ACOUSTIC APPLICATION OF PHASE RECONSTRUCTION ALGORITHMS IN OPTICS

Tomoki Kobayashi, Tomoro Tanaka, Kohei Yatabe, Yasuhiro Oikawa

Department of Intermedia Art and Science, Waseda University, Tokyo, Japan

ABSTRACT

Phase reconstruction from amplitude spectrograms has attracted attention in recent acoustics because of its potential applications in speech synthesis and enhancement. The most well-known algorithm in acoustics is based on alternating projection and called Griffin–Lim algorithm (GLA). At the same time, GLA is known as the Gerchberg–Saxton algorithm in optics, and a lot of its variants have been proposed independently of those in acoustics. In this paper, we propose to apply phase reconstruction algorithms developed in the optics community to acoustic applications and evaluate them using acoustical metrics. Specifically, we propose to apply the averaged alternating reflections (AAR), relaxed AAR (RAAR), and hybrid input-output (HIO) algorithms to acoustic signals. Our experimental results suggested that RAAR has enough potential for acoustic applications because it clearly outperformed GLA.

Index Terms— Phase retrieval, short-time Fourier transform, Griffin–Lim algorithm, alternating projection, reflection operator.

1. INTRODUCTION

Phase reconstruction is a technology of recovering a complex-valued signal only from its magnitude [1]. It has recently attracted attentions in acoustics because of its potential applications in speech synthesis [2–5] and enhancement [6–11]. Furthermore, some recent combination with a deep neural network demonstrated promising applications of phase reconstruction in acoustics, including sound synthesis [12] and audio inpainting [13]. Hence, investigating a new phase reconstruction algorithm is important for extending the possibility of acoustic signal processing. In this paper, we focus on the iterative algorithms that recover phase only from a given magnitude.

Among iterative phase reconstruction methods, the most famous and popular algorithm is the Griffin–Lim algorithm (GLA) [14–16]. It is an alternating projection method that only requires computation of the short-time Fourier transform (STFT) and the inverse STFT (ISTFT) for each iteration. The implementation of GLA is extremely simple if a toolbox for STFT is available, which should be the reason of its popularity. However, the performance of GLA is not excellent, and hence several improvements have been made. For example, the fast GLA (FGLA) applied the Nesterov-like acceleration technique to speed up GLA [17]. The alternating direction method of multipliers (ADMM) has also been applied in order to realize an improved version of GLA [18]. These algorithms have successfully achieved a good phase reconstruction performance while the simplicity of GLA is maintained to some extent.

Independently of these algorithms in acoustics, a lot of phase reconstruction methods have been proposed in the literature of optics. In optics, the algorithm similar to GLA is well-known as the Gerchberg–Saxton algorithm proposed around 1970 [19] that is more than 10 years before the proposal of GLA. Since then, many algorithms have been developed [17, 18]. One interesting direction was to use a reflection operator (or reflected resolvent) in place of the projection operator of the algorithm [20–25]. Such methods should be able to speed up an iterative algorithm in general, but they have not been applied to acoustic signals to the best of our knowledge.

In this paper, we propose to apply phase reconstruction algorithms developed in the optics community to acoustic applications. We chose three typical algorithms using the reflection operator: the averaged alternating reflections (AAR) [20, 21], relaxed AAR (RAAR) [22], and hybrid input-output (HIO) [23–25] algorithms. Their performance was experimentally investigated in terms of acoustical metrics to see their potential in acoustics. The experimental results showed that RAAR performed better in average over various conditions, and hence it is promising in acoustics.

2. ACOUSTIC PHASE RECONSTRUCTION

In acoustics, phase reconstruction is considered together with a specific time-frequency transformation. In this paper, we focus on the phase reconstruction from the magnitude of STFT coefficients.

2.1. Phase Reconstruction from STFT Magnitude

STFT (or discrete Gabor transform) of a signal $\mathbf{x} \in \mathbb{R}^L$ is given by

$$X[m,n] = \sum_{l=0}^{K-1} x[l+an]\, w[l]\, \mathrm{e}^{-2\pi \mathrm{i} ml/M}, \tag{1}$$

where $\mathbf{X} \in \mathbb{C}^{M\times N}$ is the corresponding spectrogram, $\mathbf{w} \in \mathbb{R}^K$ is a window, K is the window length, a is the time-shifting step, n and m are the time and frequency indices, respectively, the index of the signal $l + an$ is understood modulo L, and i is the imaginary unit. Acoustic phase reconstruction aims at recovering the time-domain signal $\mathbf{x}$ only from the magnitude of its spectrogram $|\mathbf{X}|$, where $|\cdot|$ represents the element-wise absolute value.

Given a magnitude spectrogram $\mathbf{A} \in \mathbb{R}_+^{M\times N}$, a phase reconstruction problem is formulated as the following feasibility problem:

$$\text{Find}\;\; \mathbf{X} \;\; \text{s.t.}\;\; \mathbf{X} \in \mathcal{A} \cap \mathcal{C}, \tag{2}$$

where $\mathcal{A}$ is the set of spectrograms whose magnitude is $\mathbf{A}$,

$$\mathcal{A} = \left\{ \mathbf{X} \in \mathbb{C}^{M\times N} \mid |\mathbf{X}| = \mathbf{A} \right\}, \tag{3}$$

$\mathcal{C}$ is the set of spectrograms that are in the image of STFT,

$$\mathcal{C} = \left\{ \mathbf{X} \in \mathbb{C}^{M\times N} \mid \exists \mathbf{x} \in \mathbb{R}^L \;\text{s.t.}\; \mathrm{STFT}_{\mathbf{w}}(\mathbf{x}) = \mathbf{X} \right\}, \tag{4}$$

and $\mathrm{STFT}_{\mathbf{w}}(\cdot)$ denotes the operator of STFT. Since directly solving this feasibility problem is impossible in general, the usual strategy is to apply an iterative algorithm to approximately solve it. The aim of this paper is to investigate the performance of such iterative algorithms in terms of acoustical metrics.

2.2. Griffin–Lim Algorithm (GLA)

A typical method for solving a feasibility problem is the alternating projection algorithm. Let $P_{\mathcal{S}}$ be a projector onto a set $\mathcal{S}$:

$$P_{\mathcal{S}}(\mathbf{x}) = \arg\min_{\mathbf{y}\in\mathcal{S}} \|\mathbf{x}-\mathbf{y}\|, \tag{5}$$

where $\|\cdot\|$ is the Euclidean norm. Then, an alternating projection method for the phase reconstruction problem in Eq. (2) is given by

$$\mathbf{X}^{[i+1]} = P_{\mathcal{C}}(P_{\mathcal{A}}(\mathbf{X}^{[i]})), \tag{6}$$

where i is the iteration count. This algorithm is called GLA. The two projection operators in Eq. (6) can be easily computed as follows:

$$P_{\mathcal{C}}(\mathbf{X}) = \mathrm{STFT}_{\mathbf{w}}(\mathrm{ISTFT}_{\widetilde{\mathbf{w}}}(\mathbf{X})), \tag{7}$$

$$P_{\mathcal{A}}(\mathbf{X}) = \mathbf{A}\odot\mathbf{X}\oslash|\mathbf{X}|, \tag{8}$$

where $\mathrm{ISTFT}_{\widetilde{\mathbf{w}}}(\cdot)$ denotes ISTFT using $\widetilde{\mathbf{w}}$ as the synthesis window, $\widetilde{\mathbf{w}}$ is the canonical dual window of $\mathbf{w}$ that gives the perfect reconstruction property to STFT, and $\odot$ and $\oslash$ represents the element-wise multiplication and division, respectively. Since only difficult parts are STFT and ISTFT, GLA can be easily implemented, which should be the reason of its popularity in acoustics.

2.3. Fast Griffin–Lim Algorithm (FGLA)

A practically important drawback of GLA is its slow convergence. To improve its performance, several variants have been proposed. FGLA is a simple but powerful variant that applies the Nesterov-like acceleration to GLA [17]. That is, after each iteration of Eq. (6),

$$\mathbf{X}^{[i+1]} = \mathbf{X}^{[i+1]} + \alpha(\mathbf{X}^{[i+1]} - \mathbf{X}^{[i]}) \tag{9}$$

is performed, where $\alpha > 0$ controls the convergence speed. It has been empirically shown that this additional step not only speed up the algorithm but also improves the quality of reconstruction. This should be because the acceleration technique helps the algorithm to escape from poor local minima.

2.4. Alternating Direction Method of Multipliers (ADMM)

The ADMM-based algorithm proposed in [18] is another algorithm using the two projection operators. ADMM is a popular algorithm for solving general optimization problems. Its fast convergence in practice has been empirically known, and hence it has been applied to acoustic phase reconstruction. The ADMM-based phase reconstruction algorithm is given in the bottom of Table 1, together with GLA and FGLA for comparison. This algorithm performs better than the other methods when the number of iterations is small [18]. The algorithmic parameter $\rho \geq 0$ controls robustness for noisy observation. I denotes the identity operator, and the parentheses are used as $(P_{\mathcal{C}} + \rho I)(\mathbf{X}) = P_{\mathcal{C}}(\mathbf{X}) + \rho I(\mathbf{X}) = P_{\mathcal{C}}(\mathbf{X}) + \rho\mathbf{X}$.

Table 1. Phase reconstruction algorithms known in acoustics

Griffin–Lim algorithm (GLA)
$\mathbf{X}^{[i+1]} = P_{\mathcal{C}}(P_{\mathcal{A}}(\mathbf{X}^{[i]}))$
Fast Griffin–Lim algorithm (FGLA)
$\mathbf{X}^{[i+1]} = P_{\mathcal{C}}(P_{\mathcal{A}}(\mathbf{Y}^{[i]}))$ $\mathbf{Y}^{[i+1]} = \mathbf{X}^{[i+1]} + \alpha(\mathbf{X}^{[i+1]} - \mathbf{X}^{[i]})$
Alternating direction method of multipliers (ADMM)
$\mathbf{Z}^{[i+1]} = P_{\mathcal{A}}(\mathbf{X}^{[i]} - \mathbf{Y}^{[i]})$ $\mathbf{X}^{[i+1]} = \frac{1}{\rho+1}(P_{\mathcal{C}} + \rho I)(\mathbf{Z}^{[i+1]} + \mathbf{Y}^{[i]})$ $\mathbf{Y}^{[i+1]} = \mathbf{Y}^{[i]} + \mathbf{Z}^{[i+1]} - \mathbf{X}^{[i+1]}$

3. PHASE RECONSTRUCTION ALGORITHMS IN OPTICS

Since the phase reconstruction problem originated in optics, a lot of algorithms have been proposed in the context of optical imaging. However, those algorithms have seldom used for acoustic applications. This should be because of the following reasons. First, the mainstream of phase reconstruction in optics is not based on STFT but based on the two-dimensional Fourier transform. A transform similar to STFT is considered only for a specific measurement method (e.g., ptychography). Second, the technical terms in optics may not be familiar to many of acoustic engineers. A phase reconstruction method is usually introduced with an optical measurement setting that may include many unfamiliar words. Third, the projector $P_{\mathcal{C}}$ is not in the form of Eq. (7) but frequently include some additional operations (in the "time domain"). For example, compactness of the support and non-negativity are often enforced. Such time-domain constraints have almost never been used in acoustics, and hence direct application of optical phase reconstruction algorithms seems unsuitable for acoustics.

In this paper, we borrow the structure of the algorithms in optics and propose to apply them to acoustic applications. Three algorithms are introduced and applied *without* the operations specific to optics. That is, some operations of the original algorithms may be omitted, and the projection operator $P_{\mathcal{C}}$ in Eq. (7) is used instead[1].

3.1. Iterative Algorithms using Reflection Operator

For convex feasibility problems, it is known that iterative algorithms using the reflection operator usually outperform the alternating projection algorithm. By applying such algorithms to the feasibility problem of phase reconstruction in Eq. (2), several algorithms have been proposed in the literature.

The reflection operators considered in this papers are

$$R_{\mathcal{C}}(\mathbf{X}) = (2P_{\mathcal{C}} - I)(\mathbf{X}) \qquad (= 2P_{\mathcal{C}}(\mathbf{X}) - \mathbf{X}), \tag{10}$$

$$R_{\mathcal{A}}(\mathbf{X}) = (2P_{\mathcal{A}} - I)(\mathbf{X}) \qquad (= 2P_{\mathcal{A}}(\mathbf{X}) - \mathbf{X}), \tag{11}$$

where the projection operators $P_{\mathcal{C}}$ and $P_{\mathcal{A}}$ are given in Eqs. (7) and (8), respectively. By definition, $P_{\mathcal{C}}(\mathbf{X})$ is the midpoint between $\mathbf{X}$ and $R_{\mathcal{C}}(\mathbf{X})$, and the same is true for $P_{\mathcal{A}}$ and $R_{\mathcal{A}}$. That is, the reflector R moves a point $\mathbf{X}$ twice as much as the corresponding projector P. Therefore, it can be expected that a proper combination of the reflection operators obtains an approximate solution to the feasibility problem with a smaller number of iterations.

In this paper, we propose to apply the following three algorithms to acoustic phase reconstruction [26]: AAR [20, 21], RAAR [22], and HIO [23–25]. These algorithms are summarized in Table 2.

[1]Note that the name of the algorithms introduced in this section might be inappropriate because of the modification we made for applying them to the acoustic problems. Even so, we use the names popular in optics to make the connection between the algorithms clearer.

Table 2. Phase reconstruction algorithms known in optics

Averaged alternating reflections (AAR)
$\mathbf{X}^{[i+1]} = \frac{1}{2}(I + R_{\mathcal{C}}R_{\mathcal{A}})(\mathbf{X}^{[i]})$
Relaxed averaged alternating reflections (RAAR)
$\mathbf{X}^{[i+1]} = \big(\frac{\beta}{2}(I + R_{\mathcal{C}}R_{\mathcal{A}}) + (1-\beta)P_{\mathcal{A}}\big)(\mathbf{X}^{[i]})$
Hybrid input-output algorithm (HIO)
$\mathbf{X}^{[i+1]} = \frac{1}{2}\big(I + R_{\mathcal{C}}(R_{\mathcal{A}} + (\beta-1)P_{\mathcal{A}}) + (1-\beta)P_{\mathcal{A}}\big)(\mathbf{X}^{[i]})$

AAR given here is an application of the Douglas–Rachford algorithm, which is closely related to ADMM, to the feasibility problem in Eq. (2). RAAR is a relaxed version of AAR specified by a relaxation parameter $0 < \beta < 1$. HIO is further relaxed with different use of the relaxation parameter. Note that RAAR and HIO are equivalent to AAR when $\beta = 1$. In this sense, both RAAR and HIO are extension of AAR.

These algorithms have extensively used in optics, and their effectiveness has been confirmed in many applications. However, their effectiveness for acoustic phase reconstruction is not understood because, to the best of our knowledge, they have not applied yet. Therefore, these three algorithms are experimentally investigated in the next section by using acoustical metrics. In fact, as later shown by the experimental results, even algorithms that work well in optics do not always work well in acoustical applications.

4. EXPERIMENTS

To investigating the performance in acoustics, the above-mentioned algorithms were applied to speech signals. For the speech signals, 200 utterances (100 males and 100 females) [27] from the TIMIT database were used. They were converted to spectrograms by STFT, and the absolute value is calculated. The amplitude spectrograms were inputted to the six algorithms. The reconstructed speech signals were evaluated by three acoustic objective measures: PESQ [28], STOI [29, 30], and PEMO-Q [31].

To imitate imperfection of synthesized amplitude spectrograms in speech synthesis, degraded signals were also used. To generate the signals, Gaussian noise was added to the speech signals so that the signal-to-noise ration (SNR) becomes 0 dB. Then, the oracle Wiener filter was applied to obtain the degraded amplitude spectrograms. The quality of reconstructed speech signals were evaluated by comparison with the original clean signals.

In all conditions, initial phases were set to uniform random numbers between 0 and $2\,\pi$ in the time-frequency domain. STFT was implemented with the Hann window whose length was 32 ms, and the overlap length was either 4 or 16 ms. The algorithmic parameters of FGLA α and ADMM ρ were set to $\alpha = 0.99$ [17] and $\rho = 0.3$ [18].

4.1. Experiment 1: Performance for each iteration

First, the performance at each iteration was investigated using clean speech signals. Since RAAR and HIO have the tuning parameter β, nine variations of them ($\beta = 0.1, 0.2, \cdots, 0.9$) were tested in addition to the other four algorithms. Note that we did not include $\beta = 1$ because RAAR and HIO coincide with AAR in that case.

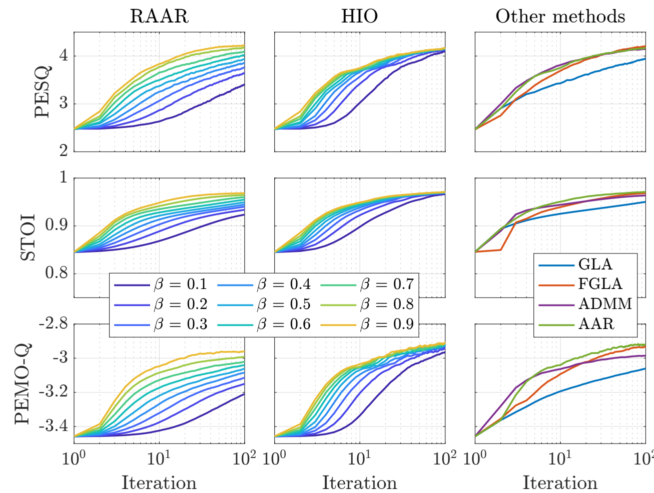

Fig. 1. Averaged scores of PESQ, STOI, and PEMO-Q at each iteration. The clean amplitude spectrograms were used as the input data, and the overlap length of STFT was 16 ms.

The scores at each iteration averaged for all signals are shown in Fig. 1. From top to bottom, the scores of PESQ, STOI, and PEMO-Q are shown. The left column shows the results of RAAR, the center column shows those of HIO, and the right column summarizes those of the other methods.

From the figures in the left column, it can be seen that the performance of RAAR improved as the parameter β increased. There are no crossing of the curves and all three metrics shows the similar tendency. The best parameter among its nine variations was $\beta = 0.9$ regardless of the number of iterations. While these results are almost the same for HIO, some difference from those of RAAR can also be seen. Although the performance of RAAR at the 100th iteration highly depends on β, that of HIO less depends on in and converged to the similar values within 100 iterations.

For the other methods, it is interesting to see that FGLA, ADMM and AAR boosted from the second or third iteration while GLA constantly improved the performance with iteration. The trajectories the scores were different for all methods, and hence the performance was different at a specific iteration.

4.2. Experiment 2: Reconstruction from clean amplitude

To see the scores of Fig. 1 in more details, we drew the box plots. The results at the 10th and 100th iterations are summarized in Figs. 2 and 3, respectively. The boxes represent the first and third quartiles, and the vertical axes were individually adjusted to enhance the visibility. The parameters of RAAR and HIO were set to $\beta = 0.9$ according to the previous experiment.

As in the results, the performance of all methods was better when the shift width was smaller. When the number of iterations was small (Fig. 2), ADMM outperformed GLA and FGLA for all metrics. AAR, RAAR and HIO were comparable to ADMM except the top right figure. According to PESQ and STOI, RAAR was slightly better than ADMM, AAR and HIO. However, PEMO-Q showed that AAR and HIO were slightly better than RAAR. When the number of iterations was large (Fig. 3), FGLA outperformed ADMM and GLA for all metrics. RAAR was comparable to FGLA except the left top figure. AAR and HIO followed them from slightly below.

These results suggest that the phase reconstruction algorithms proposed in acoustics perform differently to the number of iterations.

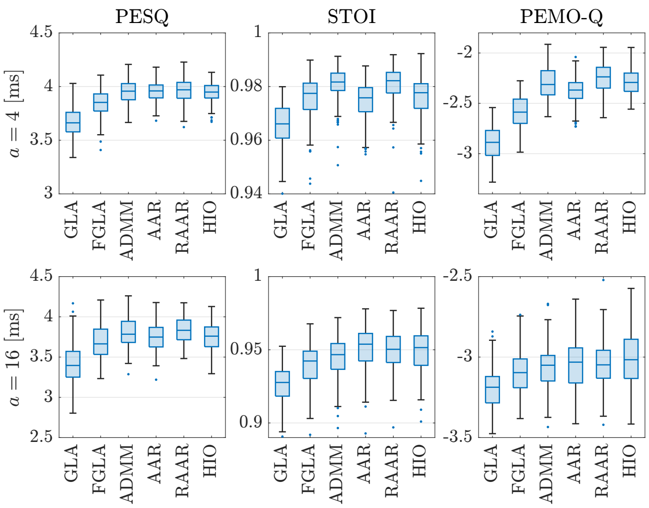

Fig. 2. Scores of reconstruction from clean amplitude (10th iter.).

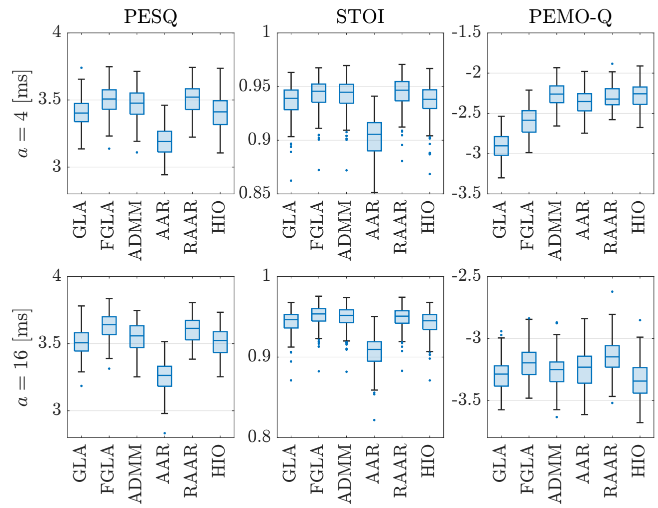

Fig. 4. Scores of reconstruction from noisy amplitude (10th iter.).

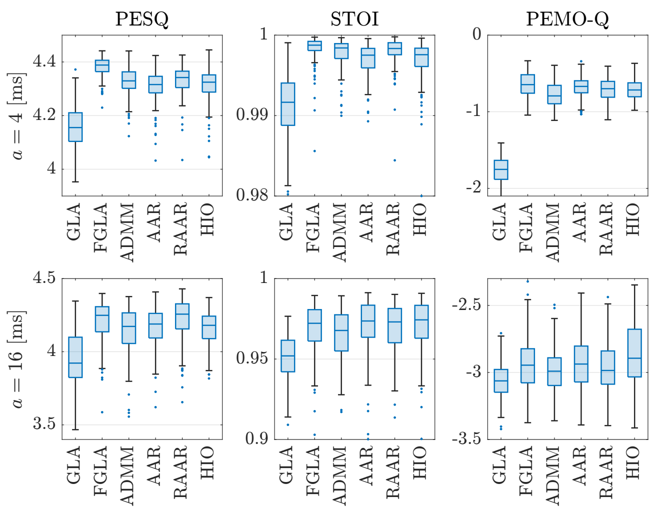

Fig. 3. Scores of reconstruction from clean amplitude (100th iter.).

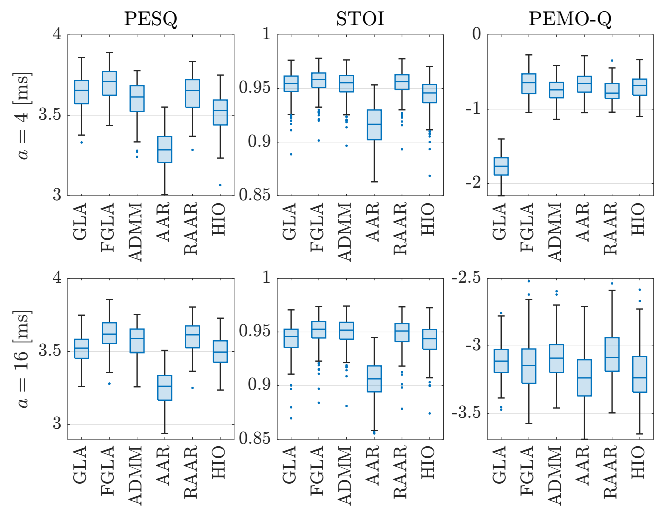

Fig. 5. Scores of reconstruction from noisy amplitude (100th iter.).

When the number of iterations is small, ADMM is a good choice, but FGLA outperforms it in the later iterations. In contrast, the phase reconstruction algorithms proposed in optics consistently performs well for both small and large iterations. In particular, RAAR performed slightly better than the other algorithms.

4.3. Experiment 3: Reconstruction from degraded amplitude

Using the same settings as those in Section 4.2, the performance of phase reconstruction from degraded amplitude spectrograms were evaluated. The results at the 10th and 100th iterations are summarized in Figs. 4 and 5, respectively. Interestingly, as oppose to the clean case, the performance with smaller shift width was not always better than that with larger shift width.

In all cases, AAR performed poorly, which should be because AAR is not reliable when the existence of the solution is not guaranteed. On the contrary, FGLA performed better than the other methods. Importantly, the performance of GLA was mildly less than that of FGLA, which indicates that the amount of degradation was severe. In this case, HIO performed similar or worse than GLA. In constrast, the performance of RAAR was comparable to FGLA and ADMM, and RAAR slightly performed better than them for some situations. This should be because the relaxation of RAAR is appropriate even for phase reconstruction of acoustic signals.

In summary, the phase reconstruction algorithms in optics perform well when the inputted amplitude spectrograms are clean. AAR and HIO are sensitive to error of amplitude, and hence they are not practical in acoustic applications. In contrast, RAAR is robust to noise and performs well for both small and large number of iterations. Therefore, RAAR is promising for acoustic applications.

5. CONCLUSIONS

In this paper, we propose to apply algorithms in optics to acosutic phase reconstructions. We experimentally investigated their performance using acoustical metrics and found that RAAR performs well in various conditions. Since some acceleration techniques including that used in FGLA might be possible to improve the performance of RAAR, such combination for realizing a fast and well-performing algorithm should be investigated in the future works.

6. REFERENCES

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

[2] Y. Wang, R. J. Skerry-Ryan, D. Stanton, Y. Wu, R. J. Weiss, N. Jaitly, Z. Yang, Y. Xiao, Z. Chen, S. Bengio, Q. Le, Y. Agiomyrgiannakis, R. Clark, and R. A. Saurous, "Tacotron: A fully end-to-end text-to-speech synthesis model," in *Proc. INTERSPEECH*, Aug 2017, pp. 3389–3393.

[3] S. Takaki, H. Kameoka, and J. Yamagishi, "Direct modeling of frequency spectra and waveform generation based on phase recovery for DNN-based speech synthesis," in *Proc. INTER-SPEECH*, Aug 2017, pp. 1128–1132.

[4] T. Kaneko, S. Takaki, H. Kameoka, and J. Yamagishi, "Generative adversarial network-based postfilter for stft spectrograms," in *Proc. INTERSPEECH*, Aug 2017.

[5] Y. Saito, S. Takamichi, and H. Saruwatari, "Text-to-speech synthesis using STFT spectra based on low-/multi-resolution generative adversarial networks," in *IEEE Int. Conf. Acoust. Speech Signal Process. (ICASSP)*, Apr 2018, pp. 5299–5303.

[6] K. Paliwal, K. Wójcicki, and B. Shannon, "The importance of phase in speech enhancement," *Speech Commun.*, vol. 53, no. 4, pp. 465–494, Apr 2011.

[7] 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.

[8] 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. 1283–1294, May 2015.

[9] 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, Jun 2015.

[10] Y. Wakabayashi, T. Fukumori, M. Nakayama, T. Nishiura, and Y. Yamashita, "Phase reconstruction method based on time-frequency domain harmonic structure for speech enhancement," in *IEEE Int. Conf. Acoust. Speech Signal Process. (ICASSP)*, Mar 2017, pp. 5560–5564.

[11] 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, Apr 2018.

[12] A. Marafioti, N. Perraudin, N. Holighaus, and P. Majdak, "Adversarial generation of time-frequency features with application in audio synthesis," in *Proc. Int. Conf. Mach. Learn. (ICML)*. Jun 2019, pp. 4352–4362, PMLR.

[13] A. Marafioti, P. Majdak, N. Holighaus, and N. Perraudin, "GACELA: A generative adversarial context encoder for long audio inpainting of music," *IEEE J. Sel. Top. Signal Process.*, vol. 15, no. 1, pp. 120–131, 2021.

[14] S. Marchesini, "A unified evaluation of iterative projection algorithms for phase retrieval," *Rev. Sci. Instrum.*, vol. 78, no. 3, pp. 011301, 2007.

[15] D. W. Griffin and J. S. Lim, "Signal estimation from modified short-time Fourier transform," *IEEE Trans. Acoust. Speech Signal Process.*, vol. 8, pp. 804–807, Apr 1983.

[16] D. Griffin and Jae Lim, "Signal estimation from modified short-time Fourier transform," *IEEE Trans. Acoust. Speech Signal Process.*, vol. 32, no. 2, pp. 236–243, Apr 1984.

[17] N. Perraudin, P. Balazs, and P.L. Søndergaard, "A fast Griffin–Lim algorithm," in *2013 IEEE Workshop Appl. Signal Process. Audio Acoust.*, Oct 2013, pp. 1–4.

[18] Y. Masuyama, K. Yatabe, and Y. Oikawa, "Griffin–Lim like phase recovery via alternating direction method of multipliers," *IEEE Signal Process. Lett.*, vol. 26, no. 1, pp. 184–188, Jan 2019.

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

[20] H.H. Bauschke, P.L. Combettes, and D.R. Luke, "Phase retrieval, error reduction algorithm, and Fienup variants: a view from convex optimization," *J. Opt. Soc. Am. A*, vol. 19, no. 7, pp. 1334–1345, Jul 2002.

[21] H.H. Bauschke, P.L. Combettes, and D.R. Luke, "Finding best approximation pairs relative to two closed convex sets in Hilbert spaces," *J. Approx. Theory*, vol. 127, no. 2, pp. 178–192, Apr 2004.

[22] D.R. Luke, "Relaxed averaged alternating reflections for diffraction imaging," *Inverse Probl.*, vol. 21, no. 1, pp. 37–50, Nov 2004.

[23] J.R. Fienup, "Reconstruction of an object from the modulus of its Fourier transform," *Opt. Lett.*, vol. 3, no. 1, pp. 27–29, Jul 1978.

[24] J.R. Fienup, "Phase retrieval algorithms: A comparison," *Appl. Opt.*, vol. 21, no. 15, pp. 2758–2769, Aug 1982.

[25] H.H. Bauschke, P.L. Conbettes, and D.R.Luke, "Hybrid projection–reflection method for phase retreval," *J. Opt. Soc. Am. A*, vol. 20, no. 6, pp. 1025–1034, Jun 2003.

[26] A. Fannjiang and T. Strohmer, "The numerics of phase retrieval," *Acta Numer.*, vol. 29, pp. 125–228, 2020.

[27] P. Mowlaee, J.Kulmer, J.Stahl, and F.Mayer, "Single channel phase-aware signal processing in speech communication: Theory and practice.," in *Hoboken, NJ, USA: Wiley*, 2016.

[28] 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," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process.*, May 2001, vol. 19, pp. 2125–2136.

[29] C.H. Taal, R.C. Hendriks, R.Heusdens, and J.Jensen, "A short-time objective intelligibility measure for time-frequency weighted noisy speech," in *IEEE Int. Conf. Acoust. Speech Signal Process.*, Mar 2010, pp. 4214–4217.

[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, Sep 2011.

[31] K. Kondo, "On the use of objective quality measures to estimate watermarked audio quality," in *Eighth Int. Conf. Intell. Inf. Hiding Multimed. Signal Process.*, 2012, pp. 126–129.