ON RELATIONSHIPS BETWEEN AMPLITUDE AND PHASE OF SHORT-TIME FOURIER TRANSFORM

Suehiro Shimauchi[†], *Shinya Kudo*[‡], *Yuma Koizumi*[†], *and Ken'ichi Furuya*[‡]

[†] NTT Media Intelligence Laboratories, NTT Corporation, Tokyo, Japan
[‡] Oita University, Oita, Japan

ABSTRACT

The relationships between the amplitude and phase of the short-time Fourier transform (STFT) are investigated. By choosing the Gaussian window for the STFT, we reveal that the group delay and instantaneous frequency of each signal segment, both of which are derived from the phase by definition, can also be explicitly linked with the amplitude. As a result, the amplitude and phase can also be linked through the group delay or instantaneous frequency without making any assumptions for the phase property of the target signals, e.g., minimum, maximum, or linear phase. The theoretical basis is also confirmed in numerical simulations.

Index Terms— Short-time Fourier transform, amplitude, phase, group delay, instantaneous frequency

1. INTRODUCTION

The short-time Fourier transform (STFT) is a typical tool for analyzing or modifying speech or audio signals. In many cases of signal reconstruction based on the STFT analysis, modification, and synthesis, e.g., speech enhancement, only the amplitude of the complex-valued output of the STFT is mostly focused upon [1, 2, 3]. However, the importance of the phase properties of the STFT has also been recognized [4, 5, 6, 7], and several methods that take phase into account have been developed [8].

The redundancy of STFT representation with overlapped signal segments introduces some dependence between its amplitude and phase. This dependence makes it helpful to develop signal-reconstruction algorithms that modify both the amplitude and phase or complex-valued output of the STFT [9, 10, 11, 12]. However, the dependence between the amplitude and phase has not necessarily been revealed in explicit forms. Some of the algorithms are based on the iterative operation proposed by Griffin and Lim [13, 14, 15, 16]. The iterative approach, which modifies the STFT spectrogram iteratively after every synthesis and analysis operations of the STFT, does not require explicit relationships between the amplitude and phase, but it minimizes the inconsistency of STFT spectrograms before and after the synthesis and analysis iterations [15].

By focusing on the models of the signals, particular relationships between the amplitude and phase are assumed. Phase-reconstruction algorithms based on the harmonic structure of voiced speech were proposed by Krawczyk and Gerkmann [17] and Mowlaee and Kulmer [18]. Sugiyama and Miyahara [19] proposed a tapping-noise suppression algorithm based on the idea of randomizing the phase when impulsive noise is detected. The sparseness of the signals was also taken into account by Eldar et al [16].

In this paper, we explicitly introduce a set of relationships between the amplitude and phase of the STFT, which might be implicitly exploited in conventional iterative algorithms. Our goal is to find the relationships that are not limited to particular signal models and to contribute to the development of further improved signal-reconstruction algorithms. However, it is indeed difficult to find such relationships for the general STFT structure. Thus, we consider a particular case in which the Gaussian window is chosen for the STFT operation. By exploiting the property of the Gaussian window, we show that the group delay and instantaneous frequency of each signal segment, both of which are derived from the phase by definition, can also be explicitly linked with the amplitude without depending on signal models. Consequently, the amplitude and phase are also directly linked through the group delay or instantaneous frequency. Yegnanarayana et al. [20] showed an explicit relationship between the amplitude and phase through cepstral coefficients. However, this relationship holds only in the case in which the minimum or maximum phase signals are applied. Krawczyk and Gerkmann [17] took into account the linearity of the phase across time in their phase-estimation algorithm for voiced speech. Similarly, Mowlaee and Kulmer [18] took into account the smoothness of the phase across time and frequency for voiced speech enhancement. Sugiyama and Miyahara [19] also took into account the linearity of the phase across frequency in their tapping-noise suppression algorithm. In contrast to these approaches, our theoretical basis is derived without any signal models. Therefore, it is not theoretically limited to signals with a minimum, maximum, or linear phase property.

2. EXPLICIT LINK OF AMPLITUDE AND PHASE

In this section, we investigate how the amplitude and phase of the STFT are linked together due to the redundant representation of the STFT. We consider the continuous STFT:

$$X_{t,\omega} = \int_{-\infty}^{\infty} w_{\tau-t} x_\tau e^{-j\omega\left(\tau - t + \frac{T}{2}\right)} d\tau, \tag{1}$$

where x_τ is a time-domain signal, $w_{\tau-t}$ is the sliding window function that focuses on the signal around time $\tau = t$, and $X_{t,\omega}$ is the complex-valued STFT output as a function of time t and angular frequency ω. The origin of the time axis for the Fourier transform is modified as $\tau = t - \frac{T}{2}$, so it becomes compatible with the discrete time version based on the discrete Fourier transform (DFT), where the focusing range T corresponds to the DFT window length. The derivatives of $X_{t,\omega}$ along time and angular frequency can be respectively obtained as follows,

$$\frac{\partial X_{t,\omega}}{\partial t} = \int_{-\infty}^{\infty} \frac{\partial w_{\tau-t}}{\partial t} x_\tau e^{-j\omega\left(\tau-t-\frac{T}{2}\right)} d\tau + j\omega \int_{-\infty}^{\infty} w_{\tau-t} x_\tau e^{-j\omega\left(\tau-t+\frac{T}{2}\right)} d\tau = \int_{-\infty}^{\infty} \frac{\partial w_{\tau-t}}{\partial t} x_\tau e^{-j\omega\left(\tau-t+\frac{T}{2}\right)} d\tau + j\omega X_{t,\omega}, \quad (2)$$

$$\frac{\partial X_{t,\omega}}{\partial \omega} = -j\int_{-\infty}^{\infty} \left(\tau - t - \frac{T}{2}\right) w_{\tau-t} x_\tau e^{-j\omega\left(\tau-t-\frac{T}{2}\right)} d\tau = -j\int_{-\infty}^{\infty} (\tau - t)\, w_{\tau-t} x_\tau e^{-j\omega\left(\tau-t-\frac{T}{2}\right)} d\tau - j\frac{T}{2} X_{t,\omega}. \quad (3)$$

To eliminate the integral terms of x_τ from Eqs. (2) and (3), we choose the Gaussian window as $w_{\tau-t}$,

$$w_{\tau-t} = e^{-\frac{(\tau-t)^2}{2\sigma^2}}. \quad (4)$$

Then its derivative along time becomes

$$\frac{\partial w_{\tau-t}}{\partial t} = \frac{(\tau - t)}{\sigma^2} e^{-\frac{(\tau-t)^2}{2\sigma^2}}, \quad (5)$$

where σ is a standard deviation. By substituting Eq. (4) into Eq. (3) and Eq. (5) into Eq. (2), the following equation is obtained,

$$\sigma^2 \frac{1}{X_{t,\omega}} \frac{\partial X_{t,\omega}}{\partial t} - j\frac{1}{X_{t,\omega}} \frac{\partial X_{t,\omega}}{\partial \omega} - j\sigma^2\omega + \frac{T}{2} = 0. \quad (6)$$

Now, the log of $X_{t,\omega}$ and its derivatives along time and angular frequency can also be expressed respectively by using the amplitude $A_{t,\omega}$ and phase $\phi_{t,\omega}$ of $X_{t,\omega}$ as

$$\log X_{t,\omega} = \log A_{t,\omega} + j\phi_{t,\omega}, \quad (7)$$

$$\frac{\partial \log X_{t,\omega}}{\partial t} = \frac{1}{X_{t,\omega}} \frac{\partial X_{t,\omega}}{\partial t} = \frac{\partial \log A_{t,\omega}}{\partial t} + j\frac{\partial \phi_{t,\omega}}{\partial t}, \quad (8)$$

$$\frac{\partial \log X_{t,\omega}}{\partial \omega} = \frac{1}{X_{t,\omega}} \frac{\partial X_{t,\omega}}{\partial \omega} = \frac{\partial \log A_{t,\omega}}{\partial \omega} + j\frac{\partial \phi_{t,\omega}}{\partial \omega}. \quad (9)$$

By using Eqs. (8) and (9), Eq. (6) becomes

$$\underbrace{\left[\sigma^2 \frac{\partial \log A_{t,\omega}}{\partial t} + \frac{\partial \phi_{t,\omega}}{\partial \omega} + \frac{T}{2}\right]}_{\text{real part}} - j\underbrace{\left[\frac{\partial \log A_{t,\omega}}{\partial \omega} - \sigma^2 \frac{\partial \phi_{t,\omega}}{\partial t} + \sigma^2\omega\right]}_{\text{imaginary part}} = 0. \quad (10)$$

From the real and imaginary parts of Eq. (10), the group delay $\mathrm{GD}_{t,\omega}$ and instantaneous frequency $\mathrm{IF}_{t,\omega}$ are respectively given as

$$\mathrm{GD}_{t,\omega} = -\frac{\partial \phi_{t,\omega}}{\partial \omega} = \sigma^2 \frac{\partial \log A_{t,\omega}}{\partial t} + \frac{T}{2}, \quad (11)$$

$$\mathrm{IF}_{t,\omega} = \frac{1}{2\pi} \frac{\partial \phi_{t,\omega}}{\partial t} = \frac{1}{2\pi\sigma^2} \frac{\partial \log A_{t,\omega}}{\partial \omega} + \frac{\omega}{2\pi}. \quad (12)$$

Equations (11) and (12) show that the group delay and instantaneous frequency, both of which are linked with the phase by definition, are also explicitly linked with the amplitude. Therefore, the group delay and instantaneous frequency can be calculated from amplitude $A_{t,\omega}$ without obtaining unwrapped phase $\phi_{t,\omega}$. Note that, since the redundant property of the STFT representation is not limited to the case in which the Gaussian window in Eq. (4) is chosen, some links other than Eqs. (11) and (12) would be obtained in other cases, though they might be more complicated than Eqs. (11) and (12). Consequently, the following links between phase and amplitude are obtained from Eqs. (11) and (12). The property of phase $\phi_{t,\omega}$ can be estimated from the information of amplitude $A_{t,\omega}$ by using

$$\phi_{t,\omega} = -\sigma^2 \int \frac{\partial \log A_{t,\omega}}{\partial t} d\omega - \frac{T}{2}\omega + C_1, \quad (13)$$

or

$$\phi_{t,\omega} = \frac{1}{\sigma^2} \int \frac{\partial \log A_{t,\omega}}{\partial \omega} dt + \omega t + C_2. \quad (14)$$

On the contrary, the property of amplitude $A_{t,\omega}$ can be estimated from the information of phase $\phi_{t,\omega}$ by using

$$\log A_{t,\omega} = -\frac{1}{\sigma^2} \int \frac{\partial \phi_{t,\omega}}{\partial \omega} dt - \frac{Tt}{2\sigma^2} + C_3, \quad (15)$$

or

$$\log A_{t,\omega} = \sigma^2 \int \frac{\partial \phi_{t,\omega}}{\partial t} d\omega - \frac{\sigma^2\omega^2}{2} + C_4, \quad (16)$$

though phase $\phi_{t,\omega}$ is required to be unwrapped.

To confirm the above relationships in practice, their discrete time versions are necessary, and some boundary conditions are required to determine the constants of the integrals C_1, C_2, C_3, and C_4. We discuss these issues in the following sections.

3. APPROXIMATION FOR DISCRETE STFT

According to the relationships derived in the previous section, we now give examples of their discrete time versions. As mentioned in the previous section, the estimation of amplitude $A_{t,\omega}$ based on Eq. (15) or (16) requires phase $\phi_{t,\omega}$ to be unwrapped, and this unwrapping is a nonunique process. Therefore, we concentrate here on approximating amplitude-based estimation processes for group delay $GD_{t,\omega}$ with Eq. (11), instantaneous frequency $IF_{t,\omega}$ with Eq. (12), and phase $\phi_{t,\omega}$ with Eq. (13) or (14).

Time t and angular frequency ω are discretized by the discrete time index n and discrete angular frequency index k,

$$t = \frac{n}{f_s}, \quad (17)$$

$$\omega = \frac{2\pi k f_s}{N}, \quad (18)$$

where f_s is the sampling frequency, and N is the number of finite length window samples and also corresponds to the point number of DFT. The focusing range T in Eq. (1) is used as the window truncation range,

$$T = \frac{N}{f_s}. \tag{19}$$

Since our aim is to numerically confirm the theoretical basis, we consider a case in which the DFT analysis is applied to every one-sample-shifted segment, i.e., $N-1$ samples are overlapped.

· **Group delay estimated from amplitude**

In Eq. (11), the derivative of $\log A_{t,\omega}$ is approximated by the symmetric difference quotient. As a result, an estimate of group delay $\mathrm{GD}_{n,k}$ can be calculated as

$$\widehat{\mathrm{GD}}_{n,k} = \frac{\sigma^2 f_s}{2} \log \frac{A_{n+1,k}+\delta}{A_{n-1,k}+\delta} + \frac{N}{2f_s}, \tag{20}$$

where δ is a small positive constant introduced for numerical stability.

· **Instantaneous frequency estimated from amplitude**

For Eq. (12), an estimate of instantaneous frequency $\mathrm{IF}_{n,k}$ can be calculated as

$$\widehat{\mathrm{IF}}_{n,k} = \frac{N}{8\pi^2\sigma^2 f_s} \log \frac{A_{n,k+1}+\delta}{A_{n,k-1}+\delta} + \frac{kf_s}{N}. \tag{21}$$

· **Phase estimated from amplitude**

In Eq. (13), the derivative of $\log A_{t,\omega}$ is approximated by the symmetric difference quotient, and the integral of the derivative is approximated by the trapezoidal rule. As a result, an estimate of phase at n can be calculated by the recursive equation of k,

$$\hat{\phi}_{n,k} = \hat{\phi}_{n,k-1} - \frac{\pi\sigma^2 f_s^2}{2N} \log \frac{A_{n+1,k}A_{n+1,k-1}+\delta}{A_{n-1,k}A_{n-1,k-1}+\delta} - \pi. \tag{22}$$

Since we assume that x_n is the real-valued signal, the initial value $\hat{\phi}_{n,0}$ is 0 or π. We give information on how to determine initial value $\hat{\phi}_{n,0}$ in the next section. When the recursive operation starts from $k=0$, the errors that accumulate during the estimation may increase in higher frequency bands.

Similarly, Eq. (14) can be approximated as follows. In this case, an estimate of phase at k can be calculated by the following recursive equation of n,

$$\hat{\phi}_{n,k} = \hat{\phi}_{n-1,k} + \frac{N}{8\pi\sigma^2 f_s^2} \log \frac{A_{n,k+1}A_{n-1,k+1}+\delta}{A_{n,k-1}A_{n-1,k-1}+\delta} + \frac{2\pi k}{N}. \tag{23}$$

Initial value $\hat{\phi}_{0,k}$ can be set to 0 if there is no signal at the starting point. However, the errors that accumulate during the estimation may become severer than those with Eq. (22), because time index n can be much larger than the frequency index k.

Note that, although sampling frequency f_s appears in Eqs. (22) and (23), the phase estimate $\hat{\phi}_{n,k}$ does not depend on f_s. This is because σ^2 is chosen to be in inverse proportion to f_s^2.

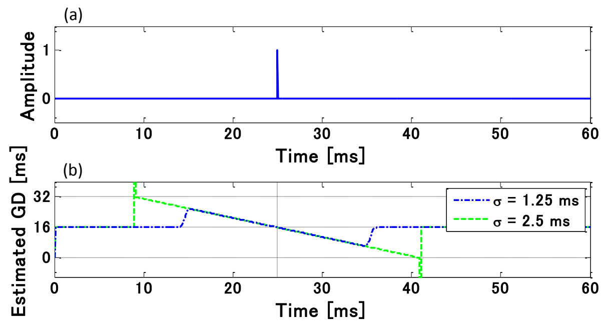

Fig. 1. Results of group-delay estimation. (a) Given impulse signal. (b) Estimated group-delay properties.

4. NUMERICAL SIMULATIONS

In this section, we explain some simulation results to demonstrate the numerical behavior of the theoretical relationships in Eqs. (11) to (14) by using their approximated discrete time versions obtained in Sec. 3. Although there are several factors that affect the approximation accuracy of the discrete versions, we here show the dependence on the window shape of $\sigma = 1.25$ or 2.5 ms for a fixed window length of $N = 512$. The shift size of the analysis segment was fixed to one sample. The sampling frequency was chosen as $f_s = 16$ kHz throughout the simulations.

· **Group-delay estimation by amplitude**

Figure 1 shows the group-delay properties estimated from the amplitude of the STFT of the impulse signal by using Eq. (20). The estimated group delays did not depend on frequency index k due to the amplitude property of the impulse signal. When the window covered the impulse signal, the values of the estimated group delay almost linearly decreased as expected. However, there were unexpected errors for $\sigma = 2.5$ ms due to negligible window edges. Even during the period when the window covered no signal, the group delay was uniquely calculated as 16 ms, which corresponds to the expectation of the delay time of the Gaussian window of $N = 512$ with $f_s = 16$ kHz.

· **Instantaneous-frequency estimation by amplitude**

Figure 2 shows the instantaneous-frequency properties estimated from the amplitude of the STFT by using Eq. (21). The applied signal was generated as

$$x_n = \sum_{m=1}^{7} \sin\left(\frac{2\pi 1000 mn}{f_s}\right). \tag{24}$$

Each frequency component was properly detected at the corresponding DFT bins. In this case, unexpected errors were observed for $\sigma = 1.25$ ms.

· **Phase estimation by amplitude**

The phase properties were estimated using Eq. (22) or (23) for the given amplitude spectrograms obtained respectively

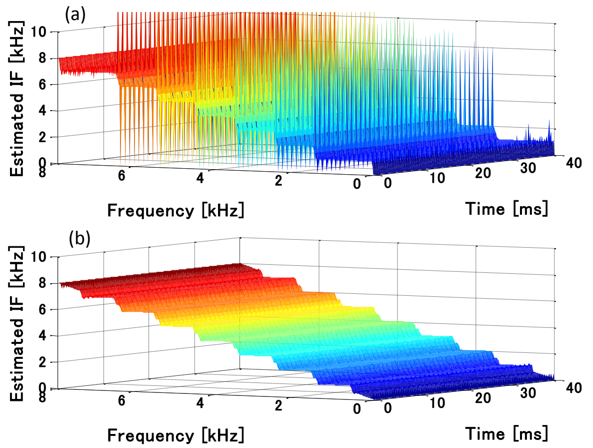

Fig. 2. Results of instantaneous-frequency estimation. (a) $\sigma = 1.25$ ms. (b) $\sigma = 2.5$ ms.

Table 1. SDRs of reconstructed signals with true amplitude and estimated phase.

	with Eq. (22)		with Eq. (23)	
	Male	Female	Male	Female
$\sigma = 1.25$ ms	19.4 dB	15.1 dB	-0.69 dB	-0.24 dB
$\sigma = 2.5$ ms	10.9 dB	7.65 dB	-0.56 dB	-0.69 dB

from the actual male and female speech signals, and the signals were reconstructed using the given amplitude and estimated phase. In the simulations, initial value $\hat{\phi}_{n,0}$ was set to the same value of the true phase, in order to concentrate on evaluating the potential performance of the theoretical basis. However, we give a practical clue of how to detect the phase alternations between 0 and π. As shown in Fig. 3, the alternations of the true phase were highly synchronized with spikes of the estimated group delay.

Table 1 lists the evaluation results of the signal-to-distortion ratios (SDRs) for the reconstructed signals, where the SDRs were calculated following the definition in [21]. While the algorithm with Eq. (23) did not provide good estimation results on any given condition, the algorithm with Eq. (22) achieved about 20 dB SDR for the male signal with $\sigma = 1.25$ ms. Figure 4 shows an example of the true and estimated phases for the male speech with $\sigma = 1.25$ ms at lower frequency bands ($k = 0, 1, 2, 3$). where the phases are plotted after wrapping. As mentioned above, we gave true phase for the estimation of $k = 0$. The estimated phases obtained by Eq, (22) (Fig. 4 (b)) mostly fit the true phases (Fig. 4 (a)), though some spike errors were observed at the moments of the phase alternations of $k = 0$. The estimated phases obtained by Eq, (23) (Fig. 4 (c)) did not fit the true phases due to the initialization problem, and this could be the reason for the bad SDR scores. However, the relative shapes of the estimated phases were rather similar to those of the true phases, and the spike errors were hardly observed in this case.

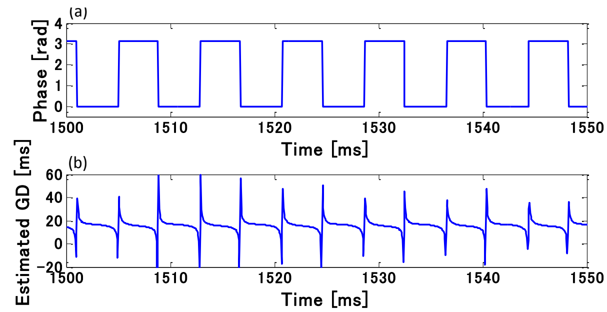

Fig. 3. Observed relationship between phase and group delay of male speech analyzed with $\sigma = 1.25$ ms.. (a) True phase for $k = 0$. (b) Estimated group delay for $k = 0$.

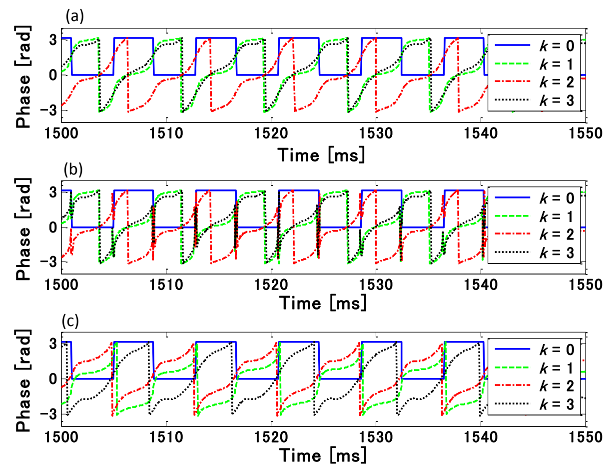

Fig. 4. Example of true and estimated phases. (a) True phase. (b) Estimated with Eq. (22). (c) Estimated with Eq. (23). (true phase was given for $k = 0$.)

5. CONCLUSION

We presented a theoretical basis to explicitly connect the group delay, instantaneous frequency, and phase to the amplitude under a particular STFT structure, which uses the Gaussian window. Since the numerical approximations require rather tight conditions, e.g., a Gaussian window with a tighter shape, or one-sample shift overlapping, it may not be easy to straightforwardly apply them to practical applications. However, we believe that the basis can provide a novel insight for further development of signal-reconstruction algorithms that take into account both the amplitude and phase properties, phase-unwrapping algorithms that exploit even amplitude properties, and so on.

6. REFERENCES

[1] D. L. Wang and J. S. Lim, “The unimportance of phase in speech enhancement,” *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 30, no. 4, pp. 679–681, Aug 1982.

[2] Y. Ephraim and D. Malah, "Speech enhancement using a minimum-mean square error short-time spectral amplitude estimator," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 32, no. 6, pp. 1109–1121, Dec 1984.

[3] J. S. Erkelens, R. C. Hendriks, and R. Heusdens, "On the estimation of complex speech dft coefficients without assuming independent real and imaginary parts," *IEEE Signal Processing Letters*, vol. 15, pp. 213–216, 2008.

[4] A. V. Oppenheim and J. S. Lim, "The importance of phase in signals," *Proceedings of the IEEE*, vol. 69, no. 5, pp. 529–541, May 1981.

[5] K. K. Paliwal and L. D. Alsteris, "On the usefulness of STFT phase spectrum in human listening tests," *Speech Communication*, vol. 45, no. 2, pp. 153–170, Feb 2005.

[6] G. Shi, M. M. Shanechi, and P. Aarabi, "On the importance of phase in human speech recognition," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 14, no. 5, pp. 1867–1874, Sept 2006.

[7] A. Maly and P. Mowlaee, "On the importance of harmonic phase modification for improved speech signal reconstruction," in *2016 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, March 2016, pp. 584–588.

[8] T. Gerkmann, M. Krawczyk-Becker, and J. Le Roux, "Phase processing for single-channel speech enhancement: History and recent advances," *IEEE Signal Processing Magazine*, vol. 32, no. 2, pp. 55–66, March 2015.

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

[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, March 2013.

[11] T. Gerkmann, "MMSE-optimal enhancement of complex speech coefficients with uncertain prior knowledge of the clean speech phase," in *2014 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, May 2014, pp. 4478–4482.

[12] D. S. Williamson, Y. Wang, and D. Wang, "Complex ratio masking for joint enhancement of magnitude and phase," in *2016 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, March 2016, pp. 5220–5224.

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

[14] X. Zhu, G. T. Beauregard, and L. L. Wyse, "Real-time signal estimation from modified short-time fourier transform magnitude spectra," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 15, no. 5, pp. 1645–1653, July 2007.

[15] J. Le Roux, Nobutaka Ono, and Shigeki Sagayama, "Explicit consistency constraints for STFT spectrograms and their application to phase reconstruction," in *ISCA Workshop Statistical Perceptual Audition (SAPA)*, 2008, pp. 23–28.

[16] Y. C. Eldar, P. Sidorenko, D. G. Mixon, S. Barel, and O. Cohen, "Sparse phase retrieval from short-time fourier measurements," *IEEE Signal Processing Letters*, vol. 22, no. 5, pp. 638–642, May 2015.

[17] M. Krawczyk and T. Gerkmann, "STFT phase reconstruction in voiced speech for an improved single-channel speech enhancement," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 22, no. 12, pp. 1931–1940, Dec 2014.

[18] P. Mowlaee and J. Kulmer, "Harmonic phase estimation in single-channel speech enhancement using phase decomposition and snr information," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 23, no. 9, pp. 1521–1532, Sept 2015.

[19] A. Sugiyama and R. Miyahara, "Tapping-noise suppression with magnitude-weighted phase-based detection," in *2013 IEEE Workshop on Applications of Signal Processing to Audio and Acoustics*, Oct 2013, pp. 1–4.

[20] B. Yegnanarayana, D. Saikia, and T. Krishnan, "Significance of group delay functions in signal reconstruction from spectral magnitude or phase," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 32, no. 3, pp. 610–623, Jun 1984.

[21] V. Emiya, E. Vincent, N. Harlander, and V. Hohmann, "Subjective and objective quality assessment of audio source separation," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 19, no. 7, pp. 2046–2057, Sept 2011.