

EXPLORING PHASE INFORMATION IN SOUND SOURCE SEPARATION APPLICATIONS

Estefanía Cano, Gerald Schuller and Christian Dittmar

Fraunhofer Institute for Digital Media Technology
Ilmenau, Germany
{cano, shl, dmr}@idmt.fraunhofer.de

ABSTRACT

Separation of instrument sounds from polyphonic music recordings is a desirable signal processing function with a wide variety of applications in music production, video games and information retrieval. In general, sound source separation algorithms attempt to exploit those characteristics of audio signals that differentiate one from the other. Many algorithms have studied spectral magnitude as a means for separation tasks. Here we propose the exploration of phase information of musical instrument signals as an alternative dimension in discriminating sound signals originating from different sources. Three cases are presented: (1) Phase contours of musical instruments notes as potential separation features. (2) Resolving overlapping harmonics using phase coupling properties of musical instruments. (3) Harmonic percussive decomposition using calculated radian ranges for each frequency bin.

1. INTRODUCTION AND PREVIOUS WORK

1.1. Phase in Music Information Retrieval

The importance of phase in signals was thoroughly described by Oppenheim and Lim in the 1980s [1]. They describe different scenarios where important features of the signal are only preserved if the spectral phase, as opposed to spectral magnitude, is retained. Among other applications, the authors present different examples of images and speech where relevant information of the signal is retained in phase-only reconstructions where the spectral magnitude is either set to unity, randomly selected or averaged over an ensemble of signals. The contrasting case where synthesis is performed by preserving spectral magnitude with zero phase, i.e., magnitude-only reconstructions, shows to preserve far less of the relevant features of the signal and decrease intelligibility. In this sense, the authors argue that many of the important features preserved in phase-only reconstructions are due to the fact that the location of events, e.g., lines and points in images, is retained. Bearing in mind that spectral magnitude of speech and images tends to fall off at high frequencies, phase-only reconstructions with unity magnitude can be interpreted as a spectral whitening process where the signal will experience a high frequency boost that will consequently accentuate lines, edges, and narrow events without modifying their location.¹ Furthermore, the authors address the cases and conditions where

some or all the magnitude information of a signal can be extracted from its phase. The minimum-phase condition where a signal can be recovered within a scale factor from its phase is discussed and iterative techniques for this purpose are presented in [2]. For the particular case of discrete time signals a set of conditions is also presented: a sequence which is zero outside the interval $0 \leq n \leq (N-1)$ is uniquely specified to within a scale factor by $(N-1)$ samples of its phase in the interval $0 < \omega < \pi$ if it has a z-transform with no zeros on the unit circle or in conjugate reciprocal pairs. More recently, Dubnov has explored the use of phase information for musical instrument characterization, modelling, coding and classification. Based on the fact that second order statistics and power spectra are phase blind, he proposes the use of Higher Order Statistics (HOS) and their associated Fourier transforms, i.e., polyspectra, to describe phase variations that cannot be revealed by regular spectral analysis. Polyspectra are the mathematical generalization of the power spectrum maintaining not only magnitude but also phase information. In [3], Dubnov uses HOS to estimate sinusoidality for quality coding of musical instruments. His system is based on the fact that the analysis of sinusoidal harmonics leads to linear or almost linear phases as opposed to the analysis of stochastic harmonics which leads to random phases. In synthetic, perfectly periodic signals, spectral components are multiple integers of the fundamental frequency and thus, frequency coupled. Similarly, the relative phases of the harmonics follow the phase of the fundamental and consequently spectral components are phase coupled. These non-linear interactions between spectral components of the signal are assessed using bicoherence as a detector of frequency coupling and kurtosis as a measure of phase coupling. These measures are used to associate a certain frequency bin as noise or harmonic content. In [4] Dubnov further explores the use of non-linear polyspectra methods in the development of a Harmonic + Noise model for analysis and synthesis of vocal and musical instrument sounds. Once again, the use of bicoherence is proposed as a sinusoidality measure in each frequency band. This is based on the fact that spectral components of certain musical instruments can show considerable sinusoidal phase deviations without actually causing the spectral peak to become immersed in noise. Phase noise variance for harmonic partials is estimated using the phase coupling measure.

Dubnov & Rodet investigate in [5] the phase coupling phenomena in the sustained portion of musical instruments sounds. It is well known that acoustical musical instruments never produce waveforms that are exactly periodic. In this sense, two different conditions are analyzed: synchronous phase deviations of proportional magnitude which preserve phase relations between partials and asynchronous deviations which do not preserve phase relations and consequently change the shape of the signal. A measure of phase coupling called Quadratic Phase Coupling (QPC) is

¹ An example of a phase-only reconstruction of a speech signal can be found in http://www.idmt.fhg.de/eng/business%20areas/analysis_audio_signals.htm.

used and its equivalency under certain assumptions to the discrete Bispectrum, i.e., 2D Fourier transform of the third order cumulant function, is presented. Phase correlation is analyzed by calculating the instantaneous frequency by means of the unwrapped phase and obtaining the fluctuations around an ideal theoretical value derived from the fundamental frequency, i.e., $f_k = k \cdot f_o$. This procedure eliminates phase deviations due to vibrato and slight pitch changes. Flute, trumpet and cello sounds are analyzed and results suggest that different instruments and possibly different instrument families have distinct phase coupling characteristics, i.e., the trumpet signal exhibits high QPC values and thus strong phase coupling among partials whereas the flute signal shows some correlation but its phase deviations are mostly uncoupled.

In [6] Cont & Dubnov expand the concept of phase coupling in musical instrument sounds to a real time multiple pitch and multiple instrument recognition system. They propose the use of the modulation spectrum presented in [7] as it is a good representation of phase coupling in musical instruments, shows both short-term and long-term information about the signal and is a non-negative representation.

In [8] Paraskevas and Chilton present an audio classification system that uses both magnitude and phase information as statistical features. The problem of phase discontinuity is addressed and two different types of such discontinuities are described: extrinsic discontinuities caused by the computation of the inverse tangent function and intrinsic discontinuities that arise from properties of the physical system producing the data and that occur due to simultaneous zero crossing of the real and imaginary components of the Fourier spectrum. An alternative method to calculate phase which overcomes both discontinuity problems and uses the z-transform of the signal is proposed. The classification system was tested using gunshot signals and results show that there is 14% performance improvement for certain classes compared to the case where only magnitude features are used. Furthermore, classification rates are also evaluated with phase information only. In general, classification rates are lower for phase-only features than for magnitude-only features; however, certain classes show to be very well characterized by their phase information.

In [9] Woodruff, Li & Wang propose the use of common amplitude modulation (CAM), pitch information and a sinusoidal signal model to resolve overlapping harmonics in monaural musical sound separation. To estimate the amplitude of the overlapping harmonic, amplitude envelopes of sinusoidal components of the same source are assumed to be correlated. This means that the unknown amplitude can be approximated from the amplitude envelopes of non-overlapped harmonics of the same source. Pitch information is used to predict phase of the overlapped harmonic by calculating the phase change of the spectral component on a frame by frame basis:

$$\Delta_{\phi_n}^{h_n}(m) = 2\pi \cdot h_n F_n(m) T \quad (1)$$

where m denotes time frame, h_n and F_n harmonic number and fundamental frequency of source n respectively and T frame shift in seconds. A least-squares solution approach is used to obtain the sinusoidal parameters. The phase change prediction error is calculated and results show that reliable estimations can be obtained for lower numbered harmonics.

1.2. Sound Source Separation

Many algorithms designed for sound source separation are solely based on analysis and processing of the spectral magnitude information of an audio file. Virtanen for example proposes in [10] a separation algorithm based on Nonnegative Matrix Factorization (NMF) of the magnitude spectrogram into a sum of components with fixed magnitude spectra and time varying gains. The system uses an iterative approach to minimize the reconstruction error and incorporates a cost function that favors temporal continuity by means of the sum of squared differences between gains in adjacent frames. A sparseness measure is also included by penalizing nonzero gains. Fitzgerald et al. [11] have extensively explored the use of Nonnegative Tensor Factorization as an extension of matrix factorization techniques for source separation. Tensors are built with magnitude spectrograms from the different audio channels and iterative techniques are used to find the different components in the mix. Shift invariance in the frequency domain has been explored and a sinusoidal shifted 2D nonnegative tensor factorization (SSNTF) algorithm has been proposed where the signal is modeled as a summation of weighted harmonically related sinusoids. In [12] Burred uses the evolution of the spectral envelope in time together with a Principal Component Analysis (PCA) approach to create a prototype curve in the timbral space which is then used as a template for grouping and separation of sinusoidal components from an audio mixture. Every and Szymanski propose in [13] a spectral filtering approach to source separation. The system detects salient spectral peaks and creates pitch trajectories over all frames. The peaks are then matched to note harmonics and filters are created to remove the individual spectrum of each note from within the mixture. Ono et al. proposed in [14] a Harmonic/Percussive separation algorithm that exploits the anisotropy of the gradient spectrograms with an auxiliary function approach to separate the mix into its constituent harmonic and percussive components.

The remainder of this paper is organized as follows: Section 2 presents three scenarios where the use of phase information is relevant and three proposed algorithms are described. Section 3 presents some conclusions and a final discussion of possible future approaches. In Sections 4 and 5 acknowledgements and used references are presented.

2. PHASE IN SOUND SOURCE SEPARATION: THREE PROPOSED ALGORITHMS

Spectral magnitude can be informative, intuitive, and numerically simple. However, working with magnitude information in source separation presents several difficulties: many separation algorithms rely on the use of instrument models or spectral envelope templates which suffer from the large diversity in terms of recordings, playing styles, register, instrument models and performers which can make them unreliable. Furthermore, some algorithms rely on assumptions about the magnitude spectrum such as spectral smoothness or harmonicity which might not always be fulfilled. Some systems also rely on the use of pitch tracking algorithms to find the evolution of the harmonic components in time. Even though solid results can be achieved, performance of such pitch-tracking algorithms will suffer under noisy conditions.

As opposed to spectral magnitude, interpreting phase information is a more challenging task as it is not visually intuitive, presents numerical discontinuities that need to be dealt with and in its pure form might not be very informative. Phase on its own might not always be sufficient to achieve solid sound separation. However, it is our belief that phase can be complementary to

magnitude information, increase robustness and enhance performance in separation algorithms. Three scenarios will be presented where phase information has been used in separation tasks.

2.1. Phase contours of musical instrument notes as separation features

In general, the principle of Common Fate states that different parts of the spectrum that change in the same way in time will probably belong to the same environmental sound [15]. In this sense, two types of changes can be studied: frequency modulation changes and amplitude modulation changes. Amplitude modulation changes in sound separation applications have been studied in [10, 11, 12]. Here we are concerned with changes in the frequency and phase of harmonic components belonging to the same source: In a mixture of sounds, any partials that change in frequency in an exactly synchronous way and whose temporal paths are parallel on a log-frequency scale are probably partials that have been derived from a single acoustic source [15]. Furthermore, we explore the importance of Micromodulations in harmonic partials as a sound separation feature. Micromodulations refer to small frequency modulations that occur naturally in the human voice and musical instruments and that have potent effects on the perceptual grouping of the component harmonics.

Four different signals are studied: (1) C5 violin note, (2) C5 trumpet note, (3) C5 clarinet note and (4) C5 piano note. All signals are monophonic tracks with a sampling frequency $F_s = 44100$ Hz taken from the University of Iowa Musical Instruments Database [16]. A Hann-window 4096 samples long is used and a hop size $T = 512$. For the different frequency bins, the Fourier phase is differentiated in time and phase increments between time frames are found. Inherent discontinuities in phase values are resolved and kept within a $[-\pi, \pi]$ range. This procedure would be equivalent to finding the instantaneous frequency if phase values were divided by the hopsize T and normalized with the sampling period. The basic assumption behind this procedure is that if there is a tonal component, some linearity within the phase values can be expected without placing any constraint in terms of pitch variations or applied vibrato. If such variations are large enough a frequency bin shifting might occur, i.e., the observed harmonic component might present itself in different frequency bins through the duration of the signal. In general, frequency bin shifting in harmonic components can be expected to be limited to adjacent bins making it easier to track changes.

The phase contours obtained for the five signals are presented in Figures 1-4. For all the cases the pitch detection algorithm described in [17] was used to detect relevant peaks in the audio track. Phase contours are presented for the fundamental frequency and the most prominent harmonic components.

It can be seen that for the violin, clarinet and trumpet notes, the micromodulations in frequency follow similar trajectories and the principle of Common Fate can be observed. However, for the piano note, micromodulations in frequency seem to be completely uncorrelated. Figure 2 shows the phase contours for the trumpet C5 note with vibrato. The large variations in the phase contours exhibit both the extent and frequency of the vibrato and how it presents itself in the different harmonics. Instead of removing vibrato as in [5], this approach sees vibrato as a potential feature for sound separation. It can be seen both in the clarinet (Figure 3) and violin (Figure 1) notes that for the attacks and decays of the notes Common Fate is not so clear and micromodulations are not so correlated. It is important to mention that for harmonic components whose magnitude is very close to zero,

phase values are completely uninformative and do not provide solid information for separation applications.

This approach presents several benefits that can be exploited: (1) No assumption has to be made regarding harmonicity of musical instruments as harmonic components can be tracked by looking for similar phase trajectories in time. (2) Common onset and decay of harmonic components can also be potentially exploited as phase values fall out of a predicted range (see Section 2.3) when a tone is not present. In contrast, this approach also presents several difficulties: Collisions of harmonic components can be misleading as it has been observed that for such cases, the phase trajectory of the harmonic component with the largest magnitude prevails, showing once more the intricate relationship between spectral phase and magnitude. Figure 5 shows an example where a clarinet C5 note and a trumpet G5 note have been mixed and a harmonic collision is present between the clarinet's second harmonic (H2) and the trumpet's first harmonic (H1). In this signal, the amplitude of the trumpet note was much higher than the clarinet note and it can be seen that the phase trajectory for such harmonic follows the trajectory of the trumpet's F0. In Section 2.2 a method to resolve harmonic collisions is presented. Particularly for higher harmonics, frequency bin shifting makes tracking a more complex task. However, by exploring alternative frequency resolutions in the time frequency transform, both overlapping between harmonics and frequency bin shifting can be minimized. Approaches like multiresolution Fourier Transforms [18] or logarithmic frequency resolution can be explored. As shown in Figure 4 for the piano note, not all instruments show properties in the phase trajectories that exhibit Common Fate and consequently the approach is instrument dependent.

2.2. Resolving overlapping harmonics using phase coupling properties of musical instruments

As discussed in Section 1, phase coupling is an important characteristic of musical instrument sounds. In general, phase coupling implies that for a triplet of harmonically related partials with harmonic numbers j , k , and h , with $h = j + k$, any deviations that occur in their respective phases ϕ_j, ϕ_k will sum up to occur identically in ϕ_h .

$$\phi_j + \phi_k - \phi_h = 0 \quad (2)$$

As presented by Dubnov & Rodet in [5], phase coupling characteristics of musical instruments differ for instrument families and types. In general, musical instruments are never perfectly phase-coupled and deviations are always expected. However, when it comes to resolving overlapped harmonics where the frequency information of the different components is hidden within the mix, we propose the use of phase coupling properties to estimate frequency information of the overlapped component. For such an estimate, one condition must be fulfilled: to be able to estimate information of harmonic h , it must be guaranteed that the information of at least two harmonics j and k which fulfill the condition $h = j + k$ is available.

Two signals were analyzed for this purpose: (1) Trumpet C5 note. The sixth harmonic H6 is reconstructed using H2 and H4. (2) Violin C5 note. The third harmonic H3 is reconstructed using H1 and H2.

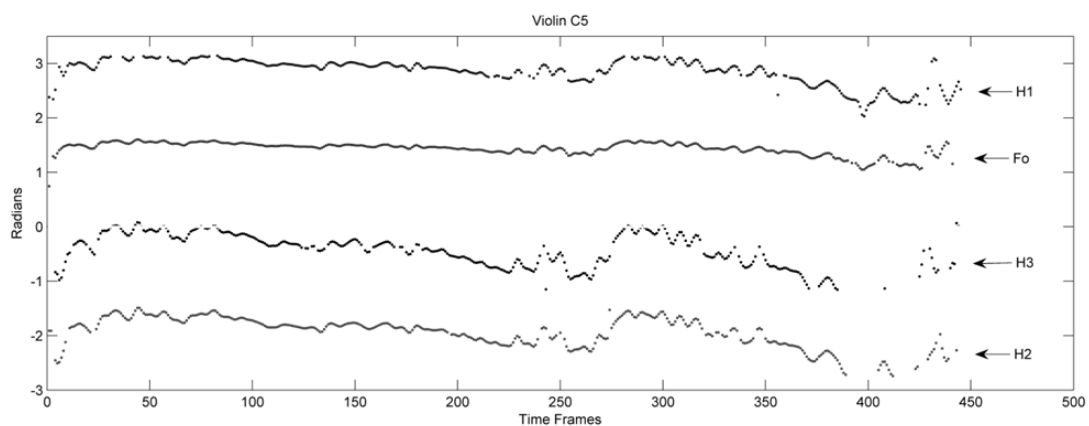


Figure 1: Phase contours obtained for a C5 violin note. The fundamental frequency F_0 and the first three harmonic components are shown.

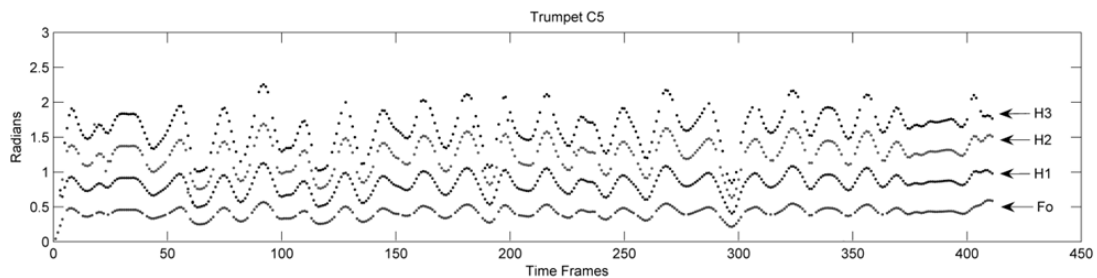


Figure 2: Phase contours obtained for a C5 trumpet note with vibrato. The fundamental frequency F_0 and the first three harmonic components are shown.

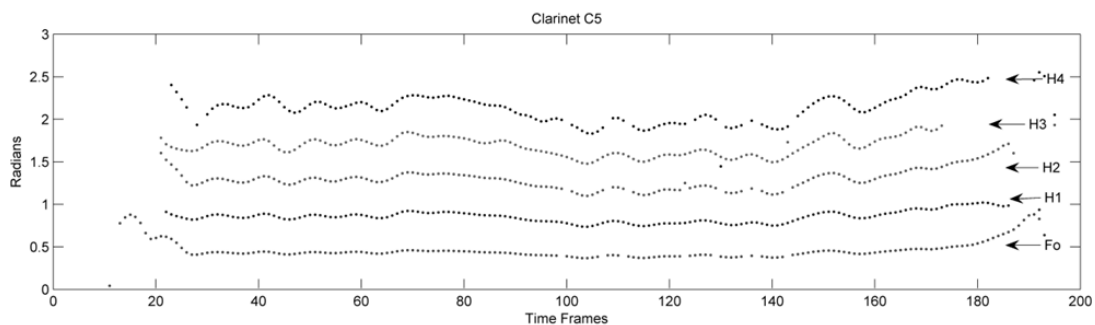


Figure 3: Phase contours obtained for a C5 clarinet note. The fundamental frequency F_0 and the first four harmonic components are shown.

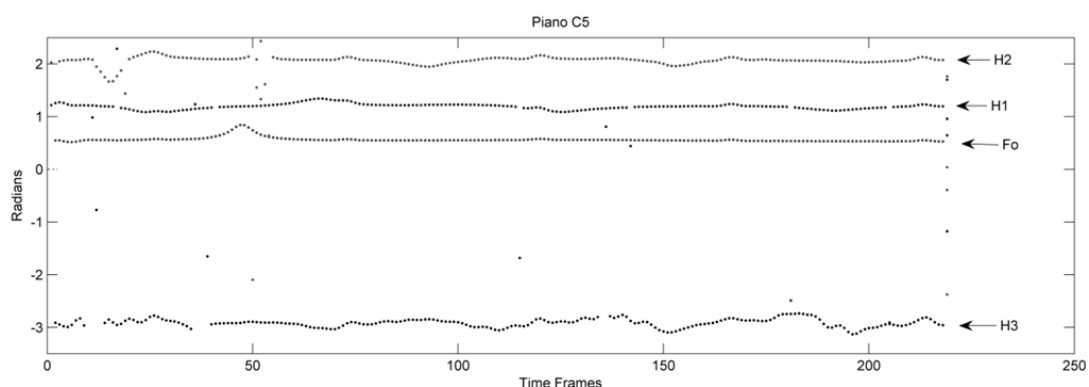


Figure 4: Phase contours obtained for a C5 piano note. The fundamental frequency F_0 and the first three harmonic components are shown.

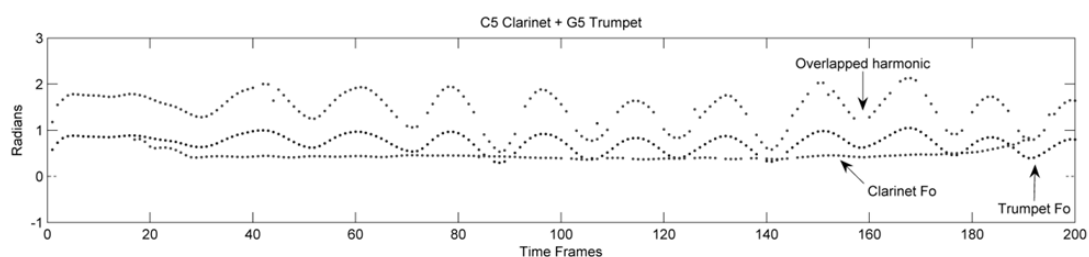


Figure 5: Phase contours obtained for a C5 clarinet note + G5 trumpet note mix. The fundamental frequencies for both instruments and an overlapped harmonic are shown.

Prediction errors of the overlapped harmonics are presented in Figure 6 and the corresponding phase contours obtained are shown in Figures 7 and 8. For visualization purposes and to avoid contour overlapping, the estimated contours in Figures 7 and 8 have been given a 0.3 vertical offset. Consequently, the upper contour in both figures represents the estimated harmonic and the lower contour represents the true harmonic. Results show that as long as condition (1) is fulfilled; reconstructing frequency information of overlapped harmonics is possible for certain instrument types by exploiting phase coupling properties of musical instruments. As far as the magnitude information, the approach used by Woodruff [9] or the iterative techniques proposed in [2] to reconstruct magnitude from phase can be explored.

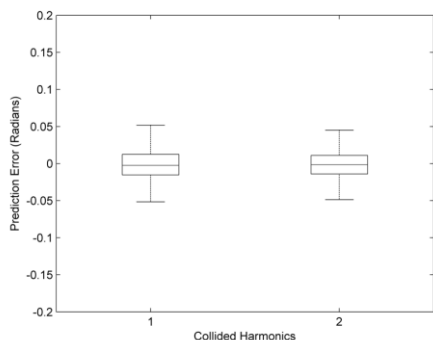


Figure 6: Prediction Error for overlapped harmonics. Column one represents the trumpet and column two the violin harmonic.

As in most source separation algorithms, being able to determine where the harmonic collisions appear is not a simple task. However if a certain number of harmonic components that exhibit similar phase trajectories as in Section 2.1 have been detected, prediction of a missing overlapped harmonic can be made using harmonicity pointers and searching the spectrogram for prominent harmonics. Furthermore, it is important to mention that phase coupling properties are different for all musical instruments and consequently performance of such a system will also be instrument dependent.

2.3. Harmonic/Percussive decomposition using calculated radian ranges for every frequency bin

In this case we exploit the fact that for a certain frequency bin, phase values of tonal components will fall within a radian range determined by the frequency band covered by the frequency bin and the hop size T of the time-frequency transform. Particularly, the condition of phase linearity is relaxed and micromodulations of frequency are allowed within the radian range of the frequency bin. Values of phase outside the calculated range are assumed non-tonal and consequently classified as percussive components. A percussive-harmonic spectral mask is created both for the phase and magnitude spectrograms and applied for synthesizing the harmonic and percussive tracks. It has been observed that when percussive and tonal components are simultaneously present in a particular time frame and frequency bin, the phase values of the percussive component prevail and they do not lay in general within the radian ranges calculated for every frequency

bin. In this sense, a strict sound separation task is not being performed as phase values outside the range imply the presence of a percussive component but not necessarily the absence of a tonal one. For such case, no estimation of the hidden tonal component is performed and the information of that frequency bin in that time frame is assumed to be percussive.

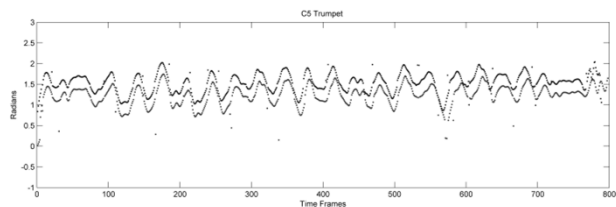


Figure 7: Estimated and true phase trajectories for the sixth harmonic H6 of a trumpet note. Top: Estimated. Bottom: True.

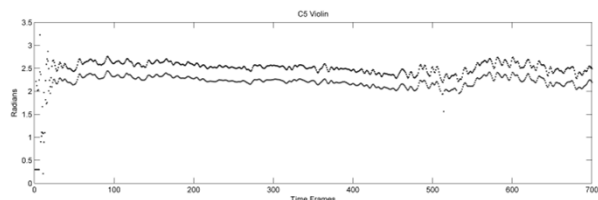


Figure 8: Estimated and true phase trajectories for the third harmonic H3 of a violin note. Top: Estimated. Bottom: True.

The algorithm is summarized as follows:

1. Calculate the STFT of the input audio signal $f(t)$.
2. For every subband k in the STFT, calculate minimum and maximum radian changes using eq. (1).
3. Create the binary spectral masks: For every time frame m and subband k , check if phase values $\phi_k(m)$ fall within the calculated radian ranges. Values within these ranges are assumed tonal and outside the ranges percussive.
4. Apply masks both to phase and magnitude spectrograms.
5. Obtain percussive $f_P(t)$ and harmonic $f_H(t)$ audio signals with inverse STFT.

To test the algorithm 3 mixtures were created from multi-track recordings available in [19]. The three tracks used for evaluation are (1) Natural Minor (Nm), (2) Seven Years of Sorrow (7Y) and (3) Wreck (WR). The algorithm was used to create independent harmonic and percussive tracks and the SISEC evaluation toolbox [20] was used to assess the algorithm's performance using the original multi-track recordings for comparison. Signal to Distortion Ratio (SDR), Signal to Artefacts Ratio (SAR) and Signal to Interference Ratio (SIR) are presented for the three signals. A thorough description of these measures and their calculations is presented in [21]. The performance measures obtained are presented in Table 1 and the audio tracks obtained can be heard in

the project's web site.² For comparison purposes the percussive and harmonic tracks obtained with Ono's [14] algorithm are also available in the web site. For Ono's algorithm the following parameters, as proposed by the authors for best performance, were used: $\alpha = 0.3$, $\gamma = 0.3$ and the maximum number of iterations was set to 50. No direct numeric comparison is presented with Ono's algorithm as the performance measures used do not correlate directly to any perceptual attribute and in certain cases particularly when the perceived loudness of interference or artifacts is much smaller than the power of the corresponding signals the numbers can be misleading. For this reason only a perceptual comparison is presented.

Especially for the harmonic tracks obtained, performance measures show good results with positive ratios in all cases. As expected, performance with percussive tracks is much lower falling in the negative values. In an auditory evaluation, the harmonic and percussive components are well separated into the respective tracks. Bass drums and singing voice are particularly challenging as in both cases, elements from each source are placed in both the harmonic and percussive tracks.

Table 1: Performance measures obtained for the three analyzed tracks.

		SDR	SAR	SIR
1.Nm	Harmonic	6.8141	12.4192	8.4533
	Percussive	-7.9250	-3.5217	-0.8490
2. 7Y	Harmonic	2.88	9.78	4.30
	Percussive	-8.02	-4.59	0.49
3.WR	Harmonic	2.1900	8.1912	4.0591
	Percussive	-5.2766	-2.2619	2.0162

3. CONCLUSIONS

Three cases have been presented where phase information has been used in sound separation problems. In all cases, phase information appears to be informative and somehow complementary to the use of magnitude information. Phase contours for musical instruments exhibit similar micromodulations in frequency for certain instruments and can be an alternative of spectral instrument templates or instrument models. For the case of overlapped harmonics, phase coupling properties can be exploited for certain instruments. For the two instruments presented, estimated harmonics show prediction errors lower than 0.05 radians. For the Harmonic-Percussive decomposition, radian ranges have been calculated for every frequency bin and by relaxing phase linearity and allowing frequency variations, tonal components have been detected. The spectral mask created allows discriminating not only magnitude but also phase information belonging to the harmonic and percussive components. Both harmonic and percussive tracks obtained can be used to facilitate transcription applications. Further studies have to be made in order to assess performance and robustness of the algorithms in more complex and demanding scenarios. A possible extension to this approach

²

http://www.idmt.fhg.de/eng/business%20areas/analysis_audio_signals.htm

is the use of Modulation Spectra as a means of exhibiting frequency variations in the different frequency bins.

4. ACKNOWLEDGMENTS

The Thuringian Ministry of Economy, Employment and Technology supported this research by granting funds of the European Fund for Regional Development to the project Songs2See, enabling transnational cooperation between Thuringian companies and their partners from other European regions.

5. REFERENCES

- [1] Oppenheim Alan V. and Lim Jae S. "The importance of phase in signals," in *Proceedings of IEEE*, pp. 529-550, 1981.
- [2] Quatieri Thomas F. and Oppenheim Alan V. "Iterative techniques for minimum phase signal reconstruction from phase or magnitude," in *IEEE Transactions on Acoustics, Speech and Signal Processing*, vol. 29, no. 6, pp. 1187-1192, 1981.
- [3] Dubnov Shlomo. "Higher order statistical estimation of sinusoidality with applications for quality coding of musical instruments," in *AES 17th International Conference on High Quality Audio Coding*, Florence, Sept. 1999.
- [4] Dubnov Shlomo. "Improved harmonic + noise model for vocal and musical instrument sounds," in *AES 22th International Conference on Virtual, Synthetic and Entertainment Audio*, Espoo, Finland, June 2002.
- [5] Dubnov Shlomo and Rodet Xavier. "Investigating the phase coupling phenomena in sustained portion of musical instruments sound," in *Journal of the Acoustical Society of America*, vol. 113, no. 1, pp. 348-359, 2003.
- [6] Cont Arshia and Dubnov Shlomo. "Real time multi-pitch and multi-instrument recognition for music signals using a sparse non-negative constraints," in *Proceedings of the 10th International Conference on Digital Audio Effects (DAFx)*, Bordeaux, France, 2007.
- [7] Sukittanon Somsak, Atlas Les E. and Pitton James W. "Modulation-scale analysis for content identification," in *IEEE Transactions on Signal Processing*, vol. 52, no. 10, pp. 3023-3035, 2004.
- [8] Paraskevas Ioannis and Chilton Edward. "Combination of magnitude and phase statistical features for audio classification," in *Acoustics Research Letters Online (ARLO)*, vol.5, no 3, pp. 111-117, 2004.
- [9] Woodruff John, Yipeng Li and Wang DeLiang. "Resolving overlapped harmonics for monaural musical sound separation using pitch and common amplitude modulation," in *International Conference on Music Information Retrieval (ISMIR)*, Philadelphia, Sept. 2008, pp. 538-543.
- [10] Virtanen Tuomas. "Monaural sound source separation by nonnegative matrix factorization with temporal continuity and sparseness criteria," in *IEEE Transactions on Audio, Speech and Language Processing*, vol. 15, no. 3, pp. 1066-1074, 2007.
- [11] Fitzgerald Derry, Cranitch Matt and Coyle Eugene. "Extended nonnegative tensor factorization models for musical sound source separation," in *Computational Intelligence and Neuroscience*. Hindawi Publishing Corporation, 2008.
- [12] Burred Juan Jose, "From Sparse Models to Timbre Learning: New Methods for Musical Sound Separation," PhD Thesis. Elektrotechnik und Informatik der Technischen Universit at Berlin, 2009.
- [13] Every Mark R. and Szymanski John E. "A spectral filtering approach to music signal separation," in *7th International Conference on Digital Audio Effects (DAFx)*, Naples, 2004.
- [14] Ono Nobutaka [et al.] "Separation of a Monaural Audio Signal into Harmonic/Percussive Components by Complementary Diffusion on Spectrogram," in *16th European Signal Processing Conference (EUSIPCO)*, Lausanne, Switzerland, Aug. 2008.
- [15] Bregman Albert S. *Auditory Scene Analysis. The perceptual organization of sound*. Cambridge : MIT Press, 1999.
- [16] University of Iowa Musical Instrument Samples. Available at <http://theremin.music.uiowa.edu/>. Accessed October 10, 2010.
- [17] Cano Estefanía and Cheng Corey. "Melody Line Detection and Source Separation in Classical Saxophone Recordings," in *Proceedings of the 12th International Conference on Digital Audio Effects (DAFx-09)*, Como, Italy, Sept 1 - 4, 2009.
- [18] Dressler Karin. "Sinusoidal Extraction using an Efficient Implementation of a Multi-Resolution FFT," in *9th International Conference on Digital Audio Effects (DAFx)*, Montreal, Canada, Sept. 2006.
- [19] Multitrack Recording. Available at <http://bass-db.gforge.inria.fr/BASS-dB/?show=browse&id=mtracks>. Accessed March 10, 2010.
- [20] SISEC Evaluation Software. Available at <http://www.irisa.fr/metiss/members/evincent/software>. Accessed March 10, 2010.
- [21] Vincent Emmanuel, Gribonval Rémi and Févotte Cédric "Performance Measurement in Blind Audio Source Separation," in *IEEE Transactions on Audio, Speech and Language*