

실내 공간 내 지향성 마이크 어레이에서의 음성 명료도 개선을 위한 벽면 흡음 처리 방법

A method of wall absorption treatment for enhancing the speech intelligibility at a directional microphone array in a room

고병윤,¹ 이정권,[†] 조완호²

(Byeong-Yun Ko,¹ Jeong-Guon Ih,[†] and Wan-Ho Cho²)

¹한국과학기술원 기계공학과, ²한국표준과학연구원

(Received September 23, 2021; revised November 1, 2021; accepted November 15, 2021)

초 록: 벽면 흡음 처리는 잔향의 제어에는 효과적이거나, 잔향이 큰 공간의 경우에는 넓은 면적에 대한 적용이 필요하며, 각 벽면의 흡음은 음성 명료도에 다르게 영향을 미친다. 본 연구에서는 실내 음성 명료도의 관점에서 빔포밍 수음장치에 대한 잔향 제어를 위해 흡음 처리가 가장 효과적인 벽을 선택하는 모사 방법을 제안한다. 고정된 위치의 빔포밍 수음 어레이에 대해, 화자 혹은 스피커에서 방출된 음향이 각 벽면과 충돌하는 변수를 이용해 벽면 중요도 계수를 정의하고, 이를 이용해 흡음 처리에 따른 수음부의 음성개선 효과를 예측하는 방법을 제시했다. 검증을 위해 체적이 107 m³, 잔향 시간이 1.1 s인 직육면체형 실내공간에 대해 모사 실험을 진행했다. 중요도 계수가 가장 높은 벽면에 헬름홀츠 흡음기를 적용할 때 수정된 음성 명료도는 500 Hz, 1 kHz에서 각각 5.1 dB와 4.8 dB, 또 음성전달 지수는 0.06만큼의 향상이 예측되었는데, 이는 상용코드의 계산 결과와 차이인지 역치(Just-Noticeable Difference, JND) 이내의 차이로 일치했다. 분석결과, 벽면 중요도가 가장 높은 곳에 흡음 처리한 결과는 면적이 가장 넓은 천장에 같은 흡음률을 적용할 때보다 음성 명료도 지수가 JND 이상으로 더 개선되는 것을 알 수 있었다.

핵심용어: 음성 명료도, 잔향, 빔포밍 마이크, 흡음 처리, 실내음향 모사

ABSTRACT: Wall absorption treatment effectively reduces reverberation, but requires a large area for a live room and each wall absorption affects speech intelligibility differently. In this study, we try to find the most effective wall for the absorption treatment using the beamforming array microphone in terms of speech intelligibility. The absorption importance factor is defined by using the collision number of reflected sounds on each wall. It allows estimating how much the speech signal will be enhanced by the absorption treatment. A cuboid room with a size of 107 m³ and a reverberation time of 1.1 s is selected for the simulation. When a Helmholtz-type absorption is treated on the wall with the most significant importance factor, the modified clarity for 500 and 1k Hz is improved by 5.1 dB and 4.8 dB respectively, and the speech transmission index is enhanced by 0.06. The difference in results between the proposed method and commercial simulation code is less than a Just-Noticeable Difference (JND). The absorption treatment on the wall with the most significant importance factor shows improvement greater than the wall with the largest area, and its difference is larger than a JND value.

Keywords: Speech intelligibility, Reverberation, Beamforming microphones, Absorption treatment, Room acoustic simulation

PACS numbers: 43.55.Dt, 43.60.Fg

†Corresponding author: Jeong-Guon Ih (J.G.Ih@kaist.ac.kr)

Department of Mechanical Engineering, Korea Advanced Institute of Science and Technology (KAIST), 291 Daehak-ro, Yuseong-gu, Daejeon 34141, Republic of Korea

(Tel: 82-42-350-3035, Fax: 82-42-350-8220)



Copyright©2021 The Acoustical Society of Korea. This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

I. 서 론

최근 들어, 실내공간에서 화상회의 프로그램과 빔포밍 마이크 등의 장비를 이용한 강의와 강연이 일반화됨에 따라 음성·화자 인식에서 발화자의 위치 변화에 적응하는 방향성 어레이 마이크의 채용이 늘고 있으며, 자동음성 인식 장치(Automatic Speech Recognition system, ASR)에 의한 기계-인간 소통이 다양하게 이루어지고 있다.

실내공간에서 음성전달 시에는 잔향에 의한 음성 신호의 시간-주파수 왜곡이 일어나게 된다. 음원과 수음자의 거리가 멀수록 기하학적 확산에 따라 직접 음의 에너지가 감소하므로, 임계거리 이상에서는 잔향음과 간섭되어 전달되는 음성의 명료도가 떨어진다.^[1] 이 때문에, 잔향 조건에서도 화자의 음성 신호를 복원할 수 있도록 지향성 스피커 및 마이크로폰,^[2] 벽면 흡음 처리와 벽체 형상 조정^[3]과 같은 다양한 방법들이 적용되고 있으며, 음성 신호의 변조 및 복원을 위한 신호처리 방법들도 다수 제안됐다.

신호처리를 이용하는 방법은 적은 비용으로 효과를 볼 수 있으며, 여러 공간에서 효과를 볼 수 있다는 장점이 있다. 신호처리를 이용한 음성개선 방법 중 대표적인 것은 음원과 마이크로폰 간의 전달 함수를 예측해 잔향을 제거하는 모드 중첩법과 주파수대역 보강법이 있다.^[4] 모드 중첩법의 경우 전달 함수가 정확히 예측되는 경우 잔향이 포함되지 않은 기존 음성 신호를 복원할 수 있으나, 모드 밀도 증가로 인하여 중·고주파수 대역에서는 전달 함수의 정확한 예측이 어렵다는 문제가 있다.^[1] 주파수대역 보강은 확산 음향장 모델과 공간 연관도를 이용해 측정된 잔향을 제거하는 방법이다.^[5] 이 방법은 계측된 음성 신호만을 이용해 잔향 요소를 제거할 수 있으나, 음원-수음자 간의 거리가 임계거리 이상일 때 공간 연관도의 참값과 추정값 차가 0.2 이상이 되면 잔향 제거 성능이 저하된다.^[6]

신호처리 기법의 성능을 향상하기 위해서는 도달되는 음성 신호의 신호대잡음비(Signal to Noise Ratio, SNR)를 높여야 하며, 이를 위해 지향성 마이크를 사용한 공간 필터링을 이용하는 것이 효과적이다.^[2] 이 방법을 이용하면 잔향음의 크기를 감소시킬 수는 있

으나, 음원과 마이크 간의 거리가 임계거리 이상이면 어레이 마이크에서 구축된 빔 방향에서의 잔향 인텐시티도 함께 증가하게 되므로, 빔 폭을 줄이기 위해 많은 수의 마이크로폰이 필요하다.

한편, 벽면에 흡음재를 부착하여 잔향을 줄이는 방법은 가장 간편하고 효과적인 방법이라고 할 수 있으며, 실내공간의 잔향시간, T 가 대략 0.5s 이상인 라이브 룸 조건에서는 높은 잔향 에너지를 줄이기 위해 일반적으로 넓은 벽면에 대한 흡음 처리를 시행한다.^[7] 그러나, 실내공간의 흡음 특성에 따라 주파수별 요구되는 흡음의 양이 다르며, 흡음 처리에 따른 비용의 증가와 실내공간의 협소화 등의 문제가 발생하므로, 효율적인 적용이 필요하다.

본 연구는 잔향이 존재하는 공간에서 음성 명료도를 개선하기 위한 흡음의 최적 배치 방법을 제안한다. 적용조건으로는, 공명이 균일하게 분포될 수 있는 공간 치수 비율인 Bolt 기준치^[3,8] 범위의 크기를 갖는 실내 음장에서, 임의의 위치에 있는 화자가 발생시키는 음성을 미리 지정된 위치에 놓인 빔포밍 어레이 마이크를 통해 측정하는 상황으로 제한했다. 이 조건은 일반 중형 세미나실, 교실 등 다수 화자가 다양한 위치에 있을 때 실내공간 전방의 거의 끝부분에 놓인 자동음성 인식 장치 및 마이크 어레이를 이용한 쌍방향 음성 교환장치에 적용될 수 있는 상황을 예시한다.

마이크로폰에 측정된 잔향 에너지에 대한 각 벽면 흡음의 영향도를 분석할 수 있다면 흡음 처리 시 변화되는 음성개선 효과를 예측할 수 있으며, 가장 잔향 제어 효과가 높은 벽면에 효율적으로 적용하는 것이 가능하다. 이를 위해, 가상 음원법을 이용하여 직육면체 형 실내공간에서 발생 되는 잔향을 공간 기하학적인 수치 기법으로 예측하고, 잔향음의 정보를 통해 각 벽면의 흡음을 변화에 따른 음성 명료도의 개선 효과를 예측하는 방법을 제안한다.

II. 이 론

2.1 대상 공간의 정의

임의의 공간에서 음원과 마이크로폰 간의 전달 함수의 시간 영역 표현은 실내 충격 응답으로 정의되며,

이를 잔향 요소들의 합으로써 표현한다. Nábělek와 Robinette^[9]은 잔향음의 시간 지연으로 인해 앞 음소가 뒤 음소를 가리는 현상을 중첩 마스킹 효과라고 정의하고, 잔향 요소의 시간 지연에 따른 음소 신호의 명료도를 분석했다. 음성 명료도의 주요 결정인자인 자음 신호의 경우, 시간 지연이 20ms 이상일 때 음성 신호의 명료도가 저하되는 것을 보인 바 있다.

음성 신호와 청각의 민감도를 주파수 분포 특성으로 나타낸 ANSI 3.5 규정에서는 표준 음성 신호의 주파수 범위를 250 Hz~8000 Hz의 옥타브 대역으로 정의한다.^[10] 음원과 수음자의 거리가 1m일 때 표준 음성 신호의 각 주파수 밴드별 음압 분포를 보면, 모음의 큰 에너지로 인해 500 Hz 영역에서 59.8 dB로 가장 높은 크기를 가지며, 자음의 영역인 1 kHz~4 kHz 영역은 주파수가 높아질수록 음압 레벨이 감소한다. 이러한 음성 신호의 주파수 분포 특성으로 인해 청각 마스킹 효과가 발생하며 음성 명료도가 감소한다. IEC 60268-16^[11]에서는 청각 마스킹의 크기를 음압이 46 dB~55 dB에서는 -40 dB/oct, 그리고 56 dB~65 dB에서는 -35 dB/oct로 정의한다. 또, 음압 레벨에 따른 청각 마스킹 효과를 따르면, 500 Hz 대역에서의 음압으로 인해 1 kHz에서는 대략 25 dB의 마스킹 효과가 발생하여 자음의 명료도가 감소함을 예측할 수 있다.

따라서 적절한 벽면 흡음 처리를 통해 음성 신호의 명료도를 효과적으로 개선하기 위해서는, 시간 지연이 20 ms 이상인 잔향 요소들의 에너지를 줄일 수 있고, 주로 저주파수 영역에 있는 모음의 높은 음압에 의해 야기되는 청각 마스킹을 개선할 방법이 필요하다.

한편, 실내공간의 잔향은 공간 조건 및 음원과 마이크의 위치에 영향을 받는다. 공간 조건과 관련해서는, 공간의 체적과 치수 비율에 의해 잔향의 주파수 특성이 일반적으로 정해짐을 고려할 수 있다. 공간의 체적 V 와 관련한 잔향의 주파수 특성으로는 $f_s = 2000 \sqrt{T/V}$ 로 정해지는 Schroeder 주파수를 고려할 수 있다.^[11] 주파수 범위가 Schroeder 주파수보다 낮으면 음성 신호가 회절과 간섭 현상에 지배받으며, $4f_s$ 이상의 주파수에서는 거의 완벽하게 기하음향학 범위에 들어가는 것으로 알려져 있다.^[12] 공간

의 치수 비율과 관련한 잔향의 주파수 특성에 관련해, 공명이 발생하는 주파수의 분포를 고려할 수 있다. Bolt의 room criteria에 따르면,^[8] 실내공간의 특정 방향 치수가 과도하게 커짐으로써 잔향 분포가 방향성을 갖는 현상을 막기 위해, 실내공간의 최대 치수 L_{max} 의 허용 범위는 다음과 같이 설정된다.^[13]

$$L_{max} < 1.9 V^{1/3}. \quad (1)$$

음원과 지향성 마이크의 위치에 따른 잔향 특성은 둘 사이의 거리에 따라 변화되는 직접음과 잔향음의 에너지 비율로 나타내어진다. 직접음과 잔향음의 인텐시티가 동일할 때의 음원과 마이크 간 거리를 나타내는 임계거리, d_c 는 c_0 가 자유음장 내의 음속이라고 할 때 다음과 같이 정의된다.^[11]

$$d_c = \sqrt{\frac{24 \ln(10) V}{16 \pi c_0 T}}. \quad (2)$$

앞에서 설명된 실내 잔향의 특성을 고려하여 본 연구의 대상을 다음과 같이 구체적으로 정의한다. Fig. 1과 같이 직육면체 실내공간이 50 m³이상의 체적을 갖고 각 치수가 Bolt의 room criteria에 적합한 비율일 때, 잔향시간이 0.5 s 이상이라고 가정한다. 참고로 해당 공간의 최적 잔향시간은 0.2 s~0.4 s로 제안된 것을 고려하면,^[7] 대상 공간은 라이브 룸 조건에 해당한다. 벽면의 파동 간섭 효과를 최소화하기 위해 벽면으로부터 ANSI 3.5 규정의 음성 신호 유효한 주파수(대략 200 Hz) 반파장 거리 이상으로 떨어뜨려 마이크와 음원을 위치시키고, 사용자의 편의를 위해 마이크 어레이는 한쪽으로 치우친 구성(r_x, r_y, r_z)에 고정되어 있다고 설정한다. 음원의 위치는 마이크 어레이 중심으로부터 $2d_c$ 이상 떨어진 임의의 위치에 있어서, 잔향장의 우세로 인해 음성 명료도가 저하된다. 이때 어레이 빔포밍을 사용하여 저하된 음성 명료도를 개선하는 상황으로 가정한다. 다만, 이러한 공간 필터링 방법의 한계로서, 주엽 방향의 잔향 요소는 그대로 남게 되는 현상을 배제할 수 없다. 상기 조건에 따라 음성 명료도를 효과적으로 개선하기 위해서는, 주엽 방향으로 전파되는 잔

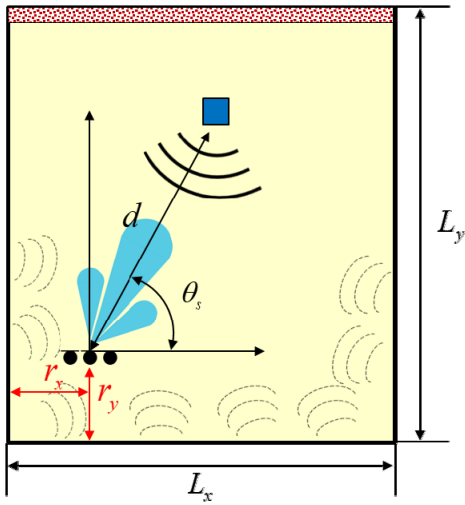


Fig. 1. (Color available online) Configuration of sound source and microphone array used for the present study: ■, sound source; ●, microphone; ▨, absorption treatment. Here, d is the distance between sound source and acoustic center of microphone array, θ_s the source angle from x axis, L_x, L_y the width and length of the room, and r_x, r_y the position of microphone array, respectively.

향음을 형성하는데 기여하는 각 벽면의 흡음 효율을 분석하고, 효율이 가장 높은 벽면에 최소한의 적절한 흡음 처리량에 대해 파악해야 한다.

2.2 음성 에너지 전달 모델

실내공간의 벽면 길이가 관심 주파수의 파장 길이보다 긴 Schroeder 주파수 이상일 때에는, 기하 음향학적인 음선을 이용할 수 있으며, 각 음선의 이동 경로와 벽면에 반사할 때 감소하는 에너지를 분석하여 최종적으로 음원과 마이크 간의 실내 충격 응답과 전달 함수를 유도할 수 있다. 이 음선법의 하나인 가상 음원 방법^[4]을 이용하게 되면, 음원에서 방사된 음성 신호가 각 벽면에 충돌하는 횟수와 방향을 통하여, 주어진 시간 구간 내에서 각 벽면이 얼마나 잔향에 기여하는지 알 수 있다. 직육면체 실내공간에서 가상 음원들의 개수 v 와 반사 충돌수 c 사이의 관계식은 다음과 같이 얻어진다.^[1]

$$v = 1.5 \times (5^c - 1). \tag{3}$$

v 의 관계식에 따라 직육면체의 실내공간에서 단일

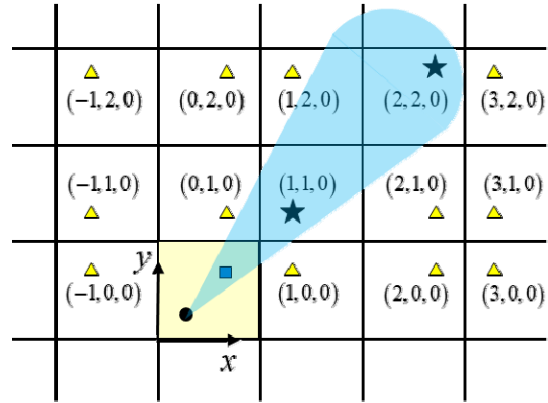


Fig. 2. (Color available online) Simulation with image sources and artificial walls: ■, original sound source; ▲, image sources; ★, image sources inside the effective beamwidth; ●, center of microphone array. (c_x, c_y, c_z) means the collision number of sound radiated from the corresponding image source in each coordinate (x, y, z) .

음원으로부터 발생한 음파가 각 벽면에 대해 반사되고 여러 경로를 거치며 충돌수가 증가하게 되어, Fig. 2에 보인 2차원 예시와 같이 수많은 가상 음원을 형성하게 된다. 이때 해당 실내공간의 형상(L_x, L_y, L_z), 음원의 위치(x_0, y_0, z_0)와 마이크의 위치(x_i, y_i, z_i)가 정해졌을 때 x, y, z 방향 충돌수(c_x, c_y, c_z)에 대한 가상 음원의 위치(x_i, y_i, z_i)는 다음의 식으로 결정된다:

$$\begin{bmatrix} x_i \\ y_i \\ z_i \end{bmatrix} = \begin{bmatrix} c_x L_x + x(-1)^{c_x} + \frac{(1 + (-1)^{|c_x+1|})L_x}{2} \\ c_y L_y + y(-1)^{c_y} + \frac{(1 + (-1)^{|c_y+1|})L_y}{2} \\ c_z L_z + z(-1)^{c_z} + \frac{(1 + (-1)^{|c_z+1|})L_z}{2} \end{bmatrix}, \tag{4}$$

여기서 $| \cdot |$ 는 반대임을 표시한다. 임의의 i 번째 가상 음원이 마이크 어레이의 음향 중심에 기여하는 에너지 $E_{IMG,i}$ 는 가상 음원과 마이크 간의 거리 $d_{i,r}$, 벽면 반사계수 R , 그리고 각 벽면의 충돌 횟수 k 에 의해 주어지는 다음의 상관관계 식에 의해 최종적으로 결정된다:

$$E_{IMG,i} = \frac{1}{d_{i,r}^2} R_{x=L_x}^{k_i(x=L_x)} R_{x=0}^{k_i(x=0)} R_{y=L_y}^{k_i(y=L_y)}$$

$$\cdot R_{y=0}^{k_i(y=0)} R_{z=L_z}^{k_i(z=L_z)} R_{z=0}^{k_i(z=0)}, \quad (5)$$

$$R_{x=L_x}^{k_i(x=L_x)} R_{x=0}^{k_i(x=0)} = \left(R_{x=L_x} R_{x=0} \right)^{\lfloor |c_x|/2 \rfloor} \cdot \left(\frac{1 + \text{sgn}(c_x)}{2} R_{x=L_x} + \frac{1 - \text{sgn}(c_x)}{2} R_{x=0} \right)^{(c_x \bmod 2)}, \quad (6)$$

여기서 좌표축의 원점이 Fig. 2와 같이 직육면체 실내공간의 한쪽 바다 모서리에 있을 때 각 벽면의 위치에 대한 예로써 $x=0$ 벽면은 +y 방향을 바라볼 때 왼쪽, $x=L_x$ 벽면은 오른쪽에 있는 벽을 의미한다. 또한, mod는 나머지 연산자를 의미하며, $\text{sgn}[c_x]$ 는 c_x 의 부호를 의미한다. 이는 가상 음원과 마이크로부터 거리가 멀어질수록, 벽면 반사 계수의 크기가 감소할수록, 벽면의 충돌 횟수가 증가할수록 가상 음원에 의해 기여되는 에너지가 감소한다는 물리적 원리와 일치한다.

III. 벽면 흡음 중요도의 정의 및 명료도 예측

3.1 벽면 흡음 중요도의 정의

Fig. 2와 같이 가상 음원이 분포된 상황에서, 빔 패턴 $B(f, \phi, \theta, \phi_s, \theta_s)$ 의 영역에 포함되어 있는 ($c_x=1, c_y=1, c_z=0$) 및 ($c_x=2, c_y=2, c_z=0$)에 관련된 잔향음 요소는 그 크기가 보존되어 마이크에 측정된다. 해당 잔향 요소들의 경우 벽면 반사로 인해 감소하는 에너지가 각각 $R_{x=L_x}, R_{x=0}, R_{y=L_y}, R_{y=0}, R_{z=L_z}, R_{z=0}$ 에 반비례하는데, $x=L_x$ 벽면 그리고 $y=L_y$ 벽면에서 공통으로 반사가 있는 것을 알 수 있다. 이에 따라, 빔 방향에 존재하는 가상 음원들의 에너지가 마이크로 전달되는 과정에 대해, 해당 벽면 흡음률이 높은 영향도를 갖는 것을 개념적으로 알 수 있다.

음원 신호가 발생한 이후 시간 t 에 따라 감소하는 실내 잔향음의 에너지는 다음과 같다.^[1]

$$E(t) = E_0 \exp[-6 \ln(10)t/T]. \quad (7)$$

$T/4$ 이후에 마이크에 도달되는 잔향음 에너지는 전체 에너지에 대해 5% 이하로 기여하므로, 앞에서 설

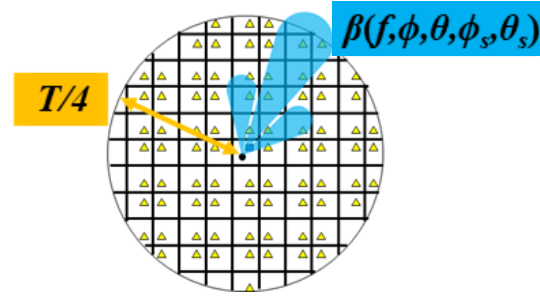


Fig. 3. (Color available online) Analysis of image sources within the reverberation radius for spatial filtering of reverberant energy: ■, original sound source; ▲, image source; ●, acoustic center of microphone array.

명한 각 벽면의 흡음에 의한 음장형성 기여도를 정량적으로 계산하기 위해 Fig. 3과 같이 전체 잔향 에너지의 95% 이상을 만드는 잔향 반경 $R=T/4$ 까지의 총 U 개의 유효 가상 음원들을 예측한다. Eq. (5)에서 정의한 각 가상 음원이 마이크로폰 어레이 음향 중심에 기여하는 에너지가 빔포밍 파워 $\beta(f, \phi, \theta, \phi_s, \theta_s)$ 에 의해 그 크기가 변하게 되며, 빔포밍 파워는 마이크로폰 어레이 빔포밍 방향이 Fig. 1과 같이 음원의 방향 ϕ_s, θ_s 에 지향되어 있을 때 주파수 f 와 가상 음원이 마이크로폰 어레이 중심에 도달되는 방향 ϕ, θ 에 따라 결정된다. U 개의 유효 가상 음원들이 마이크로폰 어레이에 기여하는 에너지를 모두 더하게 될 때 최종적으로 마이크에 측정된 잔향 에너지가 다음과 같이 구해진다:

$$E_{rev, filtered}(f) = \sum_i^U \beta(f, \phi, \theta, \phi_s, \theta_s) \frac{1}{d_{i,r}^2} \cdot R_{x=L_x}^{k_i(x=L_x)}(f) R_{x=0}^{k_i(x=0)}(f) R_{y=L_y}^{k_i(y=L_y)}(f) \cdot R_{y=0}^{k_i(y=0)}(f) R_{z=L_z}^{k_i(z=L_z)}(f) R_{z=0}^{k_i(z=0)}(f), \quad (8a)$$

또는,

$$E_{rev, filtered}(f) \approx CR_{x=L_x}^{N(x=L_x)}(f) R_{x=0}^{N(x=0)}(f) \cdot R_{y=L_y}^{N(y=L_y)}(f) R_{y=0}^{N(y=0)}(f) \cdot R_{z=L_z}^{N(z=L_z)}(f) R_{z=0}^{N(z=0)}(f), \quad (8b)$$

여기서 C 는 빔포밍 파워 $\beta(f, \phi, \theta, \phi_s, \theta_s)$ 와 음원-수음자 간 거리 $d_{i,r}$ 로 구성되는 임의의 상수이며, N 은 흡음 중요도 계수를 나타낸다. Eq. (8b)에서, 흡음 중요도 계수가 높을수록 해당 벽면에 대한 흡음 처리 시 잔향 에너지를 크게 감소시켜, 음성 신호의 명료도를 효과적으로 높일 수 있음을 알 수 있다. 흡음 중요도 계수는 각 벽면 흡음률에 따라서 변화되는 잔향 에너지를 이용해 다음과 같은 관계식으로 정리될 수 있다:

$$N(x=0, f) = \frac{\partial \log E_{rev, filtered}(f)}{\partial \log R_{x=0}(f)}, \quad (9a)$$

$$N(x=L_x, f) = \frac{\partial \log E_{rev, filtered}(f)}{\partial \log R_{x=L_x}(f)}, \quad (9b)$$

$$N(y=0, f) = \frac{\partial \log E_{rev, filtered}(f)}{\partial \log R_{y=0}(f)}, \quad (9c)$$

$$N(y=L_y, f) = \frac{\partial \log E_{rev, filtered}(f)}{\partial \log R_{y=L_y}(f)}, \quad (9d)$$

$$N(z=0, f) = \frac{\partial \log E_{rev, filtered}(f)}{\partial \log R_{z=0}(f)}, \quad (9e)$$

$$N(z=L_z, f) = \frac{\partial \log E_{rev, filtered}(f)}{\partial \log R_{z=L_z}(f)}. \quad (9f)$$

3.2 벽면 흡음에 따른 음성 명료도 예측

벽면 흡음 처리를 통해 음성 신호의 개선 여부를 판단하기 위해서는 음성 명료도 기준을 고려한 적용을 해야 한다. 흡음 중요도 계수를 이용해 음성 명료도 개선 효과를 예측하기 위해, 2장에서 설명한 잔향 제거 기준을 고려해 음성 명료도 평가 인자를 선정했다. 먼저, 주파수대역별 중심주파수 f_c 에 대해 50 ms 이내의 시간 구간 내에 존재하는 직접 입사음의 에너지 E_{dir} 와 반사되는 시간 지연이 50 ms 이내인 잔향 에너지 $E_{rev}^{t \leq 50}$ 는 신호대잡음비를 증가시키므로, 다음과 같이 잘 알려진 $C_{50}(\text{clarity})$ 을 이용할 수 있다.^[15]

$$C_{50}(f_c) = 10 \log \frac{E_{dir}(f_c) + E_{rev}^{t \leq 50}(f_c)}{E_{rev}^{t > 50}(f_c)}. \quad (10)$$

그러나, 배경소음이 작은 공간에서의 대화 상황에서는 음악과 달리 잡음에 의한 영향보다는 청음자가

음성 신호를 분명히 인지할 수 있도록, 자음의 명료도가 확실히 중요한 의미를 지니는 시간 구간인 20 ms 이내를 고려해야 한다. 이에 따라, 본 연구에서는 시간 지연이 20 ms 이내인 잔향 구간을 신호로 그 이외의 잔향 요소를 잡음으로 하여, 기존의 C_{50} 를 수정한 평가 인자 $C_{20}(\text{modified clarity})$ 를 다음과 같이 정의한다:

$$C_{20}(f_c) = 10 \log \frac{E_{dir}(f_c) + E_{rev}^{t \leq 20}(f_c)}{E_{rev}^{t > 20}(f_c)}. \quad (11)$$

신호와 잡음으로 선정된 각 잔향 구간별 에너지의 특성을 살펴볼 때 먼저 신호 요소인 시간 지연이 20 ms 이내인 잔향 구간은 실내 조건이 $V \geq 50 \text{ m}^3$ 이며 치수 비율이 Bolt의 room criteria일 때 대략 1차 또는 2차 반사 잔향 요소들이 포함되며 단일 벽면 흡음 처리로 인해 변화되는 에너지가 거의 미미하다. 잡음 요소인 시간 지연이 20 ms 이상인 잔향 구간의 에너지는 Kuttruff가 정의한 전체 잔향 에너지 E_{rev} 로부터 시간 구간에 따른 비율과 잔향 시간 T 간의 관계식에 의해 다음과 같이 정의된다.^[1]

$$E_{rev}^{t > 20}(f_c) = E_{rev}(f_c) e^{-\frac{3 \ln(10)}{25T(f_c)}}. \quad (12)$$

본 연구 대상인 $T \geq 0.5 \text{ s}$ 일 때 단일 벽면 흡음 처리로 인한 Eq. (12)의 지수함수 항의 변화가 거의 일정하다고 가정한다면 최종적으로 옥타브 밴드별 음성 명료도 인자의 변화량 ΔC_{20} 은 흡음 처리 후 잔향음 에너지 $E_{rev, TX}$ 와 흡음 처리 전 잔향음 에너지 $E_{rev, 0}$ 의 비로써 다음과 같이 근사할 수 있다:

$$\Delta C_{20}(f_c) \approx 10 \log E_{rev, TX}(f_c) / E_{rev, 0}(f_c). \quad (13)$$

이때 잔향 에너지는 Eq. (8b)에 주어진 바와 같이 각 벽면의 흡음 중요도 계수 N 과 흡음률에 따라 정해진다. 따라서, 벽면을 흡음 처리하여 변화되는 ΔC_{20} 은 흡음이 적용되는 벽면의 중요도 계수 N_{TX} , 해당 벽면의 흡음 처리 전 초기 흡음률 a_0 , 해당 벽면 흡음 처리에 사용되는 흡음재의 흡음률 a_{TX} 를 사용해 다음과

같이 예측할 수 있다:

$$\Delta C_{20}(f_c) = 10N_{TX}(f_c) \log \frac{(1 - \alpha_0(f_c))}{(1 - \alpha_{TX}(f_c))}. \quad (14)$$

두 번째 평가 인자로는 실내공간의 잔향 특성을 고려한 Speech Transmission Index(STI)를 선정했다. 잘 알려진 바와 같이, STI는 실내공간에서의 잔향으로 인해 음성 신호의 음향파위가 변조되는 정도를 음성 명료도 지표로 나타낼 수 있다.^[11] 계산된 잔향 시간을 이용해 변조 주파수 f_m 에 대한 변조 전달 함수 m 을 계산하고, 옥타브 밴드별 STI 지수에 IEC 60268-16에서 규정한 주파수 가중 계수 w_{f_c} 를 곱함으로써, STI를 다음과 같이 얻을 수 있다:

$$m(f_c) = \sqrt{1 + \left(\frac{2\pi f_m T_{60}(f_c)}{13.8} \right)^2}, \quad (15)$$

$$\text{SNR}_{\text{app}} = 10 \log \left(\frac{m(f_c)}{1 - m(f_c)} \right), \quad (16)$$

$$\text{STI} = (\overline{\text{SNR}_{\text{app}}} + 15) / 30, \quad (17)$$

$$\text{STI} = w_{125} \text{STI}_{125} + \dots + w_{8k} \text{STI}_{8k}, \quad (18)$$

여기서 w_{f_c} 는 0.5 kHz ~ 4 kHz 범위에서 0.23 이상의 값을 갖는데, 이는 자음의 주파수 범위와 사람의 청각 특성인 A-가중치에 의한 효과가 고려된 결과이다. Eqs. (15)~(18)에 보인 바와 같이, 변조 전달 함수는 잔향시간의 함수이며, 신호대잡음비의 특성에 따라서 STI가 잔향시간에 대한 로그 함수인 것을 알 수 있다. 잔향시간이 0.5 s 이상일 때의 STI값을 선형 회귀 방법을 통해 분석하면 다음과 같이 근사할 수 있다:

$$\text{STI} \approx -0.24 \log(T) + 0.59. \quad (19)$$

Eq. (19)은 잔향시간의 증가에 따라 STI가 감소함을 의미한다. 잔향시간은 다음과 같이 음속 c_0 , 공간의 체적 V , 잔향에너지 E_{rev} , 음원의 음향파위 I 와 관련

시킬 수 있다:^[11]

$$T(f_c) = 0.161 c_0 V E_{rev}(f_c) / 4II. \quad (20)$$

Eqs. (19)와 (20)을 통해 벽면을 흡음 처리하여 변화되는 옥타브 밴드별 STI의 변화량 ΔSTI 는 다음과 같이 예측할 수 있다:

$$\begin{aligned} \Delta \text{STI}(f_c) &= -0.24 \log \frac{E_{rev, TX}(f_c)}{E_{rev, 0}(f_c)} \\ &= 0.24 N_{TX}(f_c) \log \frac{(1 - \alpha_0(f_c))}{(1 - \alpha_{TX}(f_c))}. \end{aligned} \quad (21)$$

IV. 모사 실험을 통한 검증

3장에서 설명한 흡음 중요도 계수를 이용해 음성 개선 효과를 예측하는 방법을 수치 모사 실험을 통해 검증했다. 모사 실험 공간 구성으로는 2장에서 선정한 Schroeder 주파수와 Bolt의 room criteria를 고려한 6.5 m(L) × 5.0 m(W) × 3.3 m(H) ($V = 107.3 \text{ m}^3$)의 직육면체 형상을 선정했다. Fig. 1에서 벽면의 간섭 현상을 억제하기 위해 음원과 모든 마이크를 벽면으로부터 반 파장 이상 격리하고, 마이크 어레이의 음향 중심 위치를 통상적으로 쓰이는 구석 공간인 $r_x = 1.01 \text{ m}$, $r_y = 1.22 \text{ m}$ 에 배치했다. 음원과 마이크 어레이의 음향 중심 간의 거리는 2 m로 대략 $3d_c$ ($d_c \approx 0.6 \text{ m}$)의 위치이며, 음원의 방향 θ 는 80° 방향에 두었다. 음원의 높이는 서 있는 사람의 입 높이를 고려해 1.6 m에 두었으며, 마이크의 높이는 천장과 바닥으로부터의 간섭 현상을 고려해 1.3 m로 정했다. 초기 실내공간은 모든 벽면에 매끈한 벽돌^[6]이 적용된 상태이며, 이때 해당 실내공간의 잔향시간은 라이브 룸 조건인 대략 1 s 이상으로 Table 1과 같이 계산되었다.

어레이 마이크에는 고 지향성 빔포밍 방법인 Minimum Variance Distortionless Response(MVDR)^[2]를 적용하여 공간 필터를 구현했다. 마이크 어레이는 4개의 마이크로폰을 30 mm 간격으로 배치한 선형 마이크 어레이를 사용한 것으로 가정하였고, 이때 MVDR 방법이 적용되는 주파수 상한은 5.5 kHz가 된다. 본 연구에 적용되는 저주파수 하한에 가까운 500 Hz와

Table 1. Absorption coefficient α_0 of the smooth brickwork and the calculated reverberation time.

Center frequency, f_c	500 Hz	1 kHz	2 kHz	4 kHz
α_0	0.12	0.16	0.22	0.24
T, s	1.65	1.37	1.08	1.01

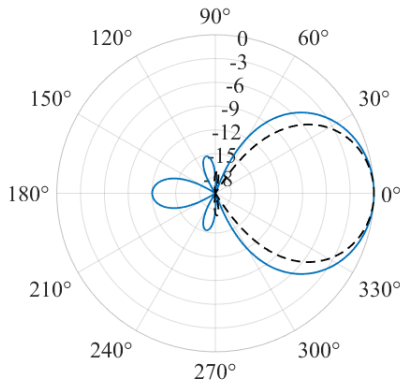


Fig. 4. (Color available online) Directivity pattern of highly directional array (MVDR)^[23] in dB scale at two frequencies: —, 500 Hz; - - -, 1 kHz.

1 kHz 주파수에서 반 전력 빔폭은 Fig. 4에서 보인 바와 같이 대략 30°로 형성된다.

가상 음원법을 사용해 계산된 공간에서 발생한 잔향음 요소들을 예측하고, 다시 이를 사용해 계산한 각 벽면 흡음률에 따른 잔향 에너지 분포를 Fig. 5에 나타내었다. 선형 회귀 분석을 통해 최종적으로 예측된 각 벽면의 흡음 중요도는 Table 2에 정리했다. 이 결과들을 통해, 주어진 구성 조건에서는 $y=L_y$ 벽면이 500 Hz와 1 kHz 대역에서 가장 높은 흡음 중요도를 갖는 것을 알 수 있다. 이는 3.1 절에서 설명한 것과 같이 주어진 실내공간에서 발생한 잔향음을 가상 음원법을 통해 분석할 때 빔포밍 방향에 분포하는 잔향음 요소 중 수음자로부터 거리가 가장 가까운 1차, 2차, 3차 반사 잔향 요소들이 해당 벽면에 대해 대부분 가장 높은 반사 횟수를 갖기 때문이다.

2장에서 언급된 잔향 제거 기준을 고려해, 흡음 처리에 적용할 흡음재로는 Table 3과 같이 500 Hz와 1 kHz 대역에서 0.63과 0.60의 흡음률을 가지는 헬름홀츠 흡음기(Helmholtz resonance absorber)^[17]를 사용했다. 일반적인 미세 섬유로 이루어진 흡음재는 실내에 적용할 만한 두께에 대해 중주파수의 흡음률이 높지 않으며 이를 고려해 넓은 면적에 적용 시 고주

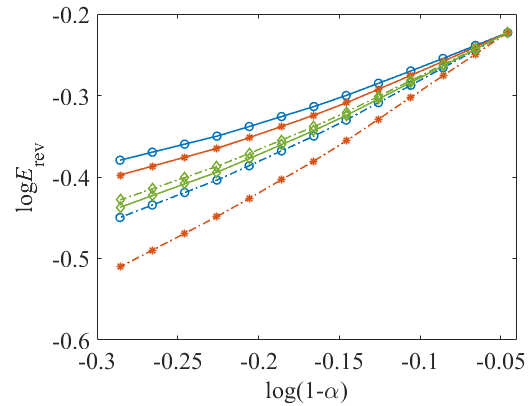


Fig. 5. (Color available online) Filtered reverberation energy depending on the absorption treatment at 500 Hz. Positions of absorbing walls are: —○—, at $x = 0$; —○—, at $x = L_x$; —*—, at $y = 0$; —*—, at $y = L_y$; —◇—, at $z = 0$; —◇—, at $z = L_z$.

Table 2. Importance factor of wall absorption at 500 Hz and 1 kHz.

Wall	$x = 0$	$x = L_x$	$y = 0$	$y = L_y$	$z = 0$	$z = L_z$
500 Hz	0.60	0.94	0.72	1.21	0.89	0.85
1 kHz	0.63	0.98	0.69	1.28	0.83	0.82

Table 3. Absorption coefficient of the employed Helmholtz type absorber.

Center frequency, f_c	500 Hz	1000 Hz	2000 Hz	4000 Hz
Helmholtz absorber, α_0	0.63	0.60	0.32	0.12

파수의 음향을 과도하게 감쇠시키므로 적절하지 않다. 헬름홀츠 흡음기를 이용해 흡음 처리할 때 변화되는 음성개선 효과를 분석하기 위해 건축 음향 시뮬레이션 코드(Odeon 13.04 Combined)를 이용했다. 모사 실험의 정확도를 위해 요구되는 음선의 최소 개수 n 은 벽면 총면적 S , 계산 시간 범위 t 를 고려해 결정했다($n \geq 8\pi c_0^2 t^2 / S$).^[18] 이에 따라 본 연구에서는 $n=20,000$ 개의 음선을 이용해 모사 실험을 수행했으며, 해당 결과에서 $y=L_y$ 벽면에 흡음 처리 시 가장 낮은 잔향시간으로 중심주파수가 500 Hz에서 1.18 s, 1 kHz에서 0.91 s, 2 kHz에서 0.82 s, 4 kHz에서 0.75 s가 나타났다.

계산된 초기 상태인 전체 벽돌면 조건에서 $C_{20}(500) = 1.2$ dB, $C_{20}(1k) = 1.5$ dB 그리고 $STI = 0.67$ 이었는데, 헬름홀츠 흡음기를 어느 한 벽면에 적용할 때에 예측되는 음성 명료도의 향상치는 Table 4에 요약해 나

Table 4. Estimation of the enhancement of speech intelligibility varying the wall position of absorption treatment.

Treatment	$\Delta C_{20}(500)$ (dB)	$\Delta C_{20}(1k)$ (dB)	Δ STI
$x = 0$ wall	+2.3	+2.0	+0.02
$x = L_x$ wall	+3.5	+3.2	+0.04
$y = 0$ wall	+2.7	+2.2	+0.03
$y = L_y$ wall	+4.6	+4.1	+0.06
$z = 0$ floor	+3.3	+2.7	+0.04
$z = L_z$ ceiling	+3.2	+2.6	+0.03

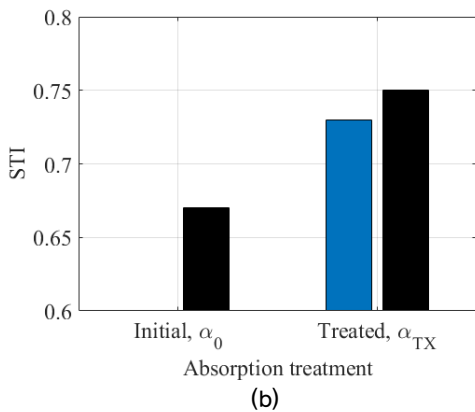
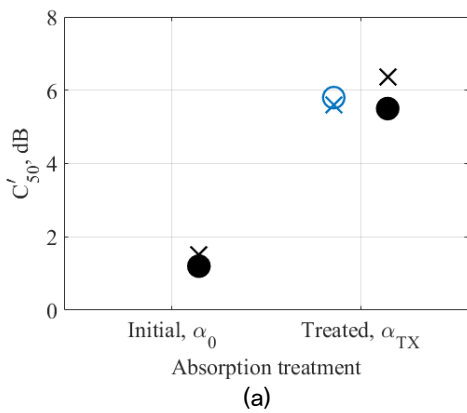


Fig. 6. (Color available online) Comparison of performance index for speech intelligibility: \bigcirc , \times , \blacksquare , estimated using importance factor; \bullet , \times , \blacksquare , calculated using image source method. (a) \bigcirc , \bullet , $C_{20}(500)$; \times , \times , $C_{20}(1k)$, (b) STI.

타내었다.

C_{80} 의 JND가 1.0 dB(500 Hz~1 kHz)^[15]이고, C_{50} 의 JND가 1.1 dB^[19]인 점을 고려하면, C_{50} 보다 C_{20} 의 JND 허용 변화량이 더 높은 것을 추정할 수 있으나, 이에 대한 실험치가 아직 없으므로 C_{20} 의 JND를 C_{50}

의 JND로써 그대로 사용한다. Fig. 6에는 $y = L_y$ 벽면에 대해 흡음 처리할 때 변화되는 명료도를 중요도 계수를 통해 예측한 값과 수치 모사 프로그램을 이용해 계산한 결과를 비교했다. C_{20} 의 JND = 1.1 dB, STI의 JND=0.03인 사실을 고려하면, 예측치 간의 차이는 JND 이하로 발생하여, 두 방법의 결과는 유효 범위 내에서 같은 값을 갖는다고 할 수 있다.

중요도가 2배 이상 차이가 나는 $y = L_y$ 벽면과 $x = 0$ 벽면을 비교했을 때, $y = L_y$ 벽면에 흡음 처리할 때 변화되는 C_{20} 가 $x = 0$ 의 흡음 처리 결과보다 2배 정도 개선되는 것을 알 수 있다. 이는 흡음 처리되는 벽면의 중요도와 음성 명료도의 개선이 강한 상관관계를 갖는다는 것을 의미한다. 또한, 면적이 가장 넓은 천장과 흡음 처리 결과를 비교해보면, 오히려 천장보다 C_{20} , STI 값들이 모두 JND 이상으로 명료도가 더 개선된 것을 알 수 있어서, 벽면 흡음 중요도 값의 유용성을 보인다.

V. 결 론

본 연구에서는 음원과 마이크론의 거리가 임계거리의 두 배 이상이고 실내 잔향시간이 0.5s 이상인 라이브 룸 조건에서, 지향성 마이크론 어레이에 입사되는 직접음과 잔향음을 처리하여 음성 인식률을 높이기 위한 벽면 흡음 처리 방법의 규칙에 관해 연구하고 개선 효과를 예측하는 방법을 다루었다. Schroeder 주파수 이상의 주파수 영역에서 실내공간에서 발생한 잔향은 마이크론 어레이의 빔포밍 방향에 존재하는 가상 음원들의 음향 에너지 기여가 특정 벽면의 흡음에 대해 민감한 것을 알 수 있었다. 벽면별 잔향의 충돌수를 기반으로 흡음 중요도를 정의하였고, 벽면 흡음률에 따른 잔향음 에너지 변화량을 예측함으로써 흡음 중요도를 유도했다. 벽면 중요도를 이용한 음성개선 예측법의 입증을 위해 라이브 룸 조건의 직육면체 실내공간을 구성하여 모사 실험을 진행했다. 흡음 중요도가 가장 높은 벽면에 헬름홀츠 흡음기를 적용할 때, 벽면 흡음 중요도를 이용해 계산한 C_{20} 및 STI 값들이 수치 모사 프로그램을 통해 구한 결과와 JND 이하로 차이가 발생함을 확인했다. 또한, 중요도가 가장 높은 곳에 흡음 처리

한 결과가 면적이 가장 넓은 천장에 같은 흡음재를 적용할 때보다 음성 명료도 지수가 JND 이상으로 더 개선되는 것을 보였고, 이를 통해 벽면 흡음 중요도를 이용한 흡음 처리법이 주어진 공간과 화자-마이크로폰 특성 조건에 있어 음성 명료도를 경제적으로 개선할 수 있음을 알 수 있다.

감사의 글

이 연구는 한국연구재단(NRF-2020R111A2066751) 및 한국표준과학연구원(KRISS -2021-GP2021-0002)의 일부 지원을 받았음. 연구에 있어 건설적인 토론을 통해 연구질을 높게 하는데 기여한 김다영 박사, 정인지 박사, 이기호 박사에게 감사를 포함합니다.

References

1. H. Kuttruff, *Room Acoustics* (CRC Press, Florida, 2016), Chaps. 4-5.
2. M. Brandstein, *Microphone Arrays: Signal Processing Techniques and Applications* (Springer, Berlin, 2001), Chaps. 1-2.
3. M. Kleiner and J. Tichy, *Acoustics of Small Rooms* (CRC Press, Florida, 2014), Chaps. 1-4.
4. J. Benesty, S. Makino, and J. Chen, *Speech Enhancement (Signals and Communication Technology)* (Springer, Berlin, 2005), Chaps. 1, 11.
5. O. Thiergart, G. Del Galdo, and E. A. Habets, "On the spatial coherence in mixed sound fields and its application to signal-to-diffuse ratio estimation," *J. Acoust. Soc. Am.* **132**, 2337-2346 (2012).
6. A. Schwarz and W. Kellermann, "Coherent-to-diffuse power ratio estimation for dereverberation," *IEEE/ACM Trans. on Audio, Speech, and Lang. Process.* **23**, 1006-1018 (2015).
7. L. L. Beranek and T. Mellow, *Acoustics: Sound Fields and Transducers* (Academic Press, London, 2012), Chap. 10.
8. R. H. Bolt, "Note on normal frequency statistics for rectangular rooms," *J. Acoust. Soc. Am.* **18**, 130-133 (1946).
9. A. K. Nábèlek and L. Robinette, "Influence of the precedence effect on word identification by normally hearing and hearing-impaired subjects," *J. Acoust. Soc. Am.* **63**, 187-194 (1978).
10. ANSI S3.5-1997, *Methods for Calculation of the Speech Intelligibility Index*, 1997.
11. IEC 60268-16, *Sound System Equipment-Part 16: Objective Rating of Speech Intelligibility by Speech Transmission Index*, ed. 4., 2011.
12. C. H. Jeong, J. G. Ih, and J. H. Rindel, "An approximate treatment of reflection coefficient in the phased beam tracing method for the simulation of enclosed sound fields at medium frequencies," *Applied Acoustics*, **69**, 601-613, (2008).
13. ISO 354: 2003, *Acoustics--Measurement of Sound Absorption in a Reverberation Room*, 2003.
14. J. B. Allen and D. A. Berkley, "Image method for efficiently simulating small-room acoustics," *J. Acoust. Soc. Am.* **65**, 943-950 (1979).
15. ISO 3382-1: 2009, *Acoustics--Measurement of Room Acoustic Parameters-Part 1: Performance Spaces*, 2009.
16. C. L. Christensen, "Odeon room acoustics program," v. 13.0., User Manual, 2013.
17. T. J. Cox and P. D'antonio, *Acoustic Absorbers and Diffusers: Theory, Design and Application* (CRC Press, London, 2009), Chap. 7.
18. J. H. Rindel, "Computer simulation techniques for acoustical design of rooms," *Acoustics Australia* **23**, 81-86 (1995).
19. J. S. Bradley, R. Reich, and S. G. Norcross, "A just noticeable difference in C50 for speech," *Applied Acoustics*, **58**, 99-108 (1999).

저자 약력

▶ 고 병 윤 (Byeong-Yun Ko)



2019년 2월 : 인하대학교 조선해양공학과 (공학사)
2021년 2월 : KAIST 기계공학과(공학석사)
2021년 2월 ~ 현재 : KAIST 기계공학과 박사과정

▶ 이 정 권 (Jeong-Guon Ih)



1985년 2월 : KAIST 기계공학과(공학박사)
1979년 7월 ~ 1990년 7월 : 대우자동차 기술연구소 NVH담당
1990년 8월 ~ 현재 : KAIST 기계공학과 교수
2018년 1월 ~ 현재 : Indonesia 반둥공대 (ITB) 공학물리과 겸직교수
2018년 6월 ~ 현재 : Denmark 덴마크공대 (DTU) 전기공학과 겸직교수
저서: MR Bai, JG Ih, and J Benesty, *Acoustic Array Systems: Theory, Implementation, and Application*, (John Wiley & Sons, Singapore, 2013).

▶ 조 완 호 (Wan-Ho Cho)



2002년 2월 : KAIST 기계공학과(공학사)
2004년 2월 : KAIST 기계공학과(공학석사)
2008년 2월 : KAIST 기계공학과(공학박사)
2008년 9월 ~ 2010년 2월 : KAIST 기계기
술연구소 박사후연구원
2010년 4월 ~ 2012년 2월 : Chuo University,
Japan, 임기제교교
2012년 3월 ~ 현재 : 한국표준과학연구원
책임 연구원