

# MEMS 마이크로폰 설치공간의 주파수응답 보상을 통한 음성인식 개선

김현갑\*, 김규식\*\*

## Improved Speech Recognition by Compensating for Frequency Response in the MEMS Microphone Installation Space

Hyun-Kab Kim\*, Gyu-Sik Kim\*\*

### 요 약

최근 통신 및 멀티미디어 기기에는 MEMS 마이크로폰이 탑재된다. 작은 크기 대비 우수한 음향 특성으로 폭 넓게 사용되고 있으며, 음성인식 기능을 구현하는데 있어 중요한 요소이다. 그러나 기존의 연구들은 실제 공간 특성에 대한 효과를 고려하는 것이 부족하다. 마이크로폰 센서는 제품 내부의 작은 챔버에 설치되는데, 이 챔버의 공간특성에 대한 파악이 필요하다. 본 연구에서는 설치되는 챔버의 공간 특성으로 인해 변화되는 주파수응답을 파악하고 이를 보상함으로써 음성인식을 개선하고자 했다. 이를 위해 용적 설계 실험을 통해 주파수 응답 필터를 산출하고, 딥 러닝 AI 알고리즘 API를 사용해 문장단위의 음성을 Speech to text 방식으로 인식하여 필터 적용에 대한 음성인식 개선을 확인하였다. 이 결과를 통해 환경에 대한 필터 보상이 음성 인식 개선에 추가적인 도움을 줄 수 있는 것을 파악할 수 있다.

### Abstract

Recently, communication and multimedia devices are equipped with MEMS microphones. It is widely used due to its excellent acoustic characteristics compared to its small size, and is an important factor in implementing the voice recognition function. However, existing studies are insufficient to consider the effect on actual spatial characteristics. The microphone sensor is installed in a small chamber inside the product, and it is necessary to understand the spatial characteristics of this chamber. In this study, we attempted to improve speech recognition by identifying and compensating for the frequency response that changes due to the spatial characteristics of the installed chamber. To this end, a frequency response filter was calculated through a volumetric design experiment, and speech recognition improvement for filter application was confirmed by recognizing sentence-level speech in a speech to text method using a deep learning AI algorithm API. From this result, it can be seen that filter compensation for the environment can additionally help improve speech recognition.

### Keywords

acoustic control, microphone, sound sensor, system control, voice recognition

---

\* 서울시립대학교 전자전기컴퓨터공학과 박사과정 · Received: Nov. 16, 2020, Revised: Dec. 22, 2020, Accepted: Dec. 25, 2020  
- ORCID: <https://orcid.org/0000-0002-4098-0504> · Corresponding Author: Gyu-Sik Kim  
\*\* 서울시립대학교 전자전기컴퓨터공학과 교수 Dept. of Electrical and Computer Engineering, University of Seoul,  
(교신저자) 163, Seoulsiripdae-ro, Dongdaemun-gu, Seoul, Korea,  
- ORCID: <https://orcid.org/0000-0001-7868-4527> Tel.: +82-2-6490-2324, Email: gskim318@uos.ac.kr

## 1. 서 론

최신 디지털 시스템 기술의 발전은 다양한 영역에서 센서를 적극적으로 활용하고 있다. 음향 센서는 마이크로폰이라고 불리는 음향 부품으로 음성인식 및 환경 감지, 통신, 멀티미디어 등에서 적극적으로 사용되고 있다. 특히 인공지능 및 신호처리 기술의 발전으로 스마트폰이나 태블릿 같은 모바일 기기뿐만 아니라, TV, 에어컨, 셋톱박스와 같은 생활 가전까지 IoT를 기반으로 하는 음성인식 기술이 보급되고 있다. MEMS(Micro Electro Mechanical System) 마이크로폰은 수 mm 정도의 작은 크기이기 때문에 물리적으로 작은 공간만 있으면 설치할 수 있다. 또한, 작은 크기 대비 높은 감도와 우수한 동작 범위를 가지고 있어 고품질의 입력신호를 확보하기에 좋다[1].

현재 기존에 많이 사용해 오던 다이내믹(Dynamic) 마이크로폰이나 ECM(Electrets Condenser Microphone) 마이크로폰은 MEMS 마이크로폰으로 점차 교체되어 가고 있다. 그러나 MEMS 마이크로폰에 대한 세부적인 연구는 부족한 편이다. 기존에 이루어진 MEMS 마이크로폰의 특성 관련 연구는 대부분 유닛 자체의 음향 특성을 개선하는 쪽으로 초점이 맞추어지고 있다[2]. 실제로 부품이 탑재되는 공간에 대한 고려를 하지 않고 있기 때문에, 마이크로폰 부품을 제품 내부에 탑재하게 되었을 때의 공간 특성을 반영하지 못한다.

본 논문에서는 다음과 같은 방법으로 MEMS 마이크로폰의 공간 정보에 대한 파악과 그에 대한 보상을 연구하였다. 먼저 마이크로폰 유닛이 설치되는 공간에 대하여 조사하여 용적을 모사하였다. 모사한 공간 용적을 음원 파장이 입사되는 hole과 연결하여 실험용 챔버 블록을 제작하였다. 제작한 챔버 블록을 사용하여 실험을 통해 공간의 주파수 응답을 산출하였다[3]. 다음으로 산출된 주파수 응답을 사용하여 일반적인 DSP 기준에 맞는 보상 특성을 디지털 필터로 제작했다. 이 필터를 실제 음원에 적용하여 AI 음성인식 시스템에서의 인식을 개선도를 평가하였다.

이 연구를 통해 MEMS 마이크로폰의 공간 환경

을 산출해줌으로써 실제 하드웨어의 공간 환경을 시뮬레이션할 수 있도록 제안할 수 있으며, 음향 부품의 개별 성능으로는 검출할 수 없었던 환경적인 성능 감소 요인을 추출하여 최적설계에 도움을 줄 수 있을 것으로 예상된다.

## II. 관련 연구 및 특성 분석

### 2.1 MEMS 마이크로폰의 특성

그림 1은 정전용량 방식의 MEMS 마이크로폰을 간략화한 기본 구조이다. 제조사에 따라 소리가 입사되는 입력 포트의 위치나 센서 부품의 세부적인 형태가 달라지지만, 전체적인 작동 구조는 크게 달라지지 않는다. 입력 포트를 통해 소리가 센서 부분으로 입사되면, 백 플레이트 뒤의 얇은 진동박막이 진동하면서 연결되어 있는 전극의 정전용량의 변화를 ASIC(Application Specific Integrated Circuit)으로 전달한다. ASIC은 전달받은 신호를 증폭하여 아날로그 오디오 신호로 기본 출력하는데, 디지털 방식의 출력이 필요한 경우 컨버터를 통해 아날로그 오디오 신호를 다시 PDM(Pulse Density Modulation)으로 변환하여 출력한다.

이처럼 MEMS 마이크로폰이란 음압을 입력 받아 감지하는 진동 박막을 비롯한 기계적인 센서 시스템을, 반도체 공정을 통한 미세공정의 실리콘 웨이퍼를 통해 구현화한 제품이다[4]. MEMS 마이크로폰은 크게 정전용량 방식과 압전소자 방식으로 나누어지는데, 여기에서는 가장 일반적인 방식인 정전용량 방식을 대상으로 한다.

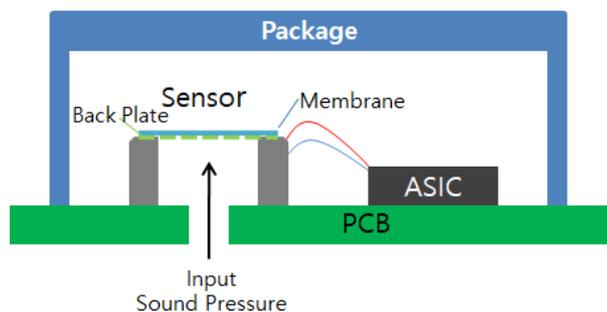


그림 1. MEMS 마이크로폰의 기본 구조  
Fig. 1. Basic structure of a MEMS microphone

일반적으로 12V ~ 48V의 구동전압이 필요한 기존의 마이크로폰 유닛에 비해 3V 내외의 낮은 구동전압을 가지고 있어 구동에 유리하면서도 감도가 높아 배터리를 사용하는 장치에 적합하다[5]. 그러나 설치되는 제품들의 크기가 점점 줄어들고 내부용적이 치밀해짐에 따라 하드웨어의 구조적인 영향을 크게 받게 되어, 제조사들에게 성능을 유지하기 위한 최적화 작업을 어렵게 만들고 있다.

## 2.2 설치공간의 특성

그림 2는 PC를 통해 마이크로폰 설치공간 용적을 묘사한 챔버 블록을 모델링한 뒤, 3D 프린터를 사용하여 출력한 실제 실험용 샘플이다. 챔버 블록의 형상은 스마트폰 및 노트북 등의 얇고 가벼운 제품에 적용되는 환경을 고려하여 얇고 작게 설계하였다. MEMS 마이크로폰이 장착되는 하단 패널을 부착하면 20mm×20mm×3.5mm의 공간이 형성되고 홀의 사이즈는 직경 1mm, 1.5mm, 2mm의 3종류로 제작하였다.



그림 2. 마이크로폰 설치 공간의 용적 재현  
Fig. 2. Reproduction of microphone installation space

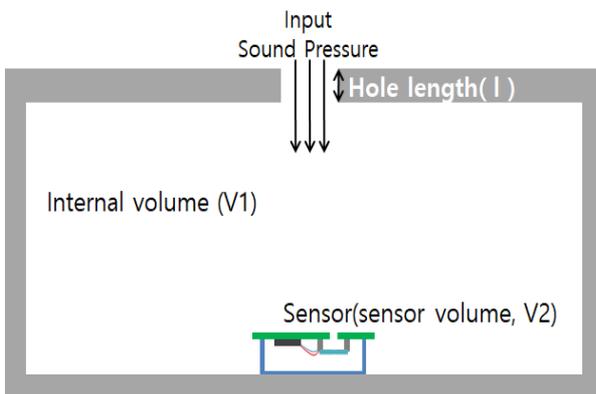


그림 3. 실험용 챔버 블록의 설치 구조도  
Fig. 3. Installation structure diagram of experimental chamber block

그림 3은 실험용 챔버 블록의 구조도이다. 실제로 MEMS 마이크로폰은 위의 구조와 비슷한 환경으로 설치되는데, 이 구조는 좁은 홀과 내부의 균일한 체적을 지니는 헬름홀츠 공명 구조와 흡사한 형태이다. 때문에 헬름홀츠 공명 구조에서의 공진 주파수를 구하는 방식을 응용하면, 해당 Chamber 공간의 최대 공진 주파수를 산출하여 실측한 실험 결과와 비교하여 유효성을 검증할 수 있다[6].

수치 산출을 위해 그림 3의 구조를 입력 홀의 단면적을 A로 홀의 길이를 l로, 내부의 체적을  $V(V=V1 - V2)$ 로 정리한다. 여기서 체적 V로 초당 유입되는 공기질량(Qkg/s)에 의해 변화하는 홀 밖의 압력 변화 값을  $P_o^{(t)}$ 라 하고, 체적 안에서 변화하는 압력변화 값을  $P_i^{(t)}$ 라 한다. 이 관계에서 홀이 공진할 때 입력부의 압력변화는 아주 작고 공기 유출입량이 최대가 된다. 따라서 홀 입구의 압력변화인  $P_o^{(t)}$ 가 최소가 되는 주파수를 찾는 것이 최대 공진지점 된다.

$$\omega_n = \frac{c}{2\pi} \sqrt{\frac{A}{Vl}} \quad (1)$$

음향 물리량의 변화를 정리하면 위의 식 (1)이 되며,

$$f = \frac{c}{2\pi} \sqrt{\frac{A}{V(V1 - V2)l}} \quad (2)$$

그림 3의 구조도를 반영하면 식 (2)가 되는데, 홀 입구의 압력변화가 최소가 되는 주파수를 구하는 것이다. 위의 식에서 c는 음속을 나타내는데, 이 실험에서 음속은 실험 환경에 대한 변수로 나타나기 때문에 보정 값으로 계산해야 한다.

$$c = \sqrt{\frac{\gamma RT}{M}} = 331.5 + (0.6 \times \theta) \text{ m/s} \quad (3)$$

식 (3)은 음속에 대한 보정 값을 정리한 식이다. 이를 통해 실험실의 공조장치를 통해 설정된 22°C의 공진 주파수 값을 산출해낼 수 있다. 산출한 음속을 챔버 블록의 구조식에 대입하면, 홀 직경에 따른 공진 주파수를 예측 산출 할 수 있다.

### III. 실험 및 평가

#### 3.1 주파수 응답 실험 및 산출

그림 4는 시험에 사용된 MEMS 마이크로폰 시료이다. 하단 포트 타입 정전용량 아날로그 마이크로폰 유닛이며, 구동 전압으로 3V를 공급하였다. 실험을 위해 AudioPrecision APx555 신호 분석기, B&K 4191 레퍼런스 마이크로폰, Focal CHORUS-V 806V 레퍼런스 스피커, B&K 2690 컨디셔닝 앰프, Crown XLi 2500 파워 앰프, Agilent E3632A DC 전원 공급기가 사용되었다.

그림 3의 구조와 같이 MEMS 마이크로폰 시료를 설치하여 마이크로폰의 주파수 응답을 측정하였다. 주파수 응답 측정을 위해 무향 음장을 형성하는 ISO 3745 규격의 완전 무향실에서 공조기를 사용해 목표 온도 22 °C를 유지하여 실험하였다. 측정 방법은 표준 규격 IEC 60268-4 기반의 레퍼런스 마이크로폰을 이용한 마이크로폰 측정법을 이용하였다[7][8].

그림 5가 홀의 직경에 따른 챔버 블록의 공간효과가 나타나는 주파수 응답 결과이다. 실선이 1mm, 굵은 점선이 1.5mm, 가는 점선이 2mm 홀의 응답 특성이다. 각 그래프의 공진점이 공간 특성의 수식

을 통해 산출한 값과 비슷한 대역에 위치한 것을 확인 할 수 있다. 챔버 블록의 제작상태에 문제가 있거나 MEMS 마이크로폰이 오작동 하는 경우 산출된 주파수와 실측 주파수가 크게 차이 나게 되어, 설치 공간의 유효성을 확인 할 수 있다.

#### 3.2 음성 인식 평가

MEMS 마이크로폰의 공간을 모사하는 챔버 블록의 공간 효과에 의해 변화한 주파수 응답 특성 영향과 이에 대한 보상 필터를 적용하여 음성인식을 평가하기위해 시험 신호의 입력력 경로에 디지털 필터를 적용한다. 신호의 출력단은 공간 특성을 나타내는 주파수 응답 필터를 적용하고, 입력단은 중요 대역을 역상으로 보정해주는 보상 필터를 적용한다.

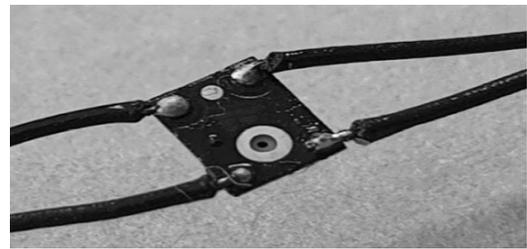


그림 4. 실험용 샘플  
Fig. 4. Test sample

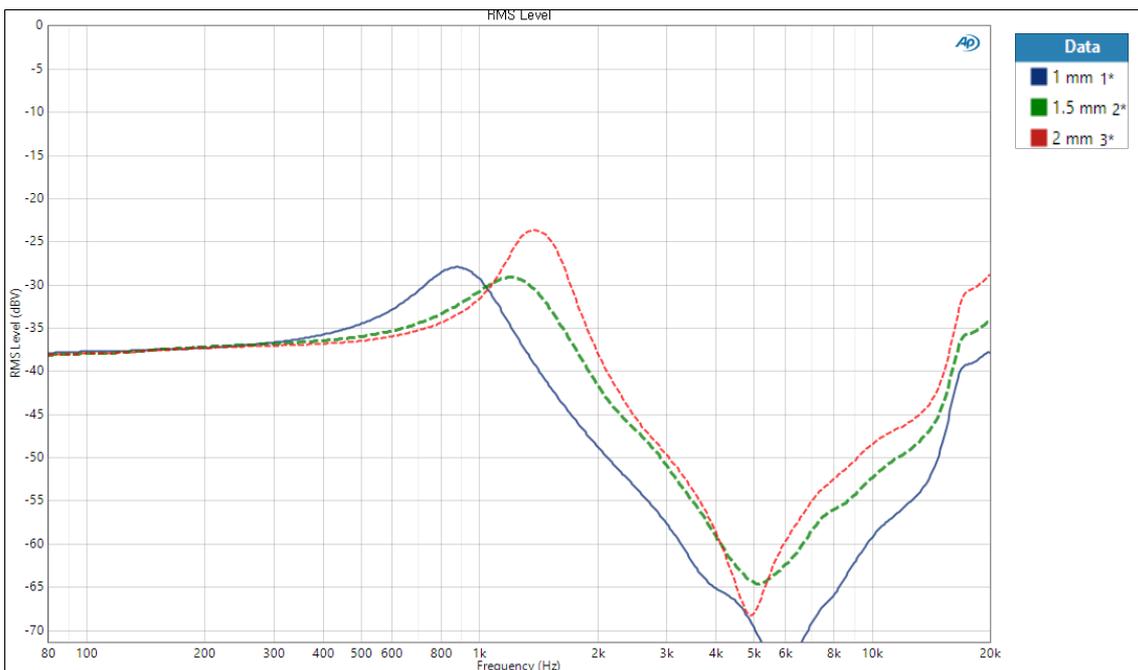


그림 5. 챔버 블록에 따른 공간효과 주파수 응답  
Fig. 5. Frequency response according to chamber block

그림 6이 중요 대역에 대한 주파수 응답을 보상해 주는 디지털 필터이다. DSP 적용 범위와 대역폭을 고려하여 제작되었다. 실선이 1mm, 굵은 점선이 1.5mm, 가는 점선이 2mm 홀에 대한 보상 필터이다. 이 보상 필터의 적용 유무에 따라 AI 음성인식 시스템에서 음성 인식 개선 정도를 확인해 볼 수 있을 것이다.

AI를 사용한 음성인식 결과를 산출하기 위해 구글(Google)에서 제공하는 Speech to text API 모듈을 사용하였다. 평가를 위한 음원은 한국어 날씨 뉴스에서 문장 단위로 추출하였으며, 총 25개의 문장을 평가에 사용하였다. 뉴스 음원은 단어 표현과 발음이 표준적이기 때문에 평가용 음원으로 적합하다.

산업 표준에 가까운 평가 결과 도출을 위해 MEMS 마이크론 제조사에서 제공한 일반적인 마이크론의 평균 홀 사이즈를 반영하여, 공간 효과 및 주파수 보상 필터는 가장 일반적인 2mm 홀 환경을 선택하였다. Speech to text는 음원에서 인식한 정보를 텍스트로 표시하는 것으로 음성 인식을 계산하기에 적당하다. 구글의 음성 인식은 기본적으로 심층신경망 이론을 기반으로 있으며, 자연어 처리 기법을 지원하고 있다. 이것은 유튜브의 자동 자막 생성 기능과 동일한 형태의 기능이다.

표 1은 시험 음원의 음성 인식 결과를 텍스트로 산출하기 위해 데이터를 요청한 코드이다. 파이썬을

사용하여 요청 코드를 보내면, 구글 서버에서 빅데이터와 연계하여 결과를 산출하여 보내준다. 보상 필터의 적용 유/무에 따라 2개의 음원에서 음성 인식 결과를 산출하여 원본과 비교해 볼 수 있다. 음원의 종류가 짧은 오디오 클립인 것을 감안하여 적용 학습 모델은 command\_and\_search를 사용하였다.

25개의 문장에 대해 공간 보상 필터의 적용 유/무에 따라 50개의 학습결과를 도출하였다. 이 결과를 원본 음원의 텍스트와 비교해 글자별로 오인식된 수를 산출하여, 문장단위 및 전체 음원 단위에 대한 오류 글자를 기술 통계로 처리하여 음성 인식의 개선 여부를 정량화 하였다.

표 1. 음성 인식 API 요청 코드  
Table 1. Speech recognition API request code

```
{
  "audio": {
    "content": "/* NEWS */"
  },
  "config": {
    "enableAutomaticPunctuation": true,
    "encoding": "LINEAR16",
    "languageCode": "ko-KR",
    "model": "command_and_search"
  }
}
```

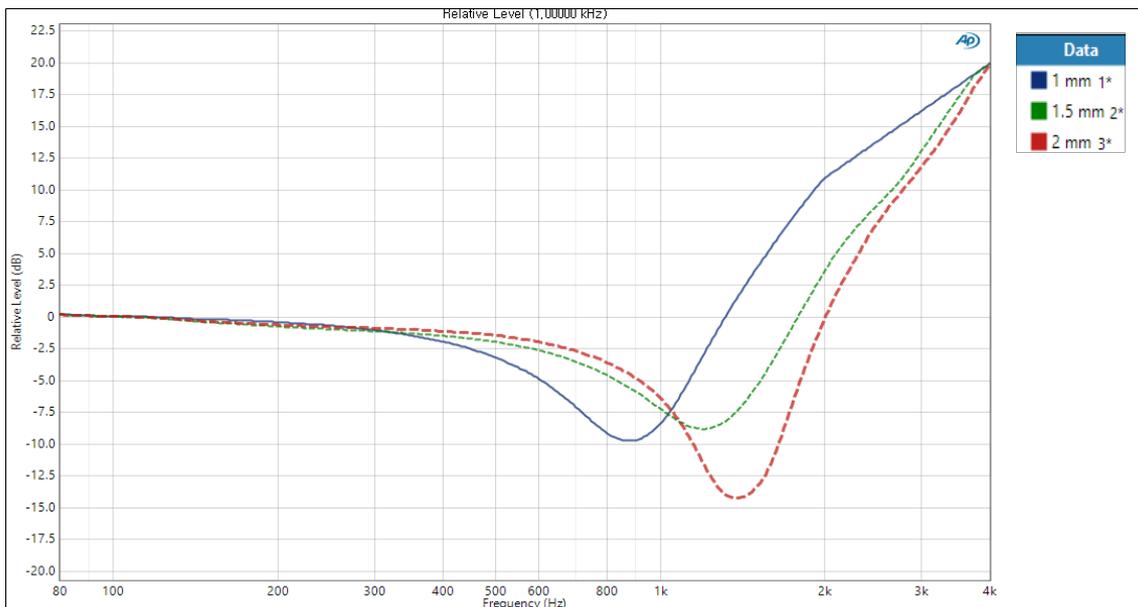


그림 6. 공간 효과 보상 필터  
Fig. 6. Spatial effect compensation filter

표 2가 학습 모듈을 통해 음성 인식된 결과를 텍스트로 산출한 다음 원본과 대조하여 문장별로 오인식된 글자를 산출한 것이다. 25개 문장에 대하여 공간 보상 필터의 적용에 유/무에 따른 오인식 글자를 Descriptive Statistics를 사용해 통계 분석(소수점 둘째자리까지 표기)하면 다음의 표 3과 같다.

표 2. 문장별 음성인식 결과  
Table 2. Speech recognition result by sentence

No.	Error letter		No.	Error letter	
	No filter	Apply filter		No filter	Apply filter
1	3	0	14	5	4
2	3	3	15	3	2
3	8	6	16	0	0
4	6	6	17	0	0
5	0	1	18	11	6
6	6	3	19	6	5
7	0	0	20	0	0
8	9	3	21	4	3
9	7	6	22	5	2
10	10	5	23	0	1
11	5	1	24	2	1
12	5	2	25	3	2
13	0	0			

표 3. 인식 오류 글자에 대한 기술 통계  
Table 3. Descriptive statistics on error letter

Statistics	No filter	Apply filter
Sum	101	62
Mean	4.04	2.48
Standard error	0.67	0.43
Median	4	2
Standard deviation	3.36	2.16
Sample variance	11.29	4.67
Skewness	0.37	0.47
Range	11	6
Confidence level (95.0%)	1.38	0.89

표 3의 통계 내용을 살펴보면, 필터의 적용으로 인해 인식 오류가 발생하는 문자의 총합이 101개에서 62개로 감소하였으며, 평균값 역시 4.04에서 2.48로 감소하였다. 표준 오차 및 표준 편차 역시 감소하였고 중앙값의 경우 절반으로 줄어들었다. 신뢰수준은 정규분포 95% 기준에서 1.38에서 0.89로 줄어들어 평균값의 신뢰 구간이 좀 더 모이게 되었다. 수치 전반적으로 오류가 줄어들어 인식률이 상승했

다는 사실을 정량적으로 확인할 수 있다. 평균값만 보더라도 약 38.61% 정도의 인식률 상승을 나타내고 있고 문장 개별로도 인식 증가를 식별할 수 있다.

이 결과는 다중 배열 센서를 활용하는 하드웨어적인 신호처리 방법과 다르게, 설계 레벨에서 산출할 수 있는 설치 공간의 특성을 보상해 주는 필터로도 음성 인식률을 보정해 줄 수 있음을 확인해준다. 다중 배열을 위해서는 더 많은 숫자의 MEMS 마이크론 유닛이 필요한데, 대부분의 보급형 제품 설계에 있어서는 사용하기 어려운 방법이다. 또한 제품 설계 완료 후 인식 저해 요소를 줄이는 최적화 과정에 적용할 경우 효과가 높다고 할 수 있다.

#### IV. 결론 및 향후 과제

지금까지 MEMS 마이크론이 하드웨어적인 구조물에 설치되었을 때, 공간 특성으로 인한 주파수 변화를 디지털 필터를 통해 보상하여 명료도를 개선하는 것에 대하여 정리하였다. 본 연구의 결과는 앞서 서술한 바와 같이 크게 2가지의 기술적인 의미를 나타내고 있다. 첫 번째는 MEMS 마이크론 유닛의 단품 측정과 분리하여 유닛 단품을 하나의 덩어리로 처리한 뒤, 설치 공간을 단순한 헬름홀츠 공명기의 형태로 간략화 하여 산출함으로써 대략적인 형태의 챔버 블록 구조를 쉽게 설계할 수 있는 점이다. 두 번째는 민감하게 반응하는 대역의 주파수를 파악하고 디지털 필터를 통해 신호 처리해 줌으로써 음성 인식의 정확도를 보조해 줄 수 있다는 점이다.

현재 MEMS 마이크론은 IoT 기술의 확산에 따라 모바일 단말에서부터 각종 생활가전에 이르기까지 대부분의 전자제어 시스템을 사용하는 실용 제품에 음성 인식을 위한 부품으로 적용되고 있다. 그러나 아직 음성 인식을 위한 MEMS 마이크론의 물리적인 환경 조정에 관한 연구가 부족하여, 음성 인식 성능을 최적하기 위한 많은 연구가 필요하다. 때문에 본 연구결과를 활용하여 제품 개발단계에서 참고자료로 사용한다면, 부품 설계 단계에서 음향 센서 성능을 확보할 수 있어 AI 음성인식과 같은 고성능 음향 입력이 필요한 기능에 도움이 될 수 있을 것으로 기대한다.

차기 연구로는 좀 더 다양한 형태의 공간 용적을 사용하여 자동으로 형상과 특성을 추정해내는 프로그램 개발하는 것을 준비하고 있으며, 음성 인식에 사용되는 공간 학습 모델의 DB를 구축해보고자 한다.

### References

[1] P. R. Scheeper, B. Nordstrand, J. O. Gullov, B. Liu, T. Clausen, L. Midjord and T. Storgaard-Larsen, "A New Measurement Microphone Based on MEMS Technology", Journal of microelectromechanical systems, Vol. 12, No. 6, pp. 880-891, Jun. 2003.

[2] J. Liu, D. T. Martin, T. Nishida, L. N. Cattafesta, M. Sheplak, and B. P. Mann, "Harmonic Balance Nonlinear Identification of a Capacitive Dual-Backplate MEMS Microphone", Journal of microelectromechanical systems, Vol. 17, No. 3, pp. 698-708, Mar. 2008.

[3] H. Takahashi, A. Suzuki, E. Iwase, K. Matsumoto, and I. Shimoyama, "MEMS microphone with a micro Helmholtz resonator", Journal of Micromechanics and Microengineering, Vol. 22, No. 8, pp. 085019, Aug. 2012.

[4] Sin Hur, Yeong Hwa Lee, Hongsu Choi, and Junsik Park, "Design and Fabrication of Miniature MEMS microphone", Korean Society for Precision Engineering Academic Presentation Conference, pp. 381-382, Nov. 2010.

[5] Hyu-sang Kwon and Kwang-Cheol Lee, "Design and fabrication of condenser microphone with rigid backplate and vertical acoustic holes using DRIE and wafer bonding technology", J. of the Korean Sensors Society, Vol. 16, No. 1, pp. 62-67, Jan. 2007.

[6] Hokyung Ryu, Seong Jin Chung, and Jin Woo Lee, "Design of a Helmholtz Resonator for Noise Reduction in a Duct Considering Geometry Information: Additional Relationship Equation and Experiment", Korean Soc. Mech. Eng. A, Vol. 38,

No. 4, pp. 459-468, Apr. 2014

[7] ISO 3745, Acoustics - Determination of sound power levels and sound energy levels of noise sources using sound pressure - Precision methods for anechoic rooms and hemi-anechoic rooms, ISO/TC 43/SC 1 Noise, International Organization for Standardization, 2012.

[8] IEC 60268-4, Sound system equipment - Part 4: Microphones, TC 100/TA 20 - Analogue and digital audio, International Electrotechnical Commission, 2010.

### 저자소개

김 현 갑 (Hyun-Kab Kim)



2018년 2월 : 서울시립대학교  
전자전기컴퓨터공학과(공학석사)  
2020년 2월 : 서울시립대학교  
전자전기컴퓨터공학과(박사수료)  
2014년 4월 ~ 현재 : 한국전자  
기술연구원 정보미디어 연구센터  
연구원

관심분야 : 시스템제어, 음향제어, 센서네트워크

김 규 식 (Gyu-Sik Kim)



1981년 2월 : 서울대학교  
전자공학과(공학사)  
1983년 2월 : 서울대학교  
제어계측공학과(공학석사)  
1990년 2월 : 서울대학교  
제어계측공학과(공학박사)  
1988년 ~ 1992년 : 대우중공업

중앙연구소 선임 연구원

1993년 ~ 현재 : 서울시립대학교 전자전기컴퓨터공학부  
교수

관심분야 : 센서네트워크, 비선형제어, 에너지 변환