

# Harmonic-Percussive Source Separation of Polyphonic Music by Suppressing Impulsive Noise Events

Gurunath Reddy M, K. Sreenivasa Rao, Partha Pratim Das

Indian Institute of Technology, Kharagpur, India

{mgurunathreddy, ksrao}@sit.iitkgp.ernet.in, ppd@cse.iitkgp.ernet.in

## Abstract

In recent years, harmonic-percussive source separation methods are gaining importance because of their potential applications in many music information retrieval tasks. The goal of the decomposition methods is to achieve near real-time separation, distortion and artifact free component spectrograms and their equivalent time domain signals for potential music applications. In this paper, we propose a decomposition method based on filtering/suppressing the impulsive interference of percussive source on the harmonic components and impulsive interference of the harmonic source on the percussive components by modified moving average filter in the Fourier frequency domain. The significant advantage of the proposed method is that it minimizes the artifacts in the separated signal spectrograms. In this work, we have proposed Affine and Gain masking methods to separate the harmonic and percussive components to achieve minimal spectral leakage. The objective measures and separated spectrograms showed that the proposed method is better than the existing rank-order filtering based harmonic-percussive separation methods.

**Index Terms:** Harmonic, Mixture, Mask, Percussion, Polyphonic, Separation.

## 1. Introduction

The components in a polyphonic music signal can be broadly classified into harmonic and percussive sources. The harmonic sources such as violin, piano and so on are pitched sources contain fundamental frequency and higher harmonics, which can be modeled with finite number of sinusoids, manifests as horizontal ridges in the magnitude spectra of the short-time-Fourier-transform (STFT). The percussive sources such as castanets and many drums exhibits impulsive like nature, difficult to model by a finite number of sinusoids results in a wideband spectral energy or a vertical ridge in the magnitude Fourier spectrum. Thus, the harmonics of the pitched sources results in impulsive like noise along the frequency bins of the Fourier spectrum where the percussion source exhibits uniform energy across frequency bins. Similarly, the percussion sources results in impulsive like noise for harmonic sources across the spectral frames where they exhibit temporal continuity along the time. Hence, in this paper, the impulsive noise like nature of percussion and harmonic sources across the spectral frames and frequency bins are suppressed to enhance the percussion and harmonic sources along the frequency bins and spectral frames respectively.

The well separated sources can be used as input for many music related applications [1]. The harmonic source can be used in multipitch extraction [2, 3], automatic pitched source transcription [4, 5], melody extraction [6, 7, 8, 9, 10], singing voice separation [11] and so on. Similarly, the percussion source can be used in onset detection [12], beat tracking [13], automatic

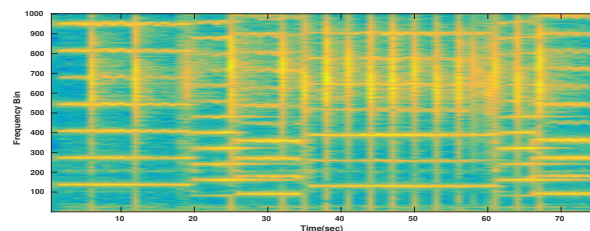


Figure 1: Complex spectrogram of the polyphonic music.

transcription of drums [14, 15], rhythm analysis [16], tempo estimation [17, 18], since these applications require signals which is free from harmonic sources and rich in percussion components.

We can find several harmonic-percussive source separation methods in the literature. In [19], the noisy phase behavior of the percussion in the input signal is exploited to separate the harmonic and percussive components in the music signal. An iterative spectrogram diffusion algorithm is proposed in [20]. The method involves diffusing the spectrogram in horizontal and vertical directions to enhance the harmonic and percussive components in the mixture spectrogram, which is based on the observation that the harmonic sources tend to exhibit themselves as horizontal ridges and percussion sources as vertical ridges in the magnitude spectrogram. The complex iterative diffusion method is replaced by a much simpler median filtering based method in [21] to separate the mixture signal into harmonic-percussive sources. The median filtering based approach [21] is extended in [22] to separate the composite signal into harmonic, percussive and residual components. Optimization based methods such as non-negative matrix factorization [23] and kernel additive modeling [24] are proposed for harmonic percussive source separation.

In this paper, we propose a modified moving average filtering based method which is capable of filtering/suppressing the impulsive like events in the spectrogram to decompose into harmonic and percussive sources. The significant advantage of the proposed method is that it minimizes the artifacts in the separated signal spectrograms. In this work, we also propose and evaluate several masking methods to separate the harmonic and percussive components with minimal leakage. Finally, we evaluate our proposed method based on objective measures and separated spectrograms.

## 2. Harmonic-Percussive Separation

The polyphonic music, in general is a mixture of harmonic and percussive sources. The harmonic sources are deterministic signals exhibiting horizontal lines in the magnitude Fourier spectrogram, whereas, percussive sources are non-deterministic impul-

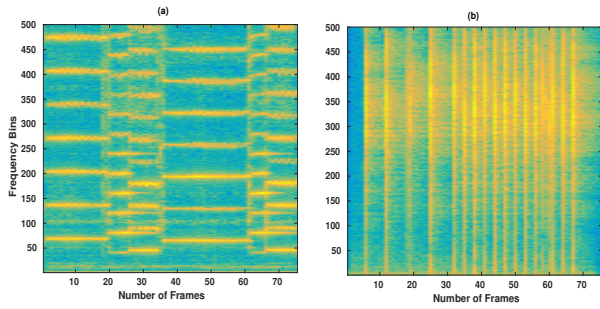


Figure 2: Spectrograms of the harmonic and percussive sources.

sive like events forms vertical lines in the Fourier spectrogram. An example spectrogram of the composite music signal consisting of violin as harmonic and castanets as percussive source is shown in Fig. 1. From Fig. 1, we can observe two distinct patterns which are orthogonal to each other in the magnitude spectrogram i.e., horizontal and vertical ridges in the spectrogram. The spectrograms of the individual sources are shown in Fig. 2. The spectrogram of the harmonic source (violin) is shown in Fig. 2(a). From Fig. 2(a), we can observe that the harmonics of the violin forms horizontal ridges in the spectrogram. Similarly, Fig 2(b) shows the spectrogram of the percussive source (castanets). From Fig 2(b), we can observe that the impulsive events of the castanets forms vertical ridges in the spectrogram. Furthermore, careful observation of the spectrograms in Figs. 1 and 2, we can conclude that harmonic peaks of the pitched (harmonic) sources forms outliers in a spectral frame where percussive sources have uniform energy. Similarly, percussive events forms outliers in a frequency bin of the spectrogram where the harmonic sources mostly have equal energy. We propose using a modified moving average smoothing filter to suppress the harmonic spectral peak outliers in the spectral frames to enhance the percussion components, and to suppress the percussion source outliers in the frequency band of the spectrogram to enhance the harmonic sources.

Traditionally, moving average filter is used to suppress the high frequency noise in the input signal. The amount of suppressed noise depends on the length of the moving average filter given by

$$M(i) = \frac{1}{N} \sum_{k=i}^{i+(N-1)} s(k) \quad (1)$$

where  $s(k)$  is the noisy input signal,  $N$  is the filter length and  $M(i)$  is the noise-suppressed signal. The frequency response of the moving average filter is given by

$$H(\omega) = \frac{1}{N} \frac{(1 - e^{-j\omega N})}{(1 - e^{-j\omega})} \quad (2)$$

Though the frequency response  $H(\omega)$  has lowpass filter characteristics, its high frequency attenuation capability is much weaker. Since as  $N$ , the length of the filter increases, height of the side lobes of the frequency response of the filter increases resulting in poor attenuation of the impulse like events. Hence, the moving average filter cannot be used for suppressing impulsive like spectral peaks in the spectrum. An example of moving average filter applied on a synthetic signal consisting of impulsive noise (blue contour) and filtered signal (red contour) are shown in Fig. 3. From Fig. 3, we can observe that the moving average filter fails to significantly attenuate the impulsive noise, which is not the desired filter characteristic required to

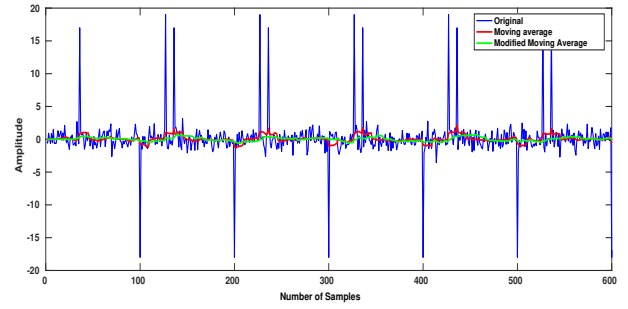


Figure 3: Comparison of impulsive noise smoothing capabilities of moving average and MMAF.

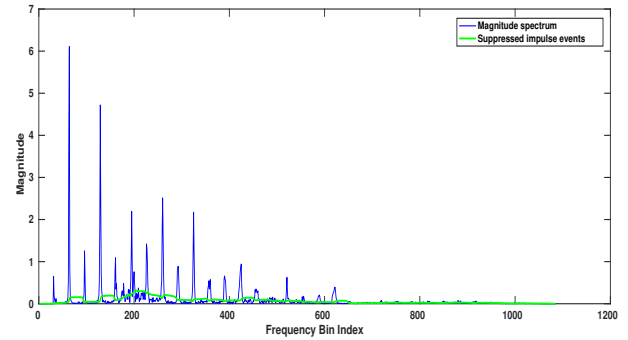


Figure 4: Impulsive noise like interference of harmonic source suppression to enhance the percussion of the spectral frame.

remove impulsive like interference in the spectrogram to separate harmonic and percussive sources. In order to overcome the limitations of the moving average filter, the modified moving average filter (MMAF) [25] is proposed to strongly attenuate the impulsive noise events in the signal. The impulsive noise smoothing MMAF is given by

$$A(i) = M(i) + \frac{(P_{os} - N_{eg}) \cdot |D_{total}|}{N^2} \quad (3)$$

where  $M(i)$  is the moving average filter given in Eq. 1,  $P_{os}$  is the total number of samples above the mean,  $N_{eg}$  is the total number of samples below the mean in  $N$  signal samples.  $|D_{total}|$  is the cumulative absolute deviation of the samples from the mean  $M(i)$  and  $N$  is the length of the filter or samples considered for smoothing.  $A(i)$  is the impulse smoothed signal. The second term in Eq. 3 acts as a correction factor to the moving average filter result  $M(i)$  to strongly attenuate the impulsive noise in the signal. An example showing the strong impulsive event attenuating characteristics of the MMAF is shown in Fig. 3. The green contour is the smoothed signal which is obtained after applying MMAF. From Fig 3, we can observe that the MMAF has high impulsive noise attenuating capability than moving average filter. Also, we can observe that the MMAF filtered signal is much more smoother than the signal obtained by averaging filter (red contour) at the impulsive events.

The plot in Fig. 4 shows the frame of a magnitude spectrogram where the harmonic peaks of the violin acts as impulsive noise to the castanets percussion source which shows noisy behavior. From Fig. 4, we can observe that MMAF mostly attenuates the impulsive noise interference of the harmonics to enhance the percussion source, which is shown as green contour. Similarly, the MMAF across the spectral bin attenuates the impulsive interference of the percussion to enhance the harmonic sources.

The aim is to decompose the given music signal  $s$  into harmonic  $s_h$  and percussive  $s_p$  sources such that  $s \sim s_h + s_p$  i.e., when the components are combined back either in spectral or time domain. Further, the combination should yield original music signal without much distortion. The input music signal  $s$  is transformed to spectral domain by applying STFT given by

$$S(l, k) = \sum_{n=0}^{N-1} s(n + lH)w(n)e^{-j2\pi kn/N} \quad (4)$$

where  $l = [0, \dots, L-1]$ ,  $k = [0, \dots, N/2]$ ,  $L$  is the total number of frames,  $N$  is the Fourier frequency bins,  $w$  is the Hamming window and  $H$  is the hop size.

The harmonic components in the magnitude spectrogram  $F(l, k) = |S(l, k)|$  is enhanced by suppressing the impulsive interference of the percussion in each frequency band (bin) given by

$$H(l, k) = \mathcal{M}\{F(l - t_h, k), \dots, F(l + t_h, k)\} \quad (5)$$

Similarly, the percussion source in a spectral frame is enhanced by suppressing the impulsive harmonic source given by

$$P(l, k) = \mathcal{M}\{F(l, k - t_p), \dots, F(l, k + t_p)\} \quad (6)$$

where  $\mathcal{M}$  is the MMAF,  $2t_h + 1$  and  $2t_p + 1$  are the MMAF filter lengths for percussion and harmonic event suppression.

The resulting enhanced harmonic  $H(l, k)$  and percussion  $P(l, k)$  spectrograms are used for generating binary masks, which are then applied on the original spectrogram  $S(l, k)$  to obtain the complex spectrograms of harmonic and percussive sources. In this paper, two new masking methods are added to the existing ones proposed in [21, 22] resulted in total five masking methods. In which, two methods are non-parametric, where the user has no control over the inter spectral leakage i.e., harmonic components leaking into percussion spectrogram and vice versa. The remaining three are the parametric masking methods, which controls the amount of spectral leakage with the help of separation parameters, discussed later. We have evaluated all five masking methods to analyze the tight spectral separation capability to minimize the inter spectral leakage due to masking.

The non-parametric methods include simple Binary threshold and Wiener filter. The Binary threshold is a hard threshold on the enhanced spectrograms to obtain the harmonic and percussive masks given by

$$H_M(l, k) = \begin{cases} 1 & \text{if } H(l, k) > P(l, k) \\ 0 & \text{otherwise} \end{cases} \quad (7)$$

$$P_M(l, k) = \begin{cases} 1 & \text{if } P(l, k) \geq H(l, k) \\ 0 & \text{otherwise} \end{cases} \quad (8)$$

The Wiener filtering results in a smooth binary mask given by

$$H_M(l, k) = \frac{H^\gamma(l, k)}{H^\gamma(l, k) + P^\gamma(l, k)} \quad (9)$$

$$P_M(l, k) = \frac{P^\gamma(l, k)}{H^\gamma(l, k) + P^\gamma(l, k)} \quad (10)$$

where  $\gamma$  is a power to which each spectral value is raised. Here, the value of  $\gamma$  is set to 2.

The parametric methods include Relative [22], Gain, and Affine masking methods which have two independent parameters  $\beta_h$  and  $\beta_p$  decides the extent of separation of the desired source from the input signal.

The relative masking method is given by

$$H_M(l, k) = \frac{H(l, k)}{P(l, k) + \epsilon} > \beta_h \quad (11)$$

$$P_M(l, k) = \frac{P(l, k)}{H(l, k) + \epsilon} \geq \beta_p \quad (12)$$

where  $\epsilon$  is a tiny constant to avoid division by zero error, the operators  $>$  and  $\geq$  results in binary values  $\{0, 1\}$ .

The gain masking method is given by

$$H_M(l, k) = H^2(l, k) > (\beta_h \cdot P^2(l, k)) \quad (13)$$

$$P_M(l, k) = P^2(l, k) \geq (\beta_p \cdot H^2(l, k)) \quad (14)$$

The affine masking method is given by

$$H_M(l, k) = ((1 - \beta_h) \cdot H(l, k)) > (\beta_h \cdot P(l, k)) \quad (15)$$

$$P_M(l, k) = ((1 - \beta_p) \cdot P(l, k)) \geq (\beta_p \cdot H(l, k)) \quad (16)$$

Here, the independent parameters  $\beta_h$  and  $\beta_p$  imposes the tight constraint on the separation process. Depending on the value of the parameter  $\beta_h$ ,  $H_M(l, k)$  will results in a binary mask mostly contains the signatures of the harmonic content. Similarly, the parameter  $\beta_p$  minimizes the leakage of the harmonic signatures into  $P_M(l, k)$ .

The binary masks  $H_M(l, k)$  and  $P_M(l, k)$  are multiplied with the original complex spectrogram  $S(l, k)$  to obtain the harmonic and percussive spectrograms

$$S_H(l, k) = S(l, k) \cdot H_M(l, k) \quad (17)$$

$$S_P(l, k) = S(l, k) \cdot P_M(l, k) \quad (18)$$

where  $\cdot$  is the element wise product. The inverse STFT is applied on the  $S_H(l, k)$  and  $S_P(l, k)$  to obtain the time domain harmonic and percussive signals  $s_h$  and  $s_p$  respectively.

### 3. Evaluation and Discussion

The separation quality of the proposed method is evaluated by computing the source to distortion ratio (SDR), source to interference ratio (SIR) and source to artifact ratio (SAR) [26, 27]. The mixture signal is obtained by adding the harmonic (vocals + harmonic instruments or harmonic instruments alone) and percussive instruments from *freesound.org*. We have collected Vocals (male and female), Flute, Cello, Violin as harmonic instruments, Snare drum, Tabla, Castanets, Hit-Hat as percussion instruments to create the mixture, drawn three instruments at a time, resulted in 84 mixture samples. All five masking methods are evaluated objectively to analyze the tight spectral separation property of each method to minimizing the inter spectral leakage due to masking. The objective measures SDR, SIR and SAR with respect to separation parameters  $\beta_h$ (beta\_H) and  $\beta_p$ (beta\_P) are shown in Figs. 5 and 6 respectively. Fig. 5 shows the objective measures for Binary threshold, Wiener, Relative and Gain masking methods for varying separation parameters  $\beta_h$  and  $\beta_p$ . Since Binary threshold and Wiener methods are non-parametric methods, the measures SDR, SIR and SAR are independent of separation parameters, shown as a horizontal plots in Fig. 5. Also, we can observe that the plots for non-parametric methods are well below the plots of Relative and Gain methods. This is because Binary threshold and Wiener methods being non-parametric, provides no explicit control over the leakage of inter spectral components, results in poor objective measures. The Relative and Gain masking methods show a similar evaluation results, but close observation of

Table 1: Objective evaluation measures in dB.

SDR			SIR			SAR		
HP	HPR-IO	P	HP	HPR-IO	P	HP	HPR-IO	P
-3.83	6.89	9.08	-3.08	19.95	22.76	6.83	7.45	9.41

the plots reveals that the Relative method needs precise setting of the separation parameters  $\beta_h$  and  $\beta_p$  to achieve good separation, whereas, Gain method gives more flexibility in choosing the parameters since the objective measures remain constant for a range of parameter values which can be observed from Fig. 5. Also, from Fig. 5, we can observe that the plots for Relative and Gain methods remain above the non-parametric masking methods, this can be attributed to the tight decomposition imposed by the separation parameters resulting in reduced inter spectral leakage. The objective measures for the Affine masking method is plotted separately in Fig. 6 because the range of separation parameters for this method is between 0 and 1 i.e.,  $0 < \beta_h < 1$  and  $0 < \beta_p < 1$ . Unlike Relative and Gain masking methods, Affine masking method is a relative weighting method, proportionately weights both harmonic and percussive enhanced spectrograms for the different values of  $\beta_h$  and  $\beta_p$ . Since Affine method weights the spectrograms relatively, for an optimal value of separation parameters, results in a more smoother, distortion free and tight separation of the harmonic and percussive components. Unlike other methods, the search range for the optimal separation parameters in Affine method is between 0 and 1. Hence, it significantly reduces the search time for finding optimal  $\beta_h$  and  $\beta_p$  for tight and smooth separation. Also, we observed that the Affine masking method results in the best separation with minimal spectral distortion in the separated sources for optimal separation parameters. The spectrograms of the separated harmonic and percussive sources of a mixture of violin (harmonic) and castanets (percussive) by the proposed method is shown in 7(a) and 7(b) and the decomposed results by the state-of-the-art iterative median filtering based method (HPR-IO) [22] is shown in 7(c) and 7(d). In Fig. 7, the proposed method uses the Affine masking method with optimal separation parameters  $\beta_h = 0.8$  and  $\beta_p = 0.4$  obtained from the plots where the measures SDR and SAR just started to meet in Fig. 6. The spectrograms for HPR-IO are plotted from the separated sources available at [28] for the authors best parameter settings. From Fig. 7, we can observe that the proposed method clearly preserves the characteristics of the harmonic and percussive sources i.e., horizontal and vertical ridges in the spectrogram without introducing much distortion, whereas the HPR-IO introduces significant artifacts in the spectrograms of the separated sources which can be clearly observed in the spectrograms shown in Figs. 7(c) and 7(d).

The proposed (P) method is compared with the harmonic-percussive source separation proposed by Fitzgerald (HP) [21] and iterative harmonic-percussive-residual separation (HPR-IO) [22] shown in Table 1. The proposed method uses Affine masking method with optimal separation parameters  $\beta_h = 0.8$  and  $\beta_p = 0.4$  discussed previously with Fourier frequency bins set to 4096 and MMAF filter length  $N = 50$  along time and frequency directions. The authors best parameters are chosen for HP and HPR-IO given in [21] and [22] respectively. From Table 1, we can observe that the objective measures for the proposed method is significantly better than the HP and HPR-IO methods. This can be attributed to the high impulsive noise suppression property of MMAF and the relative weighting property of the Affine masking method strongly preserves the spectral properties of the harmonic and percussive sources. In future,

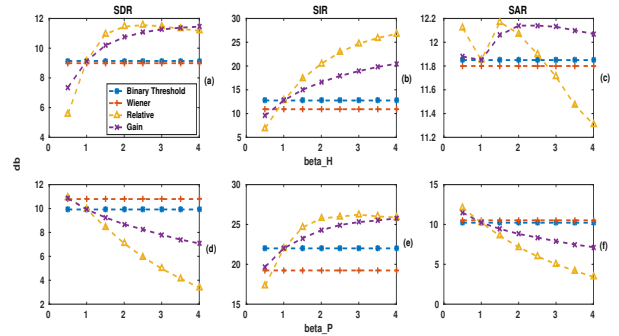


Figure 5: Performance comparison of Binary threshold, Wiener, Relative and Gain masking methods.

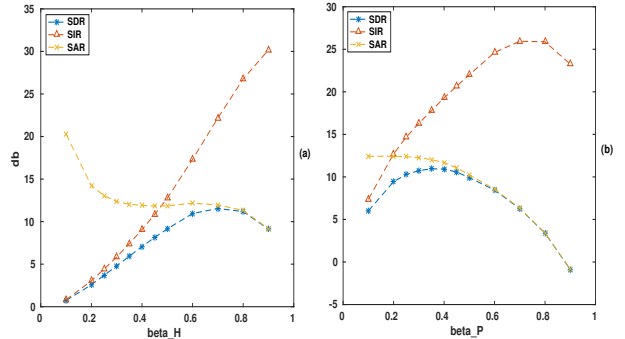


Figure 6: Objective measure of Affine masking method.

we would like to conduct a rigorous subjective evaluation test to better understand the perceptual quality of the separated signals. We would also like to use the separated harmonic source to detect the vocal and non-vocal regions in the polyphonic music signal and also to extract the vocal melody from the separated vocal regions. The decomposed signals are made available at <https://github.com/mgurunathreddy/Harmonic-Percussive.git>

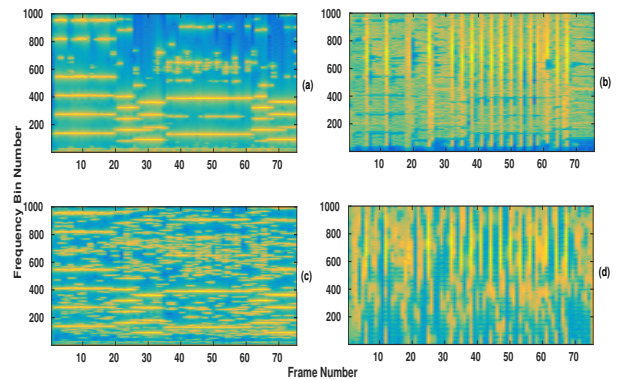


Figure 7: Separated spectrograms of the proposed and HPR-IO.

#### 4. Acknowledgements

The authors would like to thank Google for supporting first author PhD under Google India PhD Fellowship program.

## 5. References

- [1] N. Ono, K. Miyamoto, H. Kameoka, J. Le Roux, Y. Uchiyama, E. Tsunoo, T. Nishimoto, and S. Sagayama, "Harmonic and percussive sound separation and its application to mir-related tasks," in *Advances in music information retrieval*. Springer, 2010, pp. 213–236.
- [2] P. Fernandez-Cid and F. J. Casajus-Quiros, "Multi-pitch estimation for polyphonic musical signals," in *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, vol. 6, 1998, pp. 3565–3568.
- [3] R. Badeau, V. Emiya, and B. David, "Expectation-maximization algorithm for multi-pitch estimation and separation of overlapping harmonic spectra," in *Acoustics, Speech and Signal Processing, 2009. ICASSP 2009. IEEE International Conference on*. IEEE, 2009, pp. 3073–3076.
- [4] G. E. Poliner, D. P. Ellis, A. F. Ehmann, E. Gómez, S. Streich, and B. Ong, "Melody transcription from music audio: Approaches and evaluation," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 15, no. 4, pp. 1247–1256, 2007.
- [5] A. Klapuri and A. Eronen, "Automatic transcription of music," in *Proceedings of the Stockholm Music Acoustics Conference*, 1998, pp. 6–9.
- [6] H. Tachibana, T. Ono, N. Ono, and S. Sagayama, "Melody line estimation in homophonic music audio signals based on temporal-variability of melodic source," in *Acoustics speech and signal processing (icassp), 2010 ieee international conference on*. IEEE, 2010, pp. 425–428.
- [7] J. Salamon and E. Gómez, "Melody extraction from polyphonic music signals using pitch contour characteristics," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 20, no. 6, pp. 1759–1770, 2012.
- [8] M. G. Reddy and K. S. Rao, "Predominant melody extraction from vocal polyphonic music signal by combined spectro-temporal method," in *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2016, pp. 455–459.
- [9] G. Reddy and K. S. Rao, "Enhanced harmonic content and vocal note based predominant melody extraction from vocal polyphonic music signals," in *INTERSPEECH*, 2016, pp. 3309–3313.
- [10] —, "Predominant vocal melody extraction from enhanced partial harmonic content," in *European Signal Processing Conference (EUSIPCO)*, 2017, pp. 1016–1020.
- [11] Y. Li and D. Wang, "Separation of singing voice from music accompaniment for monaural recordings," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 15, no. 4, pp. 1475–1487, 2007.
- [12] J. P. Bello, L. Daudet, S. Abdallah, C. Duxbury, M. Davies, and M. B. Sandler, "A tutorial on onset detection in music signals," *IEEE Transactions on speech and audio processing*, vol. 13, no. 5, pp. 1035–1047, 2005.
- [13] M. Goto, "An audio-based real-time beat tracking system for music with or without drum-sounds," *Journal of New Music Research*, vol. 30, no. 2, pp. 159–171, 2001.
- [14] D. FitzGerald, R. Lawlor, and E. Coyle, "Sub-band independent subspace analysis for drum transcription," 2002.
- [15] O. Gillet and G. Richard, "Transcription and separation of drum signals from polyphonic music," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 16, no. 3, pp. 529–540, 2008.
- [16] J. Foote and S. Uchihashi, "The beat spectrum: A new approach to rhythm analysis," in *IEEE International Conference on Multimedia and Expo (ICME)*, 2001, pp. 881–884.
- [17] M. A. Alonso, G. Richard, and B. David, "Tempo and beat estimation of musical signals," in *ISMIR*, 2004.
- [18] M. E. Davies and M. D. Plumbley, "Exploring the effect of rhythmic style classification on automatic tempo estimation," in *Signal Processing Conference, 2008 16th European*. IEEE, 2008, pp. 1–5.
- [19] C. Duxbury, M. Davies, and M. Sandler, "Separation of transient information in musical audio using multiresolution analysis techniques," in *Digital Audio Effects (DAFX)*, 2001.
- [20] N. Ono, K. Miyamoto, J. Le Roux, H. Kameoka, and S. Sagayama, "Separation of a monaural audio signal into harmonic/percussive components by complementary diffusion on spectrogram," in *European Signal Processing Conference (EUSIPCO)*, 2008, pp. 1–4.
- [21] D. Fitzgerald, "Harmonic/percussive separation using median filtering," in *Digital Audio Effects (DAFX)*, 2010.
- [22] J. Driedger, M. Müller, and S. Disch, "Extending harmonic-percussive separation of audio signals," in *ISMIR*, 2014, pp. 611–616.
- [23] F. Canadas-Quesada, D. Fitzgerald, P. Vera-Candeas, and N. Ruiz-Reyes, "Harmonic-percussive sound separation using rhythmic information from non-negative matrix factorization in single-channel music recordings," in *Digital Audio Effects (DAFX)*, 2017.
- [24] D. FitzGerald, A. Liukus, Z. Rafii, B. Pardo, and L. Daudet, "Harmonic/percussive separation using kernel additive modelling," 2014.
- [25] B. Dvorak, "Software filter boosts signal-measurement stability, precision," *ELECTRONIC DESIGN-NEW YORK THEN HASBROUCK HEIGHTS*, vol. 51, no. 3, pp. 60–60, 2003.
- [26] E. Vincent, R. Gribonval, and C. Févotte, "Performance measurement in blind audio source separation," *IEEE transactions on audio, speech, and language processing*, vol. 14, no. 4, pp. 1462–1469, 2006.
- [27] C. Févotte, R. Gribonval, and E. Vincent, "Bss\_eval toolbox user guide—revision 2.0," 2005.
- [28] J. Driedger, M. Müller, and S. Disch, "Extending harmonic-percussive separation of audio signals," <https://www.audiolabs-erlangen.de/resources/2014-ISMIR-ExtHPSep/>.