

## 음성존재확률을 이용한 행렬식 기반 2채널 잡음제거기법

박진욱<sup>1</sup> · 홍정표<sup>2\*</sup>

### Determinant-based two-channel noise reduction method using speech presence probability

Jinuk Park<sup>1</sup> · Jungpyo Hong<sup>2\*</sup>

<sup>1</sup>Ph.D. student, School of the Electrical Engineering, Korea Advanced Institute of Science and Technology, Daejeon, 34141 Korea

<sup>2\*</sup>Assistant Professor, Department of Information and Communication Engineering, Changwon National University, Changwon, 51140 Korea

#### 요약

본 논문에서는 음성존재확률을 활용한 2채널 입력신호 상관행렬의 행렬식 기반 잡음제거 기법을 제안하였다. 제안한 기법은 음성존재확률을 이용해 기존의 행렬식 기반 2채널 잡음제거 기법의 위너 필터 이득을 음성과 잡음구간에 따라 적응적으로 조절함으로써 잡음제거 성능을 더욱 향상시키고자 하였다. 제안한 기법은 잡음 종류, 반향 조건, 신호 대잡음비, 잡음원의 개수와 방향이 다양한 모의 환경에서 객관적 평가 방법을 통해 평가되었다. 실험 결과는 대부분의 실험환경에서 행렬식 기반의 기법들이 위상차를 기반으로 한 기법들보다 성능이 우수했고 특히 제안한 음성존재확률을 이용한 행렬식 기반 잡음제거기법이 음성 신호 왜곡을 최소화하면서 가장 우수한 잡음 제거 성능을 보였다.

#### ABSTRACT

In this paper, a determinant-based two-channel noise reduction method which utilizes speech presence probability (SPP) is proposed. The proposed method improves noise reduction performance from the conventional determinant-based two-channel noise reduction method in [7] by applying SPP to the Wiener filter gain. Consequently, the proposed method adaptively controls the amount of noise reduction depending on the SPP. For performance evaluation, the segmental signal-to-noise ratio (SNR), the perceptual evaluation of speech quality, the short time objective intelligibility, and the log spectral distance were measured in the simulated noisy environments considered various types of noise, reverberation, SNR, and the direction and number of noise sources. The experimental results presented that determinant-based methods outperform phase difference-based methods in most cases. In particular, the proposed method achieved the best noise reduction performance maintaining minimum speech distortion.

**키워드** : 잡음제거, 행렬식, 음성 존재 확률, 2-채널

**Keywords** : Noise reduction, Determinant, Speech presence probability, Two-channel

Received 22 February 2022, Revised 1 March 2022, Accepted 12 March 2022

\* Corresponding Author Jungpyo Hong(E-mail: hansin@changwon.ac.kr, Tel:+82-55-213-3838)

Assistant Professor, Department of Information and Communication Engineering, Changwon University, Changwon, 51140 Korea

Open Access <http://doi.org/10.6109/jkiice.2022.26.5.649>

print ISSN: 2234-4772 online ISSN: 2288-4165

© This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License(<http://creativecommons.org/licenses/by-nc/3.0/>) which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.  
Copyright © The Korea Institute of Information and Communication Engineering.

## I. 서 론

음성 신호는 현대사회에서 사용자가 자동차나 모바일 환경 등에서 다양한 기능을 제어하기 위해 널리 사용되는 수단이다. 특히 모바일 시장이 기하급수적으로 확장됨에 따라 음성 신호를 활용하는 분야가 빠르게 증가하고 있다. 일반적으로 음성 신호는 주변 잡음에 의해 쉽게 오염되는데 이 경우 음성 인터페이스 전체 성능 하락에 치명적인 원인으로 작용하게 된다. 따라서 잡음 제거 기술은 음성인터페이스의 성능향상을 위해 필수적인 전처리 기술 중 하나로 예로부터 많은 연구가 이루어져왔다 [1].

우리가 일상에서 접하게 되는 잡음은 크게 정상성 잡음과 비정상성 잡음으로 나뉜다. 정상성 잡음은 시간에 따라 신호의 특성이 크게 변하지 않는데, 그렇기 때문에 워너 필터나 칼만 필터와 같은 단채널 잡음 제거 기술을 통하여 쉽게 제거가 가능하다[1]. 하지만 잡음원이 음성 또는 음악과 같은 비정상성 잡음일 경우 특성이 빠르게 변화하기 때문에 단채널 잡음 제거 기법으로는 제거가 어렵다. 이를 해결하기 위해 빔형성과 같은 다채널 잡음 제거기법이 연구되었다 [2-4]. 다채널 잡음제거 기법은 일반적으로 채널이 증가할수록 잡음제거 성능이 향상되나 센서의 수를 증가시키는 데 제약이 있다.

최근에는 딥러닝이 빠르게 발달하여 이를 적용한 잡음 제거 기법이 기존의 잡음 제거 기법에 비해 좋은 성능을 보여주었다[5-7]. 하지만 딥러닝 기반 잡음 제거 기법은 학습을 하기 위해 많은 데이터와 계산량이 필요하여 하드웨어의 제약으로 인해 적은 연산량이 요구되는 임베디드 시스템, 보청기 등에는 적용하기 어려운 한계점이 있다. 따라서, 이러한 경우에 적용하기 위한 2채널 잡음 제거 기법에 대한 연구가 활발하다[8-10].

2채널 잡음제거 기법 중 가장 대표적인 방법은 다음과 같다. 먼저, 위상오차필터(phase error filter, PEF) 기법은 2채널 입력 신호가 목적 신호 방향으로 조향(steering)했을 경우 입사한 2채널 입력 신호 간 위상차(phase difference)는 잡음에 의해 기인된다는 가정하에 제안된 잡음 제거 기법이다 [8]. 이후 PEF와 유사한 접근 방식으로 지향성 목적신호와 비목적신호의 비(Target-to-Non-target directional signal Ratio, TNR) 기반의 2채널 잡음제거 기법이 제안되었다 [9]. TNR은 고정 빔형성(Fixed BeamForming, FBF)의 전달함수와 차

단 행렬(Blocking Matrix, BM) 전달함수의 비로 도출될 수 있으며 이를 활용해 음성구간검출(Speech Activity Detection, SAD)이나 신호대잡음비(Signal-to-Noise Ratio, SNR)를 계산할 수 있다. 마지막으로 최근에 입력신호 상관행렬의 행렬식 분석을 통해 도출된 수식을 활용한 음성구간 검출 및 잡음제거 기법이 제안되었고 위 두가지 기법보다 개선된 잡음제거 성능을 보였다 [10].

따라서, 본 논문에서는 행렬식 기반 잡음제거 기법에 음성 존재 확률을 적용하여 음성 및 잡음구간에 따라 가변적으로 잡음제거도를 조절함으로써 잡음제거 성능을 더욱 향상시키고자 하였다. 제안한 잡음 제거 기법의 우수성을 검증하기 위해 잡음 종류, 반향 조건, 신호대잡음비, 잡음원의 개수 및 방향을 고려한 다양한 모의 실험 환경에서 평가하였다.

## II. 본 론

### 2.1. 신호 모델

본 논문에서는  $t$  번째 프레임의 2채널 입력 신호에 고속 푸리에 변환(Fast Fourier Transform, FFT)을 통하여 구한 스펙트럼 중  $k$  번째 성분을 각각  $X_1(t, k)$ 와  $X_2(t, k)$ 로 정의하면

$$X_1(t, k) = H_1(t, k)S(t, k) + N_1(t, k) \quad (1)$$

$$X_2(t, k) = H_2(t, k)S(t, k) + N_2(t, k) \quad (2)$$

와 같이 나타낼 수 있다. 여기에서  $S(t, k)$ 는 목적 신호의 스펙트럼,  $H_1(t, k)$ 와  $H_2(t, k)$ 는 음원과 마이크로폰 간의 전달함수, 그리고  $N_1(t, k)$ 와  $N_2(t, k)$ 는 각각 마이크로폰에 입력된 잡음의 스펙트럼을 말한다.

### 2.2. PEF 기법

PEF 기법은 2채널 입력 신호의 위상차를 활용하는 기법으로 2채널 배열이 목적신호방향으로 조향 되어있을 때, 위상차는 잡음에 의해 발생한다는 가정을 바탕으로 한다. 따라서 추정된 위상차는 신호대잡음비의 역과 비례한다고 간주하고 위상차를 통해 신호대잡음비 추정이 가능하다. 추정된 위상차를 이용하여  $t$ 프레임의  $k$  번째 PEF이득을 구하면

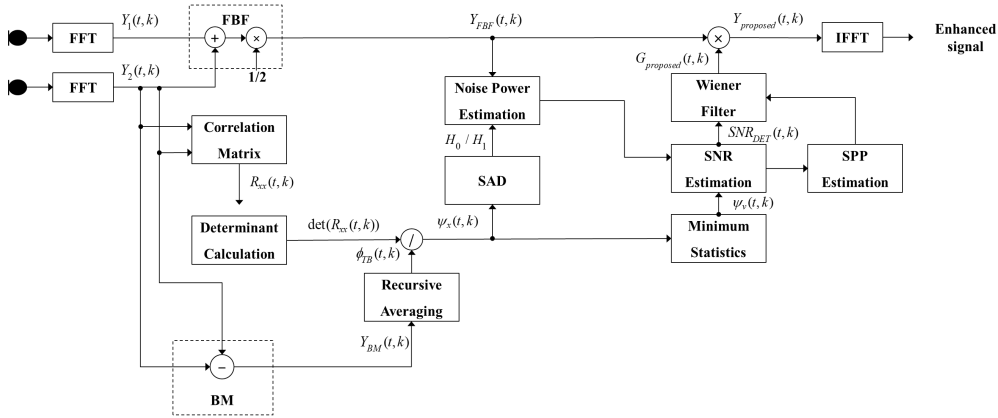


Fig. 1 Block diagram of proposed method

$$G^{PEF}(t,k) = \frac{1}{1 + r\phi^2(t,k)} \quad (3)$$

과 같다. 여기서  $r$ 은 잡음 제거도를 조절하는 변수로 값이 크면 잡음제거를 많이 하고 값이 0에 가까울수록 잡음제거를 적게 한다. 그리고 Eq. (3)의 위상차  $\phi(t,k)$ 는

$$\phi(t,k) = \angle X_1(t,k) - \angle X_2(t,k) \quad (4)$$

로 표현된다. 여기서  $K$ 는 FFT 크기의 절반,  $\angle$ 는 위상을 나타내는 기호이다.

### 2.3. TNR 기법

TNR은 지향성 목적 신호와 비목적 신호의 비로 2채널 배열이 목적신호 방향으로 조향 되었을 때 FBF의 전달 함수와 BM의 전달 함수의 비를 통해 계산할 수 있다.  $t$ 번째 프레임의  $k$ 번째 TNR은

$$\eta(t,k) = \frac{G^{EBF}(t,k)}{G^{BM}(t,k)} = \left| \frac{0.5(1 + \exp(j\phi^{norm}(t,k)))}{1 - \exp(j\phi^{norm}(t,k))} \right|^2 = \frac{1}{4} \frac{1 + \cos(j\phi^{norm}(t,k))}{1 - \cos(j\phi^{norm}(t,k))} \quad (5)$$

로 나타낼 수 있다. 여기서  $\phi^{norm}(t,k)$ 는

$$\phi^{norm}(t,k) = \phi(t,k) \times c / (d \times \omega_k) \quad (6)$$

와 같이 Eq. (4)를 정규화하여 얻을 수 있다. 또한,  $c$ 는 음속,  $d$ 는 마이크로폰 간의 거리,  $\omega_k$ 는  $k$ 번째 각 주파수를 의미한다. Eq. (5)의 TNR은 그 자체로서 음성구간검

출에 활용가능하고 다음과 같이

$$G^{TNR}(t,k) = \frac{\eta(t,k)}{\eta(t,k) + 1} \quad (7)$$

위너필터 이득의 입력으로 사용될 수 있다.

### 2.4. DET 기법

DET 기법은 입력신호 상관행렬의 행렬식을 통해서 유도된 수식을 활용하여 음성구간 검출 및 잡음 제거를 수행한다[7]. 입력신호 상관행렬  $R_{xx}(t,k)$ 의 행렬식은

$$\det(R_{xx}(t,k)) = P_S(t,k)P_{TB}(t,k) + \det(R_{nn}(t,k)) \quad (8)$$

과 같이 표현된다. 이 때  $R_{nn}(t,k)$ 는 잡음의 상관행렬이며  $P_S(t,k)$ 는 평균 음성 파워,  $P_{TB}(t,k)$ 는 평균 타겟 블록 파워를 의미하고 각각

$$P_S(t,k) = E[|S(t,k)|^2] \quad (9)$$

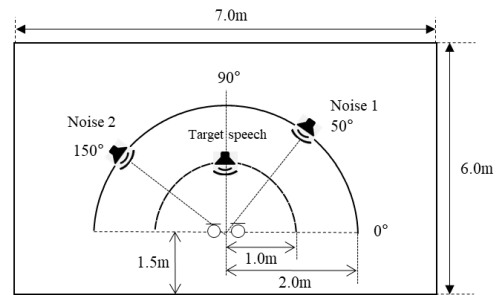


Fig. 2 The simulation room configuration

$$P_{TB}(t,k) = E[|X_1(t,k) - X_2(t,k)|^2] \quad (10)$$

와 같이 나타낼 수 있다. Eq. (8)을 활용하면 음성구간 검출을 통해 평균 음성 파워  $P_S(t,k)$ 와 평균 잡음 파워  $P_N(t,k)$ 를 구할 수 있고 이를 통해 구한 신호대잡음비와 추정된 위너필터 이득은 각각

$$SNR_{DET}(t,k) = \frac{P_S(t,k)}{P_N(t,k)} \quad (11)$$

$$G_{DET}(t,k) = \frac{SNR_{DET}(t,k)}{SNR_{DET}(t,k) + 1} \quad (12)$$

과 같이 표현된다.

### III. 제안한 방법

제안한 잡음제거 기법은 DET기법에 음성 존재 확률을 적용하여 음성 구간에서는 잡음제거도를 낮게 유지하고, 잡음 구간에서는 잡음제거도를 더 높게 가변적으로 조절하는 것을 목적으로 한다. 제안한 기법의 전체적인 블록도는 그림 1과 같다. 전반적으로 SPP 추정 블록과 위너필터 블록을 제외하고는 기존 DET기법의 블록도와 동일하다. 음성 존재 확률은 사전 신호대잡음비(*a priori* SNR)와 사후 신호대잡음비(*a posteriori* SNR)를 이용하여 얻을 수 있다 [11]. 잡음제거도를 가변적으로 조절하기 위해

$$G_{proposed}(t,k) = \frac{SNR_{DET}(t,k)}{SNR_{DET}(t,k) + \mu} \quad (13)$$

와 같이 변수 내장형 위너필터[1]를 사용하였다. 여기서  $\mu$ 값은 사용자 정의에 의해 잡음제거도를 조절하는 변수로 값이 클수록 잡음제거도가 크고 값이 0에 가까울수록 잡음제거도가 작다. 본 논문에서는  $\mu$ 값을 음성 존재확률의 역수로 할당하였고, 정리하면 최종적으로 제안하는 방법은

$$G_{proposed}(t,k) = \frac{p(t,k)SNR_{DET}(t,k)}{p(t,k)SNR_{DET}(t,k) + 1} \quad (14)$$

와 같다. 여기서  $p(t,k)$ 는  $t$ 프레임에서  $k$ 번째 음성 존재 확률을 의미한다.

## IV. 실험 환경 및 결과

### 4.1. 실험환경

본 논문에서는 제안한 잡음제거 기법을 평가하기 위해 음원과 마이크론 간의 임펄스응답 생성 기법인 Image Source method(ISM)를 이용해 실험 데이터를 합성하였다. 모의된 잡음환경은 그림 2와 같다. 방 중앙에 0.01m 간격의 마이크론 2개가 위치해있고 정면 1m 떨어진 곳에 음성 신호의 음원이 위치한다. 본 실험에서는 동시에 두 잡음원이 있는 실험 환경을 조성하였다. 마이크론과 잡음원 사이의 거리는 2m로 두 마이크론의 중심을 기준으로 각각 50도와 150도에 위치한다. 음성 신호는 TIMIT Database(DB)에서 무작위로 선정된 30문장을 사용하였고 잡음은 babble, white, factory1, volvo를 NOISEX-92 DB에서 추출하여 사용하였고 비정상적 잡음인 music에 대한 실험을 추가하기 위해 MUSAN DB를 활용하였다. 각각의 잡음 종류마다 다른 반향 조건(0ms, 150ms, 300ms)과 신호대잡음비 조건(0, 5, 10, 15, 20dB)에 대한 데이터를 합성하여 하나의 음성 문장당 75가지의 실험 데이터를 생성하였다. 실험 데이터 생성에 사용된 샘플링 주파수는 16 kHz이다.

### 4.2. 실험 결과 및 분석

성능 평가를 위해 segmental SNR[1], perceptual evaluation of speech quality (PESQ)[1], short time objective intelligibility (STOI)[12], log spectral distance(LSD)[1]의 총 네가지 객관적 평가 방법을 이용하여 제안한 기법과 본문에서 소개한 기존의 3가지 방법 및 optimally-modified multichannel wiener filter(OM-MWF)[13]와 성능을 비교·분석하였다. 표 1에는 잡음 종류별 평균 잡음 제거성능을 표로 정리하였다. 본 실험에는 2가지 정상성 잡음과 3가지 비정상성 잡음이 포함되어 있는데 두 가지 특성의 잡음에 대해

**Table. 1** Performance comparison by noise type : (a) Segmental SNR, (b) PESQ, (c) STOI, and (d) LSD

(a) Segmental SNR						
Noise type	Noisy input	PEF	TNR	DET	OM-MWF	Proposed
Babble	1.97	2.10	2.62	4.25	3.46	<b>4.62</b>
White	1.60	4.37	3.23	5.81	5.20	<b>6.19</b>
Factory1	1.82	2.15	2.36	4.58	3.61	<b>5.04</b>
Volvo	2.47	2.48	8.86	9.12	10.71	<b>11.19</b>
Music	2.43	2.69	4.07	7.00	5.80	<b>7.06</b>

(b) PESQ

Noise type	Noisy input	PEF	TNR	DET	OM-MWF	Proposed
Babble	2.55	2.62	2.70	2.85	2.45	<b>2.86</b>
White	2.34	2.53	2.56	2.78	2.68	<b>2.87</b>
Factory1	2.43	2.56	2.60	2.84	2.44	<b>2.89</b>
Volvo	3.55	3.57	3.66	3.81	3.66	<b>3.86</b>
Music	2.93	3.09	3.15	3.31	3.05	<b>3.31</b>

(c) STOI

Noise type	Noisy input	PEF	TNR	DET	OM-MWF	Proposed
Babble	0.82	0.83	0.79	0.85	0.78	<b>0.85</b>
White	0.87	0.88	0.81	0.89	0.81	<b>0.90</b>
Factory1	0.81	0.83	0.77	0.85	0.76	<b>0.85</b>
Volvo	0.94	0.94	0.92	0.95	0.95	<b>0.95</b>
Music	0.89	0.90	0.91	0.91	0.90	<b>0.91</b>

(d) LSD

Noise type	Noisy input	PEF	TNR	DET	OM-MWF	Proposed
Babble	1.24	1.07	1.07	0.94	1.09	<b>0.93</b>
White	2.51	1.85	1.59	1.62	1.54	<b>1.49</b>
Factory1	1.81	1.38	1.32	1.22	1.43	<b>1.19</b>
Volvo	0.71	0.72	0.71	0.60	0.62	<b>0.60</b>
Music	1.10	0.89	0.86	0.83	0.88	<b>0.82</b>

모두 제안된 기법의 성능이 가장 우수하였다. 특히 음성 존재확률을 활용할 경우 잡음의 특성과 무관하게 항상 잡음 제거 성능 향상에 효과적임을 확인 할 수 있었다.

**Table. 2** Performance comparison by reverberation time : (a) Segmental SNR, (b) PESQ, (c) STOI, and (d) LSD

(a) Segmental SNR

RT60 (ms)	Noisy input	PEF	TNR	DET	OM-MWF	Proposed
<b>0</b>	1.73	2.60	4.08	5.80	5.51	<b>6.66</b>
<b>150</b>	1.79	2.61	4.09	5.70	5.59	<b>6.55</b>
<b>300</b>	2.37	3.12	4.63	6.31	6.13	<b>7.06</b>

(b) PESQ

RT60 (ms)	Noisy input	PEF	TNR	DET	OM-MWF	Proposed
<b>0</b>	2.69	2.80	2.82	3.01	2.78	<b>3.05</b>
<b>150</b>	2.70	2.80	2.89	3.07	2.79	<b>3.12</b>
<b>300</b>	2.77	2.86	2.93	3.14	2.86	<b>3.20</b>

(c) STOI

RT60 (ms)	Noisy input	PEF	TNR	DET	OM-MWF	Proposed
<b>0</b>	0.87	0.88	0.83	0.89	0.83	<b>0.89</b>
<b>150</b>	0.86	0.87	0.83	0.89	0.83	<b>0.89</b>
<b>300</b>	0.85	0.86	0.82	0.88	0.81	<b>0.88</b>

(d) LSD

RT60 (ms)	Noisy input	PEF	TNR	DET	OM-MWF	Proposed
<b>0</b>	1.62	1.26	1.17	1.08	1.15	<b>1.03</b>
<b>150</b>	1.59	1.27	1.18	1.12	1.18	<b>1.06</b>
<b>300</b>	1.49	1.23	1.15	1.09	1.18	<b>1.05</b>

표 2는 3가지 반향 조건에 따른 평균 잡음 제거 성능을 나타낸다. 제안한 기법은 모든 반향 조건에서 가장 높은 성능을 보였는데 PEF나 TNR과 같은 위상차 기반의 잡음 제거 기법들은 반향이 심할수록 성능향상치가 하락하였지만 행렬식 기반의 기법들의 경우 오히려 반향이 심할 때 더 큰 성능향상을 확인할 수 있었다. 또한, 제안한 기법은 신호왜곡을 최소화하면서 잡음제거 성능을 향상시켰다.

**Table. 3** Performance comparison by SNR : (a) Segmental SNR, (b) PESQ, (c) STOI, and (d) LSD

(a) Segmental SNR

SNR (dB)	Noisy input	PEF	TNR	DET	OM-MWF	Proposed
<b>0</b>	-4.61	-3.93	-1.24	-0.64	-0.33	<b>0.54</b>
<b>5</b>	-1.90	-1.06	1.41	2.60	2.00	<b>3.55</b>
<b>10</b>	1.51	2.40	4.16	5.81	5.46	<b>6.61</b>
<b>15</b>	5.32	6.19	6.93	9.14	8.92	<b>9.82</b>
<b>20</b>	9.50	10.29	10.07	12.78	12.69	<b>13.27</b>

(b) PESQ

SNR (dB)	Noisy input	PEF	TNR	DET	OM-MWF	Proposed
<b>0</b>	2.07	2.18	2.12	2.39	2.08	<b>2.40</b>
<b>5</b>	2.38	2.49	2.54	2.76	2.49	<b>2.76</b>
<b>10</b>	2.72	2.82	2.92	3.10	2.87	<b>3.10</b>
<b>15</b>	3.05	3.15	3.26	3.41	3.17	<b>3.41</b>
<b>20</b>	3.38	3.46	3.56	3.68	3.44	<b>3.69</b>

(c) STOI

SNR (dB)	Noisy input	PEF	TNR	DET	OM-MWF	Proposed
0	0.70	0.72	0.65	0.75	0.59	<b>0.75</b>
5	0.80	0.82	0.76	0.84	0.74	<b>0.84</b>
10	0.88	0.89	0.85	0.91	0.87	<b>0.91</b>
15	0.94	0.94	0.91	0.95	0.94	<b>0.95</b>
20	0.97	0.97	0.95	0.98	0.98	<b>0.98</b>

(d) LSD

SNR (dB)	Noisy input	PEF	TNR	DET	OM-MWF	Proposed
0	2.06	1.70	1.57	1.55	1.62	<b>1.44</b>
5	1.85	1.48	1.33	1.28	1.38	<b>1.21</b>
10	1.58	1.24	1.13	1.06	1.13	<b>1.01</b>
15	1.30	1.02	0.98	0.87	0.93	<b>0.85</b>
20	1.05	0.83	0.84	0.72	0.78	<b>0.71</b>

표 3에서는 SNR 별 평균 잡음 제거 성능을 정리하였다. 모든 SNR 조건에서 제안한 기법의 성능이 가장 우수하였다. 특히 낮은 SNR 환경에서 DET와 제안한 기법의 성능 차이가 큰 것을 확인할 수 있었다. Segmental SNR, PESQ, STOI를 통해 잡음의 파워가 클수록 제안한 잡음 제거 기법의 효과가 높음을 알 수 있었고 LSD 결과를 통해 동시에 신호왜곡은 줄일 수 있음을 확인하였다.

## V. 결론

본 논문에서는 음성존재확률을 입력신호 상관행렬의 행렬식 기반 2채널 잡음제거 기술에 활용한 새로운 기법을 제안하였다. 실험 결과 행렬식 기반의 잡음제거 기법은 정상성 잡음과 비정상성 잡음에서 모두 위상차 기반 기법인 PEF나 TNR보다 우수한 잡음 제거 성능을 보였고, 또한 낮은 SNR 환경이나 반향이 심한 열악한 잡음 환경에서 더 높은 성능향상치를 보여주었다. 또한 제안한 기법은 잡음 환경이 열악할 때 음성왜곡을 더욱 줄이면서 DET보다 높은 잡음제거 성능을 보였다는 점에서 의미가 크다고 할 수 있다. 향후에는 행렬식 기반 잡음 제거 기법을 다채널로 확장하는 연구를 수행할 예정이다.

## ACKNOWLEDGEMENT

This research was supported by Changwon National University in 2021~2022.

## REFERENCES

- [ 1 ] P. C. Loizou, *Speech Enhancement*, Boca Raton, FL, USA: CRC Press, 2007.
- [ 2 ] O. Schwartz, S. Gannot, and E. A. P. Habets, "Multispeaker LCMV beamformer and postfilter for source separation and noise reduction," *IEEE Transaction Audio Speech Language Processing*, vol. 25, no. 5, pp. 940-951, May. 2017.
- [ 3 ] Y. Kubo, T. Nakatani, M. Delcroix, K. Kinoshita, and S. Araki, "Mask-based MVDR beamformer for noisy multisource environments: Introduction of time-varying spatial covariance model," in *Proceeding of the IEEE International Conference on Acoustics, Speech and Signal Processing*, Brighton, UK, pp. 6855-6859, 2019.
- [ 4 ] P. Rakesh, S. S. Priyanka, and T. K. Kumar, "Performance evaluation of beamforming techniques for speech Enhancement," in *Fourth International Conference on Signal Processing Communication and Networking*, Chennai, India, pp. 1-5, 2017.
- [ 5 ] J. Kim and M. Hahn, "Speech enhancement using a two-stage network for an efficient boosting strategy," *IEEE Signal Processing Letter*, vol. 26, no. 5, pp. 770-774, May. 2019.
- [ 6 ] J. Lee and H. -G. Kang, "A joint learning algorithm for complex-valued T-F masks in deep Learning-based single-channel speech enhancement systems," *IEEE/ACM Transaction*, vol. 27, no. 6, pp. 1098-1108, Jun. 2019.
- [ 7 ] D. Wang and J. Chen, "Supervised speech separation based on deep learning: An overview," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 26, no. 10, pp. 1702-1726, Oct. 2018.
- [ 8 ] P. Aarabi and G. Shi, "Phase-based dual-microphone robust speech enhancement," *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)*, vol. 34, no. 4, pp. 1763-1773, Aug. 2004.
- [ 9 ] S. M. Kim and H. K. Kim, "Direction-of-arrival based SNR estimation for dual-microphone speech enhancement," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 22, no. 12, pp. 2207-2217, Dec. 2014.

- [10] J. Hong, S. Park, S. Jeong, and M. Hahn, "Dual-microphone noise reduction in car environments with determinant analysis of input correlation matrix," *IEEE Sensors Journal*, vol. 16, no. 9, pp. 3131-3140, May. 2016.
- [11] I. Cohen, "Optimal speech enhancement under signal presence uncertainty using log-spectral amplitude estimator," *IEEE Signal Processing Letters*, vol. 9, no. 4, pp. 113-116, Apr. 2002.
- [12] C. H. Taal, R. C. Hendriks, R. Heusdens, and J. Jensen, "A short-time objective intelligibility measure for time-frequency weighted noisy speech," in *Proceeding of the IEEE International Conference on Acoustics, Speech and Signal Processing*, Dallas, TX, USA, pp. 4214 - 4217, 2010.
- [13] S. Jeong and Y. Kim, "An Optimally-Modified Multichannel Wiener Filter Using Speech Presence Probability," *Smart Media Journal*, vol. 7, no. 3, pp. 9-15, Sep, 2018.



**박진욱(Jinuk Park)**

2015년 ~ 현재 한국과학기술원 전기및전자공학부 석·박사 통합 과정  
※관심분야: 음성 신호 처리, 음향 이벤트 분류, 인공지능



**홍정표(Jungpyo Hong)**

2016년 한국과학기술원 전기및전자공학부 공학박사  
2016년 ~ 2020년 국방과학연구소 선임연구원  
2020년 ~ 현재 창원대학교 정보통신공학과 조교수  
※관심분야: 마이크로폰 배열 기반 음성 향상, 소나 신호 처리, 압축 센싱, 능동 소음 제거, 인공 지능