Multi-microphone system for sound detection based on wavelength constant matrix

Kazuharu Kuroiwa

Faculty of Engineering, Oita University,
700, Dannoharu, Oita, 870-1124 Japan

(Received 5 September 1997)

Keywords: Sound detection, Multi-microphone, Sound intensity, Wavelength constant, Microphone array
PACS number: 43. 60. BF

1. Introduction

One problem with the traditional method of measuring sound intensity is that it cannot decide and produce the sound intensity of interest in a multi-sound field. The objective of this study is to develop a sound detection system that allows greater facility in a sound intensity detection for a multi-sound field. As for sound detection, the linear microphone array system in conjunction with an adaptive signal processing based on LMS criterion using a newly developed error value¹⁾ or that based on an extension of constrained LMS criterion (AMNOR)²⁾ has been presented and shown to have an excellent performance. However, it is somewhat difficult for it to detect sounds at precisely specified two points at the same time to have sound intensity.

This paper outlines a sound detection system using an arbitrarily spaced multiple microphones in conjunction with an adaptive signal processing based on wavelength constant matrix. A wavelength constant matrix exactly gives a relation between the microphone outputs and the desired sounds in a multi-sound field. In the same manner as a procedure of so-called inverse problem, the weighting factors of sound detection FIR filter are decided with the inverse of the newly proposed wavelength constant matrix. Then, feeding the microphone outputs to the filters, the system can issue the desired sounds at the specified points of our choice, under the condition that the number of sounds and their arriving directions are registered into this system. Previous to the sound detection, the signal processing based on LMS criterion can estimate the sound directions by using an error value between the system outputs and a reference microphone output. Therefore, the number of detection filters should be equal to the sound number plus one. In addition, this system cannot issue the number of sounds and the distances of sound sources. At present, this system is available for sound fields where smaller than five sounds exist, with the noise reduction level being 19 dB for four sounds.

2. The theory of sound detection

An explanatory diagram for this sound detection system is illustrated in Fig. 1. In the multi-sound field,

the n -th sound source S_n , shown with '●', at $(r_n, \theta_n, \varphi_n)$ with $n=1 \sim N$ in a polar co-ordinates system radiates its original sound $p_n(t)$ towards a multi-microphone. The variables r_n , θ_n and φ_n imply a distance, zenith and bearing angles of S_n , respectively. The multi-microphone is composed of M microphones shown with '○'. The m -th microphone, M_m , at (d_m, ξ_m, η_m) receives the sum of all arriving sounds, named as $q_m(t)$, which alters in phase depending upon the mutual situation given by the spatial vector A_m of M_m , B_n of S_n , and A_v of a specified point M_v at (d_v, ξ_v, η_v) indicated with '■'. By "specified" is meant the position of our choice at which we intend to detect a certain sound. The proposed system is composed of a sound detection sub-system and a sub-system that estimates the sound direction, as illustrated in Fig. 1. Feeding $q_m(t)$'s into this system, a set of sound detection FIR filters, denoted as $h_{v,mn}$'s, provides the estimate of the desired sound, $\hat{p}_n(k)$, at M_v with other sounds being removed. Previous to the sound detection, the sound angle estimation with transfer functions, $H_{v,mn}$'s that are converted from the $h_{v,mn}$'s, finds the optimal sound angles, $(\hat{\theta}_n, \hat{\varphi}_n)$'s, by basing it on an LMS criterion with consideration of a reference sound $q_s(t)$ from the microphone S_s at (d_s, ξ_s, η_s) . Then, the sound detection filters receives $(\hat{\theta}_n, \hat{\varphi}_n)$'s with which their parameters are adaptively optimized. For this system, there is no need for any information about the sound field except for the total number N of sounds. In addition, the total number of microphones M should be equal to N for sound detection, but be $N+1$ for the estimation of sound direction.³⁾

Letting the k -th sample of the original sound $p_n(t)$ assumed at M_v and that of the microphone signal $q_m(t)$ be $p_n(k)$ and $q_m(k)$, and also letting the i -th frequency component of the DFT spectrum be $P_n(i)$ and $Q_n(i)$, then we may introduce column vectors defined as $P_v = (P_1(i) P_2(i) \cdots P_N(i))^T$ for the original sound and $Q = (Q_1(i) Q_2(i) \cdots Q_N(i))^T$ for the microphone signal, for which the upper case 'T' denotes the transposition operator. To formulate the relation in terms of P_v and Q , we may newly introduce a wavelength constant matrix C_v with the frequency, having a size of $N \times N$ and the form of

$$C_v = (c_{v,mn}) \quad (1)$$

$$c_{v,mn} = \exp \left(j \frac{2\pi i (\mathbf{A}_m - \mathbf{A}_v) \cdot (\mathbf{B}_n - \mathbf{A}_v)}{Tv |\mathbf{B}_n - \mathbf{A}_v|} \right) \quad (2)$$

where the raised dot denotes the inner product of two spatial vectors, v and T imply sound velocity and the observed time in conjunction with the length of the FIR filter and the sampling rate. Matrix C_v has been formulated by taking into account the wavelength constant property so that C_v plays a role in translating $\mathbf{P}_v$ into $\mathbf{Q}$. Particularly, the m -th column and n -th row element, $c_{v,mn}$, in matrix C_v plays a role in translating the n -th desired sound spectrum, $P_n(i)$, which is assumed at M_v into one of the components in the m -th microphone spectrum $Q_m(i)$. Therefore, $Q_m(i)$ can be formulated by summing up the quantity of $c_{v,mn} P_n(i)$ from $n=1$ upto N . In addition, the element $c_{v,mn}$ is not a func-

tion of sound source distance r_n but an implicit function only of the unknown angle (θ_n, φ_n) and given angles (ζ_m, η_m) and (ζ_v, η_v) . The estimation of (θ_n, φ_n) will be given later. With the matrix notation under the condition that (θ_n, φ_n) 's are known, the following relation holds:

$$\mathbf{Q} = C_v \mathbf{P}_v \quad (3)$$

In the same manner as the inverse problem in acoustics, it is easy to find the solution for $\mathbf{P}_v$ to Eq. (3). Namely, it has the form of

$$\mathbf{P}_v = C_v^{-1} \mathbf{Q} = \mathbf{H}_v \mathbf{Q} \quad (4)$$

$$\mathbf{H}_v = (\mathbf{H}_{v,mn}) = C_v^{-1} \quad (5)$$

where C_v^{-1} implies the inverse matrix of C_v and then is newly replaced with alternative matrix $\mathbf{H}_v$ as shown in Eq. (5). Noting that a product of two spectra in the

Fig. 1 Schematic diagram of the sound detection system.

frequency domain is equivalent to a convolution between the respective inverse Fourier transformation in the time domain, Eq. (4) can be rewritten as

$$\hat{p}_n(k) = \sum_{m=1}^N \sum_{j=0}^{L-1} h_{v,nm}(j) q_m(k-j) \quad (6)$$

where $h_{v,nm}(k)$ denotes the inverse discrete Fourier transformation of $H_{v,nm}(i)$ and is an impulse response, namely, the weighting factor for the FIR filter with tap number L . Equation (6) implies that passing the microphone signal $q_m(k)$'s through the corresponding filter $h_{v,nm}(k)$, and successively summing up their outputs from $m=1$ to N , the n -th sound, $\hat{p}_n(k)$, can be estimated according to the choice of sound number, with other sounds being removed. In addition, to distinguish the estimated from the original, the former has been expressed by '^' here. The procedure for Eq. (6) is implemented as shown on the bottom and left side of Fig. 1.

To find sound source angles (θ_n, φ_n) , we adopted one more microphone M_s , indicated by '□' and mounted at (d_s, ξ_s, η_s) , to get its discrete output $q_s(k)$ as a reference signal. Then, we express the i -th frequency component of the DFT transformation for samples of the n -th original sound $p_{sn}(k)$ at M_s as $P_{sn}(i)$ and also that of $q_s(k)$ as $Q_s(i)$. Under this situation, the same procedure, as shown in Eqs. (3)~(5), gives the vector of the original sound spectrum $\mathbf{P}_s = (P_{s1}(i) P_{s2}(i) \cdots P_{sN}(i))^T$ at M_s by simply replacing $\mathbf{A}_v$ with the spatial vector $\mathbf{A}_s$ for M_s . Eventually, the following relation holds:

$$\mathbf{P}_s = \mathbf{C}_s^{-1} \mathbf{Q} \quad (7)$$

where wavelength constant matrix $\mathbf{C}_s$ has the form:

$$\mathbf{C}_s = (c_{s,mn}) \quad (8)$$

$$c_{s,mn} = \exp\left(j \frac{2\pi i (\mathbf{A}_m - \mathbf{A}_s) \cdot (\mathbf{B}_n - \mathbf{A}_s)}{T_v |\mathbf{B}_n - \mathbf{A}_s|}\right) \quad (9)$$

Matrix $\mathbf{C}_s$ has been formulated by taking into account the wavelength constant property so that $\mathbf{C}_s$ plays a role in translating $\mathbf{P}_s$ into $\mathbf{Q}$. Particularly, the m -th column and n -th row element, $c_{s,mn}$, in matrix $\mathbf{C}_s$ plays a role in translating the n -th desired sound spectrum, $P_n(i)$, which is assumed at M_s into one of the components in the m -th microphone spectrum $Q_m(i)$. Therefore, $Q_m(i)$ can be formulated by summing up the quantity of $c_{s,mn} P_{sn}(i)$ from $n=1$ upto N . In addition, it is also an implicit function of the unknown angle (θ_n, φ_n) and given angles (ξ_m, η_m) and (ξ_s, η_s) .

Noting that the sum of all estimated sound spectrum $P_n(i)$ derived from Eq. (7) with $n=1 \sim N$ will approach the reference spectrum $Q_s(i)$ for the respective frequency component if the source angles have been exactly substituted into Eq. (9), it is possible to seek the optimal source angles in the sense of the least mean square error, by using the following criterion, e , formed in terms of implicit variables (θ_n, φ_n) as

$$e = \sum_{i=I_1}^{I_2} \left| \frac{Q_s(i) - \sum_{n=1}^N \sum_{m=1}^M H_{s,mn}(i) Q_m(i)}{Q_s(i)} \right|^2 \quad (10)$$

$$(H_{s,mn}) = \mathbf{C}_s^{-1} \quad (11)$$

where I_1 and I_2 denote the lower and upper numbers of the cutoff frequency that give the available frequency range and are related to the mutual relation between the spacing of two adjacent microphones and the frequency. Error function, e , implies the power difference between the reference sound spectrum $Q_s(i)$ at M_s and the estimated spectrum given by the sum of $H_{s,mn}(i) Q_m(i)$ with $m=1 \sim M$. It is also normalized by the former. The procedure of Eq. (10) is implemented as shown on the bottom and right side of Fig. 1.

3. The experimental performance

The condition of the computer simulation for sound detection is as follows: In the multi-sound field, first two loudspeakers at (1.5 m, 60°, 40°) and (1.5 m, 145°, 80°) radiated digits $p_1(k)/ni/$ ('two' in Japanese) and $p_2(k)/zero/$, respectively. Two other loudspeakers at (1.5 m, 150°, 110°) and (1.5 m, 105°, 130°) simultaneous-

Fig. 2 Experimental performance of the sound detection system. Sampling frequency is 10 kHz.

ly radiated band limited noise $p_3(k)$ with a center frequency of 0.5 kHz and $p_4(k)$ with 1 kHz. On the other hand, four microphones M_n 's with $n=1\sim 4$ were mounted on the X, Y, Z and $-X$ axes with an equal length of 2 cm from the origin. Moreover, the specified point M_v and the reference microphone M_s were set at the origin. In addition, the sampling frequency is 10 kHz.

Estimation of the bearing angles φ_n for four arriving sounds in the case stated above gave a feasible accuracy with the mean error being less than 0.94° under the condition that zenith angle θ_n are known. However, the estimation accuracy appeared to rapidly diminish in such case that two sounds arrive in closer mutual direction under 14° . The available number of the sound direction for the present system is limited to four angles so far.

The experimental performance for sound detection is illustrated in Fig. 2. The original sounds, $p_1(k) \sim p_4(k)$, are shown in the first four rows for reference. One of the microphone outputs, denoted as $q_1(k)$, is also shown in the fifth row, which implies the measured sound with this system off, namely, the sum of $p_n(k)$ from $n=1$ to N . Sound waves in the bottom four rows denote the detected sounds, $\hat{p}_1(k) \sim \hat{p}_4(k)$, obtained with this system on, respectively. There is no observable appearance of the original sound in $q_1(k)$. On the other hand, comparison of the detected sounds, $\hat{p}_1(k) \sim \hat{p}_4(k)$, with the original sounds, $p_1(k) \sim p_4(k)$,

one by one, shows a significant similarity between them, in other words, the waveform recovery of the original from $q_1(k)$. In this experiment, noise reduction level in digits /ni/, for example, reached 19 dB. Additionally, the coherency of recovered digits with the original increased to 0.93, while it was 0.01 before using this system.

In conclusion, it is proved that the sound detection system basing it on the wavelength constant matrix and using multiple microphones is fundamentally valid as long as the number of sounds is under five. It is also shown here that this system is capable of detecting sound at the points of choice. However, the overall performance in sound direction estimation which seriously reflects on sound detection, is thus far not satisfactory.

References

- 1) M. Abe, Y. Nagata, and K. Kido, "A new method to locate sound sources by searching the minimum value of error function," Proc. IEEE-IECEJ Int. Conf. Acoust. Speech Signal Process., 18B.2.1-18B.2.4 (1986).
- 2) Y. Kaneda and J. Ohga, "Adaptive microphone array system for noise reduction," IEEE Trans. Acoust. Speech Signal Process. **ASSP-34**, 1391-1400 (1986).
- 3) K. Kuroiwa, "Sound pickup for a reflective sound field using a microphone array," Proc. 15th Int. Cong. Acoustics, Vol. 4, 107-110 (1996).