Phase-Encoded Speech Spectrograms

Chandra Sekhar Seelamantula

Department of Electrical Engineering
Indian Institute of Science
Bangalore - 560 012, India
chandra.sekhar@ieee.org

Abstract

Spectrograms of speech and audio signals are time-frequency densities, and by construction, they are non-negative and do not have phase associated with them. Under certain conditions on the amount of overlap between consecutive frames and frequency sampling, it is possible to reconstruct the signal from the spectrogram. Deviating from this requirement, we develop a new technique to incorporate the phase of the signal in the spectrogram by satisfying what we call as the *delta dominance condition*, which in general is different from the well known minimum-phase condition. In fact, there are signals that are delta dominant but not minimum-phase and vice versa. The delta dominance condition can be satisfied in multiple ways, for example by placing a Kronecker impulse of the right amplitude or by choosing a suitable window function. A direct consequence of this novel way of constructing the spectrograms is that the phase of the signal is directly encoded or embedded in the spectrogram. We also develop a reconstruction methodology that takes such phase-encoded spectrograms and obtains the signal using the discrete Fourier transform (DFT). It is envisaged that the new class of phase-encoded spectrogram representations would find applications in various speech processing tasks such as analysis, synthesis, enhancement, and recognition.

Index Terms: magnitude spectrum, speech reconstruction, cepstrum, causal delta-dominant signal, minimum-phase signal.

1. Introduction

The Fourier spectrum of a discrete-time sequence is complex-valued in general and carries information both in magnitude and phase functions. In general, given a discrete-time sequence it is not possible to reconstruct it either from the magnitude spectrum or phase spectrum alone. However, sequences that can be modelled as outputs of rational transfer functions with poles and zeros inside the unit circle, are accurately specified by either magnitude spectrum or the phase spectrum. Such systems are said to be minimum-phase and they are associated with the minimum group delay out of all systems that have a specified magnitude spectrum. Therefore, for minimum-phase signals, either magnitude or phase spectrum become redundant.

In the case of speech signals, the minimum-phase property is seldom applicable. Even for voiced sounds, which are modelled as outputs of a stable autoregressive/all-pole transfer function in the context of linear prediction, the minimum-phase property is only an approximation. Therefore, in gen-

eral, speech signals cannot be reconstructed from magnitude or phase components of their Fourier spectrum alone. At best, one can only recover the minimum-phase component of a speech signal if one were to reconstruct the signal purely from the magnitude spectrum specification.

1.1. Prior art

The problem of reconstructing signals from their magnitude spectra has received considerable attention in the past, particularly during the 1980s and goes by the name of *Phase Retrieval*. It has its roots in optical imaging research, where the goal is to reconstruct a complex-valued object starting from non-negative intensity measurements. The initial contributions were made by Gerchberg and Saxton [1], and Fienup [2–4], who proposed iterative phase retrieval algorithms, which can be considered as special cases of alternating projection algorithms, from a convex optimization perspective [5, 6]. In the signal processing community, the initial contributions were made by Oppenheim and co-workers in a series of papers [7–10]. The most significant result in phase retrieval is the magnitude spectrum characterization of minimum-phase sequences and the associated Hilbert transform relation between the log-magnitude and phase spectra [11]. Minimum-phase sequences are associated with rational transfer functions whose poles and zeros lie within the unit circle in the complex z-plane, inherently having stability and invertibility guarantees. Hayes et al. [12] solved the phase retrieval problem for real sequences whose $\mathcal{Z}$-transforms do not contain reciprocal pole-zero pairs, with poles inside the unit circle and zeros outside the unit circle. Quatieri et al. [13] proposed iterative techniques for signal reconstruction from magnitude spectrum. The iterative algorithms are accurate for minimum-phase signals. Yegnanarayana et al. [14] considered minimum-phase signals and showed that group-delay functions play a significant role in recovering signals from their magnitude Fourier spectra [15]. Alsteris and Paliwal presented a review of various methods related to the short-time phase spectrum in speech processing [16].

In the past decade, the problem of phase retrieval has been vigorously pursued from a compressed sensing or sparsity perspective. Some notable contributions are the works of Moravec et al. [17], who developed a compressive phase retrieval algorithm, Ohlsson and Eldar [18], who developed sufficient conditions on the measurement ensemble to guarantee unique reconstruction for sparse signals, Netrapalli et al. [19], who presented theoretical convergence guarantees, Shechtman et al. [20], who proposed a greedy sparse phase retrieval algorithm, Eldar et al. [21], who considered the reconstruction problem from the short-time Fourier transform (STFT) magnitude, Lu and Vetterli [22], who solved a sparse spectral factorization problem as

a solution to the phase retrieval problem, Candès et al. [23], who developed the *PhaseLift* approach to solve the phase retrieval problem using semi-definite programming, etc. This is merely a representative list and by no means an exhaustive collection of the works in the area of phase retrieval. In most of these papers, applications related to image processing have been considered.

As far as our own contributions in this field are concerned, we developed a sparse counterpart of the Fienup algorithm in which only a k-sparse representation of the signal is computed in each iteration [24, 25]. We also proposed a non-iterative phase retrieval algorithm for minimum-phase signals based on the annihilating filter that is frequently used in spectral estimation [26]. An extension to a class of 2-D parametric signals was reported in [27].

1.2. Our contribution

This paper is motivated by one of our recent publications [28], in which we established some exact results for phase retrieval of continuous-domain functions that lie in a shift-invariant space spanned by a generator kernel and its integer-shifted versions. The weights attached to the kernel and its shifted versions constitute the discrete representation of the function in that space. Contrary to the common understanding that continuous-time functions that can be reconstructed from magnitude spectra must be minimum-phase and hence causal, we showed that exact reconstruction is possible even when the continuous-time functions are not causal (although the corresponding sequences in the shift-invariant representation must be), but the discrete-time sequence representations need not be minimum-phase. We introduced a new class of sequences known as causal, delta-dominant (CDD) sequences, which may or may not be minimum-phase, but allow for signal reconstruction from magnitude spectrum only. An interesting result that we proved in [28] is that *finite-length* CDD sequences are minimum-phase! Consequently, any finite-length sequence can be converted to a minimum-phase sequence by making it CDD, which essentially boils down to adding an impulse of the "right" magnitude at the origin, without having to directly worry about the locations of the zeros. Thus, one could *encode* the phase in the magnitude spectrum and retrieve it during reconstruction. The key contribution of this paper is the proof that phase encoding is possible. By considering application to speech signals, we show how one could construct *phase-encoded spectrograms*, from which it is possible to reconstruct the signal exactly.

In Section 2, we present the key result of this paper. We show that although any given segment of speech cannot be reconstructed exactly from its magnitude spectrum, by slightly modifying the signal, one can achieve exact reconstruction. The modification is in terms of adding a Kronecker impulse of appropriate strength at the origin, which allows the phase of the signal to be *encoded* in the magnitude spectrum. The amplitude of the impulse is specified depending on the values of the signal within that frame. As it will be apparent from the proof, the cepstrum plays a significant role in enabling signal reconstruction. The reconstruction technique is nonlinear, noniterative, and exact. In Section 3, we present some examples on synthesized as well as real speech data to validate the theoretical findings.

2. The key result: Phase encoding is possible

Consider a causal finite-length sequence $x[n], 1 \leq n \leq N$. Such a sequence is obtained, for example, by windowing a speech signal. The discrete-time Fourier transform (DTFT) of $x[n]$ is given by $X(e^{j\omega}) = \sum_{n=1}^{N} x[n] e^{-j\omega n}$, which is a 2π-periodic function in ω.

We construct a sequence $\tilde{x}[n] = \alpha\,\delta[n] + x[n]$, which differs from $x[n]$ only in the value at the origin ($n{=}0$). Without loss of generality, let α be a positive real-valued constant. The DTFT of $\tilde{x}[n]$ is given as $\tilde{X}(e^{j\omega}) = \sum_{n=0}^{N} \tilde{x}[n] e^{-j\omega n} = \alpha + X(e^{j\omega})$, which differs from $X(e^{j\omega})$ only by a constant offset across all frequencies. The claim is that, if $\alpha > |X(e^{j\omega})|, \forall \omega \in [-\pi, \pi]$, then $\tilde{x}[n]$ and therefore $x[n]$ can be exactly recovered from $\left| \tilde{X}(e^{j\omega}) \right|$.

In order to prove the above claim, consider the logarithm of the magnitude-squared spectrum:

$$
\begin{aligned}
\log \left| \tilde{X}(e^{j\omega}) \right|^2 &= \log \tilde{X}(e^{j\omega}) + \log \tilde{X}^*(e^{j\omega}) \\
&= \log \left(\alpha + X(e^{j\omega}) \right) + \log \left(\alpha + X^*(e^{j\omega}) \right).
\end{aligned}
$$

The first term on the right-hand side of the above equation is

$$
\log \left(\alpha + X(e^{j\omega}) \right) = \log \alpha + \log \left(1 + \frac{1}{\alpha} X(e^{j\omega}) \right). \quad (1)
$$

Since $\alpha > |X(e^{j\omega})|, \forall \omega \in [-\pi, \pi]$, we can invoke the Taylor-series expansion:

$$
\log \left(1 + \frac{1}{\alpha} X(e^{j\omega}) \right) = \sum_{m=1}^{\infty} \frac{(-1)^{m-1}}{m} \frac{X^m(e^{j\omega})}{\alpha^m}. \quad (2)
$$

The expansion consists of higher-order powers of $X(e^{j\omega})$. As m increases, the contribution of the higher-order powers decreases because $\left| \frac{1}{\alpha} X(e^{j\omega}) \right| < 1$. Recall that $x[n]$ and $X(e^{j\omega})$ form a DTFT pair. Therefore, the time-domain sequence corresponding to $X^2(e^{j\omega})$ is $(x * x)[n]$, which is the convolution of $x[n]$ with itself: $(x * x)[n] \xleftrightarrow{\text{DTFT}} X^2(e^{j\omega})$. Since $x[n]$ is causal, $(x * x)[n]$ is also causal. Extending this argument to higher values of m, we note that $X^m(e^{j\omega})$ is the DTFT of the m-fold convolution of $x[n]$: $\underbrace{(x * x * \cdots * x)}_{m \text{ times}}[n] \xleftrightarrow{\text{DTFT}} X^m(e^{j\omega})$. The m-fold convolution of $x[n]$ is also causal. Therefore, every term in the expansion on the right-hand side of (2) is the DTFT of a causal sequence. The expansion as a whole, which equals $\log \left(1 + \frac{1}{\alpha} X(e^{j\omega}) \right)$, is the DTFT of a causal sequence.

Now, consider the second term on the right-hand side of (1): $\log \left(\alpha + X^*(e^{j\omega}) \right)$. Since $X^*(e^{j\omega})$ is the DTFT of $x[-n]$, which is anti-causal, all higher-powers of $X^{*n}(e^{j\omega})$ also correspond to the DTFT of anti-causal sequences. As a consequence, it turns out that $\log \left(\alpha + X^*(e^{j\omega}) \right)$ is the DTFT of an anti-causal sequence.

Thus, the inverse DTFT of $\log \left| \tilde{X}(e^{j\omega}) \right|^2$, which is the cepstrum, comprises two components: a causal sequence, and an anti-causal sequence. If we retain only the causal part of the cepstrum and compute the DTFT, then we get $\log \tilde{X}(e^{j\omega})$. Thus, $\log \tilde{X}(e^{j\omega})$ can be recovered from $\log \left| \tilde{X}(e^{j\omega}) \right|^2$ by a combination of three operations: (i) inverse DTFT; (ii) selection of causal part (which is easily done by multiplying with a unit-step sequence); and (iii) DTFT of the causal sequence. Once $\log \tilde{X}(e^{j\omega})$ is obtained, $\tilde{X}(e^{j\omega})$ can be obtained by a complex exponential operation. The sequence $\tilde{x}[n]$ can be obtained by computing the inverse DTFT of $\tilde{X}(e^{j\omega})$. Since α is known, $x[n]$
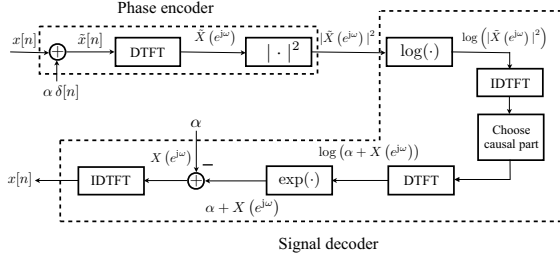
Figure 1: A block diagram of the proposed phase encoding scheme and reconstruction scheme.
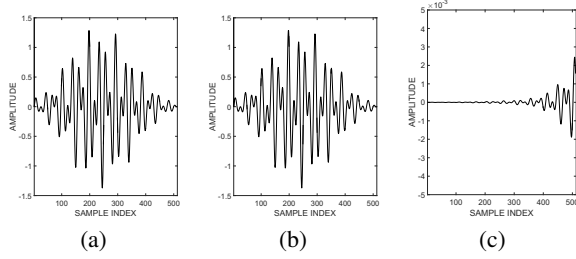

(a) (b) (c)

Figure 2: Signal reconstruction from Fourier spectrum magnitude: (a) Ground truth, (b) Reconstructed signal, and (c) Reconstruction error. Note that the y-axis in (c) is three orders of magnitude smaller than that in (a) and (b).

is readily computed from $\tilde{x}[n]$. Thus, $x[n]$ can be recovered from $\left|\tilde{X}(e^{j\omega})\right|$. The sequence of operations proposed to perform phase encoding and to perform signal decoding is shown in Figure 1.

We next address the issue of choice of α. Since α has to be chosen such that $|X(e^{j\omega})| < \alpha$, and $|X(e^{j\omega})| \leq \sum_{n=1}^{N} |x[n]|$, a reasonable choice for α is $k \sum_{n=1}^{N} |x[n]|$, where $k \gg 1$. Larger the value of α, faster is the decay of the terms in the expansion given in (2).

We have shown that by adding an impulse of appropriate strength at the origin of a causal finite-length sequence, it can be recovered only from the magnitude spectrum. The inverse DTFT of $\log\left|\tilde{X}(e^{j\omega})\right|$ is the real cepstrum associated with $\tilde{x}[n]$, and the inverse DTFT of $\log \tilde{X}(e^{j\omega})$ is the complex cepstrum of $\tilde{x}[n]$. Essentially, the signal is recovered from the magnitude spectrum by making use of the properties of the cepstrum. A similar cepstrum property was used by Drugman et al. [29] to estimate the glottal flow by decomposing the complex cepstrum of speech signal into causal and anti-causal components. The glottal flow is estimated from the anti-causal component of the complex cepstrum. In the proposed approach, we showed that the the real cepstrum of a CDD signal $\tilde{x}[n]$ can be decomposed into causal and anti-causal parts, and that, it is possible to reconstruct the original signal $x[n]$ using the causal part of the real cepstrum of the CDD signal.
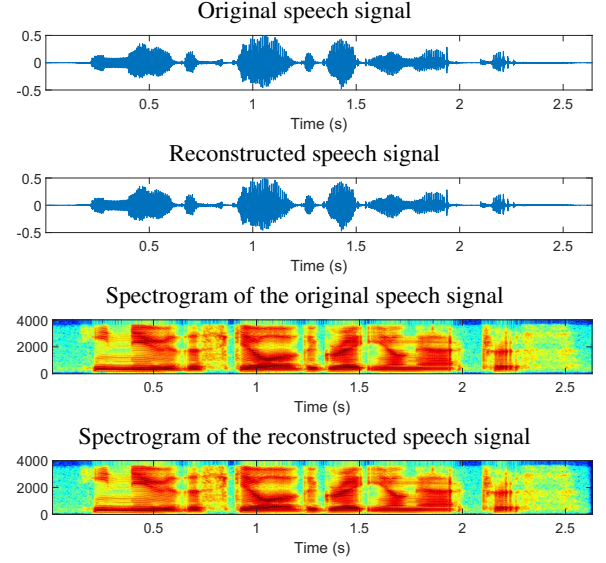

Figure 3: Original and reconstructed speech signals and their corresponding spectrograms. The spectrograms have been computed using a Hann window of duration 20 ms. The reconstruction SNR is 37 dB.

2.1. Practical issues

In practice, we cannot compute the DTFT. We can only compute the discrete Fourier transform (DFT), which is a sampled version of the DTFT. The logarithm and exponential operations described in the previous section in the context of computing the cepstrum are nonlinear and hence aliasing problems may occur. In order to suppress the aliasing error, the sequence must be zero-padded before computing the DFT. Alternatively, if a critically-sampled DFT spectrum magnitude is available to start with, one may interpolate it before performing reconstruction.

3. Phase encoding of signals

3.1. Synthesized data

We present some simulation results to validate the theoretical findings. We generated a signal $g[n] = \delta[n] + x[n]$, where $x[n] = e^{-0.8n/N} \sin(0.2n) + e^{-0.75n/N} \sin(0.33n), 0 \leq n \leq 511$. The sequence $\tilde{x}[n]$ is constructed as described in Section 2, and the reconstruction technique is applied to $|\tilde{X}(e^{j\omega})|^2$. The signal $x[n]$, the reconstructed signal $\hat{x}[n]$, and the reconstruction error $x[n] - \hat{x}[n]$ are shown in Figure 2. Although in theory, exact reconstruction is guaranteed, in practice, the reconstruction error is nonzero due to the discrete approximation to the Fourier transform. The error is small and acceptable for most practical applications.

3.2. Phase-encoded speech spectrograms

We next present results on real speech data from the NOIZEUS database [30]. We took a sentence and computed the Fourier magnitude spectra based on a 20 ms sliding window approach. Each frame is separately reconstructed using the technique described in Section 2. The ground truth signal, the reconstructed signal, and their spectrograms are shown in Figure 3. The reconstruction signal-to-noise ratio (SNR) is computed as the ra-

Original speech signal

Reconstructed speech signal

Spectrogram of the original speech signal

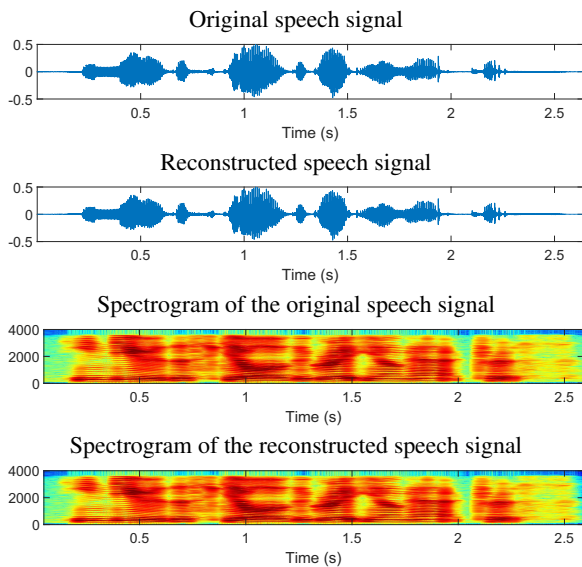Spectrogram of the reconstructed speech signal

Figure 4: Original and reconstructed speech signals and their corresponding spectrograms. The spectrograms have been computed using a Hann window of duration 100 ms. The reconstruction SNR is 42.08 dB.

Original speech signal

Reconstructed speech signal

Spectrogram of the original speech signal
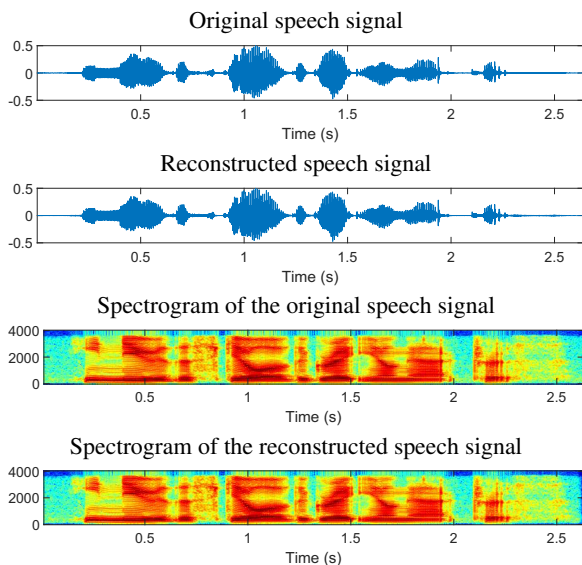
Spectrogram of the reconstructed speech signal

Figure 5: Original and reconstructed speech signals and their corresponding spectrograms. The spectrograms have been computed using a triangular window of duration 20 ms. The reconstruction SNR is 37.58 dB.

tio of reconstructed signal energy to the energy in the reconstruction error signal (difference between the original signal and the reconstructed signal), expressed as a logarithm. The spectrograms are shown to the same dynamic range of 120 dB on both subplots. From the figure, it is clear that the reconstruction is accurate. We next repeated the experiment with a window duration of 100 ms. The results are shown in Figure 4. Note that the reconstruction is accurate. This result implies that the usual choice of 20 ms window is not critical for the particular reconstruction technique presented in this paper. In other words, the reconstruction technique remains valid and in fact

Original speech signal

Reconstructed speech signal

Spectrogram of the original speech signal

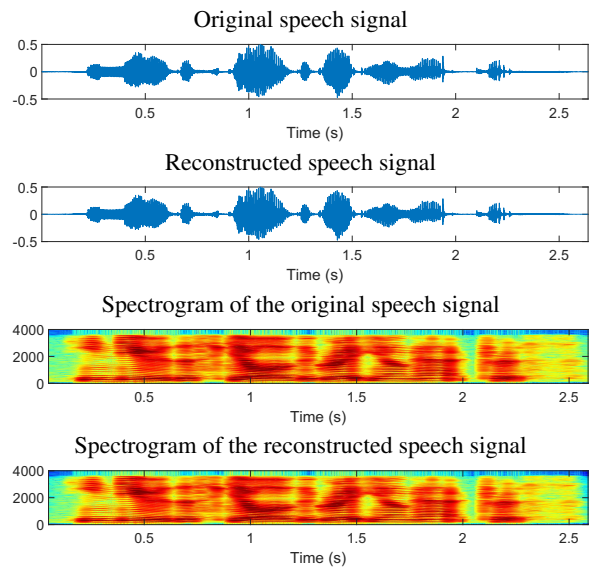Spectrogram of the reconstructed speech signal

Figure 6: Original and reconstructed speech signals and their corresponding spectrograms. The spectrograms have been computed using a triangular window of duration 100 ms. The reconstruction SNR is 42.87 dB.

better even when the analysis duration is considerably longer than the usual 20 ms. The experiments were also repeated with a different window function (triangular instead of Hann window). The results are shown in Figures 5 and 6, for window durations of 20 ms and 100 ms, respectively. In both cases, we observe that the quality of reconstruction is good. This experiment demonstrates that the technique is not critically dependent on the shape of the window function.

4. Conclusions

We addressed the problem of speech reconstruction from short-time Fourier magnitude spectrum. We showed that by adding an impulse of the right amplitude to the sequence before computing the short-time spectrum, one can ensure that exact signal reconstruction is possible from the magnitude spectrum. Essentially, the modification ensures that the signal becomes minimum-phase. Hitherto, one would check for the minimum-phase property by explicitly determining the locations of the poles and zeros with respect to the unit circle, but this is not necessary. Thus, the phase of the signal can be encoded in the magnitude spectrum and it can be retrieved exactly by following the reconstruction method presented in this paper. We showed simulation results as well as results on real speech signals to validate the claims. Quite surprisingly, it turned out that within the framework of the proposed reconstruction technique, the window shape is not critical and longer windows turned out to be better in terms of the reconstruction SNR. The signal was reconstructed with high accuracy for both Hann and triangular windows when the short-time spectral magnitudes were computed with 20 ms duration and with 100 ms duration. Thus, phase encoding in speech spectrograms is possible irrespective of whether the analysis is short-term or long-term. The proposed phase encoding approach might find potential applications in developing speech representations for speech synthesis applications and recognition.

5. References

[1] R. W. Gerchberg, "A practical algorithm for the determination of phase from image and diffraction plane pictures," *Optik*, vol. 35, pp. 237−246, 1972.

[2] J. R. Fienup, "Phase retrieval algorithms: A comparison," *Applied Optics*, vol. 21, no. 15, pp. 2758−2769, 1982.

[3] J. R. Fienup, "Reconstruction of an object from the modulus of its Fourier transform," *Optics Letters*, vol. 3, no. 1, pp. 27−29, 1978.

[4] J. R. Fienup, "Phase retrieval algorithms: A personal tour," *Applied Optics*, vol. 52, no. 1, pp. 45−56, 2013.

[5] H. H. Bauschke, P. L. Combettes, and D. R. Luke, "Phase retrieval, error reduction algorithm, and Fienup variants: A view from convex optimization," *Journal of the Optical Society of America (A)*, vol. 19, no. 7, pp. 1334−1345, 2002.

[6] H. H. Bauschke, P. L. Combettes, and D. R. Luke, "Hybrid projection-reflection method for phase retrieval," *Journal of the Optical Society of America (A)*, vol. 20, no. 6, pp. 1025−1034, 2003.

[7] A. V. Oppenheim, J. S. Lim, and S. R. Curtis, "Signal synthesis and reconstruction from partial Fourier-domain information," *J. Opt. Soc. Am.,* vol. 73, issue 11, pp. 1413−1420, 1983.

[8] M. H. Hayes, J. S. Lim, and A. V. Oppenheim, "Signal reconstruction from phase or magnitude," *IEEE Trans. Acoust. Speech Sig. Proc.,* vol. 28, no. 6, pp. 672−680, 1980.

[9] P. Van-Hove, M. H. Hayes, J. S. Lim, and A. V. Oppenheim, "Signal reconstruction from signed Fourier transform magnitude," *IEEE Trans. Acoust. Speech Sig. Proc.,* vol. ASSP-31, no. 5, pp. 1286−1293, Oct. 1983.

[10] S. H. Nawab, T. F. Quatieri, and J. S. Lim, "Signal reconstruction from short-time Fourier transform magnitude," *IEEE Trans. Acoust. Speech Sig. Proc.,* vol. ASSP-31, no. 4, pp. 986−998, Aug. 1983.

[11] A. V. Oppenheim and R. W. Schafer, *Discrete-Time Signal Processing*. Prentice-Hall, 1999.

[12] M. Hayes, J. Lim, and A. V. Oppenheim, "Signal reconstruction from phase or magnitude," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 28, no. 6, pp. 672−680, 1980.

[13] T. Quatieri and A. V. Oppenheim, "Iterative techniques for minimum-phase signal reconstruction from phase or magnitude," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 29, no. 6, pp. 1187−1193, 1981.

[14] B. Yegnanarayana and A. Dhayalan, "Noniterative techniques for minimum-phase signal reconstruction from phase or magnitude," in *Proceedings of IEEE International Conference on Acoustics, Speech, and Signal Processing*, vol. 8, 1983, pp. 639−642.

[15] B. Yegnanarayana, D. K. Saikia, and T. R. 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, 1984.

[16] L. D. Alsteris and K. K. Paliwal, "Short-time phase spectrum in speech processing: A review and some experimental results," *Digital Signal Processing*, vol. 17, issue 3, pp. 578−616, May 2007.

[17] M. L. Moravec, J. K. Romberg, and R. Baraniuk, "Compressive phase retrieval," in *Proceedings of SPIE International Symposium on Optical Science and Technology*, vol. 6701, 2007, pp. 670 120−670 120–11.

[18] H. Ohlsson and Y. C. Eldar, "On conditions for uniqueness in sparse phase retrieval," in *Proceedings of IEEE International Conference on Acoustics, Speech, and Signal Processing*, 2014, pp. 1841−1845.

[19] P. Netrapalli, P. Jain, and S. Sanghavi, "Phase retrieval using alternating minimization," in *Proceedings of Advances in Neural Information Processing Systems*, 2013, pp. 2796−2804.

[20] Y. Shechtman, A. Beck, and Y. C. Eldar, "GESPAR: Efficient phase retrieval of sparse signals," arXiv:1301.1018, 2013.

[21] 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.

[22] Y. M. Lu and M. Vetterli, "Sparse spectral factorization: Unicity and reconstruction algorithms," in *Proceedings of IEEE International Conference on Acoustics, Speech, and Signal Processing*, 2011, pp. 5976−5979.

[23] E. J. Candès, Y. C. Eldar, T. Strohmer, and V. Voroninski, "Phase retrieval via matrix completion," *SIAM Journal on Imaging Sciences*, vol. 6, no. 1, pp. 199−225, 2013.

[24] S. Mukherjee and C. S. Seelamantula, "An iterative algorithm for phase retrieval with sparsity constaints: Application to frequency-domain optical-coherence tomography," *in Proceedings of IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)* 2012, March 25−30, Kyoto, Japan.

[25] S. Mukherjee and C. S. Seelamantula, "Fienup algorithm with sparsity constraints: Application to frequency-domain optical-coherence tomography," *IEEE Transactions on Signal Processing*, vol. 62, no. 18, pp. 4659−4672, 2014.

[26] S. Mukherjee and C. S. Seelamantula, "A non-iterative phase retrieval algorithm for minimum-phase signals using the annihilating filter," *Sampling Theory in Signal and Image Processing, (Special issue on SampTA 2011)*, vol. 11, no. 2-3, pp. 165−193, 2012.

[27] B. A. Shenoy and C. S. Seelamantula, "Exact phase retrieval for a class of 2-D parametric signals," *IEEE Transactions on Signal Processing*, vol. 63, no. 1, pp. 90−103, 2015.

[28] B. A. Shenoy, S. Mulleti, and C. S. Seelamantula, "Exact phase retrieval in principal shift-invariant spaces," *IEEE Transactions on Signal Processing*, vol. 64, no. 2, pp. 406−416, 2016.

[29] T. Drugman, B. Bozkurt, and T. Dutoit, "Causal-anticausal decomposition of speech using complex cepstrum for glottal source estimation," *Speech Communication*, vol. 53, no. 6, pp. 855−866, 2011.

[30] Y. Hu, and P. Loizou, "Subjective evaluation and comparison of speech enhancement algorithms," *Speech Communication*, vol. 49, no. 7, pp. 588−601, 2007.