Proceedings of the 7th WSEAS International Conference on SIGNAL PROCESSING, ROBOTICS and AUTOMATION (ISPRA '08)
University of Cambridge, UK, February 20-22, 2008

Note separation of polyphonic music by energy split

KRISTÓF ACZÉL[1], ISTVÁN VAJK[2]
Department of Automation and Applied Informatics
Budapest University of Technology and Economics
3-9. Műegyetem rkp. , H-1111, Budapest
HUNGARY

Abstract: - A polyphonic musical recording is the superposition of many separate tracks which are downmixed to fewer channels. Decomposition of such a signal to separate instrument tracks or notes has always been a challenge. It is theoretically impossible to extract the component tracks without the information that was lost at the superposition. In this paper a new way of sound separation of mono-aural digital recordings is proposed where lost information is recovered by using a model of real instruments in order to make the separation of individual musical notes possible.

Key-Words: - sound separation, instrument print, polyphonic music, energy split

1 Introduction

A musical recording typically is a polyphonic material that is the composition of many individual notes originating from a number of instruments. The recording is made using many microphones. However their output signals are then downmixed to fewer, typically one or two channels. Once the recording is downmixed, no correction can be made to the individual notes in it. Our long term interest in sound separation is motivated by the problem of correcting these recordings if they have incorrect notes (either in frequency, volume, intonation, length etc).

We allow a reasonable amount of user input and processing time in favor of separation quality. User input involves entering the musical score (note starting/ending times, frequencies, used instruments). Although due to the nature of real-life music this input will never be 100% accurate, it can be precise enough for getting a first estimate on the note parameters in the recording, instead of relying on musical transcription ([1], [2], [3], [4], [5]) and instrument recognition ([6], [7], [8], [9]) algorithms which are inferior to the performance of a human listener.

The complexity of separation can be attributed to the fact that the information to be retrieved is actually missing from the signal. This issue made researchers approach the problem in many different ways. [10] is a sound source separation algorithm that requires no prior knowledge on the instrument notes in the recording, and performs the task of separation based purely on azimuth discrimination within the stereo field. The results are impressive. However, separating individual notes is not in the focus, only instrument groups.

[12], [13], [14] describe a method which separates harmonic sounds by applying linear models for the overtone series of the sound. The method is based on a two-stage approach: after applying a multipitch estimator to find the initial sound parameters, more accurate sinusoidal parameters are estimated in an iterative procedure. Separating the spectra of concurrent musical sounds is based on the spectral smoothness principle [11].

Beamforming techniques [15] along with the Independent Component Analysis framework offer a different way of separation. A large array of microphones is employed for recording an event. The travel time of the time signal and the difference of the recorded signals are then used in calculations that increase the receiver sensitivity in the direction of wanted signals and decrease the sensitivity in other directions. However, these methods rely on certain preliminary conditions and studio setup to achieve good results.

The following sections will guide the reader through the separation process, each section covering one important block of the algorithm. Section 2 shows an overview of these blocks and the signal flow in the system. The whole separation process is carried out in frequency domain, thus the first step is converting the signal. This step is covered in section 3 along with the inverse transformation used at the end of the separation process. Section 4 describes the instrument model that is used in the system, while section 5 covers the details of the actual separation process, the

Simplified Energy Split (SES) method. Missing information on note intonation ('playmode') and volume levels must be calculated prior to the actual separation. Section 6 deals with the most obvious solution for the problem.

2 Overview of the separation process

This section shows an overview of the separation process. Short descriptions of the building blocks are given which are discussed in detail in later sections.

First, all time-domain signals are converted to frequency domain. Windowed FFT is employed for this purpose along with Frequency Estimation ([21]) and Phase Memory ([18]) methods.

The separation is aided by stored instrument samples. These samples are stored one by one in bandogram format which is basically a spectrogram split to subbands, in which the energy is summed.

Bandograms originating from instrument notes are stored in the sample store. One instrument will have a number of samples differing by their frequency and playmode. This makes possible to select the sample that best matches the note in the recording to help the separation process. The collection of bandogram samples from the same instruments will be called an *instrument print*.

The playmode and volume detector receives the original recording, the musical score and the instrument prints. Its role is to select one sample from each instrument print that best fits the original instrument in the recording.

The Simplified Energy Splitter gets the spectrogram of the recording and the bandogram of the selected samples as input. It splits the energy in the spectrogram of the original recording to components that resemble the input bandograms.

Finally, the frequency-domain signals are converted back to time domain.

Figure 1 shows the block-diagram of the sound separation process. The following notation is used for different representations of signals:

- **Simple waveform (W)**
- **Simple FFT (S):** spectrogram storing $c_{k,t}$ amplitude and $\varphi_{k,t}$ phase for each bin.
- **Frequency estimated spectrogram (F):** $c_{k,t}$ amplitudes and $\varphi_{k,t}$ phases remain the same as those of simple FFT, but an $f_{k,t}^{true}$ true frequency value is stored in addition for each bin.
- **Bandogram (B):** A spectrogram split to subbands, in which the energy is summed. Only these sums are stored, no detailed information on bin amplitudes and no phase information either.

The following chapters will guide the reader step by step through the separation process. Each section will cover one block in detail. Due to size limitations, however, some of the blocks will not be exhaustively discussed in this paper.

3 Transformation to frequency domain

This section proposes an easy, yet powerful algorithm that is able to generate a spectrogram of the recording that is much more precise for musical analysis than the conventional FFT spectrogram.

Earlier literature [16], [17] covered different transformation methods in order to determine the best possible means for analysis of audio signals. Current research has examined the analysis of polyphonic musical signals in particular [18], [19].

In [21] a frequency estimation method is shown, that calculates true frequencies present in the original signal from subsequent phase values. For a frame starting at time t the FFT coefficients and phases are $c_{k,t}$ and $\varphi_{k,t}$, respectively. In this document the time index will be omitted in some of the equations for understandability. Two subsequent frames are needed by the algorithm for the calculation. Assuming that two subsequent frames start at t_1 and t_2, a true frequency f_{k,t_2}^{true} can be computed for each bin. The nominal frequency of the k[th] bin is

$$f_k = k\frac{samplerate}{framesize}.$$

The true frequency of each bin will deviate from this value, and can be expressed as:

$$f_{k,t_2}^{true} = f_k + \frac{\varphi_{k,t_2}^{dev}}{2\pi\cdot(t_2-t_1)},$$

with

$$\begin{aligned}\varphi_{k,t_2}^{expt} &= \varphi_{k,t_1} + (t_2-t_1)\cdot 2\pi f_k \\ \varphi_{k,t_2}^{dev} &= \varphi_{k,t_2} - \varphi_{k,t_2}^{expt} + l\cdot 2\pi\end{aligned},$$

where φ_{k,t_2}^{expt} is the phase of bin k in time t_2; φ_{k,t_2}^{expt} is the expected phase; φ_{k,t_2}^{dev} is the deviance between the expected and measured phase; f_{k,t_2}^{true} is the estimated true frequency of bin k in time t_2 and $l\in Z: -\pi < \varphi_{k,t_2}^{dev} \le +\pi$. The greater the time

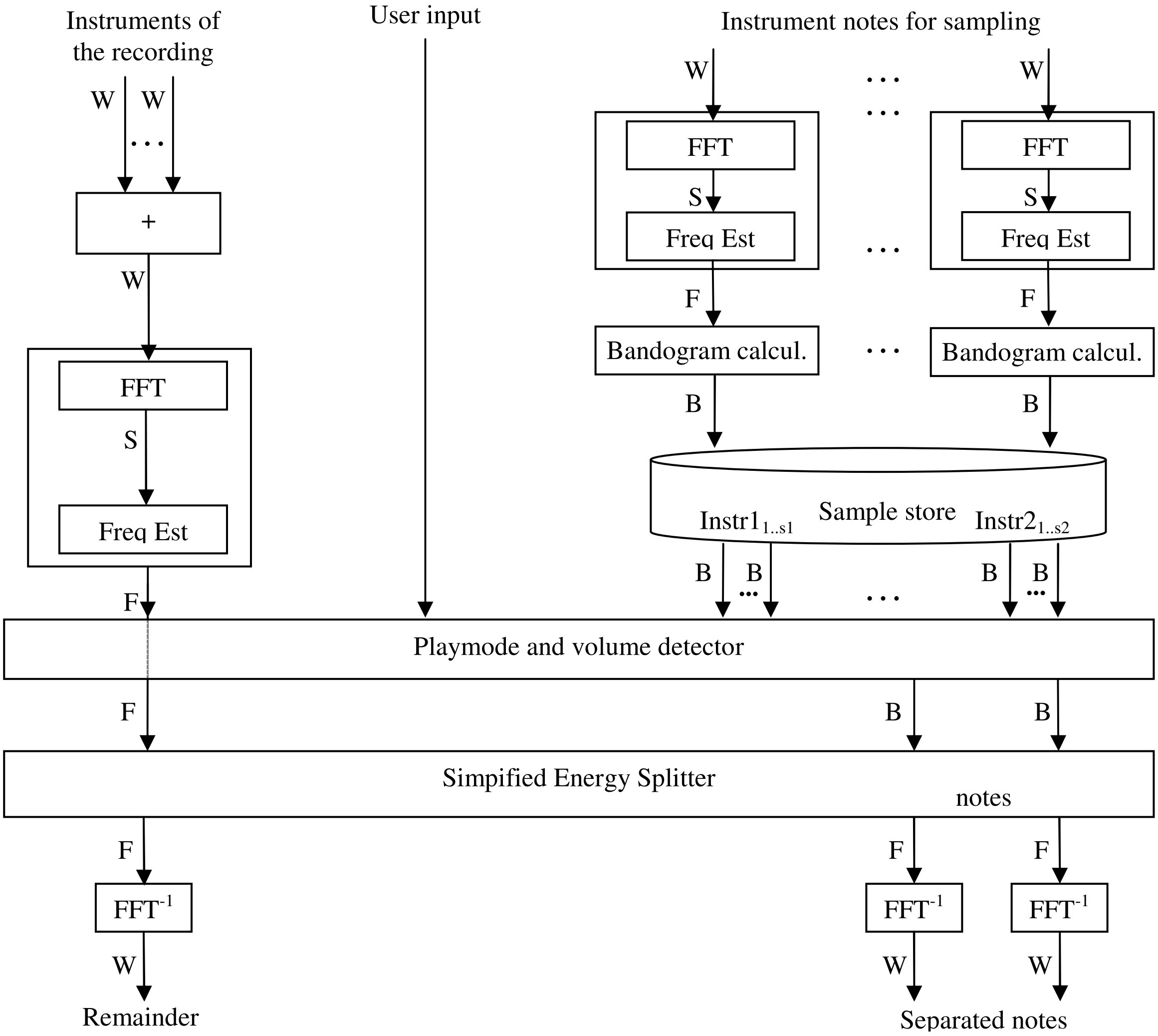

Figure 1: Signal flow and block diagram of the separation process

difference between the start of the frames the more precise the estimated value of f_{k,t_2}^{true}. On the other hand, big time differences limit the maximum detectable distance between f_{k,t_2}^{true} and f_k.

Sometimes, mainly for lower frequencies or complex signals with many components, f_{k,t_2}^{true} will fluctuate around the real value that is present in the original signal. The true frequencies for periodic waves can be found more precisely by taking the weighted average of the last *m*, current and next *m* phase deviations. This extension will be referred to as Phase Memory (PM) and the new PM estimated true frequency can be calculated as

$$\hat{f}_{k,t_2}^{true} = f_k + \frac{\hat{\varphi}_{k,t_2}^{dev}}{2\pi \cdot (t_2 - t_1)},$$

where

$$\hat{\varphi}_{k,t_x}^{dev} = \frac{\sum_{x=-\infty}^{\infty} \varphi_{k,t_x}^{dev} \cdot c_{k,t_x} \cdot \vartheta(x)}{\sum_{x=-\infty}^{\infty} c_{k,t_x} \cdot \vartheta(x)},$$

where $\vartheta(x)$ denotes the weighing function. Finding optimal $\vartheta(x)$ is out of the scope now.

It is important to mention that, although the proposed PM method is a very effective tool for signal analysis, it is not used in the transformation back to time domain. In a usual case, where the target is only the isolation of notes, not even f_{k,t_2}^{true} is needed, a simple inverse FFT using c_{k,t_2} and φ_{k,t_2} values will accomplish the task. However, in cases like pitch shifting, where the separated notes (and thus f_{k,t_2}^{true} frequencies) are altered, it may be necessary to recalculate the φ_{k,t_2} phases from the new f_{k,t_2}^{true} values before the inverse FFT.

4 Instrument prints

The main complexity of sound separation lies in the paradox that we need to regain information from a signal that does not fully contain it. At some point we will definitely have to input additional information into the separation system to complete the missing data. Human listeners, who are known to be able to do the separation in their mind, use memories of instruments and memories of the notes in the musical piece being performed. This is their source of additional information. Copying nature has been proven to be the right approach many times. This section shows a way of implementing a memory of known instruments, trying to mimic the way the human brain works.

The representation used to store instrument features will be called an instrument print. [22] presents experiments that examine the dynamic attributes of timbre evaluating the role of onsets in similarity judgments. It also gives an overview of researches pursuing the identification of the most important properties of instrument sounds that make a human listener able to distinguish between them. The instrument prints in this paper are in part based on these researches, in the sense that they contain the features that were found important in the experiments mentioned. However, separation purposes require more information on instruments than pure identification does.

An instrument print contains samples from an instrument on different frequencies and with different intonations, 'playmodes'. The term 'playmode' refers to the way the instrument was played, e.g. the hardness of a piano key hit, the blowing strength of the flute or the intonation of a saxophone note. One print can have more than one playmode dimensions, depending on the way an instrument can be played. These cannot always be defined by mathematical definitions, very often they can only be expressed by subjective terms (e.g. sharpness, warmth etc.). The instrument print is a collection of samples on different frequencies *f* and also with different values in the playmode space $\underline{\mathbf{M}} = [m_1, m_2, ...m_p]$. It can be regarded as a function

$$\underline{\mathbf{A}}(\underline{\mathbf{M}}, f_k, f_{base}, t)$$

showing how amplitudes through the frequency range change over time for a specific note at f_{base} frequency, played with a playmode $\underline{\mathbf{M}}$, with the conditions $t, m_x, f_{base} \in \mathrm{R}^+$, $0 < m_x < m_{x,\max}$, $0 \le t < \infty$ and $0 < f \le 20000 Hz$.

In reality, a sample will not store a continuous spectogram, only a bandogram which represents the sum energy characteristics in certain frequency subbands that are aligned on a logarithmical frequency scale. One sample is calculated from a sound signal containing one note as

$$A_{\mathbf{M}, f_{base}, b, t} = \sum_{f_{base} \cdot 2^{\frac{b-0,5}{R}} < \hat{f}_{k,t}^{true} < f_{base} \cdot 2^{\frac{b+0,5}{R}}} c_{k,t} ,$$

where

$$b = \left\lfloor \log_{\sqrt[R]{2}} \frac{f_{base}}{\hat{f}_{k,t}^{true}} \right\rfloor$$

identifies the specific subband, while R is an experimental value defining the resolution in frequency range, that is, the number of subbands per octave. Experiments showed that $R = 12$ provides good enough resolution in log frequency.

The number of stored instrument samples is finite both in frequency and playmode spaces. Missing samples will be interpolated from the existing ones when needed.

5 The Simplified Energy Splitter

This section describes the heart of the separation process, the Simplified Energy Splitter (SES). First, the main issue of separation will be briefly presented. Since the original decomposition problem cannot be solved due to the lack of information, further on a certain simplification will be proposed that, although lowers the quality, makes it possible to carry out the separation even under these circumstances. By applying this change the separation problem will be simplified to an energy split problem. Finally, the reader will be guided through the implementation of the split process itself.

The equation system of the original separation problem is

$$\underline{\mathbf{c}}_{r\tau} = \sum_{\forall i} \underline{\mathbf{s}}_{i,r\tau}^{orig} ,$$

where $\underline{\mathbf{c}}_{r\tau} = [c_{r\tau,n} \cdot e^{\gamma_{r\tau,n}}]$ is the mixed signal which is the input of the separation algorithm; $\underline{\hat{\mathbf{s}}}_{i,r\tau} = [s_{i,r\tau,n}^{orig} \cdot e^{\sigma_{i,r\tau,n}}]$ are the original notes. Time is now represented as $t = r\tau$, where r stands for the current frame and τ is the time difference between subsequent frames. The above undetermined system of equations cannot be solved unambiguously without other constraints to add.

Unfortunately our knowledge on the original notes is rather limited. No precise information is available on the base frequency $f_{base\_i}(r\tau)$ of the original notes, their starting/ending times and their playmode which also changes over time. Each original note will 'resemble' one instrument sample in our database to some extent, but there are no perfect matches ever. It is obvious that under these circumstances we will not be able to decompose the recording $\underline{\mathbf{c}}$ to an array of signals that are perfect replicas of the original $\underline{\mathbf{s}}_i^{orig}$ ones. The target is to decompose it to signals that resemble the original ones, or – lacking the original notes – at least the samples that are used in the separation.

The term 'resemblance' is of course an expression taken from real life rather than an exact mathematical measure. In the case of an automated algorithm, however, it must be defined in an exact manner in order to be able to interpret and validate the outcome of the separation algorithm. Due to space considerations the definition of resemblance is not discussed in this paper.

As the original separation problem cannot be solved, simplifications have to be made. The most obvious change is eliminating the unknown $\hat{\sigma}_{i,0\tau,k}$ phases from the equation system:

$$\hat{\sigma}_{i,r\tau,k} = \gamma_{r\tau,k}$$

This rephrases the original problem to

$$\underline{\mathbf{c}}_{r\tau} = \sum_{\forall i} \underline{\hat{\mathbf{s}}}_{i,r\tau} + \underline{\hat{\mathbf{c}}}_{r\tau}$$

where $\underline{\hat{\mathbf{s}}}_i$ is separated note i and $\underline{\hat{\mathbf{c}}}$ is the remainder of the recording. This perceptually motivated modification exploits the fact that the human ear does not differentiate by the phase of the heard sinusoids, we only hear magnitude differences.

Of course any modifications will have a smaller or greater impact on the quality of the separation, causing artifacts in the output. The modified equation does not handle periodic signals with closely located frequencies well. If two or more signals cancel each other, this effect will also appear in the separated notes. Experiments showed that this tradeoff is acceptable in most cases.

In the energy split step bandograms of the right samples will be used to recreate spectrograms of the notes to be separated from the remaining part of the recording. Semi-linear decomposition will be used. Assuming that we know the exact frequency, volume and playmode of a certain note that we want to separate, the following iterative algorithm can be proposed to divide the energy between the target notes. We start out with the original Frequency Estimated FFT image of the recording. In each step a fraction of the energy of the selected samples is transferred from the FFT of the recording to the FFT of the separated notes. This ensures a fair division of the energy of the recording.

Let $\underline{\hat{\mathbf{c}}}_{[0],0,r\tau} = \left[c_{[0],0,k,r\tau}\right] = \underline{\mathbf{c}}$ denote the initial energy residing in the recording. $\underline{\hat{\mathbf{s}}}_{[d],i,r\tau}$ will denote the current energy in the separated notes, being $\underline{\hat{\mathbf{s}}}_{[0],i,r\tau} = \underline{\mathbf{0}}$ initially. Each step $[d]$ contains i substeps, in which an α fraction of the energy in the reference sample A_i is transferred from the current remaining energy $\underline{\hat{\mathbf{c}}}_{[d],i,r\tau}$ to the separated note $\underline{\hat{\mathbf{s}}}_{[d+1],i,r\tau}$ if still possible, as in

$$c_{[d],i,k,r\tau} = \begin{cases} f_{base} \cdot 2^{\frac{b-0,5}{R}} < \hat{f}_{k,r\tau}^{true} < f_{base} \cdot 2^{\frac{b+0,5}{R}} : \\ \qquad \delta\left(\hat{c}_{[d],i-1,k,r\tau}\left(1 - \dfrac{\alpha_{[d],b,i,r\tau}}{D}\right)\right) \\ \text{otherwise}: \\ \qquad \hat{c}_{[d],i,k,r\tau} \end{cases}$$

with

$$\alpha_{[d],b,i,r\tau} = \frac{A_i\left(M_i, f_{base\_i}, b, r\tau - T_{i,start}\right)}{\sum\limits_{f_{base} \cdot 2^{\frac{b-0,5}{R}} < \hat{f}_{k,r\tau}^{true} < f_{base} \cdot 2^{\frac{b+0,5}{R}}} \hat{c}_{[0],0,k,r\tau}}$$

The current energy in note i (which is being isolated) can be calculated in step $[d]$ as

$$\underline{\hat{\mathbf{s}}}_{[d+1],i,r\tau} = \underline{\hat{\mathbf{s}}}_{[d],i,r\tau} + \left(\underline{\hat{\mathbf{c}}}_{[d],i-1,r\tau} - \underline{\hat{\mathbf{c}}}_{[d],i,r\tau}\right)$$

which is the starting value of the next $[d+1]$ step for separated notes, while

$$\underline{\hat{\mathbf{c}}}_{[d+1],0,r\tau} = \underline{\hat{\mathbf{c}}}_{[d],I,r\tau}$$

is the starting value of the next step for the remaining energy in the recording, where I is the number of instruments in the time frame. With D denoting the number of steps, $\underline{\mathbf{c}}_{[D],0,r\tau}$ is the remaining part of the recording after the separation and $\underline{\hat{\mathbf{s}}}_{[D],i,r\tau}$ will represent the coefficients of instrument i after the separation.

6 Playmode detection

In the previous section all note parameters were assumed to be known. However, this is not the usual scenario. While the user can input the location of the instrument notes in frequency and time, he/she may not be capable of entering either the playmode matrix **M** or the volume. Further on the automatic detection of the playmode and volume will be covered.

To carry out the energy split step an optimal playmode matrix **M** must be found. For the sake of convenience the volume will also be incorporated in **M** from now on. **M** is by definition perfect if the separated notes are the perfect replicas of the parent instrument samples that were used in the energy split, and the remaining part is zero. In general, an **M** matrix is considered good provided the energy split step that uses **M** generates notes that 'resemble' their parent sample while letting $\mathbf{c}_{[D]}$ get as close to zero as possible. All combinations of **M** matrices will be tested for separation error, and the one causing the least error will be selected. Depending on the size and possible values of **M** the number of steps needed for finding the best combination of the instrument samples may require huge computational power. If we consider the playmode space to be continuous, it is not even possible to iterate through all the combinations. Finding an algorithm faster than brute force iteration, however, is out of the scope of this article.

7 Implementation and test scenario

This section shows an example scenario of a difficult problem in the area of sound separation. The results illustrate the quality of the method proposed for isolating notes in practice.

In the test case two instruments – a piano and a saxophone – played the tune in Figure 2. As shown, the two instruments played on the same frequency, which is one of the hardest cases of sound separation. The instruments were recorded separately, one after the other. Their signals were then mixed together and fed to the separation algorithm.

Figure 2: musical score of the analysed fragment

Instrument prints were also built from these two instruments by sampling all halftones in the frequency range in question at three different playmodes ('soft', 'neutral' and 'hard').

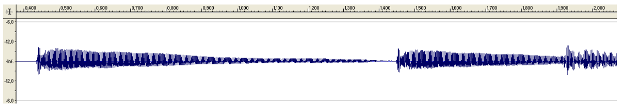

Figure 3: waveform of the original piano track

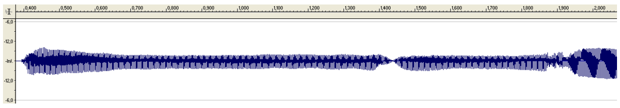

Figure 4: waveform of the original sax. track

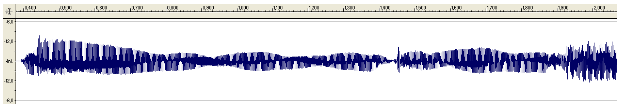

Figure 5: waveform of the mixed signal

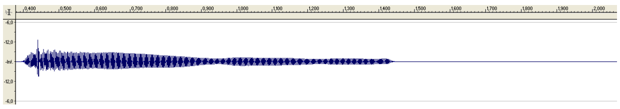

Figure 6: waveform of the first separated piano note

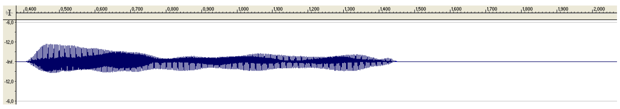

Figure 7: waveform of the first separated sax. note

Figure 3 to Figure 7 show the two input instruments, the mixed signal, and two separated notes. Human listeners confirmed that the separated notes did sound like real instruments, even if somewhat distorted.

8 Conclusion

The paper has shown a method for separating single instrument notes from a recording using pre-recorded instrument prints and the Simplified Energy Splitter algorithm. The results are quite promising. The example waveforms in Section 7 along with a few other separation samples can be downloaded from http://avalon.aut.bme.hu/~aczelkri/separation.

For recordings that only contain harmonically unrelated notes the algorithm provides very clear results. In real life, however, consonant notes with overlapping overtones are usually favored over dissonant ones. Our test results show that even in cases where some notes are located on each other's base or overtone frequencies the separation provides reasonably good results.

Acknowledgements:
This work has been supported by the fund of the Hungarian Research Fund (grant number T68370)

References:

[1] G. E. Poliner, D. P. W. Ellis, A. F. Ehmann, E. Gomez, S. Streich; B. Ong, Melody Transcription From Music Audio: Approaches and Evaluation, *IEEE Transactions on Audio, Speech and Language Processing*, Volume 15, Issue 4, May 2007, pp. 1247 – 1256

[2] H. Thornburg, R. J. Leistikow, J. Berger, Melody Extraction and Musical Onset Detection via Probabilistic Models of Framewise STFT Peak Data, *IEEE Transactions on Audio, Speech and Language Processing*, Volume 15, Issue 4, May 2007, pp. 1257 – 1272

[3] J. P. Bello, L. Daudet; M. B. Sandler, Automatic Piano Transcription Using Frequency and Time-Domain Information, *IEEE Transactions on Audio, Speech and Language Processing*, Volume 14, Issue 6, Nov. 2006, pp. 2242 – 2251

[4] A. T. Cemgil, H. J. Kappen; D. Barber, A generative model for music transcription, *IEEE Transactions on Audio, Speech and Language Processing*, Volume 14, Issue 2, March 2006, pp. 679 – 694

[5] S. A. Abdallah, M. D. Plumbley, Unsupervised analysis of polyphonic music by sparse coding, *IEEE Transactions on Neural Networks*, Volume 17, Issue 1, Jan. 2006 pp. 179 - 196

[6] A. A. Wieczorkowska, A. Czyżewski , Rough Set Based Automatic Classification of Musical Instrument Sounds, *Electronic Notes in Theoretical Computer Science*, Volume 82, Issue 4, March 2003, pp. 298-309

[7] J. C. Brown, Computer identification of musical instruments using pattern recognition with cepstral coeffcients as features, *J. Acoust. Soc. of Am.*, Vol. 105 (1999), pp. 1933-1941.

[8] P. Herrera, X. Amatriain, E. Batlle, X. Serra, Towards instrument segmentation for music content description: a critical review of instrument classification techniques, *Proc. of International Symposium on Music Information Retrieval (ISMIR 2000)*, Plymouth, MA, 2000.

[9] P. Herrera-Boyer, G. Peeters, S. Dubnov, 2003. Automatic classification of musical instrument sounds, *J. New Music Res.* Volume 32, 3–21.

[10] D. Barry, R. Lawlor, E. Coyle, Sound Source Separation: Azimuth Discrimination and Resynthesis, *Proc. of 7th International Conference on Digital Audio Effects*, DAFX 04, Naples, Italy, 2004.

[11] A. Klapuri, Multipitch estimation and sound source separation by the spectral smoothness principle, *Proc. of IEEE International Conference on Acoustics, Speech and Signal Processing*, Salt Lake City, USA, 2001.

[12] T. Virtanen, A. Klapuri, eparation of Harmonic Sounds Using Multipitch Analysis and Iterative Parameter Estimation, *Proc. of IEEE Workshop on Applications of Signal Processing to Audio and Acoustics*, New York, USA, 2001.

[13] T. Virtanen, A. Klapuri, Separation of Harmonic Sound Sources Using Sinusoidal Modeling, *Proc of. IEEE International Conference on Acoustics, Speech and Signal Processing*, Istanbul, Turkey, 2000.

[14] T. Virtanen, A. Klapuri, Separation of harmonic sounds using linear models for the overtone series, *Proc. of IEEE International Conference on Acoustics, Speech and Signal Processing*, Orlando, Fla, USA, May 2002.

[15] N. Mitianoudis, M. E. Davies, Using Beamforming in the audio source separation problem, *7th Int Symp on Signal Processing and its Applications*, Paris, July 2003

[16] R. Pintelon, J. Schoukens, *System Identification, A frequency domain approach*, ISBN 0-7803-6000-1, Wiley-IEEE Press, May 2001, pp. 33-44

[17] S. Gade, H. Herlufsen, Use Of Weighting Functions in DFT/FFT Analysis (Part I), *Brüel & Kjær Technical Review,* No. 3., 1987

[18] K. Aczél, Sz. Iváncsy, Musical source analysis with DFT, *Proc. of MicroCAD 2006 International Scientific Conference*, Miskolc, Hungary, March 2006.

[19] K. Aczél, Sz. Iváncsy, Musical source analysis: spectrogram versus cochleagram, in Press, *MicroCAD 2007 International Scientific Conference*, University of Miskolc (Miskolc, Hungary), March 2007.

[20] K. Aczél, Sz. Iváncsy, Sound separation of polyphonic music using instrument prints, *Proc of EUSIPCO 2007*, Poznan, Poland, Sept. 2007.

[21] S. M. Bernsee, Pitch Shifting Using the Fourier Transform http://www.bernsee.com/dspdimension.com/html/pshiftstft.html (11-11-2007)

[22] P. Iverson; C. L. Krumhansl, Isolating the dynamic attributes of musical timbre, The *J. Acoust. Soc. of America*, Volume 94, Issue 5, November 1993, pp.2595-2603