Interspeech 2018
2-6 September 2018, Hyderabad

An Optimization Framework for Recovery of Speech From Phase-Encoded Spectrograms

Abhilash Sainathan[1]*, Sunil Rudresh*[2]*, and Chandra Sekhar Seelamantula*[2]

[1]Department of Electrical Communication Engineering
[2]Department of Electrical Engineering
Indian Institute of Science
Bangalore - 560 012, India
{abhilash, sunilr, chandrasekhar}@iisc.ac.in

Abstract

In general, reconstruction of a speech signal from the spectrogram is non-unique because of the unavailability of the phase spectrum. Considering zero phase would result in a minimum-phase reconstruction. This limitation is overcome by computing the recently introduced *phase-encoded spectrogram*. In this approach, one modifies each frame of a speech signal to possess the causal, delta-dominant (CDD) property prior to computing the spectrogram. In an earlier publication, we showed that finite-length CDD sequences can be retrieved exactly from their magnitude spectra using a cepstrum technique. Although exactness is guaranteed in principle, practical implementations result in a limited, but high, reconstruction accuracy. In this paper, we focus on increasing the reconstruction accuracy. We formulate the reconstruction problem within an optimization framework and deploy a recently proposed iterative, alternating direction method of multipliers (ADMM) algorithm called autocorrelation retrieval–Kolmogorov factorization (CoRK). Experimental validations show that the CoRK algorithm results in a reconstruction accurate up to machine precision. We also show that both CoRK and cepstrum techniques are robust and invariant to the choice of the window duration, the amount of overlap between consecutive speech frames, the strength of the delta used to impart the CDD property, and the presence of noise.

Index Terms: Phase-encoded spectrogram, causal delta-dominant sequence, autocorrelation retrieval and Kolmogorov factorization (CoRK), cepstrum.

1. Introduction

The spectrogram is a widely used representation tool for a multitude of speech processing tasks such as synthesis, analysis, quality assessment, recognition, enhancement, etc. In general, a spectrogram is phase-blind. An attempt to reconstruct the speech signal from its spectrogram gives rise to an infinite number of solutions as each combination of the spectrogram with a chosen phase spectrum results in a different speech signal. Also, speech signals, in general, do not fall in the class of minimum-phase signals. Several experiments in the past have shown that over short durations (about 20 to 30 ms), the magnitude spectrogram is more important than the phase as far as speech perception is concerned [1–3]. However, several recent studies have shown the increasing importance of phase spectrum towards speech perception and various speech applications [4–8].

The problem of reconstructing a speech signal from its magnitude spectrum is essentially the problem of retrieval of the phase spectrum from the spectrogram. Hence, the problem directly falls within the realm of phase retrieval.

1.1. Related Work

The uncertainty in decoding the speech signal from its spectrogram is resolved by resorting to a differently constructed spectrogram called the phase-encoded spectrogram, proposed by Seelamantula [9]. Each windowed frame of a speech signal is converted into a causal delta-dominant (CDD) sequence, which has been shown to be minimum-phase. The spectrogram thus obtained could be used to reconstruct the speech signal, which establishes that phase-encoding is possible. Soni et al. [10] proposed an alternative method to obtain a phase-encoded spectrogram based on the symmetry property of the Fourier transform of the even and odd signal, which was shown to be identical in concept with [9] under some conditions. They also showed exact signal reconstruction for their proposed encoding scheme. Before formally stating the problem addressed in this paper, we review important related phase-retrieval techniques.

A classical 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]. Both iterative [12] and non-iterative [13–15] techniques have been developed, which result in exact phase retrieval for minimum-phase signals. Early methods for solving this problem include the error-reduction algorithms due to Gerchberg and Saxton [16] and Fienup [17], and numerous variants thereof. These algorithms rely on an alternating minimization strategy—essentially, one alternates between the measurement domain and the signal domain applying appropriate constraints in respective domains.

The phase retrieval problem has been extensively researched in the signal processing community [18]. In particular, Shenoy et al. [19] identified a class of signals for which exact phase retrieval is possible. The new class of signals called *causal delta-dominant (CDD) signals* is a generalization, but reduces to the well-known class of minimum-phase signals when the signal length is finite. The developments were carried out in the continuous domain in principal shift-invariant spaces spanned by a generator kernel. Subsequently, Huang et al. [20] showed that the phase retrieval problem can be solved using Kolmogorov's spectral factorization.

Recently, phase retrieval has received a lot of attention from the compressed sensing (CS) community. An early CS phase retrieval algorithm is the compressive phase retrieval (CPR) algorithm by Moravec et al. [21]. Subsequently, numerous techniques have been developed based on the sparsity criterion. In general, the sparsity-driven phase retrieval algorithms reconstruct the signal up to a global phase factor [22–24]. A sparse counterpart of the classical Fienup algo-

 10.21437/Interspeech.2018-1987

rithm was reported by Mukherjee and Seelamantula [25]. Other approaches to phase retrieval employ convex relaxation, where phase retrieval is reformulated as a rank-minimization problem [26–28], semidefinite-programming-based techniques [29, 30], and non-convex formulations based on a suitable initialization and gradient-descent updates [31, 32].

1.2. This Paper

We formulate the problem of recovering a speech signal from its phase-encoded spectrogram in an optimization framework, where one is required to solve multiple 1-D phase retrieval problems. We employ an iterative alternating direction method of multipliers (ADMM)-based technique, which is different from the techniques proposed earlier [9, 10] to solve the problem. We show that the task of recovering the speech signal falls within the framework considered in [20], which is based on autocorrelation retrieval and Kolmogorov factorization (CoRK), which can be effectively deployed to solve the problem under consideration. In Section 2, we give an overview of the construction of the phase-encoded spectrograms and formulate the speech recovery problem within the framework of phase retrieval. We provide an overview of the CoRK algorithm [20] for the retrieval of CDD sequences. Apart from CDD-based phase encoding, other variants of the phase encoding schemes are also considered. In Section 3, we show results on real speech data. We demonstrate that the CoRK algorithm converges relatively fast and is effective in recovering the speech signal with better signal-to-reconstruction-noise ratio (SRNR) compared with the cepstrum technique originally proposed in [9]. The perceptual evaluation of speech quality (PESQ) scores for both methods were found to be comparable. We also examine the effect of window duration, overlap ratio, and the strength of the Kronecker impulse (which is added to impart the CDD property to a sequence) on the performance of the recovery algorithm.

2. Phase-Encoding and Reconstruction

2.1. Phase-Encoding in Speech Spectrograms

Consider a K-length speech signal vector $\mathbf{s} = \{s_n\}_{n=1}^{K}$ and an N-length window function $\mathbf{w} = [w_1\, w_2 \cdots w_N]$, where $N < K$. The frames $\{\mathbf{y}_i\}_{i=1}^{P}$ obtained by windowing the speech signal are given by

$$
\begin{aligned}
\mathbf{y}_1 &= \mathbf{w} \circ [s_1\, s_2 \ldots s_N],\\
\mathbf{y}_2 &= \mathbf{w} \circ [s_{(N+1-l)}\, s_{(N+2-l)} \cdots s_{(2N-l)}],\\
&\ \ \vdots \\
\mathbf{y}_P &= \mathbf{w} \circ [s_{((P-1)N+1-l)}\, s_{((P-1)N+2-l)} \cdots s_{(PN-l)}],
\end{aligned} \tag{1}
$$

where l denotes the number of overlapping samples between successive frames and $\circ$ denotes the Hadamard product. To obtain a phase-encoded spectrogram, it is first required to convert each frame $\mathbf{y}_i$ into a CDD sequence [9]. In order to make the exposition self contained, we next review the definition of a CDD sequence and its minimum-phase property.

Definition 1. A sequence $\{y_n\}_{n\in\mathbb{Z}}$ is said to be causal and delta dominant (CDD) if $y_n = 0, n < 0$ and $y_0 > \sum_{n\geq 1} |y_n|$.

Lemma 1 (Shenoy et al. [19])**.** *A finite-length CDD sequence is also a minimum-phase sequence .*

Each frame $\mathbf{y}_i$ is converted into a CDD sequence by appending an α_i at the beginning of a frame as

$$\bar{\mathbf{y}}_i = [\alpha_i\ \ \mathbf{y}_i]^{\mathrm{T}}, \tag{2}$$

where $\alpha_i = \beta \sum_{n=1}^{N} \left| w_n s_{((i-1)N+n-l)} \right|$ and $\beta > 1$. The M-length discrete Fourier transform (DFT) with $M \geq 2(N+1)$ of each frame $\bar{\mathbf{y}}_i$ is given by $\mathbf{F}_M \bar{\mathbf{y}}_i$, where $\mathbf{F}_M$ is the first $(N+1)$ columns of the M-point DFT matrix. The i^{th} column of the phase-encoded spectrogram is given by $|\mathbf{F}_M \bar{\mathbf{y}}_i|^2$, where $|\cdot|^2$ is computed elementwise.

2.2. Formulation of the Cost Function

To reconstruct the speech signal $\mathbf{s}$ given its phase-encoded spectrogram, it is first necessary to recover each $(N+1)$-length CDD frame $\bar{\mathbf{y}}_i$ from $|\mathbf{F}_M \bar{\mathbf{y}}_i|^2$, which is equivalent to recovering a sequence from its Fourier magnitude spectrum. This is essentially the problem of phase retrieval of a minimum-phase signal, because a finite-length CDD sequence is also minimum phase [19]. Having retrieved the CDD frame, we recover $\mathbf{y}_i$ by removing the α_i term. The original speech signal is then recovered by appropriately accounting for the windowing operation and performing overlap-add between consecutive frames.

Formally, we seek an estimate of $\bar{\mathbf{y}}_i$ from the measurements

$$\mathbf{b}_i = |\mathbf{F}_M \bar{\mathbf{y}}_i|^2 + \mathbf{c}_i, \tag{3}$$

where $\mathbf{c}_i$ is a noise vector. The following least-squares cost function is considered:

$$\underset{\bar{\mathbf{y}}_i \in \mathbb{R}^{N+1}}{\text{minimize}} \quad \left\| \mathbf{b}_i - |\mathbf{F}_M \bar{\mathbf{y}}_i|^2 \right\|^2. \tag{4}$$

In order to obtain an estimate of $\bar{\mathbf{y}}_i$ from (4), we deploy the ADMM technique proposed by Huang et al. [20] based on CoRK, which fits in perfectly for the signal model at hand. The CoRK algorithm is a two-step process, wherein one obtains the correlation sequence corresponding to $\bar{\mathbf{y}}_i$ in the first step, and in the second, one retrieves the minimum-phase sequence $\bar{\mathbf{y}}_i$ using Kolmogorov factorization. Next, we present an outline of the CoRK algorithm.

2.3. CoRK Algorithm for Reconstruction

2.3.1. Estimation of the Autocorrelation Sequence

Let the autocorrelation sequence of the i^{th} column $\bar{\mathbf{y}}_i$ be

$$\tilde{\mathbf{r}} = [r_{-N} \ldots r_{-1}\ \ r_0\ \ r_1 \ldots r_N]^{\mathrm{T}}, \tag{5}$$

where

$$r_k = \sum_{n=\max(k,0)}^{\min(N+k,N)} \bar{\mathbf{y}}_i[n] \bar{\mathbf{y}}_i[n-k]^*,$$
$$k = -N, \ldots, -1, 0, 1, \ldots, N.$$

The autocorrelation sequence $\tilde{\mathbf{r}}$ is different for different $\bar{\mathbf{y}}_i$s. For the sake of notational brevity, we omit the subscript i in $\tilde{\mathbf{r}}_i$ and denote it as $\tilde{\mathbf{r}}$. Using the conjugate-symmetry property of the autocorrelation $r_k = r_{-k}^*$, the redundancy in (5) could be suppressed.

Considering the one-sided autocorrelation $\mathbf{r} = [r_0\ \ r_1 \ldots r_N]^{\mathrm{T}}$, Huang et al. [20] proposed a convex formulation as follows:

$$\begin{aligned} &\underset{\mathbf{r}\in\mathbb{R}^{N+1}}{\text{minimize}} \quad \left\|\mathbf{b}_i - \mathcal{R}\{\mathbf{F}_M\tilde{\mathbf{I}}\mathbf{r}\}\right\|^2 \\ &\text{subject to} \quad \mathcal{R}\{\mathbf{F}_L\tilde{\mathbf{I}}\mathbf{r}\} \geq \mathbf{0}, \end{aligned} \tag{6}$$

where $\mathcal{R}\{\cdot\}$ gives the real part of its argument, $\tilde{\mathbf{I}} = \text{diag}[1\ 2\ 2\ldots 2]$, and $\mathbf{F}_L$ is the matrix formed by the first $(N+1)$ columns of the L-point DFT matrix.

The correlation sequence $\mathbf{r}$ is retrieved by solving the optimization problem in (6) using the ADMM [20], whose update steps are given as

$$\begin{aligned} &\mathbf{r} \leftarrow \frac{1}{2}\left(\frac{1}{M}\mathbf{F}_M^{\mathrm{H}}\mathbf{b}_i + \frac{1}{L}\mathbf{F}_L^{\mathrm{H}}(\mathbf{z}-\mathbf{u})\right), \\ &\mathbf{z} \leftarrow \max\left(\mathbf{0}, \mathcal{R}\{\mathbf{F}_L\tilde{\mathbf{I}}\mathbf{r}\} + \mathbf{u}\right), \\ &\mathbf{u} \leftarrow \mathbf{u} + \mathcal{R}\{\mathbf{F}_L\tilde{\mathbf{I}}\mathbf{r}\} - \mathbf{z}, \end{aligned} \tag{7}$$

where $\mathbf{z} \in \mathbb{R}^L$ and $\mathbf{u} \in \mathbb{R}^L$ are the auxiliary variable and the Lagrange multiplier in the ADMM, respectively. To begin with, the initializations $\mathbf{z} = \mathbf{0}_{L\times 1}$ and $\mathbf{u} = \mathbf{0}_{L\times 1}$ are used in the ADMM iterations.

2.3.2. Estimation of the Minimum-Phase Signal

Since $\bar{\mathbf{y}}_i$s are finite-length CDD sequences, logarithms of their $\mathcal{Z}$-transforms are unilateral and hence, real and imaginary parts of logarithm of the $\mathcal{Z}$-transforms form Hilbert transform pairs. Exploiting this property, Kolmogorov proposed a method for recovering the minimum-phase signal $\bar{\mathbf{y}}_i$ that generates the correlation $\mathbf{r}$. The outline of the algorithm is given below.

- Real part of logarithm of $\mathcal{Z}$-transform of $\bar{\mathbf{y}}_i$: $\boldsymbol{\gamma} = \frac{1}{2}\ln\mathcal{R}\{\mathbf{F}_L\tilde{\mathbf{I}}\mathbf{r}\}$.
- Compute the Hilbert transform (approximated by DFT): $\boldsymbol{\phi} = \mathbf{F}_L\boldsymbol{\gamma}$,
$$\varphi_n = \begin{cases} 0, & n = 0, L/2, \\ -\mathrm{j}\phi_n, & n = 1, 2, \ldots, L/2-1, \\ \mathrm{j}\phi_n, & n = L/2+1, \ldots, L-1. \end{cases}$$
- Imaginary part of logarithm of $\mathcal{Z}$-transform of $\bar{\mathbf{y}}_i$: $\boldsymbol{\eta} = \frac{1}{L}\mathbf{F}_L^{\mathrm{H}}\boldsymbol{\varphi}$.
- Minimum-phase sequence: $\hat{\bar{\mathbf{y}}}_i = \frac{1}{L}\mathbf{F}_L^{\mathrm{H}}\exp(\boldsymbol{\gamma} - \mathrm{j}\boldsymbol{\eta})$,

where the first $(N+1)$ elements of $\hat{\bar{\mathbf{y}}}_i$ constitute the estimate of $\bar{\mathbf{y}}_i$. A large value of L is required to obtain a higher accuracy and is chosen to be the smallest power of 2 greater than $32(N+1)$.

2.4. Other Encoding Schemes

In order to obtain a phase-encoded spectrogram, each frame $\mathbf{y}_i$ was converted into a CDD sequence by prefixing an appropriate Kronecker impulse at the beginning of the frame as shown in (2). Alternatively, one could also append an impulse at the end of a frame as

$$\bar{\mathbf{y}}_i = [\mathbf{y}_i \;\; \alpha_i]^{\mathrm{T}}, \tag{8}$$

which can be shown to be a maximum-phase signal. This model also allows for perfect signal recovery. This model does not require any modification in the CoRK algorithm except that the first $(N+1)$ elements of $\hat{\bar{\mathbf{y}}}_i$ are reversed to get an estimate of $\bar{\mathbf{y}}_i$. In yet another variation, the Kronecker impulse could also be included with zeros inserted in between as

$$\begin{aligned} &\bar{\mathbf{y}}_i = [\alpha_i \;\; 0\ldots 0 \;\; \mathbf{y}_i]^{\mathrm{T}} \\ \text{or} \quad &\bar{\mathbf{y}}_i = [\mathbf{y}_i \;\; 0\ldots 0 \;\; \alpha_i]^{\mathrm{T}}, \end{aligned} \tag{9}$$

with the number of zeros being arbitrary. The resulting augmented frame can still be uniquely identified from its corresponding column of the phase-encoded spectrogram. We have observed that all the results that will be presented in Section 3 hold equally well for these alternative encoding schemes as well.

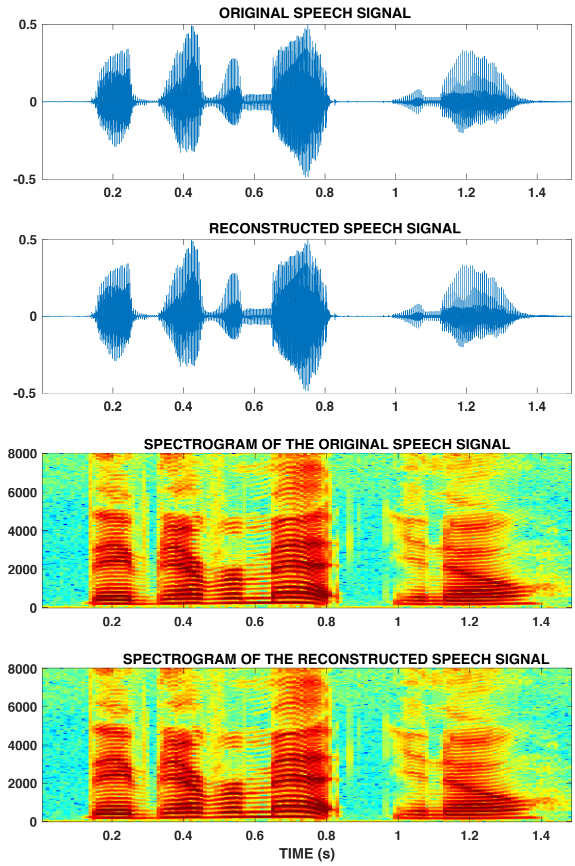

Figure 1: *Original and reconstructed speech signals and their corresponding spectrograms. The spectrograms are computed using a Hamming window of duration* 20 *ms with* 50% *overlap and* $\beta = 1.1$. *The reconstruction is accurate up to machine precision—the computed SRNR turned out to be* 280 *dB.*

3. Experiments and Results

We present results on real speech data from the TIMIT database [33]. For all the experiments, a speech signal with analysis window (Hamming) of 20 ms with 50% overlap between consecutive frames is considered to compute its spectrogram with phase encoding ($\beta = 1.1$) as explained in Section 2.1. For experiments involving noise, additive white Gaussian noise with zero mean is considered. Each frame is separately reconstructed using 50 iterations of the CoRK algorithm described in Section 2.3. The ground-truth signal, the reconstructed signal, and their spectrograms for a clean speech segment are shown in Figure 1. The SRNR is computed as the ratio of the reconstructed signal

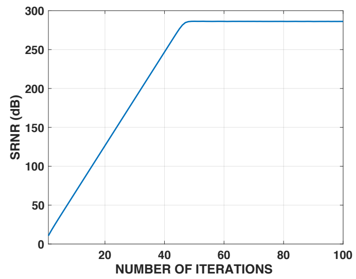

Figure 2: *SRNR of the CoRK algorithm as a function of the number of iterations.*

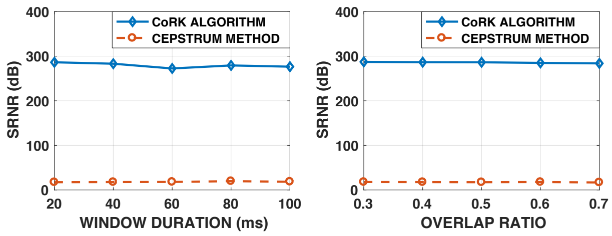

Figure 3: *A comparison of the reconstruction performance of the CoRK and cepstrum techniques using Hamming window as a function of the window duration and the fraction of overlap between consecutive frames.*

energy to the energy in the reconstruction error (which is the difference between the ground-truth and the reconstructed signal), expressed as a logarithm. The spectrograms are shown to the same dynamic range of 120 dB in both figures. We observe from the figure that the reconstruction is accurate up to machine precision. The SRNRs achieved for a clean speech signal with the iterations of the CoRK algorithm are shown in Figure 2. This shows that the CoRK algorithm can achieve SRNRs up to 280 dB, which are higher than that achieved by the cepstrum method (SRNR of 20 dB) [9].

1. Effect of window size and overlap ratio: We repeated the experiment for a clean speech signal by varying the window duration and the overlap between successive frames. Figure 3 shows that the SRNR remains almost the same irrespective of the window duration and the overlap, with the CoRK algorithm having a higher SRNR.

2. Effect of noise: Figure 4 shows the effect of input SNR on the SRNR for both CoRK and the cepstrum methods. For each input SNR, the SRNRs are averaged over 50 noise realizations. At lower input SNR, both the methods show similar performance, whereas at higher input SNR, the CoRK algorithm shows superior reconstruction with SRNR matching up to the input SNR.

3. PESQ evaluation: We performed the experiment of recovering the speech signal from its phase-encoded spectrogram for the 330 speech signals taken from the TIMIT database. The average PESQ scores of the CoRK and cepstrum methods are found to be comparable and are presented in Table 1.

4. Effect of β: Figure 5 shows that the SRNR for the cepstrum method gradually increases and remains constant after $\beta = 20$. For the CoRK algorithm, it is found that the SRNR remains as high as 250 dB even when β is varied over a wide range: 1.1 to 50. Even when β is only marginally greater than 1, the CoRK algorithm has a high recovery SRNR.

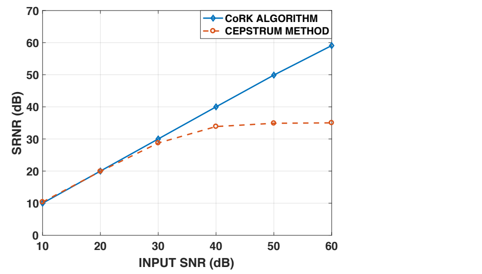

Figure 4: *Reconstruction performance of the CoRK and cepstrum techniques at different input SNR levels.*

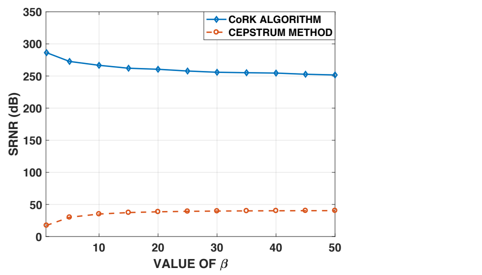

Figure 5: *Effect of the CDD conversion parameter β on the reconstruction performance.*

Table 1: *PESQ scores for the TIMIT database using CoRK and cepstrum reconstructions for different input SNR levels.*

Input SNR (dB)	Cepstrum method	CoRK algorithm
clean speech	4.4557	4.500
10	2.2839	2.2843
15	2.6508	2.6517
20	3.0063	3.0076
25	3.3499	3.3514
30	3.6896	3.6922
40	4.2763	4.2840
50	4.4431	4.4712
60	4.4549	4.4951

4. Conclusions

We considered the problem of reconstruction of speech from its phase-encoded magnitude spectrum, which turns out to be a phase retrieval problem. It is in turn equivalent to the reconstruction of a signal from its autocorrelation. We employed the property that finite-length CDD sequences can be retrieved exactly from their magnitude spectrum. We deployed a recently developed autocorrelation retrieval algorithm (CoRK), which fits in perfectly for the signal model at hand as it fulfills all the prerequisites for the CoRK algorithm to function accurately. We carried out validations on real speech signals and showed high-quality reconstruction with the SRNR as high as 280 dB. We also demonstrated that the performance of the CoRK algorithm is robust and invariant to parameters such as window duration, overlap between consecutive frames, the strength of the Kronecker impulse for the CDD-encoding, and other types of phase encoding considered in this paper. The proposed method operates on a frame-by-frame basis and can therefore handle nonstationary noise as well. Subjective assessment and comparison of the reconstruction techniques is a topic for further investigation.

5. References

[1] G. S. Ohm, "Über die Definition des Tones, nebst daran geknüpfter Theorie der Sirene und ähnlicher tonbildender Vorrichtungen," *Annalen der Physik*, vol. 135, no. 8, pp. 513–565, 1843.

[2] H. Helmholtz, *On the Sensations of Tone*. Courier Corporation, 2013.

[3] R. S. Turner, "The Ohm-Seebeck dispute, Hermann von Helmholtz, and the origins of physiological acoustics," *The British Journal for the History of Science*, vol. 10, no. 1, pp. 1–24, 1977.

[4] K. K. Paliwal and L. Alsteris, "Usefulness of phase spectrum in human speech perception," in *Proceedings of European Conference on Speech Communication and Technology (Eurospeech)*, Sep. 2003, pp. 2117–2120.

[5] H. Pobloth and W. B. Kleijn, "On phase perception in speech," in *Proceedings of IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, vol. 1, 1999, pp. 29–32.

[6] R. Schluter and H. Ney, "Using phase spectrum information for improved speech recognition performance," in *Proceedings of IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, vol. 1, 2001, pp. 133–136.

[7] 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, 2006.

[8] D.-S. Kim, "Perceptual phase quantization of speech," *IEEE Transactions on Speech and Audio Processing*, vol. 11, no. 4, pp. 355–364, 2003.

[9] C. S. Seelamantula, "Phase-encoded speech spectrograms." in *Proceedings of INTERSPEECH*, 2016, pp. 1775–1779.

[10] M. H. Soni, R. Tak, and H. A. Patil, "Novel shifted real spectrum for exact signal reconstruction," in *Proceedings of INTERSPEECH*, 2017, pp. 3112–3116.

[11] A. V. Oppenheim, *Discrete-Time Signal Processing*. Pearson Education India, 1999.

[12] T. Quatieri and A. 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.

[13] 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 (ICASSP)*, vol. 8, 1983, pp. 639–642.

[14] S. Mukherjee and C. S. Seelamantula, "A non-iterative phase retrieval algorithm for minimum-phase signals using the annihilating filter," *Sampling Theory in Signal & Image Processing*, vol. 11, no. 2–3, pp. 165–193, 2012.

[15] 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.

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

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

[18] Y. Shechtman, Y. C. Eldar, O. Cohen, H. N. Chapman, J. Miao, and M. Segev, "Phase retrieval with application to optical imaging: A contemporary overview," *IEEE Signal Processing Magazine*, vol. 32, no. 3, pp. 87–109, 2015.

[19] 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.

[20] K. Huang, Y. C. Eldar, and N. D. Sidiropoulos, "Phase retrieval from 1D Fourier measurements: Convexity, uniqueness, and algorithms," *IEEE Transactions on Signal Processing*, vol. 64, no. 23, pp. 6105–6117, 2016.

[21] M. L. Moravec, J. K. Romberg, and R. G. Baraniuk, "Compressive phase retrieval," in *Wavelets XII*, vol. 6701. International Society for Optics and Photonics, 2007, pp. 120–670.

[22] Y. Shechtman, A. Beck, and Y. C. Eldar, "GESPAR: Efficient phase retrieval of sparse signals," *IEEE Transactions on Signal Processing*, vol. 62, no. 4, pp. 928–938, 2014.

[23] H. Ohlsson, A. Yang, R. Dong, and S. Sastry, "CPRL – An extension of compressive sensing to the phase retrieval problem," in *Proceedings of Advances in Neural Information Processing Systems*, 2012, pp. 1367–1375.

[24] 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 (ICASSP)*, 2014, pp. 1841–1845.

[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] K. Jaganathan, S. Oymak, and B. Hassibi, "Phase retrieval for sparse signals using rank minimization," in *Proceedings of IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2012, pp. 3449–3452.

[27] E. J. Candès, T. Strohmer, and V. Voroninski, "PhaseLift: Exact and stable signal recovery from magnitude measurements via convex programming," *Communications on Pure and Applied Mathematics*, vol. 66, no. 8, pp. 1241–1274, 2013.

[28] I. Waldspurger, A. d'Aspremont, and S. Mallat, "Phase recovery, maxcut and complex semidefinite programming," *Mathematical Programming*, vol. 149, no. 1-2, pp. 47–81, 2015.

[29] Y. Shechtman, Y. C. Eldar, A. Szameit, and M. Segev, "Sparsity based sub-wavelength imaging with partially incoherent light via quadratic compressed sensing," *Optics Express*, vol. 19, no. 16, pp. 14 807–14 822, 2011.

[30] E. J. Candès, Y. C. Eldar, T. Strohmer, and V. Voroninski, "Phase retrieval via matrix completion," *SIAM Review*, vol. 57, no. 2, pp. 225–251, 2015.

[31] E. J. Candès, X. Li, and M. Soltanolkotabi, "Phase retrieval via Wirtinger flow: Theory and algorithms," *IEEE Transactions on Information Theory*, vol. 61, no. 4, pp. 1985–2007, 2015.

[32] G. Wang, G. B. Giannakis, and Y. C. Eldar, "Solving systems of random quadratic equations via truncated amplitude flow," *IEEE Transactions on Information Theory*, vol. 64, no. 2, pp. 773–794, 2018.

[33] V. Zue, S. Seneff, and J. Glass, "Speech database development at MIT: TIMIT and beyond," *Speech Communication*, vol. 9, no. 4, pp. 351–356, 1990.