Phase Reconstruction for Artificial Bandwidth Extension toward Musical Instrument Sound Signal

Yuya HOSODA[1†], Arata KAWAMURA[2], Youji IIGUNI[1]
[1] Graduate School of Engineering Science, Oska University, Japan
[2]Faculty of Information Science and Engineering, Kyoto Sangyo University, Japan
[†]hosoda@sip.sys.es.osaka-u.ac.jp

***Abstract*—We propose a phase reconstruction method for an artificial bandwidth extension (ABE) approach toward a musical instrument sound signal. The ABE approach reconstructs the missing upper bandwidth of 3400–8000 Hz (UB) from the existing narrow bandwidth of 300–3400 Hz (NB). The proposed method reconstructs the missing UB phase spectrum based on the short-time Fourier Transform phase improvement (STFTPI) approach that emphasizes a harmonic structure. The STFTPI approach is sensitive to the slight fundamental frequency estimation error, and thus we correct the estimated fundamental frequency using the existing NB phase difference. Besides, the proposed method discriminates whether the UB harmonic should be emphasized using the relation between the amplitude spectrum of the UB harmonic and that of the frequencies adjacent to the UB harmonic. The Experimental results show that the proposed method enhances the audio signal quality to accurately reconstruct the missing UB harmonic structure. The code of the proposed method is available at https://github.com/Yuya-Hosoda/Works.**

I. Introduction

Recording audio archives before the 1980s may suffer from audio quality deterioration due to artifacts, such as noise, packet loss, clipping, audio coding technique, and the bandwidth limitation with a low sampling rate. Researchers have proposed audio enhancement methods, such as noise reduction [1], audio inpainting [2], declipping [3], and audio restoration [4]. In this paper, we address the audio enhancement methods for the bandwidth limitation with a low sampling rate.

An artificial bandwidth extension (ABE) approach is an audio enhancement method that reconstructs the missing upper bandwidth of 3400–8000 Hz (UB) from the existing narrow bandwidth of 300–3400 Hz (NB). For a speech signal, the ABE approach using a source-filter model is well established [5], consisting of the UB excitation signal extension and the UB spectral envelope estimation. Whereas, musical instrument sound signals, such as brass instruments and woodwind instruments, are not easy to estimate the UB spectral envelope due to the complex structure [6]. The ABE approach using a spectral band replication (SBR) [7] [8] successfully reproduces the UB spectral envelope but needs to code the gain as side information in advance, which is not available for the recording audio archives. Researchers have devised the ABE approach that directly estimates the missing UB frequency components without side information [9]– [11]. While the UB amplitude spectrum is estimated using database methods, such as Deep Neural Networks [9], Convolution Neural Networks [10], and Time-Frequency Networks [11], the UB phase spectrum is generated using the simple method of duplicating the flipped NB phase spectrum.

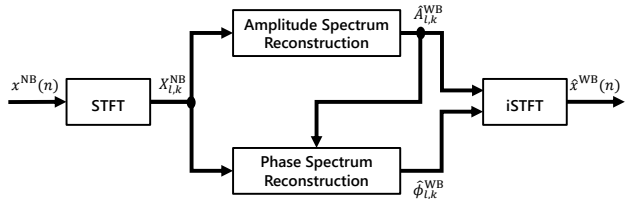

Fig. 1. Block diagram of the ABE framework.

Recent studies have suggested that the phase spectrum plays a vital role in audio signal processing [12] [13]. Also, the phase reconstruction method contributes to the ABE performance improvement [20]. Many phase reconstruction methods have been devised, such as iterative algorithms [14], geometry-based methods [15], machine learning [19], sinusoidal models [16]– [18]. The phase reconstruction methods using sinusoidal models assume that a signal consists of multiple sinusoidal waves with a harmonic structure and emphasize the harmonics with a short calculation time. Since the musical instrument sound signal has the UB harmonics, the ABE approach with the phase reconstruction method using sinusoidal models will enhance the audio signal quality more efficiently. However, the phase reconstruction methods using sinusoidal models are sensitive to the slight fundamental frequency estimation error, resulting in the significant phase spectrum reconstruction error. And, while the harmonics have attenuated as the frequency bandwidth is high, overemphasizing the UB harmonics may generate mechanical and buzzing noise [18].

This paper proposes a phase reconstruction method for the ABE approach toward a musical instrument sound signal. The proposed method uses the short-time Fourier transform phase improvement (STFTPI) approach [17] [18] as the baseline method but has the following advantages. First, we correct the fundamental frequency estimation error using the existing NB phase difference based on the sinusoidal model. Second, the proposed method discriminates whether the UB harmonics should be emphasized using the relation between the amplitude spectrum of the harmonic and that of the frequencies adjacent to the harmonic. Experiments show that the proposed method successfully reconstructs the harmonic structure and enhances the audio signal quality.

II. Artificial Bandwidth Extension

Figure 1 shows the ABE framework. The input is an NB audio signal at a sampling rate F_{s} of 8 kHz, and the output is a bandwidth extended audio signal on the wide bandwidth (WB) of 300–8000 Hz with $F_{\mathrm{s}} = 16$ kHz. Let $x^{\mathrm{NB}}(n)$ be an up-sampled NB audio signal with 16 kHz sample indices n. Given a frame length N, Short-Time Fourier Transform (STFT) at l-th frame $X_l^{\mathrm{NB}}(k)$ is defined as

$$\begin{aligned} X_{l,k}^{\mathrm{NB}} &= \sum_{n=0}^{N-1} x^{\mathrm{NB}}(lL+n)w(n)\mathrm{e}^{-j\frac{2\pi n}{N}k} \\ &= A_{l,k}^{\mathrm{NB}} \cdot \mathrm{e}^{j\phi_{l,k}^{\mathrm{NB}}}, \end{aligned} \tag{1}$$

with a frame shift L, a window function $w(n)$, frequency indices $k(k = 0, 1, \ldots, N-1)$, and $j = \sqrt{-1}$. Also, $A_{l,k}^{\mathrm{NB}}$ (≥ 0) and $\phi_{l,k}^{\mathrm{NB}}$ denote an NB amplitude spectrum and an NB phase spectrum, respectively. The ABE approach reconstructs a WB amplitude spectrum $\hat{A}_{l,k}^{\mathrm{WB}}$ and a WB phase spectrum $\hat{\phi}_{l,k}^{\mathrm{WB}}$ using $A_{l,k}^{\mathrm{NB}}$ and $\phi_{l,k}^{\mathrm{NB}}$. A bandwidth extended WB audio signal $\hat{x}^{\mathrm{WB}}(n)$ is generated using inverse STFT (iSTFT) following as

$$\hat{x}^{\mathrm{WB}}(n) = \frac{1}{N}\sum_{l}\sum_{k=0}^{N-1} w_s(n-lL)\hat{X}_{l,k}^{\mathrm{WB}}\mathrm{e}^{j\frac{2\pi k}{N}(n-ll)}, \tag{2}$$

with

$$\hat{X}_{l,k}^{\mathrm{WB}} = \hat{A}_{l,k}^{\mathrm{WB}} \cdot \mathrm{e}^{j\hat{\phi}_{l,k}^{\mathrm{WB}}}, \tag{3}$$

$$1 = \sum_{l} w(n-lL)w_s(n-lL). \tag{4}$$

In this paper, we reconstruct $\hat{\phi}_{l,k}^{\mathrm{WB}}$ using the STFTPI approach.

III. Short-Time Fourier Transform Phase Improvement

The STFTPI approach [17] [18] assumes that a signal has a harmonic structure consisting of multiple sinusoidal waves and that the fundamental frequency changes little over time. Given a fundamental frequency $f0_l$, the frequency index for the h-th harmonic k_l^h is defined as

$$k_l^h = \arg\min_k |k - \kappa_l^h|, \tag{5}$$

$$\kappa_l^h = \frac{h \cdot f0_l}{F_{\mathrm{s}}}N, \tag{6}$$

where κ_l^h is a non-integer frequency value on the frequency bin scale, and $|\cdot|$ denotes the absolute operator. The phase spectrum of the h-th harmonic $\tilde{\phi}_{l,k_l^h}^{\mathrm{WB}}$ and the phase spectrum of the frequencies adjacent to the h-th harmonic $\tilde{\phi}_{l,k_l^h+\delta}^{\mathrm{WB}}$ are then given as

$$\tilde{\phi}_{l,k_l^h}^{\mathrm{WB}} = \tilde{\phi}_{l,k_{l-1}^h}^{\mathrm{WB}} + 2\pi\frac{h \cdot f0_l}{F_{\mathrm{s}}}L, \tag{7}$$

$$\tilde{\phi}_{l,k_l^h+\delta_l}^{\mathrm{WB}} = \tilde{\phi}_{l,k_l^h}^{\mathrm{WB}} - \phi_{k_l^h-\kappa_l^h}^{W} + \phi_{k_l^h-\kappa_l^h+\delta_l}^{W}, \tag{8}$$

where ϕ_{ω}^{W} denotes the phase property of the window function with a frequency ω via discrete-time Fourier transform, and $\delta_l \in [-\kappa_l^1, \kappa_l^1]$ denotes the distance on the frequency bin scale between the harmonic and the frequency adjacent to the harmonic [17] [18].

The STFTPI approach obtains the fundamental frequency from the fundamental frequency estimation method. Let $f0_l'$ be an estimated fundamental frequency. From Eq.(7), the fundamental frequency estimation error results in the phase spectrum error $2\pi\frac{h\cdot|f0_l - f0_l'|}{F_{\mathrm{s}}}L$, which becomes larger as the harmonic is higher. Besides, since the harmonic have attenuated at higher frequency bandwidth, the STFTPI approach for the ABE approach needs to avoid overemphasizing the UB harmonic.

IV. Proposed Algorithms

First, the proposed method corrects the fundamental frequency error using the NB phase differences based on the sinusoidal model. Second, we discriminate whether the UB harmonics should be emphasized using the relation between the amplitude spectrum of the harmonic and that of the frequencies adjacent to the harmonic.

A. Fundamental Frequency Estimation Error Correction

The proposed method assumes that the existing NB has a correct harmonic structure. We define $\Phi_l^h = \phi_{l,k_l^{h+1}}^{\mathrm{NB}} - \phi_{l,k_l^h}^{\mathrm{NB}}$ $(h = 1, \ldots, H^{\mathrm{NB}} - 1)$ as the NB phase difference between the h-th harmonic and $(h-1)$-th harmonic, where H^{NB} denotes the number of the NB harmonics. From Eq.(7), the NB phase difference represents the fundamental frequency as

$$f0_l^h = \frac{(\Phi_l^h - \Phi_{l-1}^h)F_{\mathrm{s}}}{2\pi L}. \tag{9}$$

However, Eq.(9) is not easy to solve because of the phase wrapping problem. In this paper, we assume that the estimated fundamental frequency is not great different from the original fundamental frequency. The proposed method calculates the fundamental frequency as $\tilde{f0}_l^h = f0_l^h(m^*)$ with

$$m^* = \arg\min_m |f0_l' - f0_l^h(m)|, \tag{10}$$

$$f0_l^h(m) = \frac{\mathrm{mod}(\Phi_l^h - \Phi_{l-1}^h, 2\pi)\cdot F_{\mathrm{s}}}{2\pi L} + m\cdot\frac{F_{\mathrm{s}}}{L}, \tag{11}$$

where $\mathrm{mod}\,(\cdot, 2\pi)$ denotes the phase wrapping processing by the modulo operator, and m is a non-negative integer.

The NB phase spectrum obtained by the STFT may not satisfy the sinusoidal model if the frequency resolution is insufficient, and the fundamental frequency estimation error correction using the NB phase difference will give outliers. The proposed method thus corrects the fundamental frequency within a threshold. Let $\breve{f0}_l$ be the median among the fundamental frequencies obtained by the NB phase difference. Given a threshold F, we get the corrected fundamental frequency as

$$\hat{f0}_l = \begin{cases} \breve{f0}_l & |\breve{f0}_l - f0'_l| \leq F \\ f0'_l & |\breve{f0}_l - f0'_l| > F \end{cases}. \tag{12}$$

B. Harmonic Discrimination

To avoid overemphasizing the UB harmonic, the proposed method generates the UB phase spectrum using the STFTPI approach only when the UB harmonic structure exists. Figure 2 shows the NB and UB amplitude spectra for an oboe signal with the fundamental frequency of 375 Hz. While the NB amplitude spectrum peak corresponds to the harmonic, the UB amplitude spectrum of the harmonic is almost equivalent to that of the frequencies adjacent to the harmonic, and no harmonic structure exists. We, therefore, discriminate whether the harmonic should be emphasized using the relation between the amplitude spectrum of the harmonic and that of the frequencies adjacent to the harmonic.

First, we average the amplitude spectrum of the frequencies adjacent to the harmonic as

$$\bar{A}_l^h = \frac{1}{k_l^1}\Bigg(\sum_{k=k_l^h-\lfloor k_l^1/2\rfloor}^{k_l^h+\lfloor k_l^1/2\rfloor} \hat{A}_{l,k}^{\mathrm{WB}} - \hat{A}_{l,k_l^h}^{\mathrm{WB}}\Bigg), \tag{13}$$

with the floor function $\lfloor\cdot\rfloor$. Given a threshold P, the proposed method defines a binary variable

$$c_l^h = \begin{cases} 1 & \hat{A}_{l,k_l^h}^{\mathrm{WB}}/\bar{A}_l^h \geq P \\ 0 & \hat{A}_{l,k_l^h}^{\mathrm{WB}}/\bar{A}_l^h < P \end{cases}, \tag{14}$$

where $c_l^h = 1$ denotes the case where the h-th harmonic should be emphasized, and vice versa.

C. Phase Spectrum Reconstruction

The proposed method reconstructs the WB phase spectrum with the corrected fundamental frequency and the binary variable. When the h-th harmonic should not be emphasized ($c_l^h = 0$), the WB phase spectrum is calculated with the method of duplicating the flipped NB phase spectrum as well as the traditional ABE approach [9]. The flipped NB phase spectrum $\phi_{l,k}^{\mathrm{Flip}}$ is given as

$$\phi_{l,k}^{\mathrm{Flip}} = \begin{cases} \phi_{l,k}^{\mathrm{NB}} & k = 0,\ldots,N/4 \\ -\phi_{l,N/2-k}^{\mathrm{NB}} & k = N/4+1,\ldots,N/2 \end{cases}. \tag{15}$$

Equations (7) and (8) are then reformulated to get

$$\tilde{\phi}_{l,k_l^h}^{\mathrm{WB}} = \begin{cases} \tilde{\phi}_{l,k_{l-1}^h}^{\mathrm{WB}} + 2\pi\frac{h\cdot\hat{f0}_l}{F_{\mathrm{s}}}L & c_l^h = 1 \\ \phi_{l,k_l^h}^{\mathrm{Flip}} & c_l^h = 0 \end{cases}, \tag{16}$$

$$\tilde{\phi}_{l,k_l^h+\delta_l}^{\mathrm{WB}} = \begin{cases} \tilde{\phi}_{l,k_l^h}^{\mathrm{WB}} - \phi_{k_l^h-\hat{\kappa}_l^h}^{W} + \phi_{k_l^h-\hat{\kappa}_l^h+\delta_l}^{W} & c_l^h = 1 \\ \phi_{l,k_l^h}^{\mathrm{Flip}} & c_l^h = 0 \end{cases}, \tag{17}$$

with

$$\hat{\kappa}_l^h = \frac{h\cdot\hat{f0}_l}{F_{\mathrm{s}}}N. \tag{18}$$

Consequently, we get $\hat{\phi}_{l,k}^{\mathrm{WB}}$ as

$$\hat{\phi}_{l,k}^{\mathrm{WB}} = \begin{cases} \phi_{l,k}^{\mathrm{NB}} & k = 0,\ldots,N/4 \\ \tilde{\phi}_{l,k}^{\mathrm{WB}} & k = N/4+1,\ldots,N/2 \end{cases}, \tag{19}$$

where the phase spectrum is oddly symmetric such that $\hat{\phi}_{l,k}^{\mathrm{WB}} = -\hat{\phi}_{l,N-k}^{\mathrm{WB}}$ $(k = N/2+1,\ldots,N-1)$.

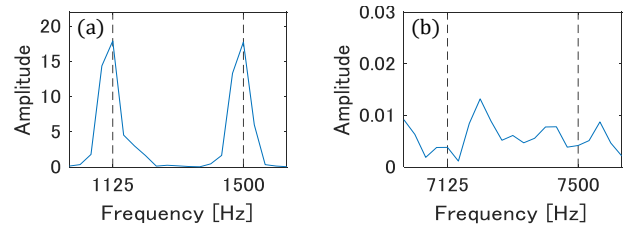

Fig. 2. Amplitude spectra for an oboe signal with the fundamental frequency of 375 Hz. (a) NB. (b) UB. Dash lines show harmonics.

V. Results

A. Experiment Setup

We conducted two experiments to validate the proposed method. In the first experiment, we verified the harmonic structure reconstruction performance. The audio signal was multiple sinusoidal signals with harmonic structures, where each fundamental frequency corresponded to the equal temperament from C5 ($f0_l = 523.251$Hz) to C6 ($f0_l = 1046.502$Hz). Here, the missing UB amplitude spectrum has been reconstructed perfectly. In the second experiment, we verified the ABE performance with the phase reconstruction methods. The audio signals were musical instrument sound signals (oboe, trombone, and flute) with $F_{\mathrm{s}} = 16$ kHz from the publicly available EBU SQAM database [21]. Since the ABE approach is challenging to reconstruct the missing UB perfectly [22] [23], we used the incomplete UB amplitude spectrum which is processed under the MP3 coding with 64 kbps and the additive white Gaussian noise at a Signal-to-Noise ratio of 40 dB. The STFT processing set the frame length 32 ms, frame shift 4 ms, and a Hamming window. In this paper, we obtained the fundamental frequency using the sub-harmonic summation method [24]. The proposed method also set the thresholds $F = f0_l'/2$ and $P = 1$. We compared with the following phase reconstruction methods:

- **FLIP**: Duplication of flipped NB phase spectrum (Eq.(15))
- **GL**: Gliffin-Lim algorithm with iteration times 100 [18]
- **STFTPI**: Sinusoidal models (Eq.(7)(8)) [17] [18]
- **Prop** (Our method): STFTPI with corrected fundamental frequency and harmonic discrimination

We evaluated the spectral distance between the original audio signal and the bandwidth extended WB audio signal using Log-spectral distortion (LSD) [25]. Given the number of analysis frames L_{Eva}, LSD [dB] is defined as

$$\mathrm{LSD} = \frac{1}{L_{\mathrm{Eva}}}\sum_{l=0}^{L_{\mathrm{Eva}}-1}\sqrt{\frac{1}{|\mathcal{K}|}\sum_{k\in\mathcal{K}}\left[10\log_{10}\frac{|X_{l,k}^{\mathrm{WB}}|^2}{|\hat{X}_{l,k}^{\mathrm{WB}}|^2}\right]^2}, \tag{20}$$

where $\mathcal{K}$ is the set of the frequency indices at the frequency bandwidth to be analyzed. We calculated LSD of 0–8 kHz (WB-LSD) and 4–8 kHz (UB-LSD) with the analysis frame length 20 ms.

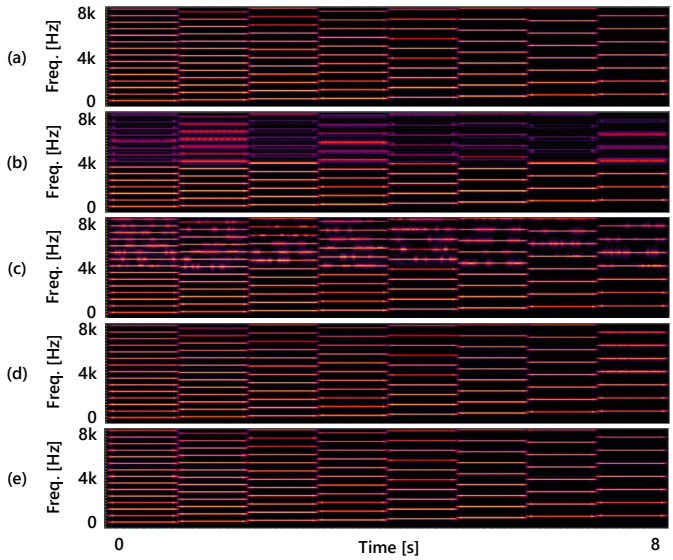

Fig. 3. Sound spectrogram. The original audio signal is multiple sinusoidal signals with harmonic structures. (a)Original audio signal. (b)**FLIP**. (c)**GL**. (d)**STFTPI**. (e)**Prop**.

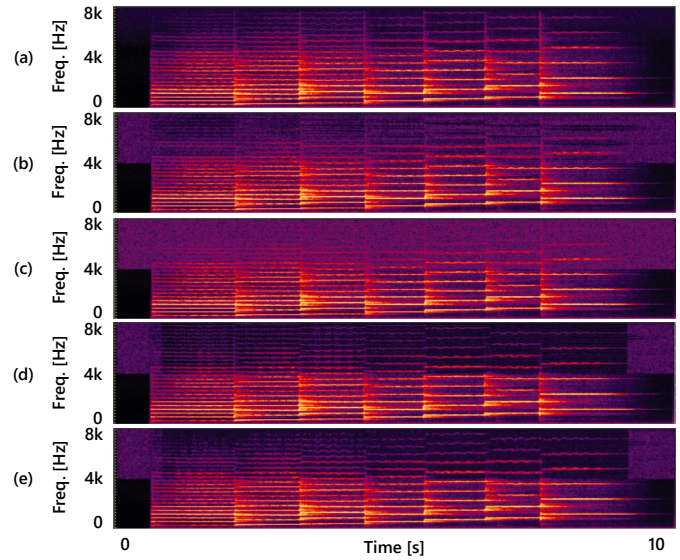

Fig. 4. Sound spectrogram. The original audio signal is an oboe signal. (a)Original audio signal. (b)**FLIP**. (c)**GL**. (d)**STFTPI**. (e)**Prop**.

TABLE I
LSD RESULTS TOWARD MULTIPLE SINUSOIDAL SIGNALS WITH HARMONIC STRUCTURES.

	FLIP	**GL**	**STFTPI**	**Prop**
WB-LSD	17.83	3.18	4.60	2.22
UB-LSD	25.13	4.36	6.39	3.06

TABLE II
LSD RESULTS TOWARD AN OBOE SIGNAL.

		FLIP	**GL**	**STFTPI**	**Prop**
Oboe	WB-LSD	8.17	9.99	8.07	7.52
	UB-LSD	11.53	14.10	11.38	10.62
Trombone	WB-LSD	10.09	12.52	9.87	9.33
	UB-LSD	14.24	17.65	13.94	13.17
Flute	WB-LSD	8.75	10.60	9.19	8.84
	UB-LSD	10.63	14.95	12.97	12.47

B. Experiment 1:Multiple Sinusoidal Signal with Harmonic Structure

Figure 3 shows the sound spectrograms for the original multiple sinusoidal signals with the harmonic structure and the bandwidth extended WB audio signals. While **FLIP** has not reconstructed the UB harmonic structure even when the missing UB amplitude spectrum has been reconstructed perfectly, **GL**, **STFTPI**, and **Prop** have reconstructed the UB harmonic structure. These facts confirm that the phase reconstruction method contributes to the ABE approach. Besides, for an audio signal of 8 s in length, the calculation time for **FLIP**, **GL**, **STFTPI**, and **Prop** is 0.07 s, 9.00 s, 0.60 s, and 0.80 s, respectively. The STFTPI approach is acceptable for long recording audio archives. Table I shows LSD results. **Prop** improved WB-LSD by 0.96 dB and UB-LSD by 1.30 dB compared to **STFTPI**. It can be seen that the fundamental frequency estimation error correction enhances the harmonic structure reconstruction performance more efficiently.

C. Experiment 2:Musical Instrument Sound Signal

Figure 4 shows the sound spectrograms for the original oboe signal and the bandwidth extended WB audio signals. From Fig. 4(c), **GL** has insufficiently reconstructed the harmonic structure. The phase reconstruction method using the iterative algorithm assumes that the amplitude spectrum is clean and is not suitable for the ABE approach. Figures 4(d)(e) show that the STFTPI approach has reconstructed the harmonic structure even when the UB amplitude spectrum has not been reconstructed perfectly. Besides, it can be seen that **Prop** has avoided overemphasizing the UB harmonics near the Nyquist frequency. Table II shows the LSD results. **Prop** achieved the lowest WB-LSD and UB-LSD for the oboe and trombone signals. However, **Flip** achieved the lowest WB-LSD and UB-LSD for the flute signal because the flute signal has contained the frequencies to be emphasized other than the harmonics, which is not easy to reconstruct in the proposed method. Therefore, combining the proposed method with other phase reconstruction methods, which emphasizes the frequencies other than the harmonics, will enhance the audio signal quality more efficiently.

VI. CONCLUSION

In this paper, we proposed the STFTPI approach for the ABE approach toward a musical instrument sound signal. For multiple sinusoidal signals with harmonic structures, the proposed method has reconstructed the UB harmonic structure in a short time and achieved the lowest WB-LSD of 2.22 dB and LSD of 3.06 dB. Also, the harmonic structures for the musical instrument sound signals, such as an oboe and a trombone signals, have been reconstructed using the proposed method. It is expected that the combination with other phase reconstruction method will enhance the ABE performance.

REFERENCES

[1] J. Lim and A. V. Oppenheim, "Enhancement and bandwidth compression of noisy speech," *Proc. IEEE*, vol.67, no.12, pp.1586–1604, 1979.

[2] A. Marafioti, N. Perraudin, N. Holighaus, and P. Majdak, "A context encoder for audio inpainting," *IEEE/ACM Trans. Audio, Speech, Lang. Process.*, vol.27, no.12, pp.2362–2372, 2019.

[3] B. Defraene, N. Mansour, S. De Hertogh, T. van Waterschoot, M. Diehl, and M. Moonen, "Declipping of audio signals using perceptual compressed sensing," *IEEE Trans. Audio, Speech, Lang. Process.*, vol.21, no.12, pp.2627–2637, 2013.

[4] J. Deng, B. Schuller, F. Eyben, D. Schuller, Z. Zhang, H. Francois, and E. Oh, "Exploiting time-frequency patterns with lstm-rnns for low-bitrate audio restoration," *Neural Comput. Appl.*, pp.1–13, 2019.

[5] J. Abel and T. Fingscheidt, "Artificial speech bandwidth extension using deep neural networks for wideband spectral envelope estimation," *IEEE/ACM Trans. Audio, Speech, Lang. Process.*, vol.26, no.1, pp.71–83, 2018.

[6] C. W. Wu and M. Vinton, "Blind bandwidth extension using k-means and support vector regression," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.(ICASSP)*, 2017, pp.721–725.

[7] M. Dietz, L. Liljeryd, K. Kjorling, and O. Kunz, "Spectral band replication a novel approach in audio coding," in *Proc. 112th AES Conv.*, 2002.

[8] H. W. Hsu and C. M. Liu, "Decimation-whitening filter in spectral band replication," *IEEE Trans. Audio, Speech, Lang. Process.*, vol.19, no.8, pp.2304–2313, 2011.

[9] K. Li and C. H. Lee, "A deep neural network approach to speech bandwidth expansion," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.(ICASSP)*, 2015, pp.4395–4399.

[10] M. Miron and M. E. P. Davies, "High frequency magnitude spectrogram reconstruction for music mixtures using convolutional autoencoders," In *Proc. 21st Int. Conf. Digital Audio Effects (DAFx-18)*, 2018, pp.173–180.

[11] T. Y. Lim, R. A. Yeh, Y. Xu, M. N. Do, and M. H. Johnson, "Time-frequency networks for audio super-resolution," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.(ICASSP)*, 2018, pp.646–650.

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

[13] P. Mowlaee, R. Saeidi, and Y. Stylianou, "Advances in phase-aware signal processing in speech communication," *Speech Commun.*, vol.81, pp.1–29, 2016.

[14] D. Griffin and J. Lim, "Signal estimation from modified short-time Fourier transform," *IEEE Trans. Acoust., Speech, Signal Process.*, vol.ASSP-32, no.2, pp.236–242, 1984.

[15] P. Mowlaee and R. Saeidi, "Time-frequency constraints for phase estimation in single-channel speech enhancement," In *Proc. Int. Workshop Acoust. Signal Enhance. (IWAENC)*, 2014, pp.338–342.

[16] P. Mowlaee and J. Kulmer, "Phase estimation in single-channel speech enhancement: Limits-potential," *IEEE/ACM Trans. Audio, Speech, Lang. Process.*, vol.23, no.8, pp.1–12, 2015.

[17] 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, 2014.

[18] 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.

[19] Y. Masuyama, K. Yatabe, Y. Koizumi, Y. Oikawa, and N. Harada, "Phase reconstruction based on recurrent phase unwrapping with deep neural networks," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.(ICASSP)*, 2020, pp.826–830.

[20] S. Hu, B. Zhang, B. Liang, E. Zhao, and S. Lui, "Phase-aware music super-resolution using generative adversarial networks," in *Proc. Annu. Conf. Int. Speech Commun. Assoc.(INTERSPEECH)*, 2020, pp.4074–4078.

[21] European Broadcasting Union, *"Sound Quality Assessment Material Recordings for Subjective Tests"*, 2008, [online] Available: https://tech.ebu.ch/publications/sqamcd.

[22] P. Jax and P. Vary, "An upper bound on the quality of artificial bandwidth extension of narrowband speech signals," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.(ICASSP)*, 2002, pp.237–240.

[23] M. Nilsson, H. Gustafsson, S. V. Andersen, and W. B. Kleijn, "Gaussian mixture model based mutual information estimation between frequency bands in speech," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.(ICASSP)*, 2002, pp.525–528.

[24] D. J. Hermes, "Measurement of pitch by subharmonic summation," *J. Acoust. Soc. Amer.*, vol.83, no.1, pp.257–264, 1988.

[25] L. R. Rabiner and B.-H. Juang, *Fundamentals of Speech Recognition*, New Jersey, Englewood Cliffs:Prentice-Hall, 1993.