Iterative Sinusoidal-based Partial Phase Reconstruction in Single-channel Source Separation

Mario Kaoru Watanabe and Pejman Mowlaee

Signal Processing and Speech Communication Laboratory
Graz University of Technology, Graz, Austria
mario.watanabe@student.tugraz.at pejman.mowlaee@tugraz.at

Abstract

Partial phase reconstruction based on a confidence domain has recently been shown to provide improved signal reconstruction performance in a single-channel source separation scenario. In this paper, we replace the previous binarized fixed-threshold confidence domain with a new signal-dependent one estimated by employing a sinusoidal model to be applied on the estimated magnitude spectrum of the underlying sources in the mixture. We also extend the sinusoidal-based confidence domain into Multiple Input Spectrogram Inversion (MISI) framework, and we propose to re-distribute the remixing error at each iteration on the sinusoidal-signal components. Our experiments on both oracle and estimated spectra show that the proposed method achieves improved separation results at a lower number of iterations, making it as a favorable choice for faster phase estimation.

Index Terms: Iterative signal reconstruction, partial phase estimation, sinusoidal model, single-channel source separation.

1. Introduction

Separating audio and speech mixtures is of great importance in many practical applications to name a few: to design a robust automatic speech recognition system, speech transmission in mobile telephony, and in enhancing the perceived signal quality of the desired signals recorded in adverse noise scenarios such when corrupted with other sources or the reverberation in the room.

Figure 1 shows the block diagram of a typical single-channel source separation algorithm composed of two stages: amplitude spectrum estimation followed by signal reconstruction. Wiener filter is commonly used to provide the estimated magnitude spectrum among other methods like binary mask, non-negative matrix factorization [1], mixture estimation [2–6]. The estimated magnitude spectrum from the amplitude estimation stage is then passed to the signal reconstruction stage, where a phase spectrum is required to recover the time-domain representation of each source. Many previous proposals for single-channel source separation mainly focus on deriving optimal filters to enhance the magnitude spectra from a given mixed signal (see stage 1 in Figure 1). However, recent studies [7–16], demonstrate that further improvement in separation

This work was partially funded by the European project DIRHA (FP7-ICT-2011-7-288121) and by the K-Project AAP in the COMET (Competence Centers for Excellent Technologies) programme with joint support from BMVIT, BMWFJ, Styrian Business Promotion Agency (SFG), and the Government of Styria (”Abt. 3: Wissenschaft und Forschung” as well as ”Abt. 14: Wirtschaft und Innovation”). The programme COMET is managed by the Austrian Research Promotion Agency (FFG).

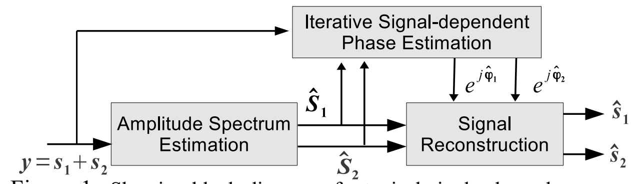

Figure 1: Showing block diagram of a typical single-channel source separation algorithm using Wiener filter for amplitude estimation (first block), and iterative signal reconstruction (second block).

performance is achievable by replacing the mixture phase spectrum by an estimated one employed for reconstructing the separated signals (see stage 2 in Figure 1).

In this paper, we focus on the signal reconstruction stage and propose to replace the fixed-threshold based confidence domain used in the partial phase reconstruction (PPR) presented in [14] with a signal-dependent one. The new signal-dependent confidence domain is estimated by applying a sinusoidal model on the estimated individual source magnitude spectrum. We explore the effectiveness of the proposed approach for two scenarios, with and without MISI. We evaluate the separation performance obtained by the proposed method versus the state-of-the-art for both oracle and quantized magnitude spectra. We further report the convergence rate and the robustness of the phase estimation algorithms versus the amount of quantization error.

2. Previous Methods for Phase Estimation for Single-channel Source Separation

2.1. Iterative Phase Reconstruction Using Griffin and Lim

From a chronological standpoint, Griffin and Lim were the first ones to present a solution to recover the time-domain signal from the modified magnitude spectrum only [17]. They suggested to iteratively estimate the STFT which is closest to the original one in a MMSE sense:

$$\hat{\mathbf{S}}^{(k)} = G(|\mathbf{S}|e^{i\angle\hat{\mathbf{S}}^{(k-1)}}), \tag{1}$$

where function $G(\mathbf{S}) = \text{STFT}[\text{STFT}^{-1}\mathbf{S}]$ produces a consistent STFT from any set of complex values, and k as iteration index.

2.2. Multiple Input Spectrogram Inversion (MISI)

In [8], Gunawan and Sen suggested to incorporate the Griffin and Lim iterative signal reconstruction approach inside a blind source separation (BSS) framework. To this end, they proposed to re-distribute the remixing error calculated at each iteration as

  25 – 29 August 2013, Lyon, France

the feedback to be used for the signal reconstruction in the next iteration. They called this approach as "Multiple Input Spectrogram Inversion (MISI)" that distributes the remixing error defined in the time-frequency domain as:

$$\mathbf{E} = \mathbf{Y} - \sum_j \hat{\mathbf{S}}_j, \tag{2}$$

where j is the index to refer to the jth underlying source and we have $j \in [1, J]$ with J as the number of sources in the observed mixture $\mathbf{Y}$. The principle used in MISI is to distribute the remixing error, $\mathbf{E}$ equally among the underlying sources, leading to the jth modified source spectrum at the kth iteration given as below

$$\mathbf{C}_j^{(k)} = \mathrm{G}(\hat{\mathbf{S}}_j^{(k-1)}) + \frac{\mathbf{E}}{J}. \tag{3}$$

The phase spectrum updated as such is associated with the Wiener filter estimate to produce the output separated signal given as below

$$\hat{\mathbf{S}}_j^{(k)} = |\hat{\mathbf{S}}_j^{(0)}| e^{i\angle \mathbf{C}_j^{(k)}} \tag{4}$$

The MISI-based phase reconstruction in [8] worked in full-band and provided high separation performance given the oracle magnitude spectra of the sources. However, according to the results presented in [7, 14], MISI leads to some degradation in performance as soon as the signal spectra are quantized because the remixing error might be distributed on the silence parts of the individual signal spectra.

2.3. Partial Phase reconstruction (PPR)

In [14], authors proposed to confine the iterative phase estimation to a limited set of signal components (called confidence domain) and suggested the following iterative signal reconstruction:

$$\hat{\mathbf{S}}_j^{(k+1)} = \begin{cases} |\alpha_j Y| e^{j\angle G(\hat{\mathbf{S}}_j^{(k)})} & \text{for} \quad (n,m) \notin \Omega_j \\ \alpha_j Y & \text{for} \quad (n,m) \in \Omega_j \end{cases} \tag{5}$$

where $\Omega_j = \{(m,n) | \alpha_j(m,n) > \tau\}$ is defined as the confidence domain with τ as a constant, with m and n as the frequency and time indices. As reported in [14], PPR achieved better separation results compared to standard Griffin and Lim.

2.4. Partial Phase Reconstruction for Informed Source Separation (ISSIR)

In [7], the authors suggested a phase estimation approach in the MISI framework for informed source separation (ISS). The idea was to extend the fixed-threshold confidence domain ($\tau = 0.8$) in [14] to a MISI framework where the re-mixing error was distributed based on a signal activity function $D(m)$. The iterative signal reconstruction at the kth iteration is given by:

$$\hat{S}_j^{(k)}(m) = \Psi_j(m) \left(G(\hat{S}_j^{(k-1)}) + \frac{E(m)}{D(m)} \right) \tag{6}$$

where Ψ_j is defined as the confidence domain assigned 1 for $\alpha_j(m) > \rho$ and otherwise 0 with ρ as a constant threshold, and $D(n,m) = \sum_j \Psi_j(n,m)$.

2.5. Consistent Wiener Filter

Finally, the authors in [9] proposed to use consistent Wiener filter which applies the consistency constraint to solve the constrained optimization in the classical Wiener filter. Estimating a new Wiener filter by imposing the consistency constraint they showed:

$$\hat{\mathbf{S}}_j^{(k+1)} \leftarrow \frac{\frac{Y(m)}{\sum_{k \neq j} \alpha_k(m)} + \gamma G(\hat{\mathbf{S}}_j^{(k)})}{\frac{1}{\alpha_j(m)} + \frac{1}{\sum_{k \neq j} \alpha_k(m)} + \gamma}. \tag{7}$$

According to the results reported in [14], consistent Wiener filter achieved improved signal-to-distortion ratio (SDR) at the cost of lower signal-to-interference ratio (SIR) compared to the partial phase reconstruction (PPR). The consistent Wiener filtering required to resort to ad-hoc adjustments on key parameter and converges at a slower rate compared to PPR.

3. Proposed Signal-dependent Partial Phase Estimation

3.1. Cramer Rao Lower bound Justification for Sinusoidal-based Confidence Domain

Instead of employing the fixed-threshold confidence domain for phase estimation, we argue that we rely on the sinusoidal components estimated from the magnitude spectra of the sources as the confidence domain. According to [18], the Cramer Rao Band (CRB) of phase estimation for one sinusoid in noise, given its fundamental frequency and known/unknown amplitude, is given by

$$CRB(\phi_l | \omega_0) = \frac{2\sigma_l^2 \mathbf{D}}{N} \tag{8}$$

where $\mathbf{D} = \mathrm{diag}\frac{1}{A_l^2}$, N is the data length, and $\mathrm{SNR}_l = \frac{A_l^2}{\sigma_l^2}$ is the local signal-to-noise ratio (SNR) for the individual sinusoid. Then the asymptotic variance for the Cramer Rao Band (CRB) is $Var\{\hat{\phi}_l\} = \frac{2\sigma_l^2}{NA_l^2}$. Hence, the phase estimation error is governed by the data length N as well as the local SNR of the l-th sinusoidal component. This estimator asymptotically is unbiased for large enough N [19], explaining improved phase estimation performance for larger window length, as was reported in [20]. On the other hand, the variance for the phase estimation error gets the lowest at components with high SNR_l, arguably the sinusoidal components that exhibit the major impact in terms of representation of the individual signals. Therefore, it suffices to rely on the noisy phase of the Wiener filtered source spectra at a set of sinusoidal components as signal-dependent confidence domain. This choice is also supported by earlier founding that the noisy phase is a reasonable estimate for the clean phase spectrum for signal components with local signal-to-noise ratio of larger than 6 decibels [21]. The frequencies of sinusoids are indexed as $l \in [1, L]$ where L is the model order of sinusoids (here assumed fixed and known). As a proof of concept, Figure 2 shows how phase difference from one frame to another follows the harmonic structure of spectrogram. As expected, at sinusoidal trajectories, due to the high SNR_l given by (8), the variance of the phase estimation error gets the lowest. Hamming window and Chebyshev windows are chosen in order to maintain a better resolution for the harmonic structure in spectrogram and phase difference plots, respectively.

3.2. Estimating Sinusoidal-based Confidence Domain

A sinusoidal model is employed to the Wiener filtered magnitude spectrum estimate of each jth source $j \in [1, J]$, at each short-time frame due to speech non-stationarity, to find the maximum likelihood sinusoidal components (of fixed model order L) which are composed of amplitude, frequency and phase. To

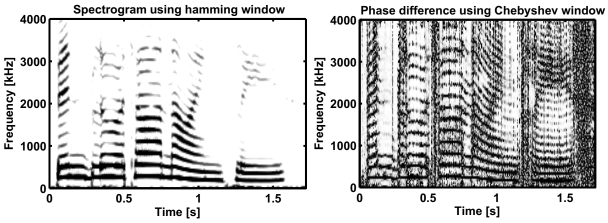

Figure 2: Spectrograms using Hamming window (left) and the phase difference spectrum between consecutive frames using Chebyshev window of dynamic range 40 dB (right) for a clean female speech signal. The results are shown for $L = 40$ number of sinusoids.

this end, we employ peak picking as proposed in [22]. Let the jth source as $s_j(n)$ consist of a set of complex sinusoids and additive noise written for $n = 0, \ldots, N-1$ as

$$s_j(n) = \sum_{l=1}^{L} a_{j,l} e^{j\omega_{j,l} n} + e_j(n), \tag{9}$$

where $e_j(n)$ is additive noise, l is an index that refers to the lth sinusoidal component characterized by $[\mathbf{A}_j, \boldsymbol{\omega}_j, \boldsymbol{\phi}_j]$ where $\mathbf{A}_j = [A_{j,1}A_{j,2}\cdots A_{j,L}]^T$, $\boldsymbol{\omega}_j = [\omega_{j,1}\omega_{j,2}\cdots\omega_{j,L}]^T$, $\boldsymbol{\phi}_j = [\phi_{j,1}\phi_{j,2}\cdots\phi_{j,L}]^T$ denote the jth source's amplitude, frequency and phase vectors, respectively, each of size $L \times 1$, and L being the model order of sinusoids for the desired signal, l refers to the lth sinusoid with $l \in [1, L]$, $\omega_{j,l}$ is the frequency of the lth sinusoid, $a_{j,l} = A_{j,l}e^{j\phi_{j,l}}$ are the complex-value amplitudes for speech for the lth sinusoid where $A_{j,l}$'s are non-zero, real amplitude and $\phi_{j,l}$'s are the phase. Note in this work we have no access to the clean phase of sinusoids for individual sources and we only have an estimation of the amplitude and frequencies of sinusoids obtained from the Wiener filtered magnitude spectra. The output of the sinusoidal-based confidence domain are $\Omega_{j,sin}$ given by the frequency of sinusoids $\boldsymbol{\omega}_j$ and $\mathbf{A}_j$.

3.3. Error Redistribution in Sinusoidal Domain

The frequencies of sinusoids given by sinusoidal model are used to define the sinusoidal-based confidence domain defined as: $\Omega_{j,sin} = \{\omega_{j,l}\}_{l=1}^{L}$. While the proposed sinusoidal-based confidence domain is employed to select where to keep the mixture phase and where to iteratively modify it, we further extend the phase estimation algorithm to the MISI framework by re-distributing the re-mixing error $\mathbf{E}$ according to some signal-activity function to be defined by sinusoidal model. To this end, we employ the amplitude of the selected sinusoids to obtain the speech presence probability (spp) contributed by the lth sinusoidal component defined as $p_{j,l} = \frac{A_{j,l}}{\sum_l^L A_{j,l}}$ and we have $0 \leq p_{j,l} \leq 1$ where $\Omega_{j,\text{sin}}$ refers to the sinusoidal confidence domain, defined as $\Psi_{j,sin}(m) = p_{j,m}$ if $(n,m) \in \Omega_{j,\text{sin}}$ and otherwise it is assigned to 0. The estimated complex spectrum at the kth iteration is given by distributing error on the sinusoidal components while keeping the phase as mixture for $(n,m) \in \Omega_{j,sin}$ and updating the phase based on G&L for $(n,m) \notin \Omega_{j,sin}$ and we have:

$$\hat{\mathbf{S}}_j^{(k+1)} = \begin{cases} G(\hat{\mathbf{S}}_j^{(k)}) + p(\boldsymbol{\Omega}_j|\mathbf{Y})\mathbf{E} & \text{for} \quad (n,m) \notin \Omega_{j,sin} \\ \alpha_j Y + p(\boldsymbol{\Omega}_j|\mathbf{Y})\mathbf{E} & \text{for} \quad (n,m) \in \Omega_{j,sin} \end{cases} \tag{10}$$

where following Bayes formula, we define $p(\boldsymbol{\Omega}_{j,\text{sin}}|\mathbf{Y})$ as the posterior probability of choosing $\boldsymbol{\Omega}_{j,\text{sin}}$ given the observation,

Algorithm 1 Iterative sinusoidal-based phase estimation

$[\hat{\mathbf{s}}_1(n), \hat{\mathbf{s}}_2(n)]$=**function**($\boldsymbol{\alpha}_2$,$\boldsymbol{\alpha}_2$,$\mathbf{Y}$)
Given the Wiener filter magnitude spectrum estimates $\hat{\mathbf{S}}_j = \boldsymbol{\alpha}_j\mathbf{Y}$ with $j \in [1, J]$, a sinusoidal model is used to estimate L sinusoidal components $\hat{\boldsymbol{\theta}}_{j,l} = \{\hat{A}_{j,l}, \hat{\omega}_{j,l}, \hat{\phi}_{j,l}\}_{l=1}^{L}$.
Sinusoidal confidence domain $\Omega_{j,sin} = \{\omega_{j,l}\}_{l=1}^{L}$
Sinusoidal-based speech presence probability $p_{j,l}$
$p(\boldsymbol{\Omega}_j|\mathbf{Y})$ defined in (11)
Calculate $\hat{\mathbf{S}}_j^{(k+1)}$ using (10)
Do iterations for $k \leq K$.

as:

$$p(\boldsymbol{\Omega}_{j,\text{sin}}|\mathbf{Y}) = \frac{p_{j,l}p(\boldsymbol{\Omega}_{j,\text{sin}})}{\sum_j p_{j,l}}, \tag{11}$$

indicating the relative contribution of each source with respect to the other one in the observed mixture. We further assume equal probability for both speaker classes, i.e., $p(\boldsymbol{\Omega}_{j,\text{sin}}) = 0.5$.

As a proof of concept, Figure 3 shows the principle of the proposed sinusoidal-based confidence domain and compares it with a fixed-threshold confidence domain as suggested by [14]. The proposed sinusoidal confidence domain is represented by the frequencies of sinusoids (middle panel). Furthermore, the amplitude of sinusoids are used to find the speech presence probability contributed by each lth sinusoid of the jth speaker as defined by $p_{j,l}$ (bottom panel). The steps required in the proposed phase estimation algorithm are presented and itemized in Algorithm 1. The iterative calculations are performed on the whole signal.

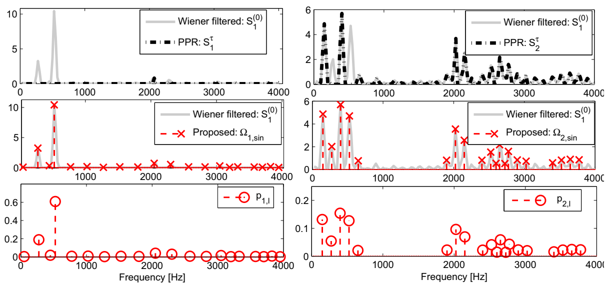

Figure 3: Showing (top) ideal Wiener filters along with the estimated magnitude spectrum using fixed threshold confidence domain in (PPR) using $\tau = 0.8$, (middle) sinusoidal confidence domain and the estimated magnitude spectrum, and (bottom) speech presence probability of sinusoidal confidence domain.

4. Experimental Results

4.1. Experimental Setup

In our experiments, we chose the GRID corpus database [5]. We randomly chose 20 speakers utterances (10 male and 10 female) and mixed them at 0 decibels signal-to-signal ratio. As our frame setup we chose 32 ms as window length and 4 ms as frameshift. As our benchmarks, we report the results of PPR [14], ISSIR [7], GS [8], GL [17], consistent Wiener filter [9] and the Wiener filter. From our experiment, we found that the recommended value of the threshold $\tau = 0.8$ and the number of iterations $K = 20$ as in [14] leads to the best performance for the speech separation scenario. For ISSIR, following the recommendation in [7], we chose $\rho = 0.01$ and $K = 100$. For the consistent Wiener filter presented in [9], we followed the same strategy of gradually increasing an initialized small γ value to impose the consistency constraint to the Wiener filter. For Griffin and Lim [17] as well as Gunawan and Sen [8], we selected 200 iterations to ensure of their convergence. As performance criteria, we report BSS-EVAL composed of SDR, SIR, and signal-to-artifact ratio (SAR) all measured in decibels.

4.2. Impact of Sinusoidal Model Order and Iterations

We studied the performance of the proposed method for a range of number iterations ($0 < K < 50$) and sinusoidal model order ($5 < L < 45$). From the results, we optimize for the model order of $L = 10$ and the iterations number of $K = 30$.

4.3. Overall Phase estimation Performance Compared to Benchmarks

4.3.1. Oracle Scenario

We report the convergence behavior of the proposed phase estimation algorithm versus others for oracle scenario. The results are shown in Figure 4. The proposed method achieves the fastest convergence among the others with slight improved performance in SDR and SAR compared to PPR [14]. The consistent Wiener filter achieves the highest SDR score, however suffers from a slow convergence before providing a high enough SIR performance. Both GS and GL provide large enough SIR but at the expense of significantly low SDR and SAR values, which is similar to the results reported for music source separation in [7, 14, 15]. Following the results reported in [10], ISSIR achieves slightly degraded performance compared to PPR in [14] for oracle scenario.

4.3.2. Quantized Scenario

To evaluate the robustness of the different phase estimation algorithm we consider the quantized magnitude spectrum scenario before producing the Wiener filters as inputs to phase estimation algorithms. The complex signal spectra are corrupted with additive white Gaussian noise at a range of $0 \leq \text{SQNR}_{in} \leq 60$ with 10 decibels stepsize (see Figure 5). The BSS EVAL results are shown in Figure 6. The proposed method shows a robust performance with slightly improved SDR and SAR scores compared to PPR (with fixed threshold confidence domain) [14]. Again, consistent Wiener filter performs best in terms of SDR but at the expense of lower SIR performance compared to the proposed method as well as with a lower convergence rate (Figure 4). For $0 \leq \text{SQNR}_{in} \leq 20$ ISSIR presents better SIR performance compared to the proposed method, PPR and consistent Wiener filter but at the expense of a lower SDR and SAR. Both GL and GS again provide very high SIR performance for large SQNR scenarios (low quantization error), but only at the expense of getting a significantly lower SDR and SAR. Informal listening tests showed that the proposed method leads to reduced trace of interference in the separated phase-estimated source signals[1].

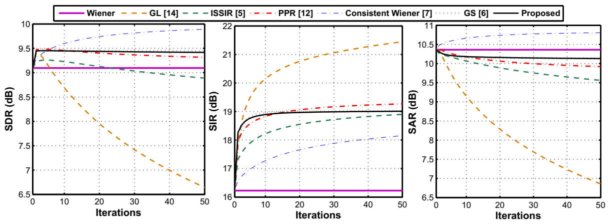

Figure 4: Showing the BSS EVAL: (left) SDR, (middle) SIR, and (right) SAR calculated for different phase estimation algorithms versus the number of iterations. The magnitude spectra are obtained by oracle Wiener filter.

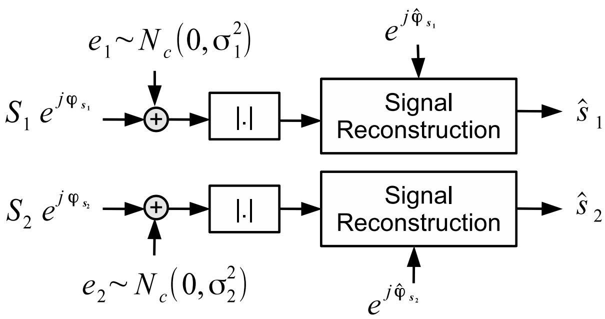

Figure 5: Simulating magnitude spectrum scenario by adding AWGN noise to each source samples.

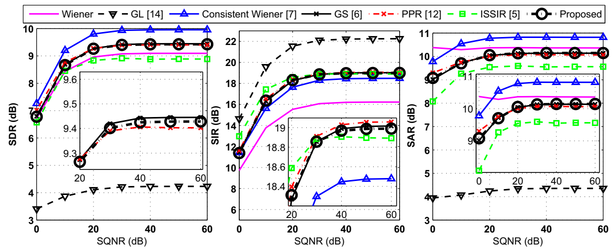

Figure 6: Showing the BSS EVAL results for quantized magnitude scenario measured by $0 \leq \text{SQNR} \leq 60$ decibels: (left) SDR, (middle) SIR, and (right) SAR all measured in decibels for different phase estimation algorithms.

5. Conclusion

In this paper, we proposed an iterative sinusoidal-based phase reconstruction algorithm for single-channel source separation. To this end, as the first step, we suggested a new signal-dependent confidence domain estimated by applying sinusoidal model on Wiener filtered magnitude spectra. Furthermore, we extended the previous iterative remixing error redistribution in fullband spectra to sinusoidal domain. The proposed method leads to an improved or at least a comparable separation performance compared to other recent state-of-the-art phase estimation methods. We argue that the choice of sinusoidal components as the confidence domain is supported by the high local signal-to-noise ratio of these spectral components, as well as a lower estimation error variance, rendering the phase estimates as a reliable one.

The proposed method requires to maintain the noisy phase at sinusoidal components only (as low as 10 sinusoids), arguably lower in dimension than the previous STFT (fullband) scenario with a fixed threshold. Therefore, in audio coding application for example, it is possible to transmit less information about the noisy phase i.e. at sinusoids only rather than the fullband STFT, when reconstructing the signal at the receiver end. The proposed method, among the others, balances a tradeoff between improved signal-to-distortion ratio (SDR) and signal-to-interference (SIR) without degradation in signal-to-artifact (SAR).

[1]Some audio wave files are available at the following link: `http://www.spsc.tugraz.at/IS2013phase`

6. References

[1] T. Virtanen, "Monaural sound source separation by nonnegative matrix factorization with temporal continuity and sparseness criteria," *IEEE Trans. Audio, Speech, and Language Process.*, vol. 15, no. 3, pp. 1066 – 1074, 2007.

[2] P. Mowlaee, R. Saeidi, M.G. Christensen, Z.-H. Tan, T. Kinnunen, P. Franti, and Jensen S.H., "A joint approach for single-channel speaker identification and speech separation," *Audio, Speech, and Language Processing, IEEE Transactions on*, vol. 20, no. 9, pp. 2586–2601, nov. 2012.

[3] P. Mowlaee, M. Christensen, and S. Jensen, "New results on single-channel speech separation using sinusoidal modeling," *IEEE Trans. Audio, Speech, and Language Process.*, vol. 19, no. 5, pp. 1265 – 1277, 2011.

[4] P. Mowlaee, *New Strategies for Single-channel Speech Separation*, Ph.D. thesis, Institut for Elektroniske Systemer, Aalborg Universitet, 2010.

[5] M. Cooke, J. R. Hershey, and S. J. Rennie, "Monaural speech separation and recognition challenge," *Elsevier Computer Speech and Language*, vol. 24, no. 1, pp. 1–15, 2010.

[6] M. Stark, M. Wohlmayr, and F. Pernkopf, "Source-filter based single channel speech separation using pitch information," *IEEE Transactions on Audio, Speech and Language Processing*, vol. 19, pp. 242 – 255, Feb. 2011.

[7] N. Sturmel and L. Daudet, "Informed source separation using iterative reconstruction," *Audio, Speech, and Language Processing, IEEE Transactions on*, vol. 21, no. 1, pp. 178 –185, jan. 2013.

[8] D. Gunawan and D. Sen, "Iterative phase estimation for the synthesis of separated sources from single-channel mixtures," *Signal Processing Letters, IEEE*, vol. 17, no. 5, pp. 421 –424, may 2010.

[9] J. Le Roux, E. Vincent, Y. Mizuno, H. Kameoka, N. Ono, and S. Sagayama, "Consistent Wiener filtering: generalized time-frequency masking respecting spectrogram consistency," in *9th Int. Conf. on Latent Variable Analysis and Signal Separation (LVA/ICA)*, 2010, pp. 89–96.

[10] J. Le Roux and E. Vincent, "Consistent Wiener filtering for audio source separation," *IEEE signal processing letters*, vol. 20, no. 3, pp. 217 – 220, 2013.

[11] P. Mowlaee, R. Saiedi, and R. Martin, "Phase estimation for signal reconstruction in single-channel speech separation," in *Proceedings of the International Conference on Spoken Language Processing*, 2012.

[12] P. Mowlaee and R. Martin, "On phase importance in parameter estimation for single-channel source separation," in *The International Workshop on Acoustic Signal Enhancement (IWAENC)*, 2012.

[13] P. Mowlaee and R. Saeidi, "On phase importance in parameter estimation in single-channel speech enhancement," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Processing*, May 2013, pp. 7462 – 7466.

[14] N. Sturmel and L. Daudet, "Iterative phase reconstruction of wiener filtered signals," in *Acoustics, Speech and Signal Processing (ICASSP), 2012 IEEE International Conference on*, march 2012, pp. 101 –104.

[15] N. Sturmel, L. Daudet, and L. Girin, "Phase-based informed source separation of music," in *Proc. 13th International Conference on Digital Audio Effects (DAFx-12)*, Sept. 2012.

[16] P. Mowlaee, M. Watanabe, and R. Saeidi, "Show & tell: Phase-aware single-channel speech enhancement," in *14th Annual Conference of the International Speech Communication Association*, 2013.

[17] D. Griffin and Jae Lim, "Signal estimation from modified short-time fourier transform," *Acoustics, Speech and Signal Processing, IEEE Transactions on*, vol. 32, no. 2, pp. 236 – 243, apr 1984.

[18] A. Nehorai and B. Porat, "Adaptive comb filtering for harmonic signal enhancement," *Acoustics, Speech and Signal Processing, IEEE Transactions on*, vol. 34, no. 5, pp. 1124 – 1138, oct 1986.

[19] S.M. Kay, *Fundamentals of Statistical Signal Processing, Volume I: Estimation Theory*, Prentice Hall, 1993.

[20] K. K. Paliwal and L. Alsteris, "Usefulness of phase spectrum in human speech perception," in *Proc. Eurospeech*, 2003, pp. 2117–2120.

[21] P. Vary, "Noise suppression by spectral magnitude estimation mechanism and theoretical limits," *Signal Processing*, vol. 8, no. 4, pp. 387 – 400, 1985.

[22] R. McAulay and T.F. Quatieri, "Speech analysis/synthesis based on a sinusoidal representation," *Acoustics, Speech and Signal Processing, IEEE Transactions on*, vol. 34, no. 4, pp. 744–754, 1986.