Analysis of moving sound sources —Separation of multiple sources—

Tsuyoshi Usagawa,* Seizi Nishimura,* Masanao Ebata,*
and Josuke Okda**

**Department of Electrical Engineering and Computer Science, Kumamoto University,
2-39-1, Kurokami, Kumamoto, 860 Japan*

***Department of Electrical Engineering, Kyushu Tokai University,
223, Ohe-Toroku, Kumamoto, 862 Japan*

(Received 7 September 1987)

This paper proposes the identification and separation method of multiple sound sources on a single moving object. The proposed method is based on the compensation of the Doppler effect for a specific sound source and uses the difference of temporal characteristics between the specific sound source and others. After deriving both of observed and compensated spectra in a general form, we discuss the frequency shift of two sound sources model. Then simple models composed of two moving sources which radiate sinusoidal waves are simulated, and the results show the effectiveness of the proposed method. The proposed method is effective only for narrow band or sinusoidal signal, because this method depends on the temporal characteristics of frequency shift due to the movement of sound sources.

PACS number: 43. 60. Gk, 43. 60. Rw, 43. 85. Kr

1. INTRODUCTION

To control the noise, we need the spectral and positional information of the noise source. Usually, it is also required to separate the specific sound source from surrounding other sources. In the previous paper,^{1,2)} we studied the method to derive the spectral and positional information of a single moving source. To analyze actual noise sources, we need to separate and identify multiple sources at once.

As a side effect of the reduction of frequency shift due to the Doppler effect, the compensation for a specific source makes the signals from other sources nonstationary as the "Doppler-like effect." This evidence suggests the possibility to separate the specific sound source under multiple sound sources condition.

In this paper, we first derive the directly observed spectrum and the compensated spectrum for the

specific source under general multiple sound sources condition, and discuss a simple model which consists of two sound sources on a single moving object. Then we propose the method to separate and identify multiple moving sources and carry out the computer simulation to evaluate the performance of the proposed method.

2. GENERAL REPRESENTATION OF OBSERVED AND COMPENSATED SPECTRA³⁾

2.1 General Model of Multiple Moving Sources

Let us assume that we observe the N sound sources $S_1 \dots S_N$ by the single sensor M , as shown in Fig. 1. Let position vector of the source S_i from the sensor M at time t_i be $r_i(t_i)$. Though the sensor is fixed, sources are either in motion or fixed. When the signal $x_i(t_i)$ is radiated by the source S_i at time t_i and it is observed by the sensor at time t , the observed signal $y(t)$ and its short term spectrum

Fig. 1 Geometrical relationship between sensor and sound sources.

$Y(f; t)$ are described as follows;

$$y(t) = \sum_{i=1}^N a_i(t_i) x_i(t_i) + n(t), \quad (1)$$

$$Y(f; t) = \sum_{i=1}^N a_i(t_i) X_i(f; t_i) + N(f; t), \quad (2)$$

where $a_i(t_i)$ is an attenuation factor which relates to the distance between the i -th source and the sensor, and $X_i(f; t_i)$ is the short term spectrum of $x_i(t_i)$ which is defined as follows;

$$X_i(f; t_i) = \frac{1}{T} \int_{t_i-T/2}^{t_i+T/2} x_i(\tau) \exp(j2\pi f\tau) d\tau, \quad 0 < T \ll 1. \quad (3)$$

Also $n(t)$ is an observation noise and it is independent of each $x_i(t_i)$. $N(f; t)$ is the spectrum of $n(t)$.

When c represents the velocity of sound, the relationship between radiated time t_i and observation time t is represented as,

$$t = t_i + \frac{|r_i(t_i)|}{c} = g_i^{-1}(t_i), \quad t_i = g_i(t). \quad (4)$$

2.2 Spectrum of Directly Observed Signal

Using the first order approximation of $g_i(t)$, $g_i(t) = g_{i0} + g_i'(t)t$, the observed spectrum $Y(f; t)$ is given as,

$$Y(f; t) = \sum_{i=1}^N \frac{a_i(g_i(t))}{g_i'(t)} X_i\left(\frac{f}{g_i'(t)}; g_i(t)\right) \cdot \exp\left(j2\pi f \frac{g_{i0}}{g_i'(t)}\right) + N(f; t). \quad (5)$$

In the summation of Eq. (5), the first component is the attenuation due to the propagation, the second is the spectrum which is affected by the

Doppler effect and the last is the phase shift.

2.3 Spectrum of Compensated Signal

Because the Doppler effect is due to the temporal change of propagation delay, we can compensate the Doppler effect by transforming the observation time into the radiated time for a specific sound source.

After the compensation is made for the k -th source's component, the compensated signal $y_k(t_k)$ is given as,

$$y_k(t_k) = a_k(t_k) x_k(t_k) + \sum_{i=1, i \neq k}^N a_i(t_i) x_i(t_i) + n(t). \quad (6)$$

Using the following relationship between t_k and t_i ,

$$t_i = g_i(t) = g_i(g_k^{-1}(t_k)) = g_{i0} + g_i'(g_k^{-1}(t_k)) g_k^{-1'}(t_k) t_k, \quad (7)$$

the spectrum of compensated signal is represented as,

$$Y_k(f; t_k) = a_k(t_k) X_k(f; t_k) + \sum_{i=1, i \neq k}^N \frac{a_i(g_i(g_k^{-1}(t_k)))}{g_i'(g_k^{-1}(t_k)) g_k^{-1'}(t_k)} \cdot X_k\left(\frac{f}{g_i'(g_k^{-1}(t_k)) g_k^{-1'}(t_k)}; g_i(g_k^{-1}(t_k))\right) \cdot \exp\left(j2\pi f \frac{g_{i0}}{g_i'(g_k^{-1}(t_k)) g_k^{-1'}(t_k)}\right) + \frac{1}{g_k^{-1'}(t_k)} \cdot N\left(\frac{f}{g_k^{-1'}(t_k)}; g_k^{-1}(t_k)\right) \cdot \exp\left(j2\pi f \frac{g_{k0}^{-1}}{g_k^{-1'}(t_k)}\right). \quad (8)$$

In contrast with Eq. (5), the spectrum component of the k -th source is the same as original spectrum except the attenuation. On the other hand, the last term which corresponds to the observation noise is affected by the Doppler-like effect.

2.4 Observed and Compensated Spectra of Two Sources Model

The amount of frequency shift is highly correlated to the relative velocity of each source, and it is minimum when the sound sources move as a single object. In this section, to discuss the difference between temporal characteristics of observed and compensated spectra, we derive the frequency shift for a simple model. The model consists of two sources which move on a straight line as a single object at

Fig. 2 Example model with two sound sources on a single moving object.

8.33 m/s as shown in Fig. 2. Figure 3 shows the frequency shift under the condition shown in Fig. 2. Figure 3(a) shows the frequency shift of the observed spectrum which is derived from Eq. (5) where $N=2$ and without observation noise. Figure 3(b) shows the frequency shift of the compensated spectrum for the moving source S_1 and it is derived from Eq. (8) where $N=2$ and $k=1$. The amount of frequency shift of S_2 in Fig. 3(b) is equal to the difference between the frequency shifts of two sources in Fig. 3(a), and depends on all of the velocity of object, the distance between sources and the trajectory of object. Because of the frequency shift, the component of S_2 observed with different sensors is less coherent than that of S_1 so that it is possible to separate and identify the component of source S_1 from the observed signal using the cross-spectral technique.^{4,5)}

3. SOURCE SEPARATION USING IMAGINARY SOUND SOURCE

In Chap. 2, the compensation of the Doppler effect is made for one of the real sound sources. However the compensation can be made for an arbitrary position relative to the estimated trajectory using the imaginary source. In this chapter we propose the separation method of multiple moving sources using the imaginary source.

Now suppose that the sound source S_1 in the previous two source model is an imaginary source and S_2 is a real one; i.e. only S_2 radiates sound wave. When the compensation of the Doppler effect is carried out for source S_1 , the component of source S_2 is affected by the Doppler-like effect and it is nonstationary. Therefore, the obtained spectrum can be suppressed by the cross-spectral technique. On the contrary, when the imaginary source S_1 overlaps with the real source S_2 , the following relationships are satisfied.

$$g_1(t) = g_2(t), \quad g_1^{-1}(t_1) = g_2^{-1}(t), \quad (9)$$

$$g_2'(g_1^{-1}(t_1))g_1'(t_1) = 1. \quad (10)$$

Thus, there is no frequency shift in the spectral component of the real source S_2 .

Using such characteristics of imaginary sound source, multiple sources can be separated by deriving the power level at arbitrary position on a moving object by the imaginary sound source. Figure 4 shows the block diagram of the proposed method to analyze multiple moving sources. The signal observed by the sensor array which is composed of

Fig. 3 Temporal change of frequency shift under the condition shown in Fig. 2. (a) Directly observed signal, (b) Compensated signal (Compensated for S_1).

Fig. 4 Configuration of the proposed method.

three microphones is first used to estimate the trajectory information of the specific sound source (Fig. 4(a)). The method of trajectory estimation consists of two parts²⁾; the position estimator using a pair of relative propagation delays, and the trajectory estimator using Kalman filter for reduction of random observation error. Then we derive the equal power level contour for the specific area on the relative coordinate fixed on the obtained trajectory as in Fig. 4(b)~(f). First, the observation points are decided at regular intervals (Fig. 4(b)), then the imaginary source is assumed at one of the observation points in order (Fig. 4(c)). The power level at the observation point is derived after the Doppler effect is compensated (Fig. 4(d) and (e)). The compensation method of the Doppler effect is based on the transformation of the time axis and it is detailed in the previous paper.¹⁾ When we carry out the compensation for the imaginary source, the accuracy in estimation of the propagation delay directly affects the results. To obtain the accurate

propagation delay, we need to consider the relationship between radiated time and observation time for the imaginary sound source, as discussed in appendix. The equal power level contour is composed using the power levels at all observation points (Fig. 4(f)).

4. COMPUTER SIMULATION

4.1 Model Setup

Figure 5 shows the geometry of simulation model, which consists of two moving sources and three sensors. The moving sound sources S_1 and S_2 are both on the same object which moves at constant speed (30 km/h = 8.33 m/s) on a line which is parallel to the sensor array. The distance between two sound sources is 2 m, and the trajectory of sources is 3 m apart from sensor array. The velocity of sound is 340 m/s.

Under the above condition, the following two models are simulated;

MODEL 1 S_1 : 500 Hz sinusoidal source,
 S_2 : 600 Hz sinusoidal source,
 (The power level of S_1 is 10 dB higher than that of S_2).

MODEL 2 S_1 : 500 Hz sinusoidal source,
 S_2 : 500 Hz sinusoidal source,
 (The power levels of S_1 and S_2 are the same).

The origin of the coordinate is fixed at the estimated trajectory of sound source S_1 , and X and Y axes represent the direction of the movement of sound source S_1 and the direction perpendicular to it, respectively. The observation points locate on each 0.05 m step to cover both sources.

4.2 Result of Simulation

Figure 6 shows the results of simulation for MODEL 1. Figure 6(a) and (b) show the contours of equal power level of 500 Hz and 600 Hz, respec-

Fig. 5 Simulation model of two moving sources as a single object.

Fig. 6 Results of simulation of MODEL 1 ($\times$: position of sources). (a) Equal power level contour at 500 Hz, (b) Equal power level contour at 600 Hz.

Fig. 7 Results of simulation of MODEL 2 ($\times$: position of sources) Equal power level contour at 500 Hz.

tively. The mark $\times$ in Fig. 6(a) is the origin and represents the location of the source S_1 , and the one in Fig. 6(b) corresponds to the source S_2 . In Fig. 6(a), the peak of contour is obtained at (0.0, 0.05) in meter, and its level is +0.6 dB. This peak corresponds to the source S_1 , and the true position of S_1 is (0, 0). In Fig. 6(b), the peak for the source S_2 is given at (-1.95, 0.0) with -11.6 dB where the true position is (-2, 0). The error of peak position is 0.05 m for both sources. Also the relative level of the source S_2 is -12.2 dB, and so the error is

2.2 dB.

Figure 7 shows the result of simulation for MODEL 2 as the contour of equal power level at 500 Hz. The coordinate is the same as in Fig. 6. Two marks $\times$ in Fig. 7 represent the location of the two sources; right and left marks correspond to S_1 and S_2 , respectively. The peaks of the obtained contour are given at (-0.05, 0.25) with +0.3 dB and at (-2.15, 0.05) with -0.7 dB. The true position of S_1 and S_2 are (0, 0) and (-2, 0) respectively, and so the error is more than that of MODEL 1. Though the radiated levels of two sources are the same, the obtained level of S_2 is 1.0 dB lower than that of S_1 .

5. CONCLUSION

This paper proposed the separation and identification method of multiple moving sound sources based on the estimated trajectory information. First, we derived the spectra of the observed and compensated signal on a multiple sources condition, and discussed a simple model which consists of two sources. Then, we proposed the separation method of multiple sources using the estimated trajectory of specific source. The results of the computer simulation show the effectiveness of the proposed method for two models both of which consist of two sinusoidal sources. The proposed method is based on the frequency shift of the observed signal, and so it is able to separate only moving sources which

radiate narrow band or sinusoidal signal.

REFERENCE

- 1) T. Usagawa, K. Matsuno, M. Ebata, and J. Okda, "Analysis of moving sound source by the cross-spectral technique—A method for reducing frequency shift by the Doppler effect—," *J. Acoust. Soc. Jpn. (E)* **6**, 281–288 (1985).
- 2) T. Usagawa, S. Nishimura, M. Ebata, and J. Okda, "Analysis of moving sound source—Orbit estimation using microphone array—," *J. Acoust. Soc. Jpn. (E)* **8**, 23–28 (1987).
- 3) T. Usagawa, S. Nishimura, M. Ebata, and J. Okda, "Analysis of moving sound sources by the cross-spectral technique—Identification of multiple moving sources—," *Tech. Rep. IECE*, EA86-26, 27–34 (1986) (in Japanese).
- 4) K. Kido, Text Book used in the 13th Lecture of *Acoust. Soc. Jpn.*, 25–46 (1976) (in Japanese).
- 5) J. Y. Chung, "Rejection of flow noise using a coherence function method," *J. Acoust. Soc. Am.* **62**, 388–395 (1977).

APPENDIX: PROPAGATION DELAY OF THE IMAGINARY SOURCES

Because the position of moving sound source is estimated by means of the cross correlation technique, the given position is not the position where the source exists at the observation time, but the one where the observed sound is radiated.

Figure A(a) shows the geometry of the sound sources S and N, and the sensor M at the observation time t . Let us assume that the sound sources move as a single object and the trajectory of source S, shown as the solid line in Fig. A, is known. Also the relative position of the source N is known as R_n . The vectors $r_s(t)$ and $r_n(t)$ represent the positional vectors from sensor M to the position of source S and N, respectively.

At time t_s , the source S radiates the signal which is observed at time t so that the propagation delay, $d_s(t)$, is represented as,

$$d_s(t) = \frac{|r_s(t_s)|}{c}. \quad (\text{A.1})$$

Note that the propagation delay $d_s(t)$ is given by the time t_s , not t . Also the positional information derived from the trajectory estimator at time t is the position of source S at time t_s , not t . The relationship between two sources and sensor at time t_s is shown in Fig. A(b).

Furthermore, the radiation time of source N, t_n , is not equal to t_s so that we cannot derive the position

Fig. A Geometry of observed trajectory and sound sources. (a) at time t (observation time), (b) at time t_s (radiated time of source S), (c) at time t_n (radiated time of source N).

of source N, $r_n(t_n)$, directly from the estimated position of source S, $r_s(t_s)$. $r_n(t_n)$ has to be estimated by the trajectory of source S.

At time t , the following relationship is held,

$$r_n(t) = r_s(t) + R_n. \quad (\text{A.2})$$

Also the positional vector $r_s(t)$ is represented as follows by the given position $r_s(t_s)$.

$$r_s(t) = r_s(t_s) + \int_{t_s}^t V(\tau) d\tau, \quad (\text{A.3})$$

where $V(t)$ is the velocity of object as a vector. Using relationship of Eq. (A.3), Eq. (A.2) can be written as,

$$r_n(t) = r_s(t_s) + \int_{t_s}^t V(\tau) d\tau + R_n. \quad (\text{A.4})$$

Also the propagation delay of source N, $d_n(t)$, is given as,

$$d_n(t) = \frac{|r_n(t_n)|}{c}, \quad (\text{A.5})$$

where the positional vector $r_n(t_n)$ satisfies the following relationship,

$$r_n(t) = r_n(t_n) + \int_{t_n}^t V(\tau) d\tau. \quad (\text{A.6})$$

Using relationship of Eqs. (A.4)~(A.6), $d_n(t)$ is derived as follows,

$$\begin{aligned} d_n(t) &= \left| r_n(t) - \int_{t_n}^t V(\tau) d\tau \right| / c \\ &= \left| r_s(t_s) + \int_{t_s}^t V(\tau) d\tau - \int_{t_n}^t V(\tau) d\tau + R_n \right| / c \\ &= \left| r_s(t_s) - \int_{t_n}^{t_s} V(\tau) d\tau + R_n \right| / c. \end{aligned} \quad (\text{A.7})$$