View metadata, citation and similar papers at core.ac.uk
brought to you by CORE
provided by OAR@UM

Filter-Based Approach for Ornamentation Detection and Recognition in Singing Folk Music

Andreas Neocleous[1,2(✉)], George Azzopardi[2,3], Christos N. Schizas[1], and Nicolai Petkov[2]

[1] Department of Computer Science, University of Cyprus, Nicosia, Cyprus
neocleous.andreas@gmail.com
[2] Johann Bernoulli Institute for Mathematics and Computer Science, University of Groningen, Groningen, The Netherlands
[3] Intelligent Computer Systems, University of Malta, Msida, Malta

Abstract. Ornamentations in music play a significant role for the emotion whi1ch a performer or a composer aims to create. The automated identification of ornamentations enhances the understanding of music, which can be used as a feature for tasks such as performer identification or mood classification. Existing methods rely on a pre-processing step that performs note segmentation. We propose an alternative method by adapting the existing two-dimensional COSFIRE filter approach to one-dimension (1D) for the automatic identification of ornamentations in monophonic folk songs. We construct a set of 1D COSFIRE filters that are selective for the 12 notes of the Western music theory. The response of a 1D COSFIRE filter is computed as the geometric mean of the differences between the fundamental frequency values in a local neighbourhood and the preferred values at the corresponding positions. We apply the proposed 1D COSFIRE filters to the pitch tracks of a song at every position along the entire signal, which in turn give response values in the range [0,1]. The 1D COSFIRE filters that we propose are effective to recognize meaningful musical information which can be transformed into symbolic representations and used for further analysis. We demonstrate the effectiveness of the proposed methodology in a new data set that we introduce, which comprises five monophonic Cypriot folk tunes consisting of 428 ornamentations. The proposed method is effective for the detection and recognition of ornamentations in singing folk music.

Keywords: Signal processing · Folk music analysis · Computational ethnomusicology · Performer classification · Mood classification · Ornamentation detection · Ornamentation recognition · COSFIRE

1 Introduction

A common technique for expressing emotions in music performance is the addition of short notes to the main melody. These notes are called ornamentations and can be arbitrarily or systematically inserted. A glissando, also known as

© Springer International Publishing Switzerland 2015
G. Azzopardi and N. Petkov (Eds.): CAIP 2015, Part I, LNCS 9256, pp. 558–569, 2015.
DOI: 10.1007/978-3-319-23192-1_47

vibrato, for instance, is a rapid alteration of a series of consecutive notes. It is one of the most frequent ornamentations. Several other ornamentations are also used such as amplitude variation called tremolo and stretching or shortening the duration of notes, among others. In Fig. 1a we present the audio signal of the song Syrinx for solo flute by Claude Debussy. The vertical lines indicate a note with a tremolo ornamentation, which is shown enlarged in Fig. 1b.

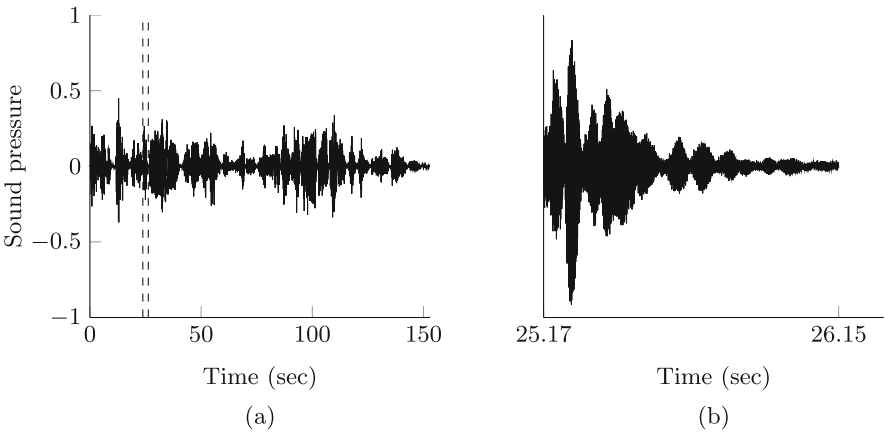

Fig. 1. (a) The sound pressure signal of the song Syrinx for solo flute by Claude Debussy. (b) Tremolo ornamentation that is characterized by amplitude modulation.

The importance of ornamentations in music has been researched and described by music theorists [1]. They are related to the feeling which a performer or a composer aims to create. In the field of music information retrieval (MIR), it has been shown that musicians have a unique way to perform their ornamentations and that it is a distinctive feature for performer identification [2–4].

The majority of these related studies are focused on applications in Western music and they mainly analyse ornamentations of musical instruments rather than singing voice [5–7]. Furthermore, the methodology of these studies require note segmentation. This creates additional computational difficulties since note segmentation is still a challenging problem in MIR.

In [8] the authors initially attempt to segment notes in traditional flute performances. Then they explore knowledge about ornamentations within a segmented note. They propose to separate ornamentations into two categories namely single-note and multi-note, which in turn are composed of two and three-sub categories, respectively.

In this work we are interested in ornamentation detection and recognition of folk music of the Eastern countries. It is generally harder to process Eastern folk songs than Western music from a signal processing point of view. Since folk music is frequently recorded in an environment such as a public place, which may

cause low quality recordings covered with background noise. Also, the singers in folk music are usually not professionals and therefore they may sing out of tune, forgetting melodies and others. One major difference between Eastern and Western music is the frequency distance between consecutive notes. In Western music this difference is strictly logarithmically equal into twelve notes per octave. In Eastern music and especially in Makam this distance between notes is not equally distributed. An automated system that will be able to capture ornamentations from an audio signal will help to accumulate additional knowledge of non-Western folk music, such as Makam in Turkish and Arabic music.

We propose a novel filter-based algorithm for the automatic identification of ornamentations in singing folk music of Cyprus. It is adapted from the two-dimensional COSFIRE approach [9], which has been demonstrated to be effective in object localization and recognition in images.

This paper is organized as follows. First we introduce the proposed type of ornamentations in Section 2. In Section 3 we describe the proposed methodology and demonstrate its effectiveness in Section 4. Finally, we discuss certain aspects of the approach that we propose and draw conclusions in Section 5.

2 Types of Ornamentations

Ornamentations in music have been precisely defined mainly in Western music since the 17th century with extensive use in the Baroque period. Since then, the composers have been annotating their desirable ornamentations in the so-called musical score. A musicological study can gather significant information from such transcriptions such as note frequency and duration, rhythm, tempo and others.

In folk music usually the composer is not known, hence there is no written score or any similar information about a song. The music is transmitted orally and mutates over time. Therefore, significant information about ornamentations is stored in the available recordings.

In this paper we make an attempt to create meaningful categories that describe the type of each ornament. We adopt the terms “single-note” and “multi-note” ornamentations from [8] and we introduce some additional sub-categories based on the music theory. We propose the following sub-categories within the single-note category: linear positive, linear negative, glissando positive and glissando negative (Fig. 2a). The major feature of this category is that only one alteration of a small note is done. A multi-note consists of alterations of more than one note and comprises two sub-categories: the vibrato positive and vibrato negative (Fig. 2b).

3 Methods

In this section we present the three main steps of the proposed methodology: feature extraction, configuration and application of the proposed 1D COSFIRE filters for ornamentation detection, followed by detection and recognition. In Fig. 3 we present the main steps of our methodology.

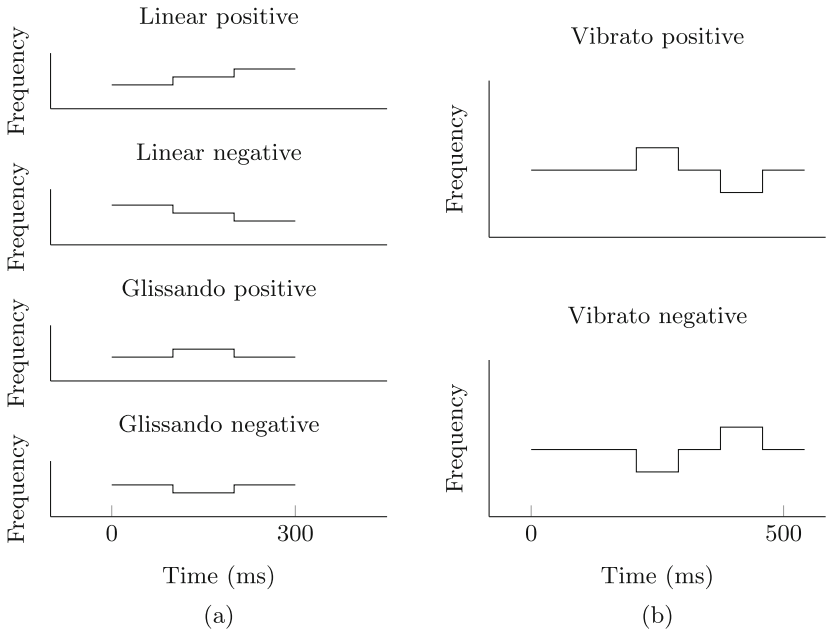

Fig. 2. Ornamentations which belong to (a) single-note and (b) multi-note subcategories. The frequencies can take any values in the range of singing voice.

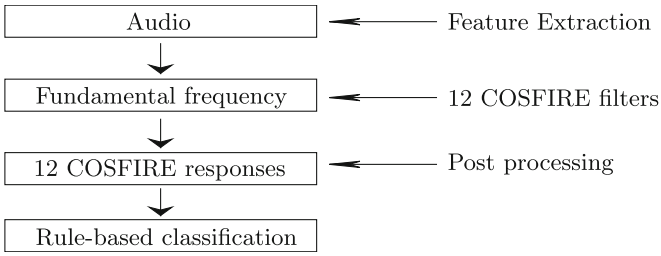

Fig. 3. The main steps of our proposed method. The audio signal is converted in to its fundamental frequency (pitch track). 12 COSFIRE filters are configured using the frequencies of the 12 western notes. The COSFIRE filters are applied to the pitch track, returning 12 responses, one for each COSFIRE filter. The responses are then binarized and post processed. Rule-based approach is used to identify and classify ornamentations.

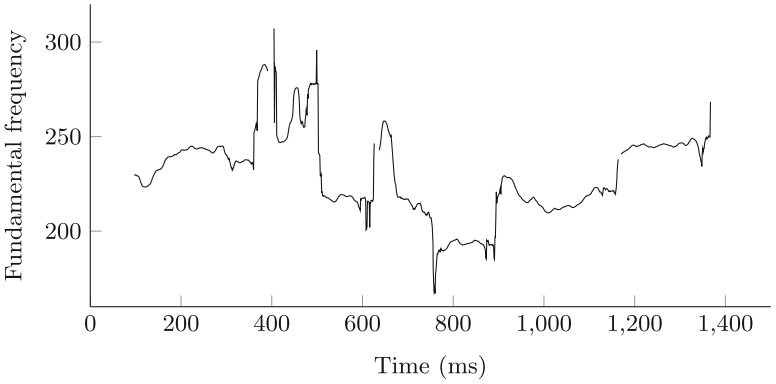

Fig. 4. Fundamental frequency as a function of time is used as input to our system.

3.1 Feature Extraction

Initially we segment the audio signal in overlapping frames of 33ms duration and 3ms overlap. Then we use the YIN algorithm to extract the fundamental frequency for each audio frame [10]. This method belongs to the time-domain based algorithms. First, the Autocorrelation function (ACF) is computed in each frame. Some of the peaks in the output of the ACF represent multiples of the period of the input signal. The repeating pattern of the audio signal is identified by choosing the highest non-zero-lag peaks of the ACF output, and process them with a number of modifications to return a candidate for the fundamental frequency. We refer to the output of the YIN algorithm as the "pitch track". In Fig. 4 we show the pitch track extracted from one of the songs in our data set. The discontinuities in the pitch track are caused by the inharmonicity of the signal at vocal pauses or other non harmonic sounds such as consonants.

We apply a post-processing step to the pitch track in order to correct some errors of YIN. The most frequent error is the so called "octave error". The algorithm erroneously chooses a candidate for a fundamental frequency that has a value in the higher or lower octave. More information about this post-processing method can be found in [11].

3.2 Configuration of a COSFIRE Filter

Originally COSFIRE filters were proposed as trainable filters for computer vision applications. Here we adopt that idea to 1D pitch tracks. A 1D COSFIRE filter uses as input the frequency values at certain positions around a specific point in time of an audio signal. The preferred frequency values and positions for which the resulting filter achieves a maximum value of 1 are determined in manual configuration process. In theory, a note has a single frequency that is constant over time. In practice however, the frequencies of a singing note vary slightly over time as illustrated by an example in Fig. 6a and the shape of the

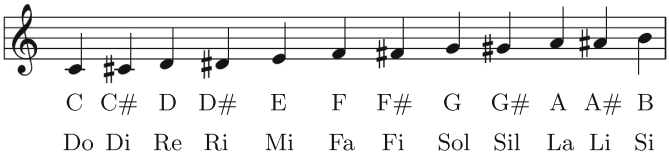

Fig. 5. The twelve notes of the Western music theory. The Western classical notation is shown first below the notes and the Spanish notation is shown second.

frequency signal is different every time the note is performed. Since one note has a specific duration and the theoretical fundamental frequency of a note is constant over time, we consider the same frequency value for a set of n points and we call this vector a prototype. The parameter n is set experimentally. We choose to configure prototypes with shorter durations as compared to the usual durations of performed notes. A COSFIRE filter which is configured with such a prototype, will return multiple strong responses along a note. This fact increases the chances of getting a strong response to the desirable position which in this case is a performed note.

We denote by $P = \{(f_i, \rho_i) \mid i = 1, \ldots, n\}$ a COSFIRE filter that is selective for a given prototype of n points. Each point i of the prototype is described by a pair (f_i, ρ_i), where f_i is the frequency of the note at position (temporal shift) ρ_i with respect to the midpoint of the prototype, where $\rho_i = i - (n+1)/2$. For instance, the COSFIRE filter that is configured by the note "A" of (n=) 5 points, results in the following set P:

$$P = \left\{ \begin{array}{l} (f_1 = 220,\ \rho_1 = \text{-2}), \\ (f_2 = 220,\ \rho_2 = \text{-1}), \\ (f_3 = 220,\ \rho_3 = 0), \\ (f_4 = 220,\ \rho_4 = 1), \\ (f_5 = 220,\ \rho_5 = 2) \end{array} \right\}$$

We configure 12 COSFIRE filters that are selective for the third octave of the 12 notes of the Western music theory as illustrated in Fig. 5. The frequencies in Hz of those 12 notes are the following: 130.8, 138.6, 146.8, 155.6, 164.8, 174.6, 185, 196, 207.7, 220, 233.1, 246.9, 261.6, 277.2, 293.7, 311.1, 329.6, 349.2.

3.3 Response of a COSFIRE Filter

We denote by $r_P(t)$ the response of a COSFIRE filter to a signal S at time t. We compute it by taking the geometric mean of the similarity values, which we obtain by a Gaussian kernel function, between the preferred fundamental frequencies f_i and the corresponding frequencies in the concerned neighbourhood.

$$r_P(t) = \left(\prod_{i=1}^{|P|} \exp \frac{-(f_i - S_{t+\rho_i})^2}{2\sigma^2} \right)^{\frac{1}{|P|}}, \sigma = \sigma_0 + \alpha(|\rho_i|) \quad (1)$$

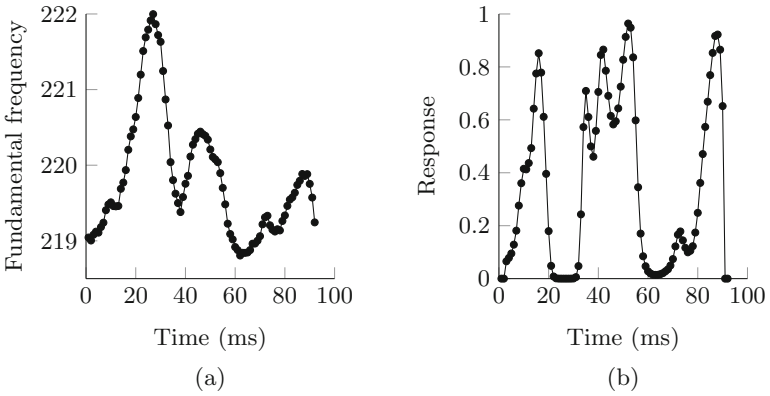

Fig. 6. (a) Fundamental frequency of an original ornamented note that is used as a test signal. (b) The response of the A-selective COSFIRE filter when applied to the test signal.

where $\sigma_0 = 0.5$ and $\alpha = 0.1$ are constant values that we set experimentally. In this way, the tolerance with respect to the preferred frequency increases with increasing distance from the support center of the COSFIRE filter at hand.

The signal in Fig. 6b shows the fundamental frequency of an original note "A" of 92 audio frames covering frequencies in the range between 218 and 222 Hz and we use it as a test signal. The COSFIRE filter P is applied in every position of the test signal and the response is shown in 6b. The maximum value of the response is achieved only at the point where the original note has the same values as the prototype. High responses are also achieved for signals that are similar to the prototype.

3.4 Post Processing

In Fig. 7a we show a part of a pitch track that was extracted from one of the songs in our data set. Below it, we illustrate the responses of three COSFIRE filters that are selective for the notes B, C and C# and are shown with thin solid lines.

Then, we binarize the responses using an absolute threshold of 0.01. We call a unit the consecutive binarized responses that have responses of 1. Such units are shown Fig. 7 (b-d) with dashed lines. In case during a time interval between two units there are no responses from other COSFIRE filters, we set all responses between those two units to 1. There is only one such an example in Fig. 7 which is obtained by the C-selective filter. If there is temporal overlap between units of different filters we only keep the unit with the longest duration.

Then we transform the binarized responses signals for an input audio signal into a symbolic representation of the sequence that every unit appears. In the example presented in Fig. 7 the symbolic sequence is [C.217, B.16, C.6, C#.13, C.6, B.11, C.5 C#.4]. The letters represent the names of the notes of the Western

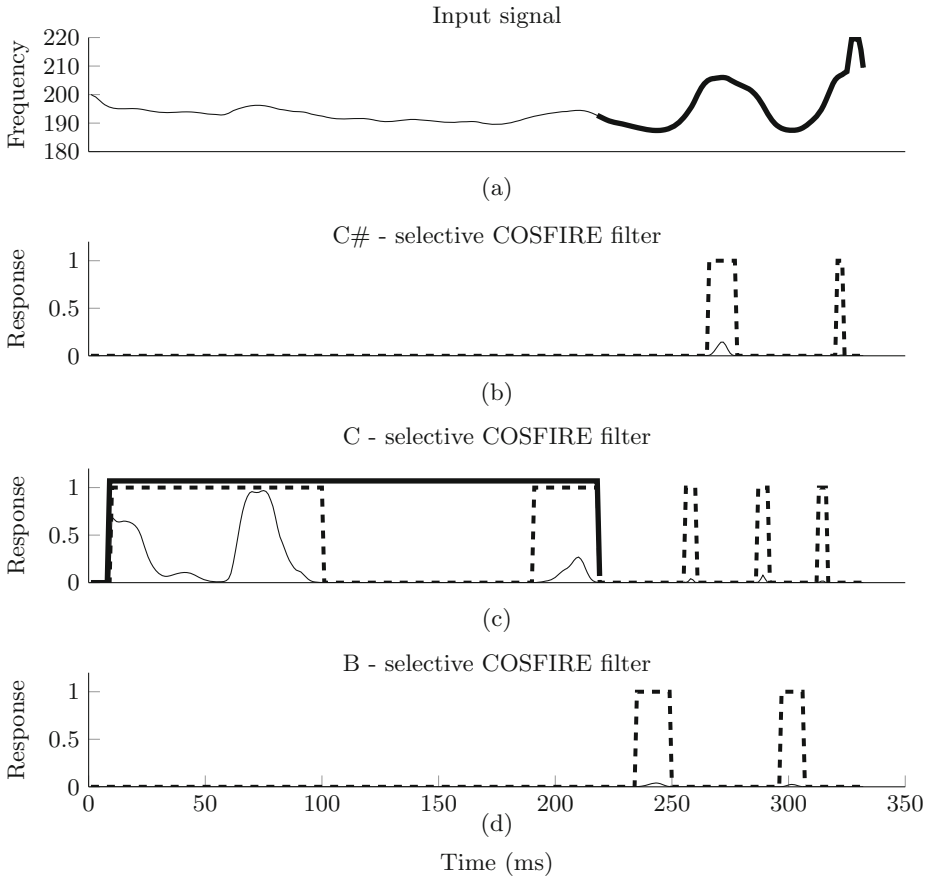

Fig. 7. Post processing procedure. (a) A part of a pitch track. The automatic detection and recognition of an ornamentation is emphasized in the second half of the signal. (b-d) The responses of three filters that were configured to be selective for the three Western notes C#, C and B. The binarized responses are shown with dashed lines. The thick line in (c) shows how consecutive units of a filter response are connected.

music and the number following the dot is the duration of each unit in number of frames. A vibrato is present at the second half with an alteration of short semitones around the note "C" and it is emphasized in Fig.7a.

3.5 Ornamentation Detection

In order to detect any type of ornamentation we first set a time threshold on the duration of every unit of the symbolic representation. The units that exceed the threshold are considered as notes and the remaining units are marked as parts of ornamentations. Typically, there is a sequence of such short units that consists of the entire ornamentation. In the example shown in Fig. 7 we present a note

whose second half is characterized by a vibrato. The evolution of the vibrato is emphasised in the pitch track. The first unit which belongs to the note "C" has a duration of 217 time frames while the remaining 7 units have duration of 16, 6, 13, 6, 11, 5 and 4 time frames. If we set a threshold of 30 time frames, the first unit will be considered as a note and the following sequence of 7 short notes will be considered to form an ornamentation.

3.6 Classification: Single-Note and Multi-note Ornamentations.

Each of the notes in Western music can be represented with numbers. For instance, the note A is always the first, hence it is represented with the number 1. Therefore, the symbolic representation shown in the example in Fig. 7 can be transformed as [4.217, 3.16, 4.6, 5.13, 4.6, 3.11, 4.5 5.4] where the number preceding the dot represents the respective note. In this example, the sequence of the notes that forms an ornamentation is: [3, 4, 5, 4, 3, 4, 5]. The classification of an ornamentation into a single-note or multi note is done with counting its peaks and its valleys. If an ornamentation has more than one peak or more than one valley then it is classified as multi-note. The ornamentation shown in Fig. 7 has two peaks and one valley and therefore it is classified as a multi-note ornamentation.

If an ornamentation is single-note, then we use four additional rules to classify it into four sub-categories shown in Fig. 2. The linear positive and the linear negative do not have peaks or valleys in the sequence and we use the sign of the derivative to decide. For instance, a linear positive ornamentation is detected by the symbolic sequence: [3, 4, 5]. The glissando positive has only one peak while the glissando negative has only one valley. The two subcategories of the multi-note are the vibrato positive and the vibrato negative. They are identified if the ornamentation starts with a positive direction or a negative direction respectively. For example, a vibrato positive is described by the symbolic sequence: [3, 4, 3].

4 Experiments and Results

4.1 Data Set

We created a data set of five Cypriot folk songs with total duration of 403 seconds containing 428 ornamentations. They are monophonic singing voice recordings encoded with 44.1kHz sampling frequency. We refer as the ground truth data the positions that are around in the middle of an ornamentation. They are manually annotated by the first author of this paper who is an experienced musician. The list of the songs used to validate our method is given in Table 1. From the 428 ornamentations, 270 are single-note and 158 are multi-note. The data set is available online[1].

[1] https://www.cs.ucy.ac.cy/projects/folk/

Table 1. The list of songs used for the validation of our method. Information about the duration, the number of ornamentations and their type is included.

	Manes	Agapisa	Panw xorio	Giallourika	Sousa	Sum
Duration (ms)	84.6	112.3	78.7	74.9	53.3	403.8
Ornamentations (#)	84	123	81	85	55	428
Single-note	56	85	46	50	33	270
Linear positive	9	24	20	10	6	50
Linear negative	7	14	6	17	6	50
Glissando positive	32	42	15	20	16	131
Glissando negative	2	5	5	3	2	17
Multi note	28	38	35	35	22	158
Vibrato positive	23	29	20	32	12	116
Vibrato neg	5	8	15	3	10	41

4.2 Results

The results are summarised in Table 2 in terms of precision, recall and F-measure. The manual annotation was done by setting a marker in the middle of every ornamentation. Then, we consider a true positive when this marker lies anywhere between the predicted start and end positions of an ornamentation. The false positives are considered when there is no pre-annotated marker between the predicted start and end positions and we count the false negatives when there is a marker with no predictions around it.

In Fig. 8 we illustrate a precision-recall plot by changing the time threshold that is used to identify ornamentations from notes. The value of the precision and recall when F-measure reaches maximum is indicated with a dot marker and it has a value of 0.76. We obtain the same results when we used COSFRE filters selective for note signals of length 5, 11, 15, 19 and 21ms.

Table 2. Results. The Ornamentation in row 1 contain results that are made of the results in subsequent lines.

	P	R	F-Score
Ornamentations	0.83	0.71	0.76
Single	0.7	0.56	0.63
Single 1	0.7	0.4	0.52
Single 2	0.63	0.62	0.62
Single 3	0.71	0.45	0.55
Single 4	0.29	0.65	0.4
Multi	0.61	0.35	0.45
Multi 1	0.61	0.48	0.54
Multi 2	0.32	0.76	0.45

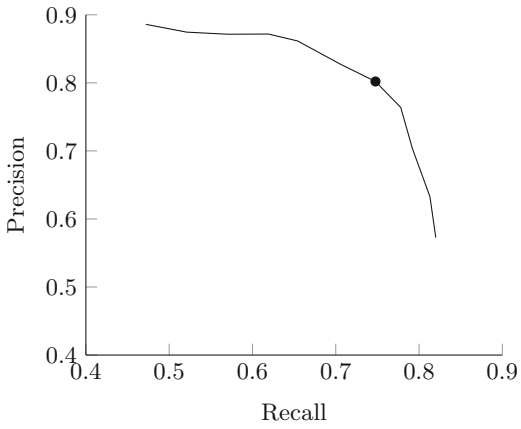

Fig. 8. Precision-recall plot obtained by varying the time threshold used to distinguish notes from parts of ornamentations.

5 Discussion and Conclusions

The COSFIRE filters are sensitive to amplitude tolerance, which is essential for non-stationary signals such as the singing voice. They are conceptually simple and easy to implement. They have been successful in pattern recognition of images in several applications including traffic sign detection and recognition [9]. They can be used for other applications such as note segmentation since they are able to capture note changes. The modelling of the pitch track with COSFIRE filters can also be used for the identification of repeating melodies.

In order to compare the proposed 1D COSFIRE filters with an already established method for the identification of similar signals, we cross correlate the same 12 prototypes with the pitch tracks of our data set. We observe that this system returns similar values for all the 12 prototypes and therefore it is not effective for ornamentation detection. Also, the results shown in Table 2 are comparable to the ones reported in [6]. Even though the database used in [6] is different than the one used in this study, we consider that our method is less complex since we avoid the note segmentation which is a challenging task itself. This study is a preliminary work on the ornamentation detection and classification. The classification of the multi ornamentations did not yield significant results, although this is due to the complexity of the problem. We aim to improve the classification stage that is currently rule-based to more sophisticated machine learning techniques. In future work we will attempt to compare our method with other databases used in previous work of other people. The results will be reported in another study. The above, demonstrate that there is a lot of work to be done in this research topic.

The contribution of our work is three-fold. First, it avoids note segmentation which is still a challenging and complex problem. Second, it uses only one feature,

the fundamental frequency and third it is effective with complex musical signals such as singing voice. The use of the pitch track as the main input feature to our method contributes to a system that is fast and less complex as compared to other methods [6–8]. The results that we obtain for the ornamentation detection and recognition are promising and a validation of additional folk music will be done in another study.

Acknowledgments. This research was funded from the Republic of Cyprus through the Cyprus research promotion foundation and also supported by the University of Cyprus by the research grant ΑΝΘΡΩΠΙΣΤΙΚΕΣ / ΑΝΘΡΩ / 0311(BE) / 19.

References

1. Taylor, E.: The AB guide to music. The Associate Board of the Royal Schools of Music (Publishing), London (1989)
2. Ramrez, R., Maestre, E., Pertusa, A., Gmez, E., Serra, X.: Performance-based Interpreter Identification in Saxophone Audio Recordings. IEEE Transactions on Circuits and Systems for Video Technology **17**, 356–364 (2007)
3. Ramrez, R., Prez A., Kersten S., Maestre E.: Performer identification in celtic violin recordings. In: International Conference on Music Information Retrieval (2008)
4. Ramrez, R., Maestre E.: A framework for performer identification in audio recordings. In: International Workshop on Machine Learning and Music - ECML-PKDD (2009)
5. Boenn, G.: Automated quantisation and transcription of musical ornaments from audio recordings. In: Proc. of the Int. Computer Music Conf. (ICMC), pp. 236–239 (2007)
6. Casey, M., Crawford, T.: Automatic location and measurement of ornaments in audio recording. In: Proc. of the 5th Int. Conf. on Music Information Retrieval (ISMIR), pp. 311–317 (2004)
7. Gainza, M., Coyle, E.: Automating ornamentation transcription. In: IEEE Int. Conf. on Acoustics, Speech and Signal Processing (2007)
8. Kokuer, M., Jancovic, P., Ali-MacLachlan, I., Athwal, C.: Automated detection of single - and multi-note ornaments in irish traditional flute playing. In: 15th International Society for Music Information Retrieval Conference (ISMIR) (2014)
9. Azzopardi, G., Petkov, N.: Trainable COSFIRE filters for keypoint detection and pattern recognition. IEEE Transactions on Pattern Analysis and Machine Intelligence **35**, 490–503 (2012)
10. Cheveigne, A., Kawahara, H.: Yin, a fundamental frequency estimator for speech and music. Journal of the Acoustical Society of America **111**(4), 1917–1930 (2002)
11. Panteli M.: Pitch Patterns of Cypriot Folk Music between Byzantine and Ottoman Influence. Master thesis. University of Pompeu Fabra (2011)

View publication stats