IEEE MELECON 2002, May 7-9, 2002, Cairo , EGYPT.

On Detecting Note Onsets in Piano Music

Matija Marolt, Alenka Kavcic, Marko Privosnik, Sasa Divjak

Faculty of Computer and Information Science, University of Ljubljana
Trzaska 25, 1000 Ljubljana, Slovenia
Phone:(386) 1 4768483, Fax: (386) 1 4264647
matija.marolt@fri.uni-lj.si

Abstract

This paper presents a brief overview of our researches in the use of connectionist systems for transcription of polyphonic piano music and concentrates on the issue of onset detection in musical signals. We propose a new technique for detecting onsets in a piano performance. The technique is based on a combination of a bank of auditory filters, a network of integrate-and-fire neurons and a multilayer perceptron. Such structure introduces several advantages over the standard peak-picking onset detection approach and we present its performance on several synthesized and real piano recordings. Results show that our approach represents a viable alternative to existing onset detection algorithms.

Keywords: music transcription, onset detection, neural networks.

1. INTRODUCTION

Transcription of polyphonic music (polyphonic pitch recognition) is a process of converting an acoustical waveform into a parametric representation, where notes, their pitches, starting times and durations are extracted from the waveform. Transcription is a difficult cognitive task and is not inherent in human perception of music. It is also a very difficult problem for current computer systems. Separating notes from a mixture of other sounds, which may include notes played by the same or different instruments or simply background noise requires robust algorithms with performance that should degrade gracefully when noise increases.

In recent years, several transcription systems have been developed. Some of them are targeted to transcription of music played on specific instruments [2-4], while others are general transcription systems [1]. Onset detection is an integral part of transcription systems, as it helps to determine exact onset times of notes in the transcribed piece. Some authors use implicit onset detection algorithms [2,3], while others, including us, chose to implement a separate onset detection algorithm to improve the accuracy of onset times.

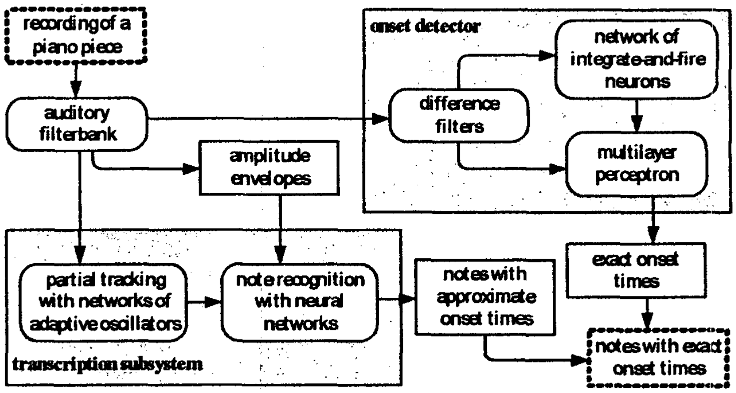
When we reviewed the structure of most current transcription systems, we were surprised by the fact that few systems use machine learning algorithms in the transcription process. Therefore, our motivation was to develop a transcription system based on neural networks, which have proved to be useful in a variety of pattern recognition tasks. We tried to avoid explicit symbolic algorithms, and instead used connectionist approaches in different parts of our system, including onset detection, which is the main topic of this paper.

2. SONIC

The name of our transcription system is SONIC. Transcription is a difficult task, so we put one major constraint on the system: it only transcribes piano music, so piano should be the only instrument in the analyzed musical signal. We didn't make any other assumptions about the signal, such as maximal polyphony, minimal note length, style of transcribed music or the type of piano used. The system takes an acoustical waveform of a piano recording (44.1 kHz sampling rate, 16 bit resolution) as its input. Stereo recordings are converted to mono. The output of the system is a MIDI file containing the transcription. Notes, their starting times, durations and loudness' are extracted from the signal.

Most current transcription systems have similar structures. First, a time-frequency representation of the input signal is calculated. Then, a partial tracking algorithm is used to discover partials of instrument tones. Finally, the found partials are associated with notes. Onset detection is used in some systems as a separate subsystem that improves the accuracy of onset times of detected notes.

SONIC has an analogous structure; its general overview is given in figure 1. The main distinction to existing approaches is that neural networks are used in partial tracking, note recognition and onset detection stages. The partial tracking and note recognition stages were presented elsewhere [5] and will not be discussed in this paper. We dedicate the next two sections to the onset detection subsystem implemented within SONIC,


recording of a piano piece
auditory filterbank
amplitude envelopes
onset detector
network of integrate-and-fire neurons
difference filters
multilayer perceptron
exact onset times
partial tracking with networks of adaptive oscillators
note recognition with neural networks
transcription subsystem
notes with approximate onset times
notes with exact onset times

and present its performance on several synthesized and real piano recordings.

3. ONSET DETECTION

3.1 Overview

Note onsets play an important role in the perception of music. Studies showed that onsets play a pivotal role in the perception of timbre, as it is much more difficult to recognize the timbre of tones with removed onsets [6]. Onsets also make it easier to detect new information in music; we can detect tones with pronounced onsets well before we can determine their pitch.

In a music transcription system, an onset detection algorithm is needed to correctly determine the starting times of notes in the transcribed signal. Several authors use an implicit onset detection scheme in their systems and make the onset time of a note equal to the time of its finding [2,3]. At first, we used a similar solution, but later abandoned it as it didn't produce accurate results, especially for notes in lower octaves, where delays of several 10ths of milliseconds were very common. Such timing deviations lead to unpleasant effects when listening to resynthesized transcriptions, and also made performance evaluation of the entire system very difficult, as one had to take such deviations into consideration. We have therefore decided to add a separate onset detection algorithm to our system.

Detection of onsets in a monophonic signal is not a difficult problem, especially if onsets are prominent, as is the case with piano tones. Onsets in a monophonic piano signal could be determined with high accuracy by simply locating peaks in the amplitude envelope of the input signal. In polyphonic music, such an approach fails, because the amplitude envelope of an entire signal reveals little of what is going on in individual frequency regions of the signal, where individual note onsets and offsets may occur. Therefore, a common characteristic of several current onset detection algorithms is that the input signal is first split into several frequency bands. Onset detection is then performed in each band separately and in the end, the would-be onsets in each band are merged into the final result.

Many researches in onset detection have been made in the field of beat and rhythm tracking [7,8]. Unfortunately, these algorithms are not accurate enough to be used in transcription systems, as they only discover very prominent onsets in a signal. Better approaches were used by Klapuri and Scheirer in their transcription systems [1,9]. As both are very similar, we will present a short overview of the onset detector used by Klapuri [1]. The algorithm first splits the input signal into 21 frequency bands with a bank of bandpass filters. Amplitude envelopes are calculated in each band with a 100 ms half-Hanning smoothing filter. Then, a relative difference function is calculated on each amplitude envelope. Peaks in the difference function correspond to possible onsets in the input signal. Peaks larger than a predetermined threshold T1 are chosen as onset candidates in each band. Because a single onset can cause many closely-spaced peaks in each band, all peaks within a 50 ms time window are merged together. Then, the remaining peaks in all frequency bands are merged together, and a clustering procedure is again used to join peaks within a 50 ms time window. A psychoacoustic loudness model is also used in the process to calculate the amplitude of each peak. In the end, all peaks that fall below a certain threshold T2 are eliminated.

We have implemented Klapuri's algorithm ourselves and found that it is very sensitive to the choice of both thresholds T1 and T2. If the values are set too low, many spurious onsets are detected and vice versa; high values produce many missing onsets. In general, it is very difficult to determine threshold values that would produce overall good results on several piano pieces and we have therefore decided to follow a different path in designing our onset detector.

3.2 Onset Detection in SONIC

The onset detector is based on a model for segmentation of speech signals, as proposed by Smith [10]. The model is founded on psychoacoustic findings and is based on a network of integrate-and-fire neurons that detects possible onsets in the input signal. We extended Smith's model with a multilayer perceptron neural network to improve the reliability of onset detection.

The first phase of the model splits the signal into several frequency bands with a bank of auditory filters, which emulate the functionality of the basilar membrane in human inner ear. Auditory filters are bandpass IIR filters, their parameters were calculated from psychoacoustic findings [11]. We use them to split the signal into 22 overlapping frequency bands, each covering half an octave.

The signal in each of the 22 resulting frequency bands is full-wave rectified and processed with the following difference filter:

$$O(t)=\int_0^t (\exp(-\frac{t-x}{f_s t_s})-\exp(-\frac{t-x}{f_s t_l}))s(x)dx \qquad (1)$$

$s(x)$ represents the signal in each frequency band, f_s the sampling rate, t_s and t_l are two time constants. The filter calculates the difference between two amplitude envelopes; one calculated with a smoothing filter with short time constant t_s (6-20 ms), and the other by smoothing the signal with a longer time constant t_l (20-40 ms). The output of the filter has positive values when the signal rises and negative otherwise.

Figure 2 shows the output of a difference filter on an excerpt taken from Glenn Gould's interpretation of Bach's Two-part Invention No. 8 (Sony 6622). The upper left part of the figure shows the acoustical waveform of the entire signal, vertical lines show note onsets. The right part of the figure shows two amplitude envelopes, calculated in the frequency band that covers the range of frequencies between pitches of notes Gb4 and B4. Envelopes were calculated with different smoothing constants (6 ms and 20 ms) and the difference in the amount of smoothing is clearly visible. Output of the difference filter is shown in the lower left part of the figure. One can see that peaks in the filter output correspond to onsets of notes that fall within the Gb4-B4 frequency band. The last note (D4) falls outside of this frequency range, so its peak is not very prominent.

The main task of the onset detector is to determine which peaks in outputs of difference filters correspond to note onsets and which are the result of various noises or beating in the signal. Our onset detector performs this task with a combination of a network of integrate-and-fire neurons and a multilayer perceptron. Outputs of all 22 difference filters are first fed into a fully connected network of integrate-and-fire neurons. Each integrate-and-fire neuron i in the network changes its activity A_i (initially set to 0) according to:

$$\frac{dA_i}{dt}=O_i(t)-\gamma A_i \qquad (2)$$

where $O_i(t)$ represents output of the i-th difference filter, and γ describes the leakiness of integration. When A_i reaches a threshold, the neuron fires (emits an output pulse), and A_i is reset to 0. After firing, there is a period of insensitivity to input, called the refractory period (50 ms in our model). Firings of neurons provide indications of amplitude growths in frequency channels. Neurons are connected to all other neurons in

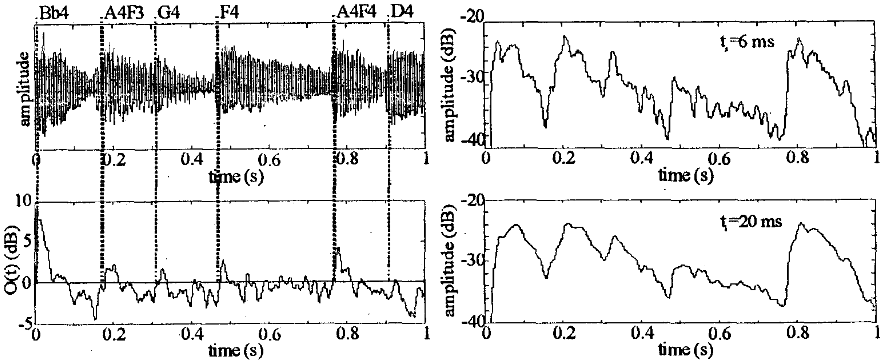


Figure 2: Input signal, amplitude envelopes and output of a difference filter

the network with excitatory connections. The firing of a neuron raises activities of all other neurons in the network and accelerates their firing, if imminent.

Onset discovery with a network of integrate-and-fire neurons provides two main advantages over classical peak-picking algorithms. Network connections cluster neuron firings, which may otherwise be dispersed in time, while at the same time the refractory period prevents neurons from generating a series of impulses at each onset. Connections also improve the detection of weak onsets, as they encourage firings of neurons that are close to the firing threshold, but would not fire without additional help.

A network of integrate-and-fire neurons outputs a series of impulses indicating the presence of onsets in the signal. Not all impulses are onsets, since various noises and beating can cause amplitude oscillations in the signal (see figure 3 in the next section). We use a multilayer perceptron (MLP) neural network to decide which impulses represent onsets. Inputs of the MLP consist of activities of integrate-and-fire neurons and several other factors, such as amplitudes of individual frequency bands. The MLP only has one output, which indicates if an onset occurred in the signal. We trained the network to recognize note onsets on a set of synthesized piano pieces and tested it on a mixture of synthesized and real piano recordings. The performance of the entire onset detection system is presented in the next section.

4. PERFORMANCE EVALUATION

We tested the algorithm on a set of synthesized and real piano pieces. The average score of the system was 98% of correctly found onsets and 2% of spurious onsets (onsets that were found; but were not present in the input signal). We present results on three real and synthesized piano pieces in table 1.

piano piece	no. of onsets	missed onsets	spurious onsets
1	4793	51 = 1.1%	3 = 0.1%
2	1305	37 = 2.8%	3 = 0.2%
3	963	10 = 1.0%	2 = 0.2%
4	786	25 = 3.1%	13 = 1.6%
5	206	13 = 6.3%	6 = 2.9%
6	556	0	8 = 1.4%

Table 1: Performance statistics on three synthesized and three real piano recordings

The synthesized pieces used are: (1) J.S. Bach, Partita no. 4, BWV828 (Fazioli piano), (2) P.I. Tchaikovsky: Miniature Overture from The Nutcracker, (Bösendorfer piano), (3) S. Joplin in S. Hayden: Kismet Rag (Steinway D40 piano). Real recordings are: (4) J.S. Bach: English suite no. 5 (BWV810), 1. movement, performer Murray Perahia (Sony Classical SK 60277), (5) F. Chopin, Nocturne Op. 9/2, performer Artur Rubinstein (RCA: 60822), (6) S. Joplin, The Entertainer, performer unknown (MCA 11836).

Results on synthesized recordings are generally better than those on real recordings. A large number of missed notes are notes played in very fast passages or in ornamentations such as thrills and fast arpeggios (most missed notes in Bach's Partita (1)). The main cause of such misses is the refractory period of integrate-and-fire neurons, which prevents them from firing and thus detecting onsets in very fast pace. The system also often misses quietly played notes, masked by louder notes or chords occurring shortly before or after the missed onset.

Poorer onset detection accuracy on real recordings is a consequence of several factors. Recordings contain

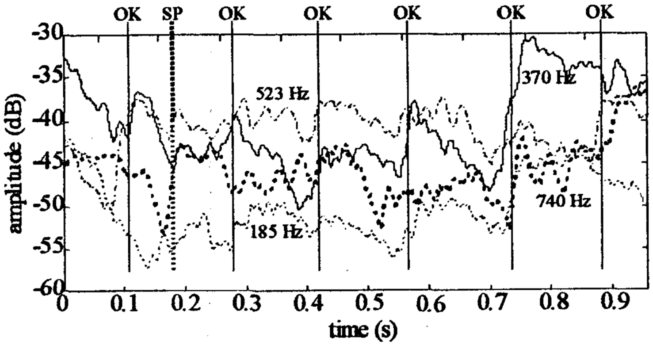


Figure 3: Detection of a spurious onset

reverberation and more noise, while the sound of real pianos includes beating and sympathetic resonance. Furthermore, performances of piano pieces are much more expressive, they contain increased dynamics, more arpeggios and pedaling. All of these factors make onset detection more difficult. Still, we are satisfied with our algorithm's performance. The causes of missed notes are similar to the ones we mentioned when looking at synthesized recordings; the increased dynamics of performances is the main factor that contributes to a larger percentage of missed notes. A good example of this is Chopin's Nocturne (5), where a distinctive melody is played over very quiet, sometimes barely audible left hand chords, which are often missed.

The larger percentage of spurious notes in real recordings is a result of more noise and piano imperfections, such as beating. An example of spurious note detection is given in figure 3. The figure represents amplitude envelopes of four frequency bands calculated on a one second excerpt of Bach's English suite (4). Vertical lines represent onsets found by the system. Six onsets were correctly found (OK), together with one spurious onset (SP). The spurious onset occurred because of a large amplitude increase in the 185 Hz frequency band for which there is no obvious explanation.

5. CONCLUSION

In this paper, we presented our approach to detection of note onsets in a polyphonic piano performance. The approach is based on a connectionist paradigm and employs a bank of auditory filters and a network of integrate-and-fire neurons, coupled with a multilayer perceptron. By using a connectionist approach to onset detection, we tried to avoid threshold problems that occur with standard "peak picking" algorithms. We presented performance statistics of our system on several synthesized and real piano recordings. Results show that connectionist approaches represent a good alternative in building onset detection systems and should be further studied. The presented onset detection algorithm brought a large improvement in the overall performance of our transcription system and we do not plan to improve it further. Our further researches will be directed to improvements of other parts of the transcription system, and will include an improved method for discovering repeated notes and an addition of a feedback mechanism to the currently strictly feed-forward transcription approach.

References

[1] Klapuri, A., Automatic Transcription of Music. M.Sc. Thesis, Tampere University of Technology, Finland, 1997.

[2] Rossi, L., Identification de Sons Polyphoniques de Piano. Ph.D. Thesis, L'Universite de Corse, France, 1998.

[3] Sterian, A.D., Model-based Segmentation of Time-Frequency Images for Musical Transcription. Ph.D. Thesis, Univesity of Michigan, 1999.

[4] Dixon, S., "On the computer recognition of solo piano music," in Proceedings of Australasian Computer Music Conference, Brisbane, Australia, 2000.

[5] Marolt, M., "Adaptive oscillator networks for partial tracking and piano music transcription", in Proceedings of the 2000 International Computer Music Conference, Berlin, Germany, 2000.

[6] Martin, K.D., Sound-Source Recognition: A Theory and Computational Model. Ph.D. Thesis, MIT, USA, 1999.

[7] Goto, M., Muraoka, Y., "Music understanding at the beat level: Real-time beat tracking system for audio signals," in Readings in Computational Auditory Scene Analysis, Mahweh, NJ: Laurence Erlbaum, 1998.

[8] Scheirer, E.D. Music-Listening Systems, Ph.D. Thesis, MIT, USA, 2000.

[9] Scheirer, E.D. Extracting expressive performance information from recorded music, M.Sc. Thesis, MIT, USA, 1995.

[10] Smith, L.S., "Onset-based Sound Segmentation," in Advances in Neural Information Processing Systems 8, Touretzky, Mozer and Haselmo (ers.), Cambridge, MA: MIT Press, 1996.

[11] R. D. Patterson, J. Holdsworth, "A functional model of neural activity patterns and auditory images," in Advances in speech, hearing and auditory images, W.A. Ainsworth (ed.), London: JAI Press, 1990.