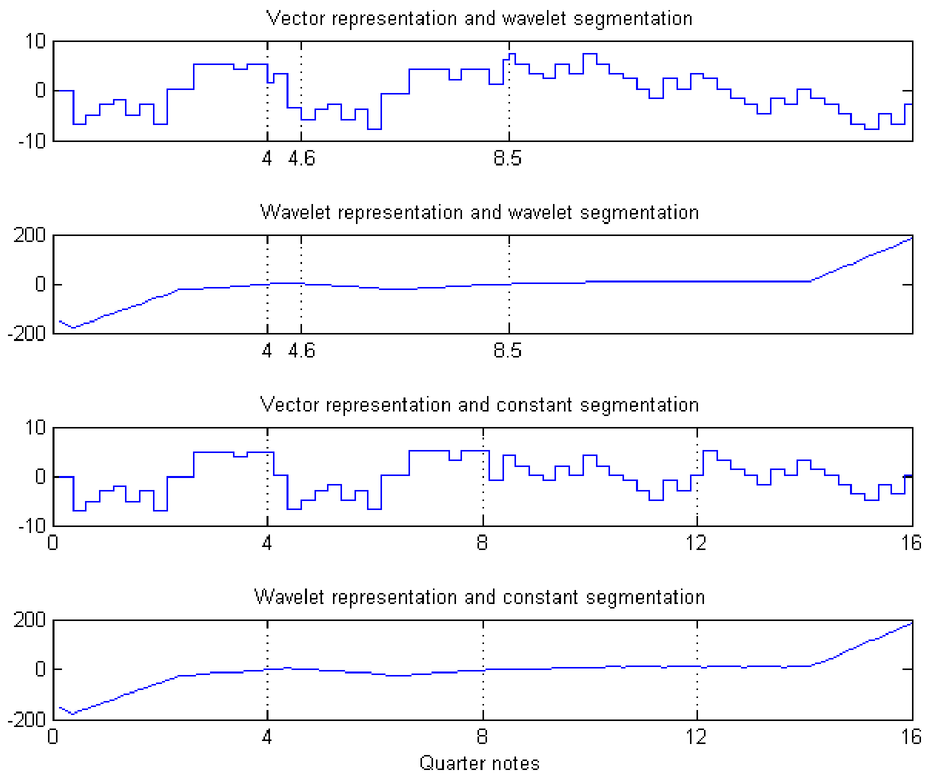

Figure 1. The first 16 quarter notes of Invention 1 upper voice in 4 combinations of representation and segmentation at a scale of 4 qn.

Figure 2 shows the visualisation of the coefficients obtained by the continuous wavelet transform with the Haar wavelet at scales from 1 to 8 quarter notes of the 16 first quarter notes of the upper voice of invention 1, in a so-called scalogram. The coefficients plotted with brighter colours correspond to greater coefficients. The scalogram produces a visible hierarchical structure, which may be useful in itself for detecting melodic patterns; however, we used only the coefficients at one selected scale for representation and segmentation in this paper.

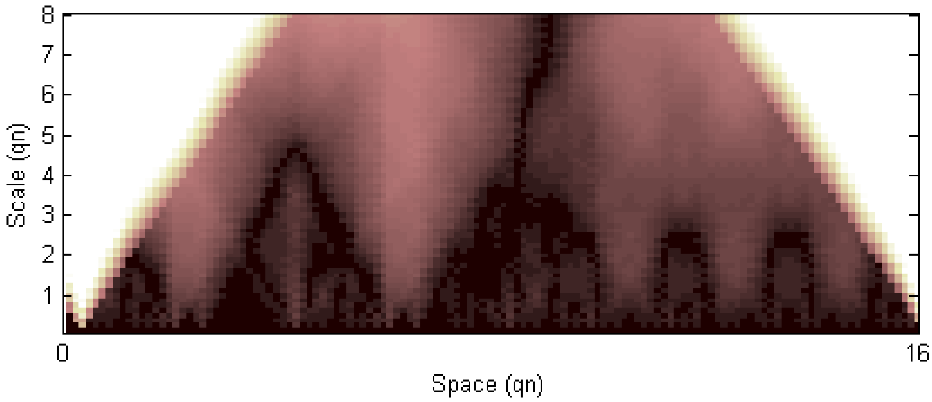

Figure 2. Visualization of the wavelet coefficients by Haar wavelet of the first 16 quarter notes of invention 1, upper voice, scales 1 to 8 quarter notes.

3.4 Classification

For classification we divided the inventions into sections. We then used the segments from two voices of one section *Sc*, labelled with their invention number, to define a classifier set *C*. We then test other sections from one invention to identify the invention they belong to by comparing their segment sets *C* to try and identify which invention they belong to with the k-nearest-neighbour method with k=1. For a segment set *Si*, composed of *n* patterns, the classifier iterates for each *si* in *Si* over all segments *sc* in *C* and determines the *sc* with the smallest distance to *si*. Two distance metrics have been selected for this task, the Euclidean distance given by

$$d_{sisc} = \sqrt{\sum_{j=1}^{n} (si_j - sc_j)(si_j - sc_j)'} \tag{4}$$

and the cityblock distance given by

$$d_{sisc} = \sum_{j=1}^{n} \left| si_j - sc_j \right| \tag{5}$$

Then, the label of *sc* is assigned to *si*. When all segments in *Si* are labelled the most frequent label of all *si* in *Si* is attached to *Si*, representing the detected source invention.

4. EXPERIMENTS

The experimental setup was designed to investigate the classification performance and the effect of segmentation and representation. We also studied the effect of including variations in the classifier, in particular inversion (flipping the pitch axis), retrograde (flipping the time axis), and retrograde inverted, which are typical contrapuntal techniques.

In order to answer these questions, we built the classifier on the expositions using both voices and tested it on different sections of the rest of the inventions, also using both voices. The classifier was built from the first 4, 8 and 16 qn, testing how much material is needed in this setup to classify successfully. After the first 16 qn each invention was divided into 3 sections of equal length, see Figure 3. We used the 4 combinations of representation and segmentation described in section 3.3. Constant segmentation and wavelet-based segmentation at scale 1 and 4 quarter notes each time. The resolution of vectors was set to 4 and 8 samples per quarter note. Finally, the nearest neighbour method was implemented testing the influence of the Euclidean and cityblock distances.

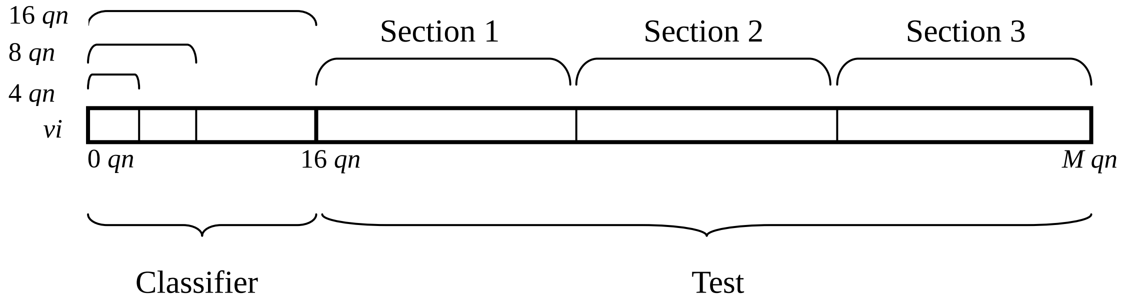

Figure 3. Scheme of classifier and test construction based on vector *vi*.

Figure 4 shows the average performance of each classifier based on the exposition considering all possible settings of the experiment, that is, resolutions, distances and scales. The best results were achieved by the combination of wavelet segmentation and wavelet representation, followed by the wavelet segmentation and pitch vector representation. In all cases, it is evident that the classifier becomes more precise as it has more information. The classifiers' performance decreases dramatically in the second and third sections when they are segmented by the constant segmentation, much lesser so by the wavelet approach.

Figure 5 is similar to Figure 4, except that this time each classifier is built using the exposition and additionally its variations. The classification rates are slightly changed by this technique. All classifiers are negatively affected, except the constant length segmentation and wavelet representation that show a slight improvement.

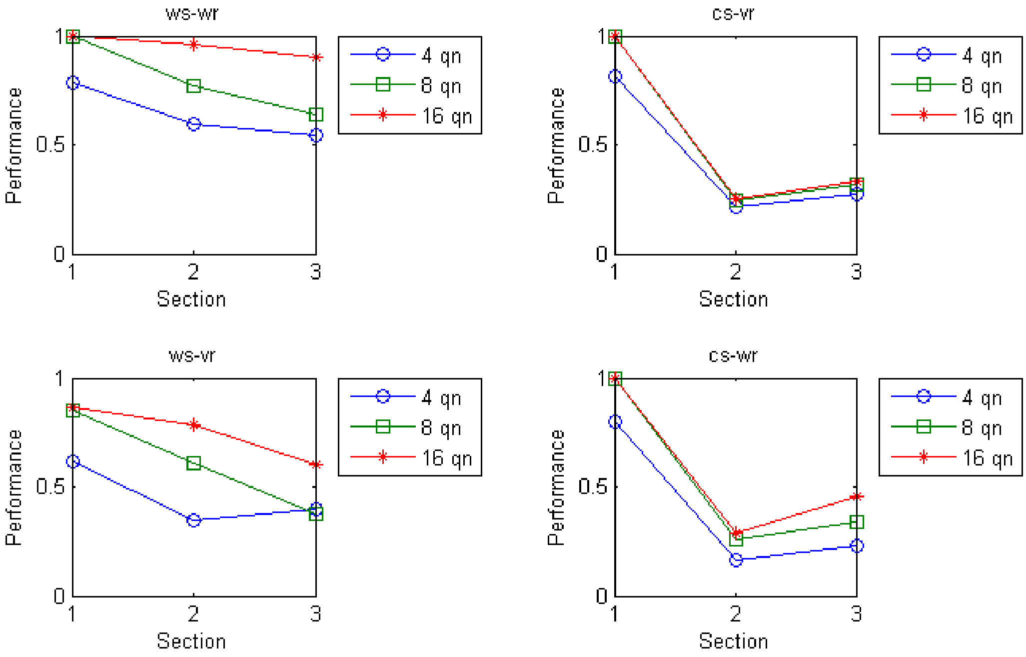

Figure 4. Performance of each section, classifier based on the exposition

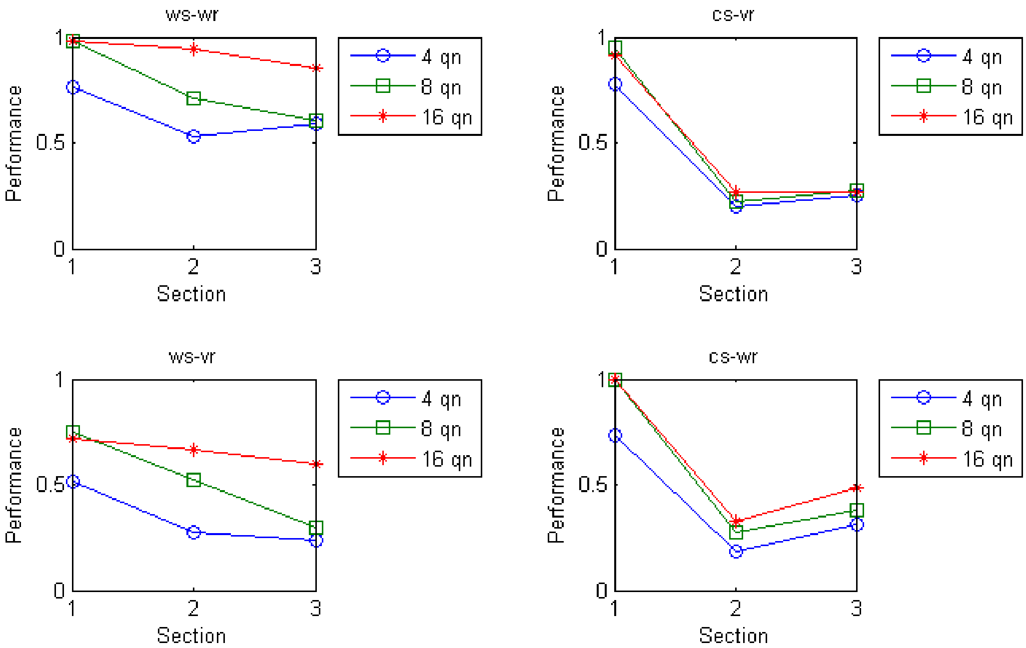

Figure 5. Performance of each section, classifier based on the exposition and its variations

The influence of scale, resolution and distance selection is presented in Tables 1 and 2. Table 1 shows the average performance of the classifier based on the exposition of the first 4, 8 and 16 qn, over the 3 sections. The results are not strongly affected by the selection of distance or resolution, but by the selection of scale. Table 2 shows the average performance of the classifier based on the exposition and its variations of the first 4, 8 and 16 qn, over the 3 sections.

Table 1. Performance resume for classifier based on the exposition. Scale in quarter notes, resolution in number of samples per quarter note and distance as metric type.

scale-resolution-distance	ws-wr %	ws-vr %	cs-vr %	cs-wr %
1-8-euclidean	94.81	64.44	54.81	57.78
1-8-cityblock	94.81	65.93	54.07	58.52
4-8-euclidean	77.78	80.74	49.63	51.11
4-8-cityblock	80.00	80.74	54.81	54.81
1-4-euclidean	88.15	38.52	48.15	54.81
1-4-cityblock	88.89	37.04	47.41	55.56
4-4-euclidean	57.78	59.26	43.70	34.81
4-4-cityblock	56.30	57.78	42.96	37.04

Table 2. Performance Resume for classifier based on the exposition and variations. Scale in quarter notes, resolution in number of samples per quarter note and distance in metric type.

scale-resolution-distance	ws-wr %	ws-vr %	cs-vr %	cs-wr %
1-8-euclidean	88.15	49.63	45.19	61.48
1-8-cityblock	91.11	50.37	46.67	59.26
4-8-euclidean	79.26	74.81	53.33	54.81
4-8-cityblock	77.04	73.33	52.59	57.04
1-4-euclidean	84.44	18.52	42.22	54.81
1-4-cityblock	83.70	18.52	42.22	54.07
4-4-euclidean	54.81	60.74	42.96	38.52
4-4-cityblock	60.74	61.48	41.48	37.04

5. DISCUSSION

The accurate classification based on the initial 16 qn with wavelet classification is better than we had expected. The relatively poor results delivered by the constant based segmentation, even when combined to wavelet representation, indicate that this segmentation approach is not appropriate to the musical structure. More sophisticated segmentation approaches may improve the results for pitch vector representation. The rather positive results provide some support that the zero crossings of the wavelet coefficients indicate musically meaningful segment boundaries. However we do not have a way to predict the appropriate scale for this purpose. Furthermore, we have currently no musically intuitive explanation why this segmentation method works as well as it does. More experiments are needed to determine whether this is specific to the dataset.

The good results of wavelet representation indicate that wavelet coefficients represent aspects of the melody that are relevant for melodic similarity and that map well onto the smallest distance, either Euclidean or cityblock. This effect may be specific to the music of Bach or the baroque style, or to the Haar wavelet or both. The strong influence of the scale parameter may be due to the scale becoming too large in relation to the length of the segments being classified.

The experiments also showed a small positive effect when including inversion, retrograde and inverted retrograde variations for constant segmentation and wavelet representation. All other configurations, showed a small deterioration with this practice.

6. CONCLUSION

This paper presents a novel approach, based on the continuous wavelet transform for melody representation and segmentation in symbolic music classification. This approach was tested on J.S. Bach's 15 two-part inventions. The classification results using wavelet-based segmentation and representation of melodic segments worked much better than the baseline pitch-vector representation and constant segmentation; thus indicating that wavelet representation is robust against melodic variations. Our method works well with simple classification and similarity models. Interestingly including contrapuntal variations led to worse results when used with the wavelet representation and segmentation.

The results indicate that wavelets do capture musically relevant properties of symbolic melodies. However, this method remains to be compared with more sophisticated similarity and segmentation approaches, on larger datasets, other classification methods and other tasks, such as genre or composer detection, to draw more general conclusions on the suitability of wavelets for representing melodies.

7. REFERENCES

[1] A. Pinto: Eigenvector-based relational motif discovery", International Conference on Music Information Retrieval, pp 207-212, Utrecht, 2010.
[2] A. Pinto, Indexing melodic sequences via wavelet transform, in Proc. ICME, pp.882-885, 2009

[3] C. Anagnostopoulou and G. Westermann: Classification in Music: A computational model for paradigmatic analysis. In: Proceedings of the International Computer Music Conference, pp 125-128, Thessaloniki, Greece, 1997.
[4] D. Conklin. Melodic analysis with segment classes. Machine Learning (2006) 65:349–360.
[5] E. Cambouropoulos and G. Widmer. Automated Motivic Analysis via Melodic Clustering. Journal of New Music Research 29(4):347-370.
[6] G. Velarde: Pattern Identification in Melody via Wavelets. Late breaking Demo, ISMIR 2010.
[7] J.S. Bach Inventions BWV 772-786, MIDI encodings ed. by Steve Rasmussen, last accessed April 2011 http://www.musedata.org/encodings/bach/rasmuss/inventio/
[8] J.S. Bach. Inventionen und Sinfonien. Ed by Ulrich Leisinger. Wiener Urtext, 2007.
[9] J.-J. Nattiez. Fondements d'une Sémiologie de la Musique. Union Généerale d'Editions, Paris. 1975.
[10] K. Adiloglu, T. Noll and K. Obermayer. A paradigmatic approach to extract he melodic structure of a musical piece. Journal of New Music Research, 35(3), pp 221–236, 2006.
[11] M. Li and R. Sleep. Melody classification using a similarity metric based on kolmogorov complexity. In: Proceedings of the Sound and Music Computing Conference (SMC'04), October 20-22, 2004, Paris, France., 2004
[12] L. Smith, and H. Honing. Time–frequency representation of musical rhythm by continuous wavelets, Journal of Mathematics and Music, Vol. 2, No. 2, pp 81–97, 2008.
[13] L. Stein: Structure and Style: The study and analysis of musical forms. Summy-Birchard Music, Miami, 1979.
[14] P.J. Ponce de León and J. M. Iñesta: Musical style classification from symbolic data: a two styles case study. In: Selected papers from the proceedings of the computer music modeling and retrieval 2003, Lecture Notes in Computer Science, vol. 2771, pp. 167-177, 2004.
[15] R. Hillewaere, B. Manderick, and D. Conklin. Global feature versus event models for folk song classification. In ISMIR 2009:10th International Society for Music Information Retrieval Conference, Kobe, Japan, 729-733, 2009.
[16] S. Mallat. A wavelet tour of signal processing. Academic Press, Third Edition, 2009.