Extension of virtual microphone technique to multiple real microphones and investigation of the impact of phase and amplitude interpolation on speech enhancement

Hanako Segawa*, Li Li†, Shoji Makino*‡, Takeshi Yamada*
*University of Tsukuba, Japan
†NTT Communication Science Laboratories, NTT Corporation, Japan
‡Waseda University, Japan

***Abstract*—A virtual microphone signal can artificially increase the number of channels of an observed signal, which leads to improved multi-channel speech enhancement performance. The conventional virtual microphone technique can only be applied to two real microphones, reducing its application to a limited range of situations. In this paper, we extend the virtual microphone technique to the interpolation of more than two real microphones. In the proposed method, the phase based on the plane wave model is linearly interpolated using an affine coefficients of the coordinates, and the amplitude is nonlinearly interpolated based on the β-divergence using coefficients inversely proportional to the distance between the virtual and real microphones. The experimental results shows that the proposed method is effective in improving speech enhancement performance. Furthermore, we investigate the impact of phase and amplitude interpolation on speech enhancement performance with two or three real microphones. The experimental results indicated that the benefits of phase interpolation based on plane wave model were limited, and that amplitude interpolation was important for improving speech enhancement performance.**

I. Introduction

Recently, speech-based systems have been promoted in many fields, and research on speech enhancement technology has been conducted. However, the performance of speech enhancement techniques using microphone arrays strongly depends on the number of microphones, limiting the situations in which can be used. For example, the performance of speech enhancement by beamforming is degraded in underdetermined situations with more sound sources than microphones.

Several methods such as time-frequency masking[1], multichannel Wiener filtering[2] and the statistical modeling of observations using latent variables[3][4] work properly in underdetermined situations. However, they tend to contain artificial noises than the methods for the determined situation. Virtual microphone technology has been proposed as a method of improving the speech enhancement performance under such underdetermined situations where the number of microphones is limited[5][6]. Virtual microphones can provide an enhanced signal with fewer artificial noises in that they use a method for the determined situation after increasing the signal virtually.

Virtual microphone technology is a technique that generates a virtual signal on the line joining two real microphones. Since the generated virtual microphone signals artificially increase the number of channels of the observed signals, we are able to improve the speech enhancement performance. The virtual microphone signal was originally proposed to be generated in the complex spectrogram domain, and generating the virtual microphone signals in time-domain has been recently proposed [7]. The virtual microphone technique can be interpolated or extrapolated. In interpolation, the phase and amplitude are obtained by linear interpolation based on the plane wave assumption and rule-based nonlinear interpolation using β-divergence. Recently, an interpolation method for amplitude using deep learning has been proposed[8][9]. However, DNN-based methods require a lot of training data and are difficult to apply to non-training environments. The interpolation method without DNN do not require training data and can be applied in any environment where the physical assumptions are considered to be met.

The virtual microphone technique has been shown to be effective in improving the speech enhancement performance of beamforming. However, the conventional virtual microphone technique can only be applied when the virtual signal is on the straight line joining the two real microphones, thus limiting the positions and situations in which it can be used. Although it is possible to place a virtual microphone signal at an arbitrary position by using the conventional method multiple times. However, it is conceivable that the speech enhancement performance may depend on the order of selection of real microphones. Furthermore, the calculation of the coefficient α in the conventional method may become complex when the number of real microphones increases and place a virtual microphone at an arbitrary coordinate.

In addition, although we confirmed that the speech enhancement performance of beamforming can be improved by using generated signals whose phase and amplitude are obtained by conventional methods, there has been insufficient investigation on how much they affect the speech enhancement performance.

In this paper, we extend the conventional method to generate

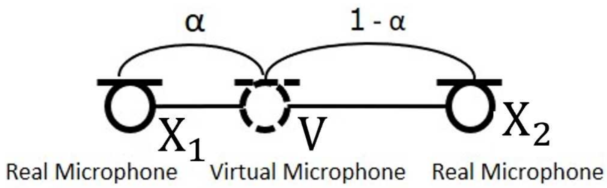

Fig. 1. Arrangement of 2 real and virtual microphones in interpolation technique

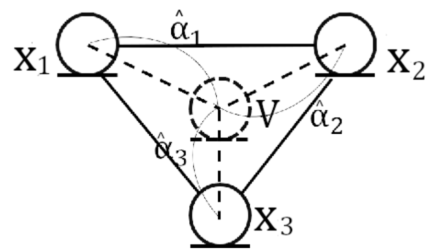

Fig. 2. Arrangement of three real microphones and virtual microphone in interpolation technique

virtual signals from arbitrary number of microphones in a single plane, and we investigate the effectiveness of the proposed method using three real microphones on the same plane. The proposed method can be easily extended to an arbitrary number of microphones in the space.

In addition, to clarify the effectiveness of phase and amplitude interpolation on speech enhancement performance, we conducted several experiments using signals from two real microphones and using three real microphones, using simulations and measured impulse responses.

II. Virtual Microphone Interpolation Using Two Microphones

In this section, we introduce the interpolation of virtual microphones using two real microphones[5][6]. In the virtual microphone technique, the virtual microphone signal $V(\omega,t,\alpha)$ is generated from the observation signals $X_i(\omega,t)$ of two real microphones. In addition, it introduces nonlinearity in the interpolation. i is the identifier of the real microphone used for generation ($i = 1, 2$), ω is the frequency bin in that microphone, and t is the time frame. α $(0 < \alpha < 1)$ is the interpolation coefficient for the virtual microphone. A virtual microphone signal is the observation estimated at the point obtained by internally dividing the line joining two real microphones in the ratio α : (1 -α). Fig. 1 shows the arrangement of real and virtual microphones.

In virtual microphone technique, the phase and amplitude are interpolated separately. The phase and amplitude of the observed signals $X_i(\omega,t)$ are respectively defined as

$$\phi_i = \angle X_i(\omega,t) = \tan^{-1}\frac{\mathrm{Im}(X_i(\omega,t))}{\mathrm{Re}(X_i(\omega,t))}, \quad (1)$$

$$A_i = |X_i(\omega,t)|. \quad (2)$$

We assume W-DO for mixed signals that each time-frequency bin is dominated by at most one sound source [1][10]. This means that the observed signal in each time-frequency bin can be regarded as a single wave. This allows the interpolation of the virtual microphone signal to be regarded as a single sound within each time-frequency bin, even when multiple sounds are arriving. Based on this assumption, the physical model of the propagating wave is approximated as that of a plane wave. The phase of the interpolated virtual microphone can be expressed by linear interpolation as

$$\begin{aligned}\phi_v &= \phi_1 + \alpha(\phi_2 - \phi_1)\\ &= (1-\alpha)\phi_1 + \alpha\phi_2. \quad (3)\end{aligned}$$

Since the observed phase has an aliasing ambiguity given by$\phi_i \pm 2n_i\pi$ with n_i, virtual microphone interpolation requires no spatial aliasing,

$$|\phi_1 - \phi_2| \le \pi. \quad (4)$$

The amplitude depends on many acoustic conditions, and it is difficult to model the actual amplitude decay faithfully. In the conventional method, the amplitude interpolation is formulated as an optimization problem in which the weighted β-divergence between the amplitudes of two real microphones is minimized, and the solution of the closed-form problem

$$A_v = \begin{cases}\exp\left((1-\alpha)\log A_1 + \alpha\log A_2\right) & (\beta = 1)\\ \left((1-\alpha)A_1^{\beta-1} + \alpha A_2^{\beta-1}\right)^{\frac{1}{\beta-1}} & (\text{otherwise})\end{cases} \quad (5)$$

is used as the interpolation rule, where β is the hyperparameter of β-divergence, which adjust the nonlinearity of amplitude interpolation. From ϕ_v and A_v obtained using equations (3) and (5) respectively, the generated virtual microphone signal can be expressed as

$$V(\omega,t,\alpha) = A_v \exp(j\phi_v). \quad (6)$$

III. Proposed Method: Interpolation of Virtual Microphones Using an Arbitrary Number of Microphones

In this section, we extend the interpolation of virtual microphones, which consists of the linear interpolation of the phase and amplitude by β-divergence, to three or more microphones on the same plane. We suppose that there are I real microphones. To simple represent a microphone at arbitrary position, we incorporate the coordinate system, and the position of the ith microphone is represented by the coordinates $p_i = [x_i, y_i, z_i]^T$, and the position of the virtual microphone signal is represented by the coordinates $p_v = [x_v, y_v, z_v]^T$. First, we consider the linear interpolation of the phase based on the plane wave assumption. When the microphone and the incoming wave are on the same plane, the coordinates of the virtual microphone signal at an arbitrary position can be represented by an affine combination of the coordinates of three real microphones on non-identical lines.

$$p_v = \lambda_1 p_1 + \lambda_2 p_2 + \lambda_3 p_3 \quad (7)$$

$$\text{s.t. } \sum_i^3 \lambda_i = 1 \quad (8)$$

Based on the plane wave assumption and sparse (W-DO), the phase interpolation can be expressed as

$$\phi_v = \sum_{i=1}^{3} \lambda_i \phi_i. \tag{9}$$

where λ_i is the linear combination coefficient that is the solution to the simultaneous linear equation.

$$\mathbf{A}\boldsymbol{\lambda} = \mathbf{c}, \tag{10}$$

where $\mathbf{A}$, $\mathbf{c}$, and λ are defined as

$$\mathbf{A} = \begin{bmatrix} x_1 & x_2 & x_3 \\ y_1 & y_2 & y_3 \\ z_1 & z_2 & z_3 \\ 1 & 1 & 1 \end{bmatrix}, \mathbf{c} = \begin{bmatrix} x_v \\ y_v \\ z_v \\ 1 \end{bmatrix}, \boldsymbol{\lambda} = \begin{bmatrix} \lambda_1 \\ \lambda_2 \\ \lambda_3 \end{bmatrix} \tag{11}$$

Since the sound source and the microphones are on the same plane, $z_i = 0$ and $z_v = 0$, it is obvious that any solution can be substituted into $\sum_i \lambda_i z_i = z_v$.

The amplitude extends the interpolation using the β-divergence at two microphones to an arbitrary number of microphones. Unlike the phase interpolation where the phase is interpolated by selected 3 real microphones, the amplitude can be interpolated by taking all I real microphones into consideration. $\hat{\alpha_i}$ is the distance from each real microphone to the coordinates of the virtual microphone and is defined as

$$\hat{\alpha_i} = \sqrt{(x_i - x_v)^2 + (y_i - y_v)^2 + (z_i - z_v)^2}. \tag{12}$$

Here, we assume that α_i, the coefficient representing the weight of the β-divergence with the amplitude of each real microphone, is inversely proportional to the distance $\hat{\alpha}$ so that the closer the real microphone is, the larger the weight. Moreover, by normalizing so that $\sum_i \alpha_i = 1$, the weight coefficient α_i is expressed as

$$\alpha_i = \frac{\prod_{k=1}^{I} \hat{\alpha}_k}{\hat{\alpha}_i \sum_{k=1}^{I} \frac{\prod_{j=1}^{I} \hat{\alpha_j}}{\hat{\alpha_k}}}. \tag{13}$$

Fig. 2 shows the relationship between α_i and the positions of the microphones. The β-divergence $D_\beta(A_v, A_i)$ between the amplitude A_v of the virtual microphone signal and the amplitude A_i of the real microphone signal of the ith channel is defined as

$$D_\beta(A_v, A_i) = \tag{14}$$
$$\begin{cases} A_v(\log A_v - \log A_i) + (A_i + A_v) & (\beta = 1) \\ \frac{A_v}{A_i} - \log \frac{\mathrm{A_v}}{\mathrm{A_i}} - 1 & (\beta = 0) \\ \frac{A_v^\beta}{\beta(\beta-1)} + \frac{A_i^\beta}{\beta} - \frac{A_v A_i^{\beta-1}}{\beta-1} & (\text{otherwise}). \end{cases}$$

Amplitude interpolation using β-divergence can be cast as an optimization problem defined as

$$A_v = \mathrm{argmin}_{A_v} \sum_{i=1}^{I} \alpha_i D_\beta(A_v, A_i). \tag{15}$$

, where the global optimum is its closed-form solution.

$$A_v = \begin{cases} \exp\{\sum_{i=1}^{I} \alpha_i \log A_i\} & (\beta = 1), \\ (\sum_{i=1}^{I} \alpha_i A_i^{\beta-1})^{\frac{1}{\beta-1}} & (\text{otherwise}). \end{cases} \tag{16}$$

It can be seen that this interpolation rule is a generalization of equation (5). Thus, the virtual microphone signal can be expressed using ϕ_v and A_v as

$$V(\omega, t, \Xi) = A_v \exp(j\phi_v), \tag{17}$$

where $\Xi = \{\alpha_i, \lambda_i\}_i$ is the set of all linear combinations of coefficients λ_i and weight coefficients α_i.

The proposed method can be easily extended to an arbitrary number of microphones in the space. In addition, the interpolation of amplitudes by the proposed method takes into account all the observations. Therefore, it can alleviate the microphone selection problem to some extent. Additionally, the conventional method requires the use of virtual microphone signals to create a virtual microphone at an arbitrary position. The validity of the virtual microphone interpolation using the virtual microphone signal is unknown. On the other hand, virtual microphone interpolations by the proposed method are based on observed signals in all, and there are no above problems.

IV. MAXSNR BEAMFORMER

In this study, we apply the virtual microphone technique to a maximum SNR beamformer to evaluate the performance of the virtual microphone technique[11][12]. A maximum SNR beamformer requires the covariance matrices of the target-active period and targetinactive period as prior information for speech enhancement. In a maximum SNR beamformer, the filter $\mathbf{w}(\omega)$ is designed to maximize the ratio $\lambda(\omega)$ of the power between the target active period and the target inactive period

$$\lambda(\omega) = \frac{\mathbf{w^H}(\omega)\mathbf{R_T}(\omega)\mathbf{w}(\omega)}{\mathbf{w^H}(\omega)\mathbf{R_I}(\omega)\mathbf{w}(\omega)} \tag{18}$$

Here, $\mathbf{R_T}(\omega)$,$\mathbf{R_I}(\omega)$ represent the covariance matrices of the target active period and target inactive period.

V. EVALUATION EXPERIMENT

A. Experimental summary

In experiments, we used four types of simulated impulse response and one type of measured impulse response in a car.

In the first experiment, virtual microphone signals were synthesized at the same coordinates for a triangular microphone array using the conventional method twice and using the proposed method. Using the synthesized signal and the three real microphone signals of the triangular microphone array, we performed underdetermined speech enhancement for four sound sources and compare the performance characteristics.

In the second experiment, the impact of the interpolation of the amplitude and phase on speech enhancement performance was investigated by comparing the virtual microphone signal in the case of two microphones with both phase and

TABLE I
POSITIONS OF SOUND SOURCES AND REVERBERATION TIMES OF SIMULATED IMPULSE RESPONSE

RT [ms]	Target	Interf 1	Interf 2	Interf 3	L
60	80°	40°	160°	130°	0.9m
120	90°	50°	145°	10°	1.2m
250	110°	20°	150°	70°	1.2m
400	70°	30°	100°	150°	1.2m

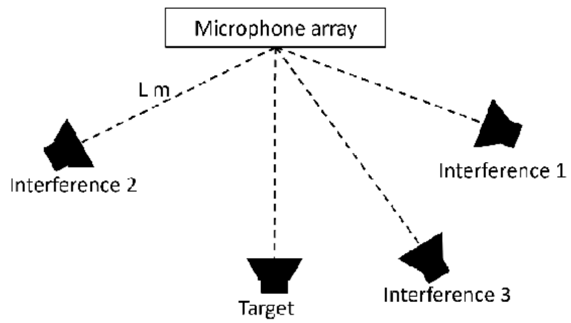

Fig. 3. Arrangement of sound sources in simulation

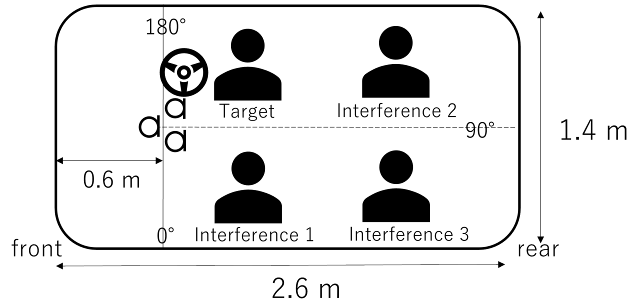

Fig. 4. Arrangement of sound sources in real car

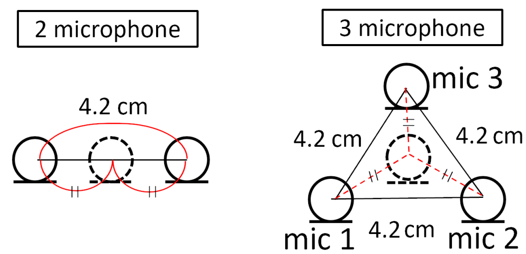

Fig. 5. Arrangement of microphones in simulation

amplitude interpolation, with only amplitude interpolation, with only phase interpolation, and without interpolation (real microphone signal).

In the final experiment, we investigated the impact of amplitude and phase interpolation on speech enhancement performance for a virtual microphone signal with three microphones in the same way.

B. Experimental conditions

We used the data of 10 speakers (six males and four females) of 503 phonemically balanced sentences in Set B of the ATR digital speech database [13]. We randomly selected 25 patterns of speaker combinations from this database and convolve four simulation impulse responses and one measured impulse response. Fig. 3 and Table I show the arrangement of the sound sources in the simulation. RT in Table I indicates the reverberation time. Fig. 4 shows the arrangement of the sources of the measured impulse response. Figs. 5 and 6 show the arrangements of the microphone array. We used Target, Interference 1, and Interference 2 in the experiment using three sound sources, and Target, Interference 1, Interference 2, and Interference 3 in the experiment using four sound sources. The hyperparameter β was 20 in the experiment with two real microphones and Conventional method with three microphones[5]. In the proposed method with three real microphones, β was 140. We used the signal-to-distortion ratio (SDR), source-to-interference ratio (SIR), and sources-to-artifacts ratio (SAR) as evaluation metrics. We used the maxSNR beamformer for speech enhancement. Table II shows the other experimental conditions.

C. Experimental results and discussion

1) Evaluation of Conventional and Proposed Methods: Table III shows the experimental results of the conventional and proposed methods for interpolating virtual microphones using three real microphones. conv((A,B),C) indicates the order of selection of real microphones in the conventional method. First, we used real microphones A and B to generate the first virtual microphone. Next, we used the generated virtual microphone and real microphone C to generate the final virtual microphone signal used for speech enhancement. In the simulation, the SDR of the conventional method was almost the same regardless of the order in which the real microphones were selected, and the proposed method showed the same SDR as the conventional method.

In the experiment using the measured impulse response, the SDR of the conventional method varied slightly depending on the order of selection of real microphones. In contrast, the proposed method showed an SDR roughly equal to the average of the results of the conventional methods. This result indicates that the proposed method eliminates the variation in speech enhancement performance due to the order of selection of real microphones in the conventional method. When the number of real microphones increases further, the number of patterns of the selection order increases further. Therefore, it is considered advantageous that a certain degree of speech enhancement performance can be expected uniquely.

2) Evaluation of Impact of Virtual Microphone Interpolation on Speech Enhancement Performance (three sound sources and two real microphones): Next, Table IV shows the experimental results of speech enhancement using two real microphone signals for three sound sources. R and V in the Phase and Amp. columns are the phase and amplitude of the signal at the virtual microphone position, denoted by R when

TABLE II
EXPERIMENTAL CONDITIONS

Sampling rate	8 kHz
Signal-to-noise ratio (SNR)	0 dB
FFT frame length	1024 samples
FFT shift	256 samples
Beamformer	maxSNR

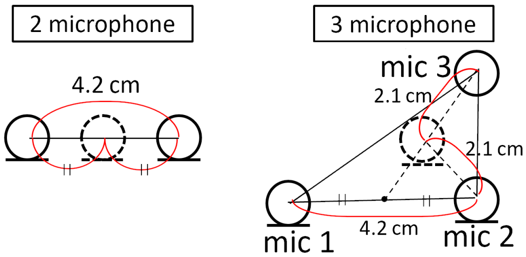

Fig. 6. Arrangement of microphones in real car

the real phase or amplitude are given and denoted by V when the virtual microphone technology interpolates the phase or amplitude.

When we use the virtual signal obtained by interpolating both the phase and the amplitude, a certain improvement in SDR was confirmed compared with the case of using only the signals from the two real microphones. When the interpolated amplitude and the real phase were given, there was almost no difference in SDR from the case where the virtual microphone technique interpolated both the phase and the amplitude. On the other hand, when we used the real amplitude and the interpolated phase, the SDR was improved by about 1.9 dB in the simulation and about 3.7 dB in the measured impulse response.

In the interpolation with two real microphones, the interpolated amplitudes were significantly different from the real amplitudes, and the amplitudes had a significant effect on the speech enhancement performance; providing the real amplitudes improved the speech enhancement performance significantly. On the other hand, there was a difference in SDR of about 11 dB in the simulation and about 6.7 dB in the measured impulse response between speech enhancement without interpolation (real microphone signal) and speech enhancement with only phase interpolation by a virtual microphone.

Therefore, it is necessary to further study the interpolation of the phase, which has not been studied much thus far. In addition, it is necessary to study factors other than the interpolation of the phase and amplitude itself, such as the consistency of the phase and amplitude.

3) Evaluation of the Impact of Virtual Microphone Interpolation on Speech Enhancement Performance (four sound sources and three real microphones): Table V shows the experimental results of speech enhancement using three real microphone signals for four sources. In interpolation with three real microphones, we confirmed that SDR was improved when using signals whose phase and amplitude were both interpolated using virtual microphone technology, compared with the case where we used only three real microphones. This result is similar to the experiment using two real microphones. On the other hand, when we use only amplitude interpolation and only phase interpolation, there was little difference in SDR from that when we interpolated both the phase and the amplitude by virtual microphone technology. This result suggests that it is necessary to consider aspects other than the interpolation of the phase and amplitude, such as the consistency of the phase and amplitude.

TABLE III
PROPOSED AND CONVENTIONAL METHODS IN SIMULATION

	Method	SDR	SIR	SAR
simu	input	-4.63	-4.63	274.57
	proposed	2.11	6.46	5.00
	conv((1,2),3)	2.10	6.44	5.01
	conv((1,3),2)	2.10	6.43	5.01
	conv((2,3),1)	2.10	6.43	5.01
real	input	-4.60	-4.60	261.54
	proposed	7.11	14.58	8.24
	conv((1,2),3)	6.91	14.29	8.04
	conv((1,3),2)	7.26	14.96	8.30
	conv((2,3),1)	7.26	14.98	8.30

TABLE IV
EXPERIMENT OF TWO MICROPHONES

	mic	Phase	Amp.	SDR	SIR	SAR
simu	input	-	-	-2.91	-2.91	273.4
	2mic	-	-	-0.94	0.44	7.95
	2mic+VM	V	V	0.46	2.62	6.61
	2mic+VM	R	V	0.70	3.01	6.51
	2mic+VM	V	R	2.46	5.05	7.30
	3mic	R	R	14.2	21.74	15.85
real	input	-	-	-2.25	-2.25	255.86
	2mic	-	-	2.63	6.07	6.37
	2mic+VM	V	V	4.81	10.45	6.86
	2mic+VM	R	V	5.52	11.51	7.34
	2mic+VM	V	R	8.75	16.07	9.97
	3mic	R	R	15.23	25.25	15.77

TABLE V
EXPERIMENT WITH THREE MICROPHONES

	mic	Phase	Amp.	SDR	SIR	SAR
simu	input	-	-	-4.21	-4.21	275.23
	3mic	-	-	1.57	5.68	4.87
	3mic+VM	V	V	2.04	6.46	5.00
	3mic+VM	R	V	2.17	6.71	5.02
	3mic+VM	V	R	2.08	6.41	5.09
	4mic	R	R	6.10	11.29	8.17
real	input	-	-	-4.60	-4.60	261.54
	3mic	-	-	5.79	12.59	7.16
	3mic+VM	V	V	7.11	14.58	8.24
	3mic+VM	R	V	7.12	14.93	8.13
	3mic+VM	V	R	7.31	15.00	8.40
	4mic	R	R	13.96	23.41	14.55

VI. CONCLUSIONS

In this paper, we proposed an extension of the virtual microphone technique to three or more real microphone signals in the same plane. The proposed method eliminates the dependence of speech enhancement performance on the order of selection of real microphones. In addition, we can easily increase the number of real microphones used for amplitude interpolation to an arbitrary number. The experimental results show that the proposed method improves speech enhancement performance in underdetermined situations. The performance of the proposed method is approximately the same as the average of the results of the conventional methods.

In addition, we investigated the effect of phase and amplitude interpolation using virtual microphone technology on speech enhancement performance. In the experiment, we found a significant difference in speech enhancement performance between the cases with and without interpolation (real microphone signal). On the basis of these results, we need to further study the interpolation method for the phase and amplitude. Other points that have not been examined also need to be examined, such as the consistency of the phase and amplitude.

VII. Acknowledgment

This work was supported by the Japan Society for the Promotion of Science (JSPS) KAKENHI through a Grant-in-Aid for Scientific Research under Grants 19H04131, and JST CREST JPMJCR19A3.

References

[1] O. Yılmaz and S. Rickard, “Blind separation of speech mixtures via time-frequency masking,” IEEE Trans. Signal Process., vol. 52, no. 7, pp. 1830–1847, 2004.

[2] N. Q. K. Duong, E. Vincent, R. Gribonval, “Under-determined reverberant audio source separation using a full-rank spatial covariance model,” IEEE Trans. Audio, Speech, and Lang. Process., vol. 18, no. 7, pp. 1830–1840, 2010.

[3] Y. Izumi, N. Ono, and S. Sagayama, “Sparseness-based 2ch BSS using the EM algorithm in reverberant environment,” Proc. WASPAA, pp. 147–150, 2007.

[4] H. Sawada, S. Araki, and S. Makino, “Underdetermined convolutive blind source separation via frequency bin-wise clustering and permutation alignment,” IEEE Trans. on Audio, Speech, and Language Processing, vol. 19, no. 3, pp. 516–527, 2011.

[5] H. Katahira, N. Ono, S. Miyabe, T. Yamada, S. Makino, “Nonlinear speech enhancement by virtual increase of channels and maximum SNR beamformer,” EURASIP Journal on Advances in Signal Processing, vol. 2016, no. 1, pp. 1–8, 2016.

[6] K. Yamaoka, S. Makino, N. Ono, T. Yamada, “Performance evaluation of nonlinear speech enhancement based on virtual increase of channels in reverberant environments,” in Proc. European Signal Processing Conference (EUSIPCO), pp. 2324–2328, 2017.

[7] T. Ochiai, M. Delcroix, T. Nakatani, R. Ikeshita, K. Kinoshita, S. Araki, ”Neural network-based virtual microphone estimator,” in Proc. ICASSP, pp. 6114-6118, 2021.

[8] K. Yamaoka, L. Li, N. Ono, S. Makino, T. Yamada, “CNN-based virtual microphone signal estimation for MPDR beamforming in underdetermined situations,“ in Proc. European Signal Processing Conference (EUSIPCO), pp. 1–5, 2019.

[9] R. Takahashi, L. Li, S. Makino, T. Yamada, “VMInNet: Interpolation of Virtual Microphones in Optimal Latent Space Explored by Autoencoder,“ in Proc. International Workshop on Nonlinear Circuits, Communications and Signal Processing(NCSP), pp. 93–96, 2021.

[10] A. Jourjine, S. Rickard, O. Yılmaz, “Blind separation of disjoint orthogonal signals: demixing n sources from 2 mixtures,” in Proc. ICASSP, pp. 2985–2988, 2000.

[11] H. L. Van Trees, “Optimum array processing,” John Wiley & Sons, 2002.

[12] S. Araki, H. Sawada, S. Makino, “Blind speech separation in a meeting situation with maximum SNR beamformers,” in Proc. IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), pp. 41–44, 2007.

[13] A. Kurematsu, K. Takeda, Y. Sagisaka, S. Katagiri, H. Kuwabara, K. Shikano, “ATR Japanese speech database as a tool of speech recognition and synthesis,” Speech Communication, vol. 9, no. 4, pp. 357–363, 1990.