SNR-Based Inter-Component Phase Estimation Using Bi-Phase Prior Statistics for Single-Channel Speech Enhancement

Siarhei Y. Barysenka, *Member, IEEE,* and Vasili I. Vorobiov

***Abstract*—The fundamental problem of phase-aware single-channel speech enhancement is the estimation of the harmonic phase of signal components from noisy observations. One approach to obtain an estimate of the harmonic phase is by smoothing the noisy harmonic phase in the time domain. Accurate identification of the smoothing regions is crucial to achieve the best enhancement results. Previous works have introduced a binary hypothesis framework for detecting these regions, which is formulated *individually* for each harmonic component based on von Mises statistics and local harmonic SNR levels. Later, smoothing frameworks have been proposed in the *inter-component phase domain*, where the inter-component phases follow the temporal constancy property for clean speech signals. However, these frameworks either do not formulate an SNR-dependent smoothing threshold or impose it empirically. The aim of this work is to formulate an SNR-dependent binary hypothesis framework for smoothing in the *inter-component phase domain*, using prior inter-component phase statistics. The framework then reconstructs the instantaneous phases of the enhanced speech signal components. This proposed framework results in superior speech intelligibility enhancement, reduces phase deviation, and does not introduce auditory buzziness artifacts compared to the previously developed framework that does not formulate an SNR-dependent smoothing threshold.**

***Index Terms*—Speech enhancement, harmonic phase, inter-component phase, bi-phase, von Mises distribution**

I. Introduction

FOR many decades of speech enhancement research, only the magnitude part of the speech spectrum was typically considered for processing, while the phase spectrum was used directly without processing to reconstruct the enhanced speech. Spectral phase processing was neglected for several reasons: i) Wang and Lim [1] reported the perceptual unimportance of phase in speech in early works, ii) Ephraim and Malah [2] showed that the noisy spectral phase is the optimal phase estimate in the minimum mean square error (MMSE) sense, iii) the difficulty in understanding patterns in spectral phase due to the phase wrapping phenomenon.

Manuscript received 21 September 2022; revised 6 February 2023 and 26 April 2023; accepted 5 June 2023. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Hakan Erdogan. *(Corresponding author: Siarhei Y. Barysenka)*

Siarhei Y. Barysenka is with Atlassian B.V., Amsterdam, The Netherlands (e-mail: siarhei.barysenka@gmail.com).

Vasili I. Vorobiov was with Belarusian State University of Informatics and Radioelectronics, Minsk, Belarus (e-mail: viv314@gmail.com).

Digital Object Identifier 10.1109/TASLP.2023.3284514

On the contrary, Oppenheim and Lim [3] conducted simple experiments on magnitude-only and phase-only speech reconstruction, which showed that phase-only (with unity magnitude) speech reconstruction is enough for speech to remain intelligible, while magnitude-only reconstruction produces completely unintelligible signal. Decades later, Paliwal *et al.* [4] showed that employing clean spectral phase in noise reduction algorithms could considerably improve the quality of the reconstructed signal if the analysis window length and segment overlap are chosen appropriately. Since the 2010s, the topic of spectral phase processing has experienced increased research interest in the speech processing domain [5] due to advances in various speech processing applications, including single-channel speech enhancement.

Over the years, the research community has proposed a variety of phase estimators, including iterative-based, geometry-based, model-based, statistics-based, and deep neural network (DNN)-based methods that employ various considerations about the signal structure to estimate spectral phase. Below, we present the key ideas of some methods in each group.

Iterative-based reconstruction methods include the Griffin-Lim method [6] and its extensions. These methods estimate the spectral phase by iteratively reconstructing the signal short-time Fourier transform (STFT) that corresponds to the closest (in the MMSE sense) STFT of the original signal. To do so, they require an estimate of the spectral magnitude.

Geometry-based reconstruction methods, introduced by Mowlaee *et al.* [7, 8], formulate the expression for the spectral phase estimate based on a vector representation of the noisy speech observation modeled as a sum of clean speech and noise vectors. However, this expression is sign-ambiguous and requires additional constraints to resolve the ambiguity. The authors used group delay deviation, instantaneous frequency deviation, and relative phase shift (RPS, by Saratxaga *et al.* [9]) constraints to resolve the sign ambiguity and provide the final phase estimate. These methods also require an estimate of the spectral magnitude.

Model-based reconstruction methods rely on considerations about the signal's harmonic structure across time and/or frequency to reconstruct the spectral phase. One example of this type of method is the short-time Fourier transform phase improvement (STFTPI, by Krawczyk and Gerkmann [10]). This method, which is suitable for voiced speech segments, predicts the phase along the time axis while considering segment-to-segment correlation in frequency bands dominated by harmonics. It then enhances the phase along the frequency

axis by compensating for the analysis window phase response for each segment. Model-based methods require an estimate of the fundamental frequency.

Statistics-based methods utilize knowledge about the prior distribution of harmonic phase in noise, expressed as a von Mises distribution. Examples of such methods include the maximum a posteriori estimator of harmonic phase (MAP, by Kulmer and Mowlaee [11]) and methods based on temporal smoothing of unwrapped phase[1] (TSUP, by Kulmer and Mowlaee [12, 13]). TSUP applies a smoothing filter to enhance the harmonic phase at frames where the von Mises concentration exceeds a certain threshold [12] or at frames detected by a SNR-dependent binary hypothesis framework [13]. Both MAP and TSUP methods require estimates of the fundamental frequency and SNR at harmonics.

DNN-based methods use deep neural networks to estimate the phase spectrogram. Due to phase sensitivity to waveform shifts, a popular approach in this field involves estimating the phase derivatives using DNNs [14–16], followed by reconstruction of the phase spectrogram. The ability to estimate the phase using short frames enables low-latency speech enhancement using DNN-based methods [17].

It is important to note that the model-based and statistics-based methods outlined above estimate harmonic phases individually, without taking into account any relations between the signal components. In recent research, the concept of *inter-component phase relations* (ICPR, by Vorobiov *et al.* [18]) has been employed in speech enhancement by two separate groups of researchers: Barysenka *et al.* [19] and Wakabayashi *et al.* [20][2]. The main concepts of both papers are outlined below.

In [19], the authors apply temporal smoothing in the inter-component phase domain to reduce the phase variance caused by noise, then restore the phase components from the enhanced ICPR trajectories. The ICPR estimator used in that work is more accurate than previous methods that estimate each phase component individually. However, the enhanced speech has a buzzy quality due to the lack of a framework to determine smoothing regions, resulting in excessive smoothing. In [20], the authors also exploit a smoothing idea, but only apply it to frequency bins where the *a priori* SNR exceeds a constant level determined empirically. This constant SNR threshold helps reduce the buzziness in the enhanced speech.

In this paper, we derive a framework to determine the smoothing regions in the ICPR domain based on local SNR levels and prior ICPR statistics, extending the speech enhancement framework from [19]. This removes the need for an empirical choice of SNR threshold levels and a voice activity detector in the speech enhancement framework from [20]. The rest of the paper is organized as follows. Section II briefly introduces ICPR and defines bi-phase. Section III presents the high-level structure of the proposed phase estimator. Section IV derives the *a priori* distribution of bi-phase, and Section V proposes a framework for detecting bi-phase smoothing regions as a binary hypothesis test. In Section VI, we present two methods for recovering harmonic phase from the bi-phase estimate. Section VII evaluates the performance of the proposed speech enhancement scheme compared to benchmarks, and Section VIII concludes the work.

[1]The term *unwrapped phase* refers to the phase obtained by removing a linear phase term from the instantaneous phase estimate, and should not be confused with *phase unwrapping*, which refers to a procedure required to mitigate the phase wrapping phenomenon.

[2]The authors of [20] do not use the term *ICPR* directly, instead, they use the term *phase distortion* (PD) introduced by Degottex and Erro in [21]. In [18, 19], it was shown that the PD represents a specific estimate of the ICPR.

II. Inter-Component Phase Relations

A. Background

The topic of inter-component phase relations (ICPR) is extensively covered in [18]. Here, we provide a brief overview of the key concepts necessary in the context of this paper.

We consider a quasipolyharmonic signal model, where the modeled signal $x(n)$ at time instant n is represented as a linear combination of quasiharmonic components with frequencies that are multiples of the fundamental frequency $F_0(n)$:

$$x(n)=\sum_{h=1}^{H}x(h,n)=\sum_{h=1}^{H}A_x(h,n)\cos(\underbrace{2\pi hF_0(n)n+\Phi_x(h,n)}_{\Psi_x(h,n)}), \tag{1}$$

where H denotes the number of quasiharmonic components, h denotes the index of a component, $A_x(h,n)$, $\Phi_x(h,n)$ and $\Psi_x(h,n)$ denote the amplitude, the harmonic phase and the instantaneous phase of a component, respectively.

The ICPR is calculated as a linear combination of the instantaneous phase functions $\Psi_x(H(p),n)$:

$$\Theta(n) = \sum_{p=1}^{P} S(p)K(p)\Psi_x(H(p),n), \tag{2a}$$

$$\text{given } \sum_{p=1}^{P} S(p)K(p)H(p) = 0, \tag{2b}$$

where P denotes the number of the instantaneous phase functions considered for the calculation; $K(p) \in \mathbb{Q}_{>0}$ denotes the constant multiplier for the instantaneous phase at index p; $H(p)$ denotes the index h of the quasiharmonic component in (1) corresponding to the instantaneous phase at index p; $S(p) \in \{-1,+1\}$ denotes the sign of the $K(p)$ multiplier. For a valid ICPR expression, the rational number multipliers $K(p)$ and signs $S(p)$ are selected freely as far as (2b) holds.

The expression (2b) ensures the cancellation of linear phase terms $2\pi hF_0(n)n$, resulting in the calculation based solely on harmonic phases. This can be demonstrated by plugging the expression of $\Psi_x(h,n)$ from (1) into (2a) and substituting the variable h with $H(p)$:

$$\begin{aligned}\Theta(n) &= \sum_{p=1}^{P} S(p)K(p)\Psi_x(H(p),n)\\ &= 2\pi F_0(n)n\underbrace{\sum_{p=1}^{P} S(p)K(p)H(p)}_{\text{equals 0 according to (2b)}}\\ &+ \sum_{p=1}^{P} S(p)K(p)\Phi_x(H(p),n) = \sum_{p=1}^{P} S(p)K(p)\Phi_x(H(p),n).\end{aligned} \tag{3}$$

The ICPR estimates can be regarded as a generalization of phase shift estimates to the case of an arbitrary number of multiple or rational frequency harmonic components. The existing phase representations in speech processing, such as relative phase shift (RPS [9]) and phase distortion (PD [21]), are specific cases of the ICPR, as shown in [18].

B. Bi-Phase

The concept of bi-phase stems from the higher-order spectra theory [22]. To provide context, we present some key definitions from the higher-order spectra theory below.

A third-order spectrum, known as a *bi-spectrum*, for a finite energy real deterministic signal $x(n)$ is defined as follows [22]:

$$M_3^x(\omega_1, \omega_2) = X(\omega_1)X(\omega_2)X^*(\omega_1 + \omega_2), \tag{4}$$

where $X(\omega)$ denotes the Fourier transform of $x(n)$, and $X^*(\omega)$ denotes the complex conjugate of $X(\omega)$.

The *bi-phase* is defined as the phase of the bi-spectrum (4): $\angle M_3^x(\omega_1, \omega_2) = \angle X(\omega_1) + \angle X(\omega_2) - \angle X(\omega_1 + \omega_2)$. Using the notation from Section II-A, it is expressed as:

$$\Theta_3(h_1, h_2, h_3, n) = \Psi_x(h_1, n) + \Psi_x(h_2, n) - \Psi_x(h_3, n), \tag{5}$$

where h_1, h_2, h_3 are integer multiples of the fundamental frequency that form a harmonically related triplet of components with frequencies:

$$\begin{cases} f_1 = h_1 F_0, & \text{where } h_1 = 1, 2, ... \\ f_2 = h_2 F_0, & \text{where } h_2 = h_1, h_1 + 1, h_1 + 2, ... \\ f_3 = h_3 F_0, & \text{where } h_3 = h_1 + h_2. \end{cases} \tag{6}$$

Under the condition that $h_1 = h_2$, the bi-phase expression in (5) simplifies from a three-component expression to a two-component expression:

$$\Theta_2(h_1, h_2 = h_1, h_3, n) = 2\Psi_x(h_1, n) - \Psi_x(h_3, n). \tag{7}$$

The derivations of the bi-phase using the ICPR expression (2a) are presented in Appendix A. Among all the possible ICPR expressions, the bi-phase is considered in this paper due to its extensive background in the higher-order spectra theory.

The cancellation of linear phase terms is a key property of bi-phase, as well as of any other ICPR. It enables the estimation of bi-phase without the need to extract the harmonic phase $\Phi_x(h, n)$ from $\Psi_x(h, n)$, and therefore eliminates the dependence on pitch estimate. This motivates further research into ICPR-based speech enhancement methods, with the goal of finding an ICPR-based spectral phase estimator that is independent of pitch estimate. The bi-phase estimation framework, which does not rely on pitch estimation, is formulated in sections IV and V. However, in this paper, the proposed bi-phase estimation framework is still embedded in a pitch-dependent speech enhancement framework, as a pitch-independent method for spectral phase recovery from bi-phase has yet to be discovered.

III. Proposed Phase Estimator

In this section we define the notation used in the rest of this paper and present the high-level structure of the proposed phase estimator.

A. Notation

Let $x(n)$ and $y(n)$ denote the clean and noisy signals, respectively, in the time domain. The noisy signal is a mixture of the clean signal and the noise: $y(n) = x(n) + \nu(n)$.

The harmonic amplitudes, harmonic phases, and instantaneous phases of the clean signal components are represented by $A_x(h, l)$, $\Phi_x(h, l)$, and $\Psi_x(h, l)$, respectively, where h is the harmonic index and l is the frame. The same symbols with $\{\cdot\}_y$ subscripts denote the corresponding quantities in the noisy signal.

Let $\Theta_x(\vec{h}, l)$ and $\Theta_y(\vec{h}, l)$ denote the bi-phase (5), (7) of the clean and noisy signal components, respectively, at harmonic indices $\vec{h} = (h_1, h_2, h_3)$ and frame l. The set of harmonic indices $\vec{h}$ satisfy equation (6). The vector representation of $\vec{h}$ is used for compact notation.

The STFT of the clean and noisy signals are denoted by $X(k, l)$ and $Y(k, l)$, respectively, where k and l denote the frequency bin and the frame, respectively. The spectral magnitude of the clean and noisy signals are represented by $|X(k, l)|$ and $|Y(k, l)|$, respectively. The spectral phase of the clean and noisy signals are represented by $\psi_x(k, l) = \angle X(k, l)$ and $\psi_y(k, l) = \angle Y(k, l)$, respectively.

The STFT of the noise $\nu(n)$ is denoted by $N(k, l)$, such that $Y(k, l) = X(k, l) + N(k, l)$. The local *a priori* and *a posteriori* signal-to-noise ratios are defined as follows:

$$\xi(k, l) = |X(k, l)|^2 / |N(k, l)|^2, \tag{8}$$

$$\zeta(k, l) = |Y(k, l)|^2 / |N(k, l)|^2. \tag{9}$$

Note that $\xi(h, l)$ and $\zeta(h, l)$ are versions of $\xi(k, l)$ and $\zeta(k, l)$, respectively, sampled at harmonic h and frame l.

B. Proposed Phase-Aware Speech Enhancement Framework

Figure 1 shows the high-level diagram of the proposed phase-aware speech enhancement framework.

The noisy speech $y(n)$ is fed into the *Pitch Estimation* block and the *STFT* block to obtain the fundamental frequency estimate $F_0(l)$, the spectral magnitude $|Y(k, l)|$, and the phase $\psi_y(k, l)$, respectively. Then, the *Instantaneous Phase Extraction* block uses the pitch-synchronous signal segmentation procedure, which is further described in Section III-C, to obtain the instantaneous phases $\Psi_y(h, l)$ from $F_0(l)$ and $\psi_y(k, l)$.

The instantaneous phases $\Psi_y(h, l)$ are supplied to the **Bi-Phase Estimator** module along with the SNR estimates obtained from the *SNR Estimation* block. These estimates are sampled at harmonics h to form $\xi(h, l)$ and $\zeta(h, l)$. The **Bi-Phase Estimator** module consists of three blocks. The *Bi-Phase Computation* block calculates the noisy bi-phase $\Theta_y(\vec{h}, l)$ from $\Psi_y(h, l)$ using equations (5) and (7). The *Smoothing Region Detector* outputs the set of frames $\mathcal{S}$ in $\Theta_y(\vec{h}, l)$ that need to be smoothed. Smoothing regions are detected based on the bi-phase statistics derived in Section IV and the SNR-based binary hypothesis presented in Section V. The *Smoothing Filter* obtains the estimated bi-phase $\hat{\Theta}_x(\vec{h}, l)$ using the smoothing procedure described in Section V-D.

The output from the **Bi-Phase Estimator** $\hat{\Theta}_x(\vec{h}, l)$ is then processed by the *Harmonic Phase Recovery* block. The recovery process requires the reference harmonic phases $\Phi_y(1, l)$

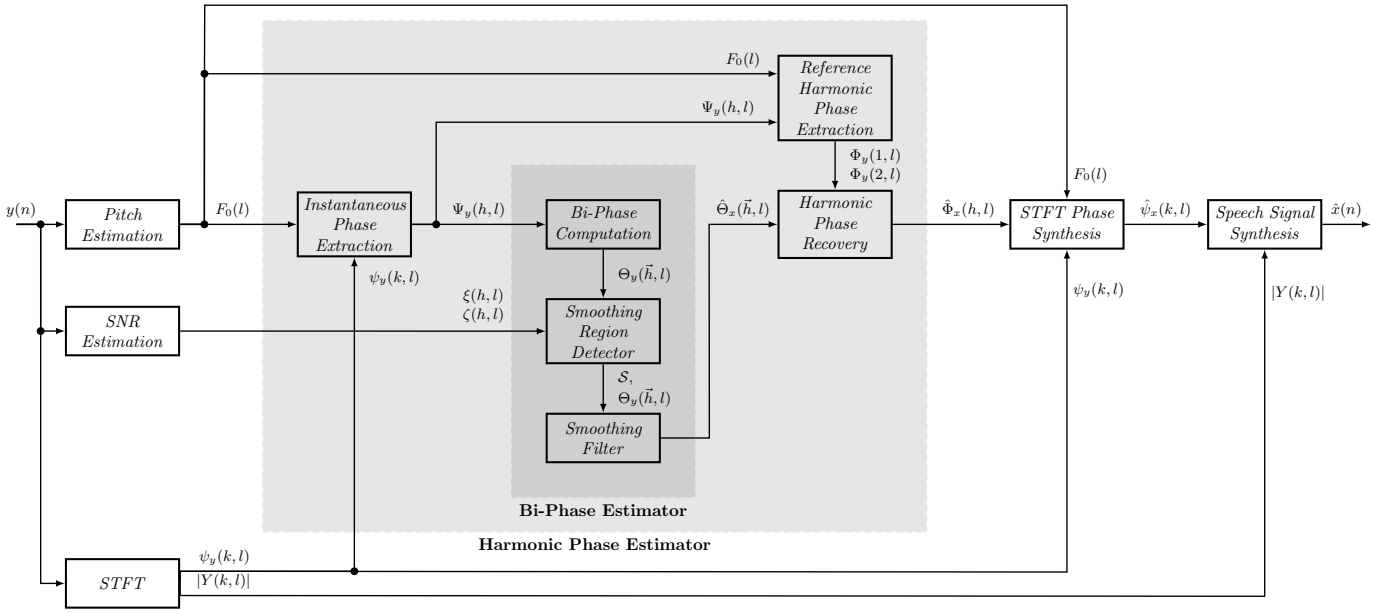


Fig. 1. Proposed phase estimation scheme in the context of the phase-aware speech enhancement framework.

and $\Phi_y(2,l)$, which are obtained in the *Reference Harmonic Phase Extraction* block by subtracting the linear phases (calculated from $F_0(l)$) from $\Psi_y(h,l)$ at $h=1,2$. Harmonic phase recovery from bi-phase is presented in Section VI.

Finally, in the *STFT Phase Synthesis* block, the estimated harmonic phases $\hat{\Phi}_x(h,l)$ are used to modify the spectral phase $\psi_y(k,l)$ and obtain the estimated spectral phase $\hat{\psi}_x(k,l)$. This modified phase is then combined with the spectral magnitude $|Y(k,l)|$ in the *Speech Signal Synthesis* block to produce the phase-enhanced speech $\hat{x}(n)$. The spectral phase modification and signal synthesis procedures are presented in Section III-D.

This framework can be used along with conventional magnitude-based speech enhancement. One possible approach, evaluated in Section VII-I, is to input the magnitude-enhanced speech signal $\hat{x}_{\text{ME}}(n)$ to the system shown in Fig. 1. Here, $y(n) \equiv \hat{x}_{\text{ME}}(n)$, and the combined magnitude-and-phase-enhanced signal $\hat{x}(n)$ is obtained at the output.

C. Pitch-Synchronous Signal Segmentation

We use the pitch-synchronous signal segmentation [21] for speech analysis and synthesis in this work, following previous studies [12, 13, 19].

The noisy speech $y(n)$ is decomposed into segments windowed by a Blackman window $w(n')$. Each segment $y_w(n',l)$ at frame l is defined as follows:

$$y_w(n',l) = y\left(n' + t(l)\right) \cdot w(n'), \tag{10}$$

where $t(l)$ denotes the time instant at each frame l, and n' denotes the STFT time index $n' \in [-(N_l-1)/2, (N_l-1)/2]$, where N_l is the analysis window length. To prevent the influence of the number of periods on the calculation of $t(l)$, the recommendation in [21] is to use four analysis instants per period for a reliable short-term phase variance:

$$t(l) = t(l-1) + 0.25/F_0(l-1) \text{ with } t(0) = 0, \tag{11}$$

where $F_0(l)$ denotes the fundamental frequency at frame l.

The noisy speech $y(n)$ is represented as the sum of segments $y_w(n',l)$, where each segment is the sum of harmonics:

$$y_w(n',l) = w(n') \sum_{h=1}^{H_l} A_y(h,l) \cos \underbrace{\left(\phi_{lin}(n',h,l) + \Phi_y(h,l)\right)}_{\Psi_y(n',h,t)}, \tag{12}$$

where H_l is the number of harmonics at frame l, $h \in [1, H_l]$ is the harmonic index, $\phi_{lin}(n',h,l) = h\omega_0(l)n'$ is the linear phase, $\omega_0(l) = 2\pi F_0(l)/f_s$ is the normalized angular fundamental frequency with f_s as the sampling frequency, and $\Psi_y(n',h,l)$ is the noisy instantaneous phase.

D. Synthesis of Phase-Enhanced Signal

The synthesis of the enhanced spectral phase, $\hat{\psi}_x(k,l)$, requires the transformation of the enhanced harmonic phase, $\hat{\Phi}_x(h,l)$, into the STFT domain. This is achieved by modifying the STFT frequency bins that are contained within the main lobe width of the analysis window [11]. The enhanced spectral phase is then given by:

$$\begin{aligned} \hat{\psi}_x(\lfloor h\omega_0(l)K \rfloor + i, l) &= \hat{\Phi}_x(h,l) + \phi_{lin}(h,l), \\ &\forall i \in [-N_p(l)/2, N_p(l)/2], \end{aligned} \tag{13}$$

where K denotes the DFT length with $k \in [0, K-1]$, and $N_p(l)$ is the minimum value of either the main lobe width of the analysis window, N_w, or frequencies close to neighboring harmonics. It is given by $N_p(l) = \min(N_w, \omega_0(l)K/(2\pi))$.

Finally, the enhanced speech signal in the STFT domain is given by $\hat{X}(k,l) = |Y(k,l)| \exp\{j\hat{\psi}_x(k,l)\}$. The time-domain signal, $\hat{x}(n)$, is obtained by the inverse DFT of $\hat{X}(k,l)$, followed by the overlap-add procedure.

IV. Prior Statistics of Bi-Phase

In this section, we derive the prior distribution of bi-phase, used later for bi-phase smoothing decision. We begin

with the von Mises distribution, commonly used as the prior distribution of harmonic phase in speech processing [13]:

$$\mathcal{VM}(\phi;\ \mu,\ \kappa) = \frac{1}{2\pi I_0(\kappa)} \exp\big\{\kappa \cos(\phi - \mu)\big\}, \tag{14}$$

where μ is the mean, κ is the concentration, and $I_\nu(\cdot)$ is the modified Bessel function of the first kind and order ν.

To derive the prior distribution of bi-phase, we consider the harmonic phase functions $\Phi(h,l)$ to be independent random variables X, Y and Z with the following von Mises statistics:

$$p(\phi \equiv X = \Phi(h_1,l)) = \mathcal{VM}(\phi;\ \mu_1(l),\ \kappa_1(l)), \tag{15}$$

$$p(\phi \equiv Y = \Phi(h_2,l)) = \mathcal{VM}(\phi;\ \mu_2(l),\ \kappa_2(l)), \tag{16}$$

$$p(\phi \equiv Z = -\Phi(h_3,l)) = \mathcal{VM}(\phi;\ -\mu_3(l),\ \kappa_3(l)). \tag{17}$$

The random variable Z corresponds to the negated harmonic phase $\Phi(h_3,l)$. The objective of this section is to derive the distribution $p(\Theta)$ of the random variable $\Theta = X + Y + Z$.

A. Three-Component Bi-Phase, $h_1 \neq h_2$

To obtain the distribution of the bi-phase, the convolution of the distributions of three random variables X, Y, and Z must be calculated. As shown in [23, Eq. 3.5.43], the convolution of the distributions of two variables X and Y is not a von Mises distribution:

$$p(\phi \equiv X+Y = \Phi(h_1,l) + \Phi(h_2,l)) = \frac{I_0(\kappa^*(\phi,l))}{2\pi I_0(\kappa_1(l)) I_0(\kappa_2(l))},$$

where $\kappa^*(\phi,l) =$

$$\sqrt{\kappa_1^2(l) + \kappa_2^2(l) + 2\kappa_1(l)\kappa_2(l)\cos\big(\phi - (\mu_1(l) + \mu_2(l))\big)}. \tag{18}$$

However, according to [23, Eq. 3.5.44], it can be approximated by a von Mises distribution. [23, Eq. 3.5.23] shows that a von Mises distribution can be approximated by the wrapped normal distribution:

$$\mathcal{VM}(\phi;\ \mu,\ \kappa) \approx \mathcal{WN}(\phi;\ \mu,\ A(\kappa)), \tag{19}$$

where

$$A(\kappa) = I_1(\kappa)/I_0(\kappa), \tag{20}$$

$$\begin{aligned}\mathcal{WN}(\phi;\ \mu,\ \rho = \exp\{-\sigma^2/2\}) \\ = \frac{1}{\sigma\sqrt{2\pi}} \sum_{k=-\infty}^{\infty} \exp\left\{\frac{-(\phi - \mu + 2\pi k)^2}{2\sigma^2}\right\}.\end{aligned} \tag{21}$$

As per [23, Eq. 3.5.67], the convolution of two wrapped normal distributions is a wrapped normal distribution:

$$\begin{aligned}&p(\phi \equiv \theta_1 + \theta_2) = \mathcal{WN}(\phi;\ \mu_1 + \mu_2,\ \rho_1\rho_2) \\ &\text{given } p(\theta_i) = \mathcal{WN}(\theta_i;\ \mu_i,\ \rho_i),\ i = 1,2.\end{aligned} \tag{22}$$

Based on (19), (20) and (22), it can be shown that the three-component bi-phase with $h_1 \neq h_2$ approximately follows a von Mises distribution:

$$p(\phi \equiv \Theta_3(h_1,h_2,h_3,l)) \approx \mathcal{VM}(\phi;\ \mu_{\Theta_3}(l),\ \kappa_{\Theta_3}(l)), \tag{23}$$

where

$$\mu_{\Theta_3}(l) = \mu_1(l) + \mu_2(l) - \mu_3(l), \tag{24}$$

$$\kappa_{\Theta_3}(l) = A^{-1}\big(A(\kappa_1(l))A(\kappa_2(l))A(\kappa_3(l))\big). \tag{25}$$

The function $A^{-1}(\cdot)$ denotes the inverse function to $A(\cdot)$.

B. Two-Component Bi-Phase, $h_1 = h_2$

Due to $h_1 = h_2$, the random variables X and Y become identical, and thus expressions (19) and (22) can no longer be used directly to approximate the distribution of $X + Y$, since X and Y are no longer independent random variables. In fact, the sum $X + Y$ represents a variable with so-called antipodal symmetry: $X + Y = 2X$. Unlike X, which represents *circular* data, $2X$ represents *axial* data. As shown in [23, Eq. 3.6.3], the distribution of $2X$ is not a von Mises distribution:

$$p(\phi \equiv 2X = 2\Phi(h_1,l)) = \frac{\cosh\left(\kappa_1(l)\cos\left(\phi - \mu_1(l)\right)\right)}{2\pi I_0(\kappa_1(l))}, \tag{26}$$

but, according to [23, Eq. 3.6.4], it can be approximated by a doubly-wrapped von Mises distribution:

$$p(\phi \equiv 2X) \approx \mathcal{VM}\big(\phi;\ \mu_{2X}(l),\ \kappa_{2X}(l)\big), \tag{27}$$

$$\mu_{2X}(l) = 2\mu_1(l), \quad \kappa_{2X}(l) = A^{-1}\big(B(\kappa_1(l))\big), \tag{28}$$

$$B(\kappa) = I_2(\kappa)/I_0(\kappa). \tag{29}$$

The expressions (19) and (22) can be used to approximate the distribution of random variable $2X + Z$, which represents the two-component bi-phase with $h_1 = h_2$:

$$p(\phi \equiv \Theta_2(h_1, h_2 = h_1, h_3, l)) \approx \mathcal{VM}(\phi;\ \mu_{\Theta_2}(l),\ \kappa_{\Theta_2}(l)), \tag{30}$$

where

$$\mu_{\Theta_2}(l) = 2\mu_1(l) - \mu_3(l), \tag{31}$$

$$\kappa_{\Theta_2}(l) = A^{-1}\big(B(\kappa_1(l))A(\kappa_3(l))\big). \tag{32}$$

C. General Expression for the Bi-Phase Distribution

Using the results from the previous sections, the bi-phase distribution can be approximated by a von Mises distribution:

$$p(\phi \equiv \Theta(\vec{h},l)) \approx \mathcal{VM}\big(\phi;\ \mu_\Theta(l),\ \kappa_\Theta(l)\big). \tag{33}$$

The parameters are computed as follows:

$$\mu_\Theta(l) = \begin{cases}\mu_1(l) + \mu_2(l) - \mu_3(l), & \text{if } h_1 \neq h_2 \\ 2\mu_1(l) - \mu_3(l), & \text{if } h_1 = h_2,\end{cases} \tag{34}$$

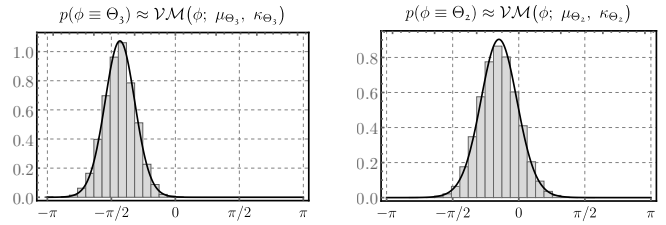
$$\kappa_\Theta(l) = \begin{cases}A^{-1}\Big(A\big(\kappa_1(l)\big)A\big(\kappa_2(l)\big)A\big(\kappa_3(l)\big)\Big), & \text{if } h_1 \neq h_2 \\ A^{-1}\Big(B\big(\kappa_1(l)\big)A\big(\kappa_3(l)\big)\Big), & \text{if } h_1 = h_2,\end{cases} \tag{35}$$

where the functions $A(\cdot)$ and $B(\cdot)$ are defined by (20), (29).

D. Numerical Simulations of Bi-Phase Prior Statistics

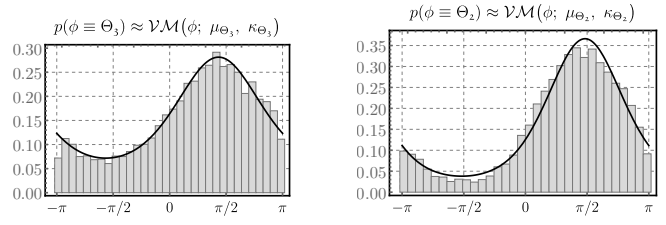
We conducted numerical simulations using Wolfram Mathematica to verify the correctness of the probability density distributions derived above. The methodology is as follows:

- Select mean (μ_1, μ_2, μ_3) and concentration $(\kappa_1, \kappa_2, \kappa_3)$.
- Generate $10,000$ pseudorandom variates for each (μ_i, κ_i) using the built-in function `RandomVariate` to obtain a realization of the random variable $\Phi(i)$ following the von Mises distribution $\mathcal{VM}(\phi; \mu_i, \kappa_i)$.
- Calculate the bi-phase as $\Theta = \Phi(1) + \Phi(2) - \Phi(3)$ (for the three-component case) or $\Theta = 2\Phi(1) - \Phi(3)$ (for the two-component case) and plot a histogram of $\mathrm{Arg}\big\{e^{j\Theta}\big\}$.



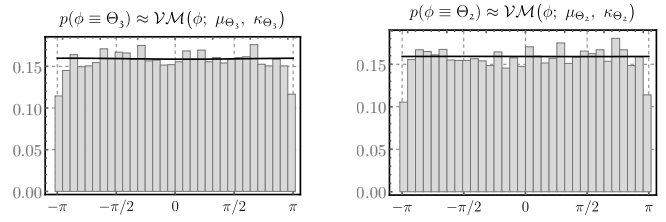
(a) $\mu_1 = -15\pi/180$, $\kappa_1 = 16$
$\mu_2 = 57\pi/180$, $\kappa_2 = 24$
$\mu_3 = 120\pi/180$, $\kappa_3 = 28$

(b) $\mu_1 = 15\pi/180$, $\kappa_1 = 24$
$\mu_3 = 57\pi/180$, $\kappa_3 = 28$



(c) $\mu_1 = 15\pi/180$, $\kappa_1 = 3$
$\mu_2 = -57\pi/180$, $\kappa_2 = 5$
$\mu_3 = -120\pi/180$, $\kappa_3 = 1$

(d) $\mu_1 = 15\pi/180$, $\kappa_1 = 4$
$\mu_3 = -57\pi/180$, $\kappa_3 = 4$



(e) $\mu_1 = 15\pi/180$, $\kappa_1 = 0.5$
$\mu_2 = -57\pi/180$, $\kappa_2 = 0.1$
$\mu_3 = 120\pi/180$, $\kappa_3 = 0.3$

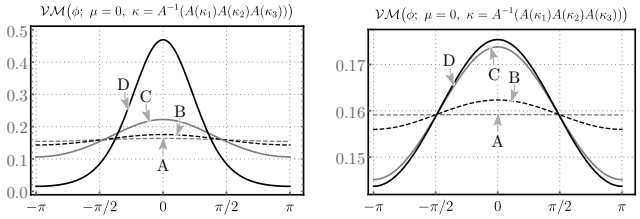
(f) $\mu_1 = -15\pi/180$, $\kappa_1 = 0.1$
$\mu_3 = 57\pi/180$, $\kappa_3 = 0.5$

Fig. 2. Histograms of the three-component bi-phase, $\Theta_3 = \Phi(1) + \Phi(2) - \Phi(3)$ (left), and two-component bi-phase, $\Theta_2 = 2\Phi(1) - \Phi(3)$ (right), computed from simulated pseudorandom von Mises data $\Phi(i)$. Solid line: the bi-phase distribution computed using the derived formulas (33)–(35).

- Compute the mean μ_Θ (34) and concentration κ_Θ (35) and plot the approximated distribution $\mathcal{VM}(\phi;\ \mu_\Theta,\ \kappa_\Theta)$ along with the histogram from the previous step.

The results of the simulations are presented in Fig. 2. The derived distribution (33) corresponds to the simulated data.

Fig. 3 shows the bi-phase probability density function approximated by the von Mises distribution under various combinations of concentration parameters of individual harmonic phases. A few observations can be made: i) from Fig. 3a, it can be seen that higher concentration values for *all* harmonic phases are required to turn the bi-phase distribution towards the Dirac delta, and ii) from Fig. 3b, it can be seen that even for very high concentration of two harmonic phases, the resulting bi-phase distribution is far from the Dirac delta if the concentration of the remaining harmonic phase remains very low. In the context of bi-phase smoothing, these observations are in line with our expectations, as we aim to prevent smoothing in regions where harmonic phases are likely to exhibit a uniform distribution. For bi-phase, having at least one harmonic phase uniformly distributed should prevent bi-phase from being smoothed, which aligns with the observed behavior of the bi-phase distribution in Fig. 3b.



(a) Distributions with various combinations of low and high κ_i:
A: $\kappa_1 = 0.5,\ \kappa_2 = 0.5,\ \kappa_3 = 0.5$
B: $\kappa_1 = 4,\ \kappa_2 = 0.5,\ \kappa_3 = 0.5$
C: $\kappa_1 = 4,\ \kappa_2 = 4,\ \kappa_3 = 0.5$
D: $\kappa_1 = 4,\ \kappa_2 = 4,\ \kappa_3 = 4$

(b) Distributions with κ_3 fixed at very low and increasing κ_1 and κ_2:
A: $\kappa_1 = 0.1,\ \kappa_2 = 0.1,\ \kappa_3 = 0.1$
B: $\kappa_1 = 1,\ \kappa_2 = 1,\ \kappa_3 = 0.1$
C: $\kappa_1 = 10,\ \kappa_2 = 10,\ \kappa_3 = 0.1$
D: $\kappa_1 = 200,\ \kappa_2 = 200,\ \kappa_3 = 0.1$

Fig. 3. Bi-phase probability density function approximated by von Mises distribution with $\mu_\Theta(l) = 0$ and various values of concentration $\kappa_\Theta(l)$.

V. Bi-Phase Smoothing Using Binary Hypothesis Test

To obtain an estimate of the bi-phase from the noisy observation, we perform smoothing in regions where a harmonic structure is present, while keeping the bi-phase unprocessed if there is no harmonic structure. The challenge is to accurately detect such regions. In the following, we present a detector of such regions based on the bi-phase prior distribution (33). This detector is incorporated into the bi-phase smoothing framework that we proposed in our earlier work [19]. In the prior work [13], a similar detector was proposed for smoothing the harmonic phase trajectories of each harmonic individually.

A. Binary Hypothesis Test Framework for Bi-Phase

Assuming the speech signal contains voiced and unvoiced regions, we formulate two hypotheses for bi-phase estimation:

$$\mathcal{H}_0 : \Theta_y(\vec{h}, l) = \Theta_\nu(\vec{h}, l), \tag{36}$$

$$\mathcal{H}_1 : \Theta_y(\vec{h}, l) = \Theta_x(\vec{h}, l) + \Theta_\nu(\vec{h}, l). \tag{37}$$

The $\mathcal{H}_0$ hypothesis represents the case of no harmonic structure, where we assume the bi-phase is uniformly distributed:

$$p(\Theta_\nu(\vec{h}, l)) = U[-\pi, \pi), \quad p(\Theta_y(\vec{h}, l) \mid \mathcal{H}_0) = \frac{1}{2\pi}. \tag{38}$$

The $\mathcal{H}_1$ hypothesis represents the case of voiced region, where the bi-phase distribution is approximated by (33):

$$p(\phi \equiv \Theta_x(\vec{h}, l) + \Theta_\nu(\vec{h}, l)) \approx \mathcal{VM}\big(\phi;\ \mu_\Theta(\vec{h}, l),\ \kappa_\Theta(\vec{h}, l)\big),$$
$$p(\Theta_y(\vec{h}, l)|\mathcal{H}_1) = \frac{\exp\big\{\kappa_\Theta(\vec{h}, l)\cos(\Theta_y(\vec{h}, l) - \mu_\Theta(\vec{h}, l))\big\}}{2\pi I_0(\kappa_\Theta(\vec{h}, l))}. \tag{39}$$

To decide $\mathcal{H}_1$, we use a detector that minimizes the probability of an erroneous decision of $\mathcal{H}_i$ when $\mathcal{H}_j$ is true [24, Ch. 3.6]:

$$\frac{p(\Theta_y(\vec{h}, l)|\mathcal{H}_1)}{p(\Theta_y(\vec{h}, l)|\mathcal{H}_0)} \underset{\mathcal{H}_0}{\overset{\mathcal{H}_1}{\gtrless}} \frac{P(\mathcal{H}_0)}{P(\mathcal{H}_1)}. \tag{40}$$

We assume that regions where the harmonic structure is present or absent are equally likely to appear in the signal, so the prior probabilities of the respective hypotheses are $P(\mathcal{H}_0) = P(\mathcal{H}_1) = 1/2$. With this consideration, after

inserting (38) and (39) into (40), the decision in (40) can be reformulated as follows:

$$\frac{\exp\left\{\kappa_\Theta(\vec{h},l)\cos(\Theta_y(\vec{h},l)-\mu_\Theta(\vec{h},l))\right\}}{I_0(\kappa_\Theta(\vec{h},l))} \underset{\mathcal{H}_0}{\overset{\mathcal{H}_1}{\gtrless}} 1. \tag{41}$$

Taking the logarithm of both sides yields:

$$\cos(\underbrace{\Theta_y(\vec{h},l)-\mu_\Theta(\vec{h},l)}_{\Theta_{\text{dev}}(\vec{h},l)}) \underset{\mathcal{H}_0}{\overset{\mathcal{H}_1}{\gtrless}} \underbrace{\frac{1}{\kappa_\Theta(\vec{h},l)}\ln I_0(\kappa_\Theta(\vec{h},l))}_{\Xi(\vec{h},l)}, \tag{42}$$

where $\Xi(\vec{h},l)$ is the decision threshold, and $\Theta_{\text{dev}}(\vec{h},l)$ is the bi-phase deviation, which represents the analog of the phase deviation discussed in [13]. In the following, we formulate $\Theta_{\text{dev}}(\vec{h},l)$ based on the *a priori* and *a posteriori* SNR.

B. Bi-Phase Deviation Based on SNR

The expression for phase deviation was derived in [13]:

$$\cos\Psi_{\text{dev}}(h,l)=\frac{\xi(h,l)+\zeta(h,l)-1}{2\sqrt{\zeta(h,l)\xi(h,l)}}, \tag{43}$$

where $\Psi_{\text{dev}}(h,l)=\Psi_y(h,l)-\Psi_x(h,l)$. The bi-phase deviation can be expressed in terms of the phase deviation of its individual phases:

$$\Theta_{\text{dev}}(\vec{h},l)=\Psi_{\text{dev}}(h_1,l)+\Psi_{\text{dev}}(h_2,l)-\Psi_{\text{dev}}(h_3,l). \tag{44}$$

By taking the cosine of both sides of (44) and applying the sum-to-product formula for cosine, we get:

$$\begin{aligned}\cos\Theta_{\text{dev}}(\vec{h},l) &= \cos\Psi_{\text{dev}}(h_1,l)\cos\Psi_{\text{dev}}(h_2,l)\cos\Psi_{\text{dev}}(h_3,l)\\ &+\cos\Psi_{\text{dev}}(h_1,l)\sin\Psi_{\text{dev}}(h_2,l)\sin\Psi_{\text{dev}}(h_3,l)\\ &+\sin\Psi_{\text{dev}}(h_1,l)\cos\Psi_{\text{dev}}(h_2,l)\sin\Psi_{\text{dev}}(h_3,l)\\ &-\sin\Psi_{\text{dev}}(h_1,l)\sin\Psi_{\text{dev}}(h_2,l)\cos\Psi_{\text{dev}}(h_3,l).\end{aligned} \tag{45}$$

Based on (43), it can be shown that

$$\begin{aligned}&\sin\Psi_{\text{dev}}(h,l)\\ &=\frac{\sqrt{2\zeta(h,l)(1+\xi(h,l))-(\xi(h,l)-1)^2-\zeta^2(h,l)}}{2\sqrt{\zeta(h,l)\xi(h,l)}}.\end{aligned} \tag{46}$$

By inserting (43) and (46) into (45), we obtain:

$$\begin{aligned}\cos\Theta_{\text{dev}}(\vec{h},l)=\frac{1}{8\alpha(\vec{h},l)}\Big[&\beta(h_1,h_2,h_3,l)+\beta(h_2,h_1,h_3,l)\\ &-\beta(h_3,h_1,h_2,l)+\gamma(\vec{h},l)\Big],\end{aligned} \tag{47}$$

where

$$\alpha(\vec{h},l)=\sqrt{\zeta(h_1,l)\xi(h_1,l)}\sqrt{\zeta(h_2,l)\xi(h_2,l)}\sqrt{\zeta(h_3,l)\xi(h_3,l)}, \tag{48}$$

$$\beta(h_1,h_2,h_3,l)=\rho(h_1,l)\chi(h_2,l)\chi(h_3,l), \tag{49}$$

$$\gamma(\vec{h},l)=\rho(h_1,l)\rho(h_2,l)\rho(h_3,l), \tag{50}$$

$$\rho(h,l)=\xi(h,l)+\zeta(h,l)-1, \tag{51}$$

$$\chi(h,l)=\sqrt{2\zeta(h,l)(1+\xi(h,l))-(\xi(h,l)-1)^2-\zeta^2(h,l)}. \tag{52}$$

C. Computation of Bi-Phase and Concentration Parameter κ_Θ

To avoid the necessity of phase unwrapping, bi-phase (5) is computed using complex exponentials as follows:

$$\Theta_y(\vec{h},l)=\angle\frac{\exp(j\Psi_y(h_1,l))\cdot\exp(j\Psi_y(h_2,l))}{\exp(j\Psi_y(h_3,l))}. \tag{53}$$

Note that it is not necessary to access the harmonic phases $\Phi_y(h,l)$ explicitly to compute bi-phase, as bi-phase is blind to linear phase terms (3). Hence, computation via instantaneous phases $\Psi_y(h,l)$ is possible. Further, $\kappa_\Theta(\vec{h},l)$ is estimated directly from the observed bi-phase data (53)[3] using a maximum likelihood estimation approach and following the computation suggestions given in [23, Ch. 5.3.1].

D. Bi-Phase Smoothing

To smooth the bi-phase across time, a symmetric moving average filtering is applied to $\Theta_y(\vec{h},l)$ within the smoothing regions:

$$\forall l\in\mathcal{S}:\quad \hat{\Theta}_x(\vec{h},l)=\angle\frac{1}{|\mathcal{W}|}\sum_{\tilde{l}\in\mathcal{W}}e^{j\Theta_y(\vec{h},\tilde{l})}, \tag{54}$$

where $\mathcal{S}$ is the set of frames within the smoothing regions, and $\mathcal{W}$ is the smoothing filter length. The phase unwrapping procedure is not required as the complex exponentials $\exp\{j\Theta_y(\vec{h},\tilde{l})\}$ are averaged instead.

The smoothing procedure (54) is similar to the one used in [19], with the difference that it now considers frames within smoothing regions. For frames outside of these regions, the noisy bi-phase observation is used directly as a bi-phase estimate:

$$\forall l\notin\mathcal{S}:\quad \hat{\Theta}_x(\vec{h},l)=\Theta_y(\vec{h},l). \tag{55}$$

The bi-phase estimation approach is summarized in Algorithm 1. This algorithm is tailored to the pitch-dependent speech enhancement framework discussed in this paper. However, with slight modifications to the variables, Algorithm 1 can also be used to estimate bi-phase across all STFT time-frequency bins, regardless of pitch. The modifications include replacing the harmonic indices $\vec{h}$ with frequency bins $\vec{k}$ (where $\vec{k}$ corresponds to the frequency relation defined by (6): $k_3=k_1+k_2$), and substituting the instantaneous phases $\Psi_y(h_i,l)$ with the spectral phase $\psi_y(k,l)$. However, as bi-phase does not contain information about linear phases, the direct recovery of the speech signal's spectral phase $\hat{\psi}_x(k,l)$ from bi-phase is a non-trivial procedure and has yet to be discovered.

VI. Harmonic Phase Recovery From Bi-Phase

To recover the harmonic phase from the estimated bi-phase, we consider two iterative algorithms. The *Barysenka-Vorobiov-Mowlaee* (BVM) algorithm, which was proposed in our prior work [19], uses only a limited set of three-component bi-phase vectors $\Theta_3(\vec{h},l)$, with $h_1=1$. The *Bartelt-Lohmann-Wirnitzer* (BLW) algorithm [25] leverages all bi-phase vectors, including three-component $\Theta_3(\vec{h},l)$ and two-component $\Theta_2(\vec{h},l)$ ones.

[3]Computing via (35) using individual $\kappa(h,l)$ produces the same results.

Algorithm 1 Bi-Phase Estimation

Initialization
$\vec{h} \leftarrow (h_1, h_2, h_3)$.
$\forall h_i \in \vec{h} : \xi(h_i, l) \leftarrow$ *a priori* SNR estimate.
$\forall h_i \in \vec{h} : \zeta(h_i, l) \leftarrow$ *a posteriori* SNR estimate.
$\forall h_i \in \vec{h} : \Psi_y(h_i, l) \leftarrow$ noisy instantaneous phase.

Smoothing Region Detection
$\cos\Theta_{\text{dev}}(\vec{h}, l) \leftarrow$ calculate cosine of bi-phase deviation from $\xi(h_i, l)$ and $\zeta(h_i, l)$ using equations (47)–(52).
$\Theta_y(\vec{h}, l) \leftarrow$ compute bi-phase from $\Psi_y(h_i, l)$ using (53).
$\kappa_\Theta(\vec{h}, l) \leftarrow$ estimate concentration parameter of von Mises distribution from $\Theta_y(\vec{h}, l)$ as in Section V-C.
$\Xi(\vec{h}, l) \leftarrow$ calculate the smoothing decision threshold from $\kappa_\Theta(\vec{h}, l)$ using the right-hand side expression in (42).
$\mathcal{S} \leftarrow$ smoothing region: a set of frames l where $\cos\Theta_{\text{dev}}(\vec{h}, l) > \Xi(\vec{h}, l)$ (42).

Smoothing
for $\forall l$ **do**
  if $l \in \mathcal{S}$ **then**
    $\hat{\Theta}_x(\vec{h}, l) \leftarrow$ perform smoothing of $\Theta_y(\vec{h}, l)$ (54).
  else
    $\hat{\Theta}_x(\vec{h}, l) \leftarrow$ assign noisy bi-phase $\Theta_y(\vec{h}, l)$ (55).
  end if
end for
return $\hat{\Theta}_x(\vec{h}, l)$.

A. Barysenka-Vorobiov-Mowlaee Iterative Recovery [19]

Assuming that H signal harmonics are available, the harmonic indices defined by (6) are given as follows:

$$\begin{aligned} &h_1 = 1, \\ &h_2 = h_1 + 1, h_1 + 2, ..., H - 1. \end{aligned} \tag{56}$$

We can define the following three-component bi-phase equations of the form (5):

$$\begin{aligned} \hat{\Theta}_x(1,2,3,l) &= \hat{\Phi}_x(1,l) + \hat{\Phi}_x(2,l) - \hat{\Phi}_x(3,l), \\ \hat{\Theta}_x(1,3,4,l) &= \hat{\Phi}_x(1,l) + \hat{\Phi}_x(3,l) - \hat{\Phi}_x(4,l), \\ &\vdots \\ \hat{\Theta}_x(1,H-1,H,l) &= \hat{\Phi}_x(1,l) + \hat{\Phi}_x(H-1,l) - \hat{\Phi}_x(H,l). \end{aligned} \tag{57}$$

After initializing with $\hat{\Phi}_x(1,l)$ and $\hat{\Phi}_x(2,l)$, all remaining harmonic phases are reconstructed recursively from the equations (57) as follows:

$$\begin{aligned} \hat{\Phi}_x(3,l) &= \hat{\Phi}_x(1,l) + \hat{\Phi}_x(2,l) - \hat{\Theta}_x(1,2,3,l), \\ \hat{\Phi}_x(4,l) &= \hat{\Phi}_x(1,l) + \hat{\Phi}_x(3,l) - \hat{\Theta}_x(1,3,4,l), \\ &\vdots \\ \hat{\Phi}_x(H,l) &= \hat{\Phi}_x(1,l) + \hat{\Phi}_x(H-1,l) - \hat{\Theta}_x(1,H-1,H,l), \end{aligned} \tag{58}$$

which requires a total of $H - 2$ iterations.

Algorithm 2 outlines the harmonic phase recovery scheme presented above. It is important to note that the algorithm does not take into account every possible bi-phase combination, and only leverages a limited set of three-component bi-phase vectors $\Theta_3(\vec{h}, l)$ with $h_1 = 1$.

Algorithm 2 Harmonic Phase Recovery: BVM [19]

Initialization
$H \leftarrow$ number of harmonic components.
$\hat{\Theta}_x(\vec{h}, l) \leftarrow$ bi-phase estimates (57).
$\hat{\Phi}_x(1,l) \leftarrow \Phi_y(1,l)$.
$\hat{\Phi}_x(2,l) \leftarrow \Phi_y(2,l)$.

Harmonic Phase Recovery
for $h = 3 \dots H$ **do**
  $\hat{\Phi}_x(h,l) \leftarrow \hat{\Phi}_x(1,l) + \hat{\Phi}_x(h-1,l) - \hat{\Theta}_x(1,h-1,h,l)$ (58).
end for
return $\hat{\Phi}_x(1 \dots H,\ l)$.

B. Bartelt-Lohmann-Wirnitzer Iterative Recovery [25]

Assuming H signal harmonics are available, the harmonic indices defined by (6) are given as follows:

$$\begin{aligned} &h_1 = 1, 2, ..., \lfloor H/2 \rfloor, \\ &h_2 = h_1, h_1 + 1, ..., H - h_1. \end{aligned} \tag{59}$$

We can define all possible three-component and two-component bi-phase equations of the form (5) and (7):

$$\begin{aligned} &\hat{\Theta}_x(1,1,2,l) = 2\hat{\Phi}_x(1,l) - \hat{\Phi}_x(2,l), \\ &\hat{\Theta}_x(1,2,3,l) = \hat{\Phi}_x(1,l) + \hat{\Phi}_x(2,l) - \hat{\Phi}_x(3,l), \\ &\hat{\Theta}_x(2,2,4,l) = 2\hat{\Phi}_x(2,l) - \hat{\Phi}_x(4,l), \\ &\hat{\Theta}_x(1,3,4,l) = \hat{\Phi}_x(1,l) + \hat{\Phi}_x(3,l) - \hat{\Phi}_x(4,l), \\ &\hat{\Theta}_x(2,3,5,l) = \hat{\Phi}_x(2,l) + \hat{\Phi}_x(3,l) - \hat{\Phi}_x(5,l), \\ &\hat{\Theta}_x(3,3,6,l) = 2\hat{\Phi}_x(3,l) - \hat{\Phi}_x(6,l), \\ &\hat{\Theta}_x(1,4,5,l) = \hat{\Phi}_x(1,l) + \hat{\Phi}_x(4,l) - \hat{\Phi}_x(5,l), \\ &\vdots \\ &\hat{\Theta}_x(1, H - \lfloor H/2 \rfloor, H - \lfloor H/2 \rfloor + 1, l) \\ &\quad = \hat{\Phi}_x(1,l) + \hat{\Phi}_x(H - \lfloor H/2 \rfloor, l) - \hat{\Phi}_x(H - \lfloor H/2 \rfloor + 1, l), \\ &\hat{\Theta}_x(2, H - \lfloor H/2 \rfloor, H - \lfloor H/2 \rfloor + 2, l) \\ &\quad = \hat{\Phi}_x(2,l) + \hat{\Phi}_x(H - \lfloor H/2 \rfloor, l) - \hat{\Phi}_x(H - \lfloor H/2 \rfloor + 2, l), \\ &\vdots \\ &\hat{\Theta}_x(\lfloor H/2 \rfloor, H - \lfloor H/2 \rfloor, H, l) \\ &\quad = \hat{\Phi}_x(\lfloor H/2 \rfloor, l) + \hat{\Phi}_x(H - \lfloor H/2 \rfloor, l) - \hat{\Phi}_x(H, l). \end{aligned} \tag{60}$$

The equations in (60) are listed in a specific order. First, all possible equations for $h_2 = 1$ are written, which results in one equation with $h_1 = 1$. Then, h_2 is incremented, and

all possible equations for $h_2 = 2$ are written, resulting in two additional equations with $h_1 = 1, 2$. This process is repeated until $h_2 = H - \lfloor H/2 \rfloor$, resulting in an additional $\lfloor H/2 \rfloor$ equations with $h_1 = 1, 2, ..., \lfloor H/2 \rfloor$.

Unlike the BVM algorithm (Section VI-A), the BLW algorithm only requires prior knowledge of $\hat{\Phi}_x(1, l)$ at the initialization stage. According to [25], various computation paths of BLW recursion may be proposed, and the algorithm can be initialized with a harmonic phase that differs from $\hat{\Phi}_x(1, l)$ based on noise conditions, availability of other reference harmonic phases, etc. However, this work focuses on the order of equations as defined in (60) and initializing the algorithm with $\hat{\Phi}_x(1, l)$, considering other scenarios to be outside the scope of the current work.

Once initialized with $\hat{\Phi}_x(1, l)$, the harmonic phases are reconstructed from the equations in (60) as follows:

$$
\begin{aligned}
&\hat{\Phi}_x(2,l) = 2\hat{\Phi}_x(1,l) - \hat{\Theta}_x(1,1,2,l),\\
&\\
&\hat{\Phi}_x(3,l) = \hat{\Phi}_x(1,l) + \hat{\Phi}_x(2,l) - \hat{\Theta}_x(1,2,3,l),\\
&\hat{\Phi}_x(4,l) = 2\hat{\Phi}_x(2,l) - \hat{\Theta}_x(2,2,4,l),\\
&\\
&\hat{\Phi}_x(4,l) = \hat{\Phi}_x(1,l) + \hat{\Phi}_x(3,l) - \hat{\Theta}_x(1,3,4,l),\\
&\hat{\Phi}_x(5,l) = \hat{\Phi}_x(2,l) + \hat{\Phi}_x(3,l) - \hat{\Theta}_x(2,3,5,l),\\
&\hat{\Phi}_x(6,l) = 2\hat{\Phi}_x(3,l) - \hat{\Theta}_x(3,3,6,l),\\
&\\
&\hat{\Phi}_x(5,l) = \hat{\Phi}_x(1,l) + \hat{\Phi}_x(4,l) - \hat{\Theta}_x(1,4,5,l),\\
&\vdots\\
&\hat{\Phi}_x(H - \lfloor H/2 \rfloor + 1, l) = \hat{\Phi}_x(1,l) + \hat{\Phi}_x(H - \lfloor H/2 \rfloor, l)\\
&\quad - \hat{\Theta}_x(1, H - \lfloor H/2 \rfloor, H - \lfloor H/2 \rfloor + 1, l),\\
&\vdots\\
&\hat{\Phi}_x(H,l) = \hat{\Phi}_x(\lfloor H/2 \rfloor, l) + \hat{\Phi}_x(H - \lfloor H/2 \rfloor, l)\\
&\quad - \hat{\Theta}_x(\lfloor H/2 \rfloor, H - \lfloor H/2 \rfloor, H, l).
\end{aligned}
\tag{61}
$$

The equations in (61) are designed so that the right-hand side expressions contain either the reference phase $\hat{\Phi}_x(1, l)$ or phases obtained from previous iterations. By evaluating these equations in the order defined in (61), all harmonic phases can be recovered.

Algorithm 3 summarizes the process for harmonic phase recovery using the BLW method. It should be noted that the algorithm provides multiple representations of $\hat{\Phi}_x(h, l)$. For instance, two representations of $\hat{\Phi}_x(4, l)$ are obtained in (61) before their first use in the calculation of $\hat{\Phi}_x(5, l) = \hat{\Phi}_x(1, l) + \hat{\Phi}_x(4, l) - \hat{\Theta}_x(1, 4, 5, l)$. According to [25], multiple representations of $\hat{\Phi}_x(h, l)$ are then averaged (as shown in Algorithm 3, Line 11). The authors of [25] consider this averaging to be an advantage, as it reduces the impact of noise. The complex exponentials $\exp\{j\hat{\Phi}_x(h, l)\}$ are averaged instead of the phases $\hat{\Phi}_x(h, l)$ to eliminate the need for phase unwrapping.

Algorithm 3 Harmonic Phase Recovery: BLW [25]

1: **Initialization**
2: $H \leftarrow$ number of harmonic components.
3: $\hat{\Theta}_x(\vec{h}, l) \leftarrow$ bi-phase estimates (60).
4: $\hat{\Phi}_x(1, l) \leftarrow \Phi_y(1, l)$.

5: **Harmonic Phase Recovery**
6: **for** $h_2 = 1 \dots H - \lfloor H/2 \rfloor$ **do**
7: **for** $h_1 = 1 \dots h_2$ **do**
8: $h_3 \leftarrow h_1 + h_2$.
9: $\hat{\Phi}^*_x(h_3, l) \leftarrow \hat{\Phi}_x(h_1, l) + \hat{\Phi}_x(h_2, l) - \hat{\Theta}_x(\vec{h}, l)$ (61).
10: **if** $\hat{\Phi}_x(h_3, l)$ is available from previous iterations **then**
11: $\hat{\Phi}_x(h_3, l) \leftarrow \angle(\exp j\hat{\Phi}_x(h_3, l) + \exp j\hat{\Phi}^*_x(h_3, l))$.
12: **else**
13: $\hat{\Phi}_x(h_3, l) \leftarrow \hat{\Phi}^*_x(h_3, l)$.
14: **end if**
15: **end for**
16: **end for**
17: **return** $\hat{\Phi}_x(1 \dots H,\ l)$.

TABLE I
SUMMARY OF HARMONIC PHASE RECOVERY ALGORITHMS

	BVM [19]	BLW [25]
Utilized bi-phase values	Three-component only: $\Theta_3(h_1, h_2, h_3, l)$, where $h_1 = 1$	All: $\Theta_3(h_1, h_2, h_3, l)$, $\Theta_2(h_1, h_1, h_3, l)$
Number of iterations	$H - 2$	$\lfloor \frac{H}{2} \rfloor (H - \lfloor \frac{H}{2} \rfloor)$
Reference phases	$\hat{\Phi}_x(1, l),\ \hat{\Phi}_x(2, l)$	$\hat{\Phi}_x(1, l)$
Averaging of multiple $\hat{\Phi}_x(h, l)$ for the same h?	No	Yes

Table I provides a brief overview of the harmonic phase recovery algorithms discussed in this section.

VII. RESULTS

In this section, we present the results of the performance evaluation of the proposed speech enhancement scheme compared to the benchmarks.

A. Databases and Evaluation Metrics

We randomly selected 50 utterances from 10 speakers, including both female and male, from the GRID corpus [26] sampled at 16 kHz. The utterances were corrupted by white noise, babble noise, factory noise, and modulated pink noise files taken from the NOISEX-92 database [27]. The SNR levels ranged from -5 to 10 dB in increments of 5 dB.

The performance of the phase estimation methods for speech enhancement is evaluated using two categories of measures, as suggested in [28]. The first category evaluates the impact of the algorithms on the speech signal structure, while the second category assesses the impact of the algorithms on noise suppression.

The first category consists of the following measures: perceptual evaluation of speech quality (PESQ) [29], short-time

objective intelligibility measure (STOI) [30], and phase deviation (PDev) [31][4]. PESQ and STOI instrumentally predict perceived speech quality and speech intelligibility, respectively. PDev quantifies the accuracy of spectral phase estimation:

$$\text{PDev} = \frac{1}{LK}\sum_{l=1}^{L}\sum_{k=1}^{K}\left(\cos\psi_{\text{dev}}(k,l) - \cos\hat{\psi}_{\text{dev}}(k,l)\right), \quad (62)$$

where $\psi_{\text{dev}}(k,l) = \psi_y(k,l) - \psi_x(k,l)$, L is the number of frames, K is the number of frequency bins in the spectral phase, and l and k are indices for frames and frequency bins.

According to [31], PDev has the highest average correlation with results from subjective listening tests for perceived quality among other phase-aware metrics. Furthermore, as reported in [32, Ch. 6.7.4], PDev does not overestimate the perceived quality of buzzy speech, which is an artifact that occurs when noisy components are excessively harmonized, resulting in a more harmonized enhanced signal compared to clean speech. PESQ, which is originally proposed for speech coding, is unable to properly capture such artifacts, that are common in phase-aware speech enhancement. Therefore, we will use PDev to quantify the level of buzziness in enhanced speech.

We report the performance of the algorithms in ΔPESQ, ΔSTOI and ΔPDev, where Δ denotes the difference in the value of the metric achieved for processed speech compared to noisy speech: $\Delta(\text{metric}) = (\text{metric})_{\text{processed}} - (\text{metric})_{\text{noisy}}$. Positive values of ΔPESQ and ΔSTOI indicate improvement in PESQ and STOI, respectively, while negative values of ΔPDev indicate improvement in reducing phase deviation.

The second category of measures includes the segmental noise attenuation (NA_{seg}) measure [33]:

$$\text{NA}_{\text{seg}} = 10\log_{10}\left(\frac{1}{L}\sum_{l=1}^{L}\frac{\sum_{n=1}^{N}\nu^2(n,l)}{\sum_{n=1}^{N}\widetilde{\nu}^2(n,l)}\right), \quad (63)$$

where L is the number of frames, N is the number of samples per frame, $\nu(n,l)$ represents the frame of noise, and $\widetilde{\nu}(n,l)$ represents the frame of attenuated noise. To estimate the attenuated noise, the spectral gain function $\hat{G}(k,l)$ is calculated from the STFT of the enhanced speech, $\hat{X}(k,l)$, and the noisy speech, $Y(k,l)$, as described in [33]:

$$\hat{G}(k,l) = \min\left(\frac{|\hat{X}(k,l)|}{|Y(k,l)|},\ 1\right)\exp\left\{j\angle\frac{\hat{X}(k,l)}{Y(k,l)}\right\}. \quad (64)$$

The attenuated noise, $\widetilde{\nu}(n,l)$, is obtained by multiplying $\hat{G}(k,l)$ by the noise STFTs, $N(k,l)$, and transforming the result back to the time domain.

B. Signal Segmentation, Pitch and SNR Estimation Settings

In this paper, the PEFAC algorithm [34] is selected as the fundamental frequency (F_0) evaluation method due to its robustness in high-noise conditions. PEFAC is configured to provide F_0 estimates with a 2 ms frame increment.

For speech parameters estimation, the pitch-synchronous segmentation is applied, as described in Section III-C. The frames are windowed using a zero-phase Blackman window of length 24 ms and rounded to the closest multiple of the fundamental period provided by PEFAC. The frame shift, dependent on F_0, is used as specified in (11).

To estimate the *a priori* and *a posteriori* SNRs, the decision-directed approach [2] with the Log-MMSE noise power spectral density estimator, as implemented in [35], is used.

C. Algorithms: Benchmarks and Proposed

We evaluate 4 harmonic phase estimation schemes, which are built from combinations of two bi-phase smoothing schemes and two phase recovery schemes. The two bi-phase smoothing schemes are the *Smooth Everywhere* (SE, proposed in our prior work [19]) and the *Binary Hypothesis* (BH, proposed in Section V). The two harmonic phase recovery schemes are the BVM and the BLW outlined in Section VI. Their combinations result in 4 harmonic phase estimation schemes: SE+BVM (*benchmark*), SE+BLW (*benchmark*), BH+BVM (*proposed*) and BH+BLW (*proposed*). The SE+BVM scheme represents the framework proposed in our prior work [19], while the other schemes are novel.

D. Proof-of-Concept: Bi-Phase Smoothing

In this section, we demonstrate a typical bi-phase reconstruction procedure in the BH scheme. Fig. 4a shows a fragment of a noisy spectrogram of a single utterance contaminated by modulated pink noise at SNR = 5 dB. The semi-transparent strokes depict frequency tracks of the 1st, 4th, and 5th multiples of the pitch. The dotted stroke in Fig. 4c shows the noisy (unprocessed) bi-phase trajectory $\Theta_y(1,4,5,l)$ calculated on components $(1,4,5)$. The abrupt changes in some areas corresponding to voiced regions are a result of degradation introduced by noise.

The dark regions in Fig. 4d represent the solution to (42) where $\mathcal{H}_1$ is decided, meaning that the bi-phase $\Theta_y(1,4,5,l)$ needs to be smoothed. After smoothing, the bi-phase trajectory $\hat{\Theta}_x(1,4,5,l)$ in the reconstructed speech is shown as the black stroke in Fig. 4c. It can be seen that the bi-phase trajectory now varies slowly in the majority of voiced regions.

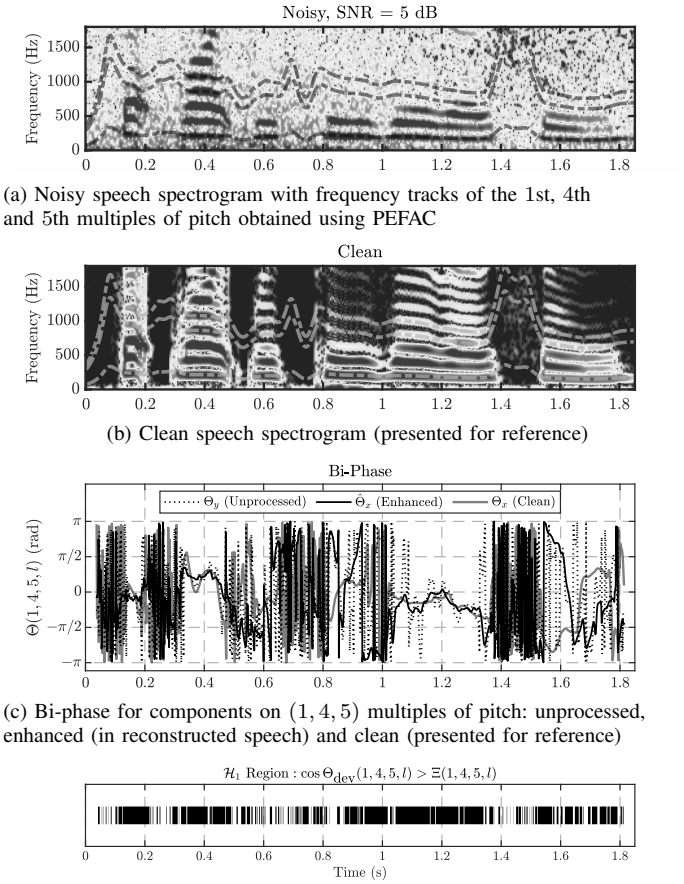
For reference, the trajectory of the clean bi-phase $\Theta_x(1,4,5,l)$ and the clean speech spectrogram are shown as the gray stroke in Fig. 4c and Fig. 4b, respectively. The enhanced bi-phase trajectory (black stroke in Fig. 4c) has improved and is closer to the clean bi-phase trajectory. The bi-phase smoothing region (Fig. 4d) closely corresponds to the areas in the clean speech spectrogram where the tracks of pitch multiples match the harmonics.

For illustration, the results shown in Fig. 4 are obtained in an oracle setup, where F_0 and SNR are known. In a blind setup, the smoothing decision is more conservative and suggests applying smoothing in fewer areas of the bi-phase trajectory. This limits the performance of the proposed framework in real-world scenarios, but prevents the introduction of non-existent harmonics that may be perceived as buzzy speech.

E. Proof-of-Concept: Spectrographic Analysis

In this section, we refer to Fig. 5 for spectrographic analysis of a single utterance contaminated by modulated pink noise

[4]To disambiguate from the term *phase distortion* (PD) discussed in Sections I, II, we use the abbreviation PDev to refer to the *phase deviation* metric.



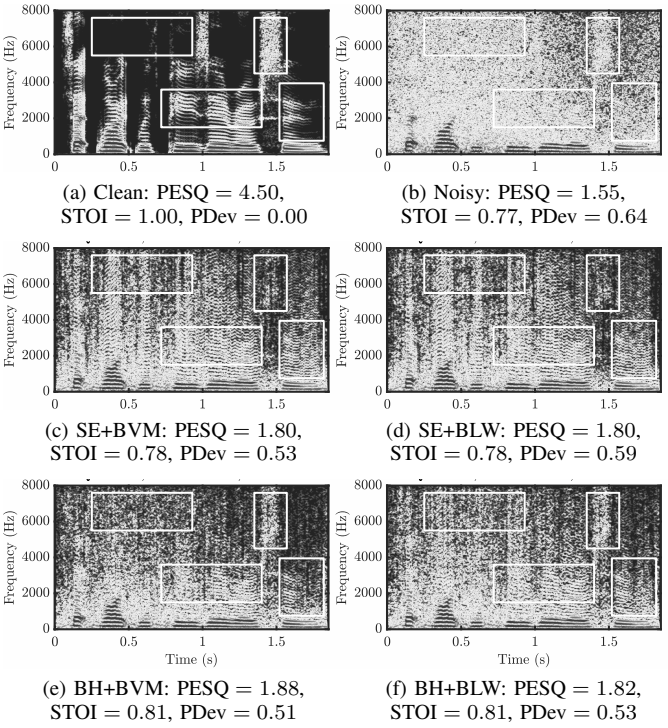
(a) Noisy speech spectrogram with frequency tracks of the 1st, 4th and 5th multiples of pitch obtained using PEFAC

(b) Clean speech spectrogram (presented for reference)

(c) Bi-phase for components on $(1, 4, 5)$ multiples of pitch: unprocessed, enhanced (in reconstructed speech) and clean (presented for reference)

(d) Bi-phase smoothing region estimated from noisy observation

Fig. 4. Typical bi-phase smoothing procedure. Female speaker. The utterance *"Set white with P zero soon"* from the GRID corpus [26] is mixed with modulated pink noise from the NOISEX-92 database [27], SNR $= 5$ dB.



(a) Clean: PESQ $= 4.50$, STOI $= 1.00$, PDev $= 0.00$

(b) Noisy: PESQ $= 1.55$, STOI $= 0.77$, PDev $= 0.64$

(c) SE+BVM: PESQ $= 1.80$, STOI $= 0.78$, PDev $= 0.53$

(d) SE+BLW: PESQ $= 1.80$, STOI $= 0.78$, PDev $= 0.59$

(e) BH+BVM: PESQ $= 1.88$, STOI $= 0.81$, PDev $= 0.51$

(f) BH+BLW: PESQ $= 1.82$, STOI $= 0.81$, PDev $= 0.53$

Fig. 5. Spectrographic analysis of phase-only speech enhancement in blind setup. Female speaker. The utterance *"Set white with P zero soon"* from the GRID corpus [26] is mixed with modulated pink noise from the NOISEX-92 database [27], SNR $= 5$ dB.

with SNR $= 5$ dB and processed by SE and BH algorithms. The phase processing is performed using the smoothing window length $\mathcal{W} = 100$ ms, with reference harmonic phases pre-processed by the TSUP method [12], following the settings of the original implementation of the SE+BVM algorithm [19].

The BH scheme has an advantage over the SE scheme as it does not restore harmonics excessively in regions where they do not exist in the original speech, as observed in the highlighted white rectangular areas in Fig. 5. This aligns with the findings in Section IV-D that the proposed framework restricts bi-phase smoothing in areas where bi-phase trajectories are likely to follow a uniform distribution, reducing the appearance of non-existent harmonics in the enhanced signal. The BH scheme produces less buzzy sound in the enhanced speech, reflected in lower PDev scores compared to the SE scheme.

The BH scheme better preserves the non-harmonic regions present in the clean speech, whereas the SE scheme suppresses them. This can be observed by comparing the signal structure in the 4 000–8 000 Hz range around the 1.5 s mark in Fig. 5. Preserving these regions improves speech intelligibility.

F. Impact of Smoothing Filter Length

The original implementation of the SE+BVM algorithm [19] uses a filter length of 100 ms for bi-phase smoothing. In this section, we evaluate the impact of the smoothing filter length $\mathcal{W}$ (54) on the performance of the proposed algorithms. The choice of $\mathcal{W}$ balances a trade-off between the short-time stationarity of speech parameters (which holds for shorter $\mathcal{W}$) and the removal of unwanted phase fluctuations introduced by noise (which is more effective for longer $\mathcal{W}$). The results are presented in Fig. 6, where each point on a chart represents the average value of the corresponding metric across all considered noise scenarios and all global SNR levels.

The results show that using shorter $\mathcal{W}$ values maintains higher intelligibility improvement compared to longer values, but very short $\mathcal{W}$ (10 ms) restricts the achievable performance in speech quality improvement and noise reduction, making the smoothing framework less efficient. On the other hand, very long $\mathcal{W}$ (100 ms) achieves the best noise reduction score at the cost of both quality and intelligibility, due to the violation of the short-term stationarity assumption. The best reduction in phase deviation is achieved between these extreme values of $\mathcal{W}$. Therefore, we use $\mathcal{W} = 20$ ms for all further experiments in this paper, which matches the settings of the Binary Hypothesis on Individual Harmonics algorithm [13].

G. Impact of Reference Harmonic Phase Pre-Processing

The original SE+BVM algorithm [19] employs TSUP [12] as a reference harmonic phase pre-processing procedure. In this section, we evaluate the impact of different pre-processing strategies on the speech enhancement performance achieved by the BH+BLW algorithm. We selected the BH+BLW algorithm as it achieved the best intelligibility improvement at

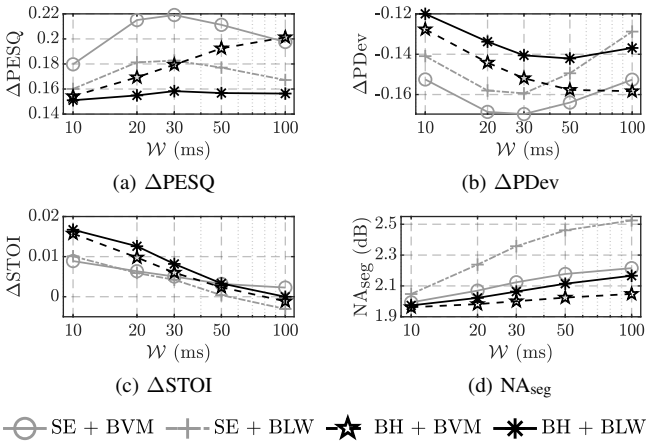


Fig. 6. Impact of smoothing filter length on phase-only enhancement performance.

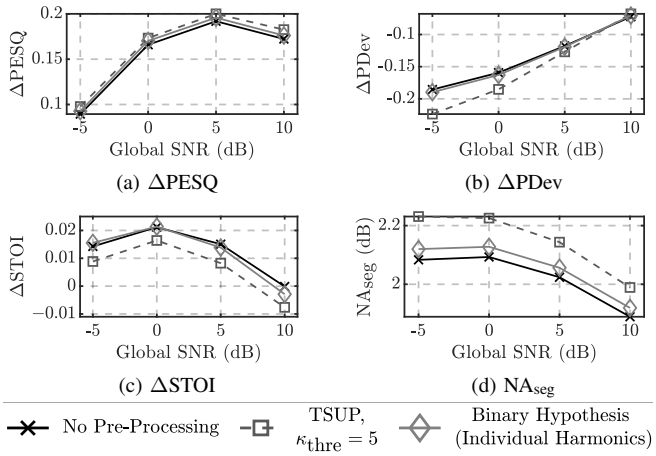


Fig. 7. Impact of reference harmonic phase pre-processing on phase-only enhancement performance of BH+BLW scheme.

$\mathcal{W} = 20$ ms among other algorithms, as shown in Fig. 6c.

Specifically, we evaluate the impact of assigning a processed version of $\Phi_y(1,l)$ as $\hat{\Phi}_x(1,l)$ in Algorithm 3, Line 4. We consider the following pre-processing algorithms: "No Pre-Processing" refers to the direct assignment of $\Phi_y(1,l)$ without processing; "TSUP" refers to pre-processing of $\Phi_y(1,l)$ using TSUP [12], where the phase is smoothed at frames where the estimated von Mises concentration κ exceeds the $\kappa_{\text{thre}} = 5$ threshold; and "Binary Hypothesis (Individual Harmonics)" refers to pre-processing of $\Phi_y(1,l)$ where the phase is smoothed at frames detected by the BH framework on individual harmonics [13]. The results are presented in Fig. 7.

The TSUP pre-processing allows for higher noise attenuation scores as well as some additional improvement in PESQ, however, at the expense of intelligibility loss. The phase deviation is lower, most notably at low SNR levels of -5 and 0 dB, and is in parity with other algorithms at higher SNR levels. "No Pre-Processing" and "Binary Hypothesis (Individual Harmonics)" perform similarly in all metrics.

While reference harmonic phase pre-processing can improve noise reduction and speech quality, keeping the reference phase unprocessed preserves the achieved intelligibility gain. Therefore, we maintain the reference phases unprocessed for all subsequent experiments in this paper.

H. Objective Evaluation of Phase-Only Enhancement Performance

In this section, we evaluate the performance of the proposed BH algorithms against SE benchmarks using the database and evaluation metrics described in Section VII-A. The smoothing window length is set to $\mathcal{W} = 20$ ms and no pre-processing of reference phases is applied. The results are shown in Fig. 8.

At a very low SNR of -5 dB, SE schemes outperform BH schemes in all metrics and noise scenarios, except for ΔSTOI in babble noise. This suggests smoothing the bi-phase trajectory everywhere leads to better results in adverse noise conditions, where the BH detector performance is limited.

Starting from SNR $= 0$ dB, BH schemes tend to demonstrate improved performance in speech intelligibility, and starting from SNR $= 5$ dB — in lowering phase deviation (with the exception of the white noise scenario). When considering ΔPESQ and ΔPDev together, both BH and SE schemes show similar results in lowering ΔPDev at SNR $= 5$ dB, but SE schemes provide higher ΔPESQ at the same time. We attribute this to the increased level of buzziness introduced by SE schemes, as indicated by the increased ΔPESQ at the expense of not lowering the phase deviation.

Averaging over multiple representations of $\hat{\Phi}_x(h,l)$ in the BLW scheme (see Section VI-B) allows for higher levels of intelligibility improvement. This is evident from the ΔSTOI charts for the BH+BVM and BH+BLW schemes, where BH+BLW consistently outperforms BH+BVM.

The performance of SE and BH schemes in noise attenuation is comparable at each SNR level.

Both SE and BH schemes have lower performance in the F_0-estimated setup in babble noise and in the SNR $= -5$ dB case for other noise types compared to the F_0-oracle setup. The performance of the algorithms in other noise types and SNR levels is similar in both F_0-estimated and F_0-oracle setups.

I. Objective Evaluation of Combined Magnitude and Phase Enhancement Performance

In this section, we evaluate the speech enhancement performance of the proposed BH phase enhancement algorithms combined with magnitude enhancement. The magnitude enhancement is done using the minimum mean-square error log-spectral amplitude (MMSE-LSA) estimator [36]. The final enhanced speech is produced by feeding the magnitude-enhanced speech into one of the phase enhancement algorithms. We consider the following algorithm combinations: MMSE-LSA + Unprocessed Phase (lower bound), MMSE-LSA + STFTPI [10], MMSE-LSA + SE+BVM [19], MMSE-LSA + Binary Hypothesis on Individual Harmonics [13], MMSE-LSA + VAD (harmonic phases are smoothed in voiced regions suggested by the magnitude-based voice activity detector from [34]), MMSE-LSA + BH+BVM (proposed), MMSE-LSA + BH+BLW (proposed), and MMSE-LSA + Clean Phase (upper bound). The smoothing window length is set to $\mathcal{W} = 20$ ms.

We use the database described in Section VII-A and present

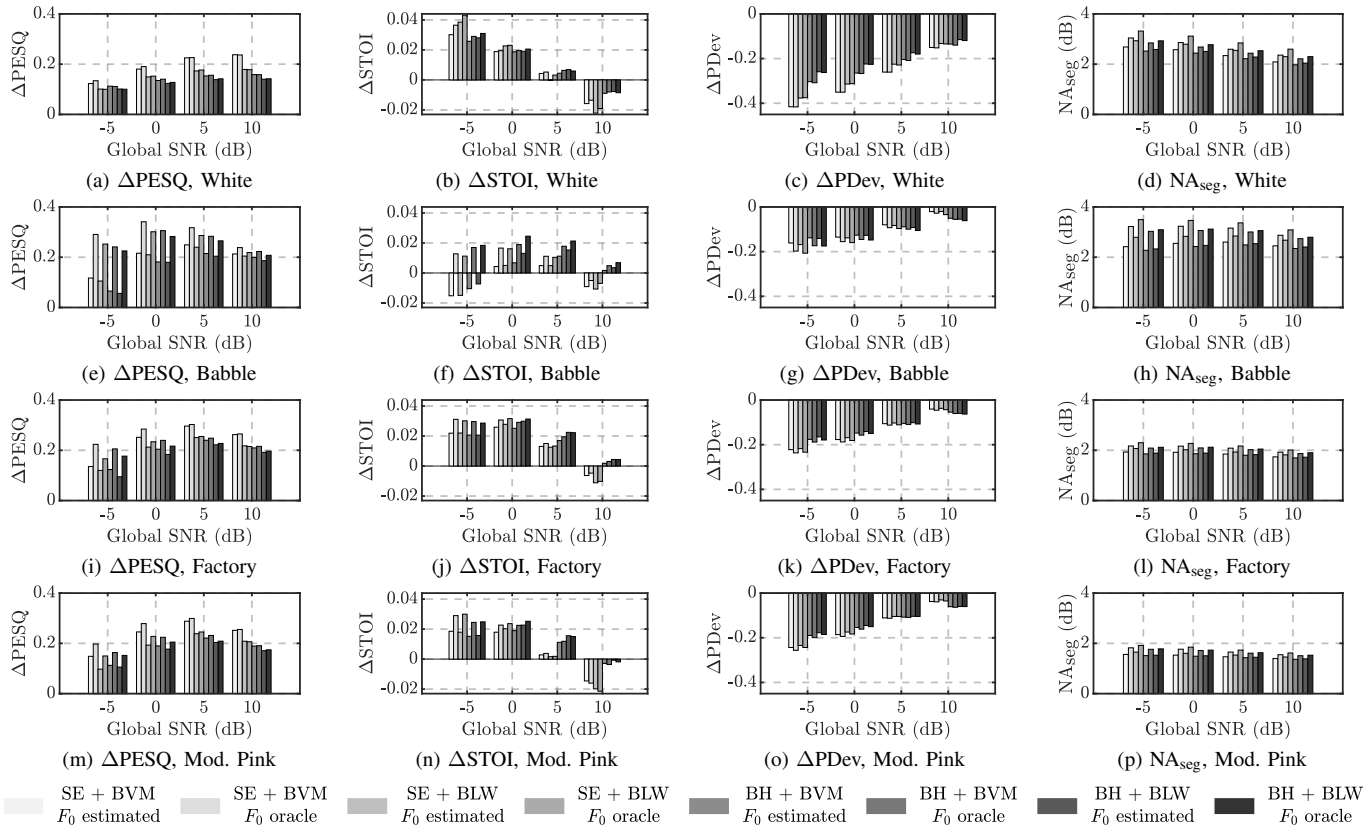


Fig. 8. Objective evaluation of phase-only speech enhancement performance for the F_0-estimated and F_0-oracle scenarios. The SNR ξ, ζ are estimated from the noisy observation.

the mean values of ΔPESQ, ΔSTOI, ΔPDev, and NA_{seg} in Table II averaged over all noise scenarios in the blind setup.

Similar to the results reported in the phase-only enhancement in Section VII-H, at very low SNR $= -5$ dB accurate identification of smoothing regions becomes less efficient compared to smoothing bi-phase everywhere, as SE+BVM attains the best ΔPESQ and ΔPDev while maintaining the same level of ΔSTOI as other algorithms.

Starting from SNR $= 0$ dB, the proposed BH schemes tend to outperform others in terms of reducing ΔPDev. At the same time, the achieved ΔPESQ remains the highest for MMSE-LSA + SE+BVM, which indicates an increased level of buzziness produced by this algorithm compared to the BH schemes, as an increase in ΔPESQ should be accompanied by a lower ΔPDev to prevent buzzing artifacts. In speech quality enhancement, the proposed bi-phase processing schemes outperform MMSE-LSA + BH (Ind. H.) and MMSE-LSA + VAD that process harmonic phases individually, as MMSE-LSA + BH+BVM and MMSE-LSA + BH+BLW achieve higher ΔPESQ and lower ΔPDev jointly at SNR of -5, 0, and 5 dB.

In speech intelligibility enhancement, MMSE-LSA + BH+BLW performs similarly to the top-performing MMSE-LSA + BH (Ind. H.) and MMSE-LSA + VAD. As in the results reported in the phase-only enhancement in Section VII-H, the BLW scheme tends to achieve higher intelligibility improvement compared to the BVM scheme due to averaging over multiple representations of $\hat{\Phi}_x(h,l)$. This is noticeable at SNR $= 0$ dB, where MMSE-LSA + BH+BLW attains ΔSTOI ≈ 0.02 compared to ΔSTOI ≈ 0.01 achieved by MMSE-LSA + BH+BVM. At SNR $= 10$ dB, all considered algorithms on average perform worse than the lower bound.

In noise attenuation, phase processing contributes to additional noise suppression (by 1 to 2 dB on average) compared to the magnitude-only enhancement. MMSE-LSA + STFTPI reaches the highest NA_{seg} due to a more aggressive replacement of noise with excessive harmonics on higher frequencies. Since STFTPI replaces the noisy harmonic phase with the constant harmonic phase at voiced intervals, we can consider this way of processing as harmonic phase smoothing using a very large smoothing window length in voiced intervals. According to Fig. 6d, an increase in the smoothing window length leads to higher levels of noise attenuation at the cost of other metrics, which is consistent with the observed performance of MMSE-LSA + STFTPI.

J. Subjective Listening Tests

In this section, we present the results of the Multi Stimulus test with Hidden Reference and Anchor (MUSHRA) [37].

1) Setup: The MUSHRA experiment consisted of 12 trials, half evaluating phase-only enhancement (P-O) and half evaluating combined magnitude and phase enhancement (M+P). We selected 12 utterances from the databases in Section VII-A and mixed them with 6 noisy conditions consisting of factory, babble, and modulated pink noise with global SNRs of 5 and 10 dB. Each condition was mixed with 2 utterances: one to be used in a P-O trial and another to be used in an M+P trial.

TABLE II
OBJECTIVE EVALUATION OF PHASE-AWARE SPEECH ENHANCEMENT PERFORMANCE COMBINED WITH ENHANCED MAGNITUDE

	MMSE-LSA [36] + Phase	Global SNR					MMSE-LSA [36] + Phase	Global SNR			
		−5 dB	0 dB	5 dB	10 dB			−5 dB	0 dB	5 dB	10 dB
ΔPESQ	Unprocessed Phase	0.23	0.40	0.49	0.50	ΔPDev	Unprocessed Phase	0.00	0.00	0.00	0.00
	STFTPI [10]	0.18	0.37	0.44	0.39		STFTPI [10]	−0.10	0.05	0.26	0.46
	SE+BVM [19]	**0.37**	**0.58**	**0.66**	**0.60**		SE+BVM [19]	**−0.37**	**−0.30**	−0.19	−0.08
	BH (Ind. H.) [13]	0.29	0.49	0.57	0.56		BH (Ind. H.) [13]	−0.33	−0.29	−0.21	**−0.12**
	VAD	0.29	0.49	0.57	0.56		VAD	−0.33	−0.29	−0.21	**−0.12**
	BH+BVM *(prop.)*	0.32	0.53	0.62	0.59		BH+BVM *(prop.)*	−0.35	**−0.30**	**−0.22**	**−0.12**
	BH+BLW *(prop.)*	0.31	0.52	0.60	0.58		BH+BLW *(prop.)*	−0.34	**−0.30**	**−0.22**	**−0.12**
	Clean Phase	0.65	0.77	0.83	0.83		Clean Phase	−0.62	−0.52	−0.39	−0.25
ΔSTOI	Unprocessed Phase	−0.01	0.01	0.01	**0.01**	NA_{seg} (dB)	Unprocessed Phase	11.9	10.9	9.92	9.09
	STFTPI [10]	−0.01	0.00	−0.01	−0.02		STFTPI [10]	**13.7**	**12.8**	**11.7**	**10.8**
	SE+BVM [19]	**0.00**	0.01	0.01	−0.01		SE+BVM [19]	13.1	12.1	11.1	10.2
	BH (Ind. H.) [13]	**0.00**	**0.02**	**0.02**	0.00		BH (Ind. H.) [13]	13.0	12.0	11.0	10.1
	VAD	**0.00**	**0.02**	**0.02**	0.00		VAD	13.0	12.0	11.0	10.1
	BH+BVM *(prop.)*	**0.00**	0.01	0.01	0.00		BH+BVM *(prop.)*	13.0	12.0	11.0	10.1
	BH+BLW *(prop.)*	**0.00**	**0.02**	0.01	0.00		BH+BLW *(prop.)*	13.1	12.0	11.0	10.2
	Clean Phase	0.07	0.08	0.07	0.04		Clean Phase	13.0	12.1	11.1	10.1

Each trial consisted of the reference (clean speech) and 5 other signals (conditions) to be rated against the reference: 1 hidden reference, 1 hidden anchor (a low-pass filtered version of the reference with a cut-off frequency of 3.5 kHz), 2 signals produced by systems under test (SE+BVM and BH+BLW in a P-O trial or MMSE-LSA+SE+BVM and MMSE-LSA+BH+BLW in an M+P trial), and 1 unprocessed noisy signal in a P-O trial or 1 MMSE-LSA magnitude-only enhanced signal in an M+P trial.

We invited 15 non-experienced listeners to volunteer for the MUSHRA experiment. At each trial, the listener was asked to subjectively evaluate the speech quality of each condition on a scale from 0 to 100, with the GUI provided by [38]. The listening experiment was conducted in a quiet room on a MacBook Pro (16-inch, 2021) laptop with the AKG K-240 Studio headphones, connected to the laptop audio output using the 3.5 mm headphone jack.

In the post-screening phase of the experiment, in accordance with ITU recommendation [37, Section 4.1.2], we excluded data from one listener who rated the hidden reference condition lower than a score of 90 for more than 15% of the test items. Among the remaining 14 listeners, whose data is analyzed further, one reported suffering from tinnitus occasionally, another reported reduced hearing due to age, and the other 12 did not report hearing impairments.

2) Analysis: Recent studies [39] indicate that MUSHRA data deviate from normality, are not independent, and are in ordinal scales. Therefore, non-parametric statistical tests are advised to analyze MUSHRA data. We report the statistical analysis results using the non-parametric Friedman test. In the Friedman test, the set of MUSHRA scores obtained in each trial by each listener is sorted in ascending order to obtain their ranks. The lowest score corresponds to rank 1, and the highest score corresponds to rank 7, which is the total number of conditions we evaluate. Finally, a test statistic is calculated from these ranks to obtain *p*-values. We use the Tukey-Kramer correction of *p*-values to reduce the probability of Type I error (false positives) for multiple comparisons.

Additionally, as suggested in [40], we supplement the *p*-value analysis with the measure of effect size, Cliff's delta:

$$d = \frac{\sum_{i=1}^{m}\sum_{j=1}^{n}[x_i > y_j] - \sum_{i=1}^{m}\sum_{j=1}^{n}[x_i < y_j]}{mn}, \quad (65)$$

where x_i and y_j are the observations of the samples of sizes m and n to be compared, and $[P]$ is the Iverson bracket, which is 1 if P is true and 0 otherwise. As in [40], we consider the effect size to be small if $0.11 \leq |d| < 0.28$, medium if $0.28 \leq |d| < 0.43$, and large if $|d| \geq 0.43$.

3) Results: Fig. 9 depicts the mean ranks of MUSHRA scores with 95% confidence intervals calculated from the Friedman test statistics. The mean MUSHRA scores, p-values and effect sizes are presented in Tables III and IV.

In phase-only enhancement, the listeners ranked SE+BVM slightly lower than the noisy speech, and BH+BLW slightly higher. This is consistent with the results of objective evaluation from Section VII-H, where for factory, babble, and modulated pink noise with SNR of 5 and 10 dB, SE+BVM outperforms BH+BLW in PESQ only without lowering PDev, thus leading to an increased level of buzziness, and a lower level of intelligibility. The SE+BVM to BH+BLW difference ($p = 0.651$) has an indistinguishable effect size $|d| = 0.08$.

In combined magnitude and phase enhancement, the listeners ranked MMSE-LSA+BH+BLW higher than both MMSE-LSA and MMSE-LSA+SE+BVM. Compared to magnitude-only enhanced speech (MMSE-LSA), MMSE-LSA+BH+BLW has a lower p-value ($p = 0.129$) and a bigger effect size ($|d| = 0.12$, small effect) than by comparing MMSE-LSA to MMSE-LSA+SE+BVM ($p = 0.805$, and an indistinguishable effect size $|d| = 0.04$). Both phase processing methods reduce the level of musical noise introduced by MMSE-LSA, thus listeners ranked both higher than MMSE-LSA. However, MMSE-LSA+BH+BLW receives the highest rank as

TABLE III
MEAN SCORES FOR THE MUSHRA EXPERIMENT

	Noisy	44.6
Phase-Only Enhancement	SE+BVM [19]	43.1
	BH+BLW *(prop.)*	46.1
Magnitude-Only Enhancement	MMSE-LSA [36]	53.5
Magnitude+Phase Enhancement	MMSE-LSA+SE+BVM [19]	54.8
	MMSE-LSA+BH+BLW *(prop.)*	57.8
	Clean	99.4

TABLE IV
p-VALUES AND EFFECT SIZES FOR THE MUSHRA EXPERIMENT

Condition	p-value	$\|d\|$ (effect size)
Phase-Only Enhancement		
Noisy — SE+BVM	0.984	0.03
Noisy — BH+BLW	0.982	0.04
SE+BVM — BH+BLW	0.651	0.08
Magnitude-Only and Magnitude+Phase Enhancement		
Noisy — MMSE-LSA	2.288×10^{-2}	0.24 (small)
Noisy — MMSE-LSA+SE+BVM	8.642×10^{-5}	0.26 (small)
Noisy — MMSE-LSA+BH+BLW	1.776×10^{-7}	0.34 (medium)
MMSE-LSA — MMSE-LSA+SE+BVM	0.805	0.04
MMSE-LSA — MMSE-LSA+BH+BLW	0.129	0.12 (small)

the buzziness artifact is not introduced in this method, which is consistent with the observations made in Section VII-I.

Among all the considered systems under test in the MUSHRA experiment, MMSE-LSA+BH+BLW achieves the highest score from the listeners (57.8 vs. 44.6 for noisy speech). Compared to noisy speech, MMSE-LSA+BH+BLW has the lowest $p = 1.776 \times 10^{-7}$ and is the only system under test that demonstrated an effect of a medium size, $|d| = 0.34$.

In [41], the reader can find listening examples, as well as references to the complete set of audio files used for the MUSHRA experiment, the data collected from 14 listeners, and the MATLAB script used for statistical analysis.

VIII. CONCLUSION

In this paper, we proposed a novel harmonic phase reconstruction framework for speech enhancement. The framework performs smoothing of the inter-component phase, captured by the bi-phase, to lower the harmonic phase variance introduced by noise. The novelty of this work is in the derivation of the *a priori* distribution of the bi-phase and its incorporation into the SNR-dependent detector of bi-phase smoothing regions.

We compared the performance of our proposed framework with a previously developed inter-component phase enhancement framework that does not formulate an SNR-dependent smoothing threshold and demonstrated an improvement in the reconstruction of speech spectral structure. The reconstructed speech does not contain excessive high-frequency harmonics and does not suffer from auditory buzziness. An objective evaluation of our proposed framework showed superior speech intelligibility enhancement and reduced phase deviation compared to the previously developed framework.

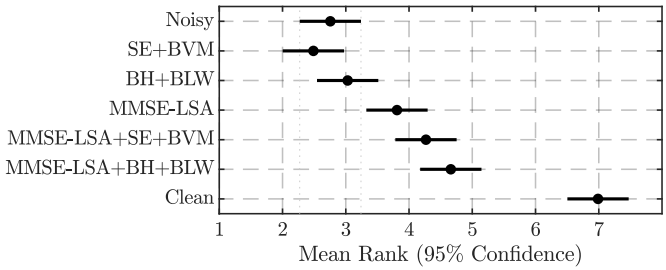


Fig. 9. Mean ranks of MUSHRA scores by Friedman test.

Like its predecessors, the proposed framework still requires the estimation of the fundamental frequency F_0. However, inter-component phase smoothing regions can be computed without prior knowledge of F_0 as the inter-component phase is blind to linear phase. Future design of spectral phase estimators based on inter-component phase estimates should aim for a framework not requiring an F_0 estimate.

APPENDIX A
DERIVATIONS OF BI-PHASE USING ICPR EXPRESSION

For the case $h_1 \neq h_2$, the bi-phase (5) takes into account the phase relation between three components. Therefore, we set $P = 3$, with $H(1) = h_1$, $H(2) = h_2$, $H(3) = h_3$, where $h_1, h_2, h_3 = h_1 + h_2$ are defined according to (6) with the exclusion of the case $h_2 = h_1$. Setting $K(1) = K(2) = K(3) = 1$ and $S(1) = 1$, $S(2) = 1$, $S(3) = -1$ will satisfy (2b). Substituting these settings into (2a), we get:

$$\begin{aligned}\Theta(n) = \sum_{p=1}^{3} S(p)K(p)\Psi_x(H(p), n) = \Psi_x(h_1, n) \\ + \Psi_x(h_2, n) - \Psi_x(h_3, n) = \Theta_3(h_1, h_2, h_3, n).\end{aligned} \tag{66}$$

For the case $h_1 = h_2$, the bi-phase (7) takes into account the phase relation between two components, therefore we set $P = 2$, and $H(1) = h_1$, $H(2) = 2h_1$. Setting $K(1) = 2$, $K(2) = 1$ and $S(1) = 1$, $S(2) = -1$ will satisfy (2b). Substituting these settings into (2a), we get:

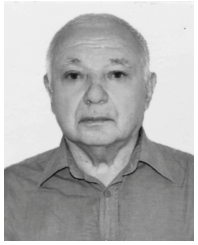
$$\begin{aligned}\Theta(n) = \sum_{p=1}^{2} S(p)K(p)\Psi_x(H(p), n) = 2\Psi_x(h_1, n) \\ - \Psi_x(2h_1, n) = \Theta_2(h_1, h_2 = h_1, h_3 = h_1 + h_2, n).\end{aligned} \tag{67}$$

REFERENCES

[1] D. Wang and J. Lim, "The unimportance of phase in speech enhancement," *IEEE Trans. Acoust., Speech, Signal Process.*, 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 Trans. Acoust., Speech, Signal Process.*, vol. 32, no. 6, pp. 1109–1121, Dec. 1984.
[3] A. V. Oppenheim and J. S. Lim, "The importance of phase in signals," *Proc. of IEEE*, vol. 69, no. 5, pp. 529–541, May 1981.
[4] K. K. Paliwal, K. K. Wójcicki, and B. J. Shannon, "The importance of phase in speech enhancement," *Speech Commun.*, vol. 53, no. 4, pp. 465–494, Apr. 2011.
[5] P. Mowlaee, R. Saeidi, and Y. Stylianou, "Advances in phase-aware signal processing in speech communication," *Speech Commun.*, vol. 81, pp. 1–29, 2016.

[6] D. Griffin and J. Lim, "Signal estimation from modified short-time fourier transform," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 32, no. 2, pp. 236–243, Apr 1984.
[7] P. Mowlaee, R. Saiedi, and R. Martin, "Phase estimation for signal reconstruction in single-channel speech separation," in *Proc. ISCA Interspeech*, Sept. 2012, pp. 1548–1551.
[8] P. Mowlaee and R. Saeidi, "Time-frequency constraints for phase estimation in single-channel speech enhancement," in *Proc. Int. Worksh. Acoust. Signal Enhanc.*, 2014, pp. 338–342.
[9] I. Saratxaga, I. Hernaez, D. Erro, E. Navas, and J. Sanchez, "Simple representation of signal phase for harmonic speech models," *Electr. Lett.*, vol. 45, no. 7, pp. 381–383, March 2009.
[10] M. Krawczyk and T. Gerkmann, "STFT phase reconstruction in voiced speech for an improved single-channel speech enhancement," *IEEE/ACM Trans. Audio, Speech, Lang. Process.*, vol. 22, no. 12, pp. 1931–1940, Dec. 2014.
[11] J. Kulmer and P. Mowlaee, "Harmonic phase estimation in single-channel speech enhancement using von Mises distribution and prior SNR," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, Apr. 2015, pp. 5063–5067.
[12] ——, "Phase estimation in single channel speech enhancement using phase decomposition," *IEEE Signal Process. Lett.*, vol. 22, no. 5, pp. 598–602, May. 2015.
[13] P. Mowlaee and J. Kulmer, "Harmonic phase estimation in single-channel speech enhancement using phase decomposition and SNR information," *IEEE/ACM Trans. Audio, Speech, Lang. Process.*, vol. 23, no. 9, pp. 1521–1532, Sept. 2015.
[14] N. Zheng and X.-L. Zhang, "Phase-aware speech enhancement based on deep neural networks," *IEEE/ACM Trans. Audio, Speech, Lang. Process.*, vol. 27, no. 1, pp. 63–76, 2019.
[15] N. B. Thien, Y. Wakabayashi, K. Iwai, and T. Nishiura, "Inter-frequency phase difference for phase reconstruction using deep neural networks and maximum likelihood," *IEEE/ACM Trans. Audio, Speech, Lang. Process.*, vol. 31, pp. 1667–1680, 2023.
[16] Y. Masuyama, K. Yatabe, K. Nagatomo, and Y. Oikawa, "Online phase reconstruction via DNN-based phase differences estimation," *IEEE/ACM Trans. Audio, Speech, Lang. Process.*, vol. 31, pp. 163–176, 2023.
[17] T. Peer and T. Gerkmann, "Phase-aware deep speech enhancement: It's all about the frame length," *JASA Exp. Lett.*, vol. 2, no. 10, p. 104802, 2022.
[18] V. I. Vorobiov, D. A. Kechik, and S. Y. Barysenka, "Inter-component phase processing of quasipolyharmonic signals," *Appl. Acoust.*, vol. 177, p. 107937, 2021.
[19] S. Y. Barysenka, V. I. Vorobiov, and P. Mowlaee, "Single-channel speech enhancement using inter-component phase relations," *Speech Commun.*, vol. 99, pp. 144–160, 2018.
[20] Y. Wakabayashi, T. Fukumori, M. Nakayama, T. Nishiura, and Y. Yamashita, "Single-channel speech enhancement with phase reconstruction based on phase distortion averaging," *IEEE/ACM Trans. Audio, Speech, Lang. Process.*, vol. 26, no. 9, pp. 1559–1569, 2018.
[21] G. Degottex and D. Erro, "A uniform phase representation for the harmonic model in speech synthesis applications," *EURASIP J. on Audio, Speech, and Music Process.*, no. 1, p. 38, Oct. 2014.
[22] C. L. Nikias and A. P. Petropulu, *Higher-order spectra analysis: a nonlinear signal processing framework*. Upper Saddle River, New Jersey: PTR Prentice Hall, 1993.
[23] K. V. Mardia and P. E. Jupp, *Directional Statistics*, ser. Wiley Series in Probability and Statistics. Wiley, 2000.
[24] S. M. Kay, *Fundamentals of Statistical Signal Processing, Volume II: Detection Theory*. Prentice Hall, 1998.
[25] H. Bartelt, A. W. Lohmann, and B. Wirnitzer, "Phase and amplitude recovery from bispectra," *Appl. Opt.*, vol. 23, no. 18, pp. 3121–3129, Sep 1984.
[26] M. Cooke, J. Barker, S. Cunningham, and X. Shao, "An audio-visual corpus for speech perception and automatic speech recognition," *J. Acoust. Soc. Am.*, vol. 120, p. 2421, 2006.
[27] A. Varga and H. J. Steeneken, "Assessment for automatic speech recognition: II. NOISEX-92: A database and an experiment to study the effect of additive noise on speech recognition systems," *Speech Commun.*, vol. 12, no. 3, pp. 247–251, 1993.
[28] S. Elshamy, N. Madhu, W. Tirry, and T. Fingscheidt, "DNN-supported speech enhancement with cepstral estimation of both excitation and envelope," *IEEE/ACM Trans. Audio, Speech, Lang. Process.*, vol. 26, no. 12, pp. 2460–2474, 2018.
[29] A. W. Rix, J. G. Beerends, M. P. Hollier, and A. P. Hekstra, "Perceptual evaluation of speech quality (PESQ) – A new method for speech quality assessment of telephone networks and codecs," *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, vol. 2, pp. 749–752, Aug. 2001.
[30] C. H. Taal, R. C. Hendriks, R. Heusdens, and J. Jensen, "An algorithm for intelligibility prediction of time-frequency weighted noisy speech," *IEEE Trans. Audio, Speech, Lang. Process.*, vol. 19, no. 7, pp. 2125–2136, Sept 2011.
[31] A. Gaich and P. Mowlaee, "On speech quality estimation of phase-aware single-channel speech enhancement," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, Apr. 2015, pp. 216–220.
[32] P. Mowlaee, J. Kulmer, J. Stahl, and F. Mayer, *Phase-Aware Signal Processing in Speech Communication: History, Theory and Practice*. John Wiley & Sons, 2016.
[33] T. Fingscheidt, S. Suhadi, and S. Stan, "Environment-optimized speech enhancement," *IEEE Trans. Audio, Speech, Lang. Process.*, vol. 16, no. 4, pp. 825–834, 2008.
[34] S. Gonzalez and M. Brookes, "PEFAC – A pitch estimation algorithm robust to high levels of noise," *IEEE/ACM Trans. Audio, Speech, Lang. Process.*, vol. 22, no. 2, pp. 518–530, Feb 2014.
[35] M. Brookes, "VOICEBOX: Speech processing toolbox for MATLAB," https://github.com/ImperialCollegeLondon/sap-voicebox, 2021.
[36] Y. Ephraim and D. Malah, "Speech enhancement using a minimum mean-square error log-spectral amplitude estimator," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 33, pp. 443–445, Apr. 1985.
[37] "ITU-R BS.1534-3. Method for the subjective assessment of intermediate quality level of audio systems – BS Series: Broadcasting service (sound)," Oct. 2015.
[38] M. Schoeffler, S. Bartoschek, F.-R. Stöter, M. Roess, S. Westphal, B. Edler, and J. Herre, "webMUSHRA – A comprehensive framework for web-based listening tests," *J. of Open Research Software*, vol. 6, p. 8, Feb 2018.
[39] C. Mendonça and S. Delikaris-Manias, "Statistical tests with MUSHRA data," in *Audio Eng. Soc. Conv.*, vol. 144, May 2018.
[40] D. Michelsanti, Z.-H. Tan, S. Sigurdsson, and J. Jensen, "Deep-learning-based audio-visual speech enhancement in presence of Lombard effect," *Speech Commun.*, vol. 115, pp. 38–50, 2019.
[41] S. Y. Barysenka and V. I. Vorobiov. SNR-based inter-component phase estimation using bi-phase prior statistics for single-channel speech enhancement: Listening examples. [Online]. Available: https://siarheibarysenka.github.io/inter-component-phase-speech-enhancement-demo/

Siarhei Y. Barysenka (S'15–M'20) received his B.Eng. degree in electrical engineering (with honors) in 2012 and M.Sc. degree in 2013, both from the Belarusian State University of Informatics and Radioelectronics in Minsk, Belarus. He has been working in the software engineering industry for mobile operating systems since 2012. Since 2019, he has been a Principal iOS Engineer at Atlassian B.V. in Amsterdam, The Netherlands. His research interests include phase-aware speech enhancement and speech parameters estimation.



Vasili I. Vorobiov received his Ph.D. in radar and radionavigation systems from the Moscow Power Engineering Institute in the USSR in 1969. During the 1970s and 1980s, he was a member of three Atlantic and three Pacific expeditions on research vessels of the USSR Academy of Sciences. Until 2021, he was a member of the research staff at the Belarusian State University of Informatics and Radioelectronics in Minsk, Belarus. He is the author of over 20 publications on radar, underwater acoustics, and speech acoustics.