<table>
<thead>
<tr>
<th>Title</th>
<th>Study on method of estimating direction of arrival using monaural modulation spectrum</th>
</tr>
</thead>
<tbody>
<tr>
<td>Author(s)</td>
<td>Ando, Masaru; Morikawa, Daisuke; Unoki, Masashi</td>
</tr>
<tr>
<td>Citation</td>
<td>Journal of Signal Processing, 18(4): 197-200</td>
</tr>
<tr>
<td>Issue Date</td>
<td>2014</td>
</tr>
<tr>
<td>Type</td>
<td>Journal Article</td>
</tr>
<tr>
<td>Text version</td>
<td>publisher</td>
</tr>
<tr>
<td>URL</td>
<td><a href="http://hdl.handle.net/10119/12893">http://hdl.handle.net/10119/12893</a></td>
</tr>
</tbody>
</table>
Abstract

Human beings can localize a target sound by using binaural cues. On the other hand, we can also localize a target sound by using monaural cues. The monaural modulation spectrum (MMS) can be regarded as an important cue of monaural sound localization. A method of estimating direction of arrival (DOA) using a machine learning scheme for classification of MMS patterns has been proposed. However, this method cannot account for the monaural DOA mechanism by using MMS patterns. To further investigate how the MMS plays an important role in monaural sound localization, we aimed to find cues in the human ability for monaural sound localization and propose a method of estimating DOA by using these cues. We investigated how the MMS of observed signals vary with the azimuth. As a result, shapes of the MMS were drawn as arcs with azimuth variations. We then proposed a method of estimating monaural DOA using these results. Simulations were carried out to verify the effectiveness of the proposed method. We found that the proposed method could estimate DOA using the MMS, except with front-back confusion discrimination.

1. Introduction

Human beings have the ability of sound localization. For example, we can easily localize the direction of an on-coming car from the noise from the car. In general, human beings use binaural cues to localize a target sound. It has been reported that humans can also localize a target sound by using monaural cues [1]. Gaining knowledge on the human ability of sound localization is important in learning more about our hearing mechanism. A method of estimating the direction of arrival (DOA) of a target sound using monaural cues can be applied to single-channel signal processing if we can apply our ability of sound localization to engineering problems.

The main cues for sound localization by using binaural hearing are interaural time difference (ITD), interaural level difference (ILD), and spectral information [2]. They are included in the head-related transfer function (HRTF), which is a transfer function between a sound source and eardrum position in each ear. In these cues, available monaural cues for sound localization can be regarded as spectral cues in the HRTF such as peaks and notches in the monaural spectral envelope. However, it is unclear how the peaks and notches in the monaural spectral envelope vary with the DOA of the sound source; therefore, these cues cannot be used to directly estimate the DOA of a sound source.

On the other hand, there have been studies on binaural modulation cues for sound localization. Thompson and Dau reported that ILD and ITD in the temporal envelope are also important cues of sound localization [3]. This report suggests that the monaural modulation spectrum (MMS) can be regarded as an important cue of monaural sound localization.

There have been related studies conducted regarding DOA with the MMS approach. Kliper et al. proposed a DOA estimation method using monaural cues in amplitude modulation patterns based on a machine learning scheme [4]. They used the MMS patterns of signals observed at the eardrum position. However, they used the machine learning scheme to classify the MMS patterns to directly estimate the azimuth of the sound source. Therefore, their method cannot account for the monaural DOA mechanism. In particular, with their method, it is still unclear how the MMS patterns can be used for monaural sound localization.

We aimed to find important monaural cues for sound localization and propose a method of estimating DOA using these cues. We investigated how the MMS of the observed signals vary with the azimuth to find monaural cues of DOA estimation. We propose a method based on the concept of the modulation transfer function (MTF).

2. Model Concept

Figure 1(a) shows a transfer function from the sound source signal to the observed signal.
source to the observed signal at the eardrum position in the
time domain, where \( y(t, \theta) \), \( h(t, \theta) \), \( x(t) \), and \( \theta \) are the ob-
served signal, head-related impulse response (HRIR), sound
source signal, and arrival direction of the sound source signal,
respectively. The observed signal is represented as

\[
y(t, \theta) = h(t, \theta) * x(t)
\]  \hspace{1cm} (1)

where \( * \) is a convolution operation. The HRIR includes
acoustic characteristics such as pinna reflection and head
diffraction.

Equation (1) can be represented in the frequency domain as

\[
Y(f, \theta) = H(f, \theta)X(f)
\]  \hspace{1cm} (2)

where \( Y(f, \theta) \), \( H(f, \theta) \), and \( X(f) \) are the spectrum of the
observed signal, HRIR, and spectrum of sound source signal,
respectively. Figure 1(b) represents a transfer function in the
modulation domain from the power envelope of the original
signal to that of the observed signal. This function is in a
different domain, as shown in Fig. 1(a), which is based on
the concept of the MTF [5], [6]. The power envelope of the
observed signal \( e_y^2(t, \theta) \) can be represented as

\[
e_y^2(t, \theta) = e_h^2(t, \theta) + e_x^2(t) \]  \hspace{1cm} (3)

where \( e_h^2(t, \theta) \) and \( e_x^2(t) \) are the power envelopes of \( h(t, \theta) \)
and \( x(t) \).

Equation (3) can be represented in the modulation-
frequency domain as

\[
E_y(f_m, \theta) = E_h(f_m, \theta)E_x(f_m) \]  \hspace{1cm} (4)

where \( E_y(f_m, \theta) \), \( E_h(f_m, \theta) \), and \( E_x(f_m) \) are MMS of
\( y(t, \theta) \), head-related MTF (HRMTF), and MMS of \( x(t) \), re-
spectively. The term \( f_m \) is the modulation frequency. Then,
HRMTF is defined as

\[
E_h(f_m, \theta) = \int_0^\infty e_h^2(t, \theta) exp(-j2\pi f_m t) dt \]  \hspace{1cm} (5)

In this study, \( e_y^2(t, \theta) \) was extracted by

\[
e_y^2(t, \theta) = \text{LPF} \left[ |y(t, \theta) + j\text{Hilbert}[y(t, \theta)]|^2 \right] \]  \hspace{1cm} (6)

where LPF[-] is a low-pass filtering and Hilbert[-] is the
Hilbert transform. This equation is based on calculation of
the instantaneous amplitude, and low-pass filtering is used to
remove the higher modulation-frequency components in the
power envelope as post-processing. We use the LPF with a
cut-off frequency of 200 Hz. Finally, \( e_y^2(t, \theta) \) is transformed to
\( E_y(f_m, \theta) \) using fast Fourier transform (FFT).

3. Monaural Modulation-Spectrum Analysis

We investigated how the MMSs of the observed signals
vary with varying azimuth; from 180 to 355 degrees, in which

Figure 2: MMS characteristics with varying azimuth by AM
signal

Figure 3: MMS characteristics with varying azimuth by AM
noise
that the effect is caused by HRIR personality. Simulations by using the other HRIRs. Therefore, we argue results in this paper, similar trends could be observed in thecies, as shown in Figs. 2 and 3. Although we omit these variations were almost the same with all modulation frequen-
ties for the right ears the left ear. The dynamic ranges of MMS shapes with the AM noise were smoother than those with the AM signal. This trend varied depending on the individual’s ears. Similar trends could be also observed for the right ears the left ear. The dynamic ranges of MMS variations were almost the same with all modulation frequen-
cies, as shown in Figs. 2 and 3. Although we omit these results in this paper, similar trends could be observed in the simulations by using the other HRIRs. Therefore, we argue that this effect is caused by HRIR personality.

These results suggest that humans may use cues based on the tendency of variation in the MMS. However, it is necessary to carry out a listening experiment.

4. Proposed Method

We propose a method of estimating DOA based on the results of MMS analysis. The flow of the proposed method is shown in Fig. 4. The MMS values are plotted with open circles in Figs. 2 and 3. These plots are approximated using second order polynomials as follows:

\[ \hat{E}_y(f_m, \theta) = p_1(f_m)\theta^2 + p_2(f_m)\theta + p_3(f_m) \]  \hspace{1cm} (7)

where \( p_1(f_m) \), \( p_2(f_m) \), and \( p_3(f_m) \) are the regression coefficients and \( \hat{E}_y(f_m, \theta) \) is the approximated value. The solid lines in Figs. 2 and 3 indicate the ideal results from Eq. (7). An inverse function is derived by using these regression curves. This can be represented as follows:

\[ \hat{\theta}(E_y) = -\frac{p_2 \pm \sqrt{p_2^2 - 4p_1(p_3 - E_y)}}{2p_1} \]  \hspace{1cm} (8)

where \( \hat{\theta} \) is the estimated azimuth. If HRIR is known, \( p_1 \), \( p_2 \), and \( p_3 \) can be calculated from the MMS of the observed signals \( y(t, \theta) \).

We assume that the input signal of the proposed method is \( y(t) \) with unknown azimuth \( \theta \). Then, the power envelope \( E^2_\gamma(t) \) is calculated using Eq. (6) and \( E_y(f_m) \) is calculated using the FFT. Finally, the unknown azimuth \( \theta \) is estimated by substituting the MMS values and regression coefficients into Eq. (8).

5. Evaluations

Simulations were carried out to verify the effectiveness of the proposed method. The AM signal, AM noise, and left ear side were used in these simulations. Figure 5 shows the simulation results. The horizontal axis indicates the azimuths of the input signals and the vertical axis indicates the estimated
azimuths. There was no effect by varying the modulation frequency, as shown in Fig. 5.

Two azimuths were estimated as positive and negative values derived by the inverse function of Eq. (8). For the positive value, estimates were correct in the back of ear position while, estimates were correct in the front of ear position for the negative value. However, in each reverse, estimates were incorrect. These false estimates were due to front-back confusion. Moreover, there were more false estimates with the AM signal than with AM noise.

These results indicate that the proposed method can correctly estimate DOA using the MMS, except with front-back confusion discrimination.

6. Conclusions

We investigated how the MMS of observed signals vary with the azimuth. The results showed that the MMS varied with the azimuth where the peak of the shape was around the ear position. We then proposed a method of estimating DOA based on our analysis results. These results indicated that the proposed method could correctly estimate DOA using the MMS, except with front-back confusion discrimination.

For future work, we will investigate how to solve the discrimination problem with regard to front-back confusion.

Acknowledgments

This work was supported by the Strategic Information and Communications R & D Promotion Programme (SCOPE: 131205001) of the Ministry of Internal Affairs and Communications (MIC), Japan. It was also supported by a Grant-in-Aid for Young Scientists (Start-up, No. 25880011).

References