SOPA version 3

SOPA project

July 22, 2015

Contents

1 Principle
   1.1 Introduction ................................................. 2
   1.2 Direction of propagation .................................. 3
   1.3 Speed of propagation ...................................... 4

2 Data configuration .......................... 8
1 Principle

1.1 Introduction

The SOPA (Stream Of Panoramic Audio) technology is grounded on the following assumption.

“The spatial audio at the given observation point can be specified by giving

- Amplitude
- Phase
- Direction of propagation
- Apparent speed of propagation

of each sine wave at the observation point.”

In general, every wave can be regarded as the sum of one or more sine waves. By applying an N-point FFT, any signal can be subdivided into sinusoids of the frequencies between \( f_0 \) Hz and \( f_{N-1} \) Hz, inclusive. Since its frequency is known, each sine wave can be specified by its amplitude and phase. The amplitude and phase are potentially contained in the monaural signal itself. A monaural signal, however, does not contain spatial information.

These sine waves are usually propagating. As long as it is propagating, the sine wave has the direction and speed of propagation. In monaural recording, however, information about the direction and speed of propagation is totally lost.

The propagation directions only in the horizontal plane were managed in SOPA version 1 [1] and the 3-dimensional propagation directions were specified in SOPA version 2 [2]. The speed of propagation, however, had not been dealt with in either SOPA version 1 or 2.

To improve the accuracy of spatial information, not only the propagation direction but also the propagation speed of each sine wave is considered in SOPA version 3.
What are conveyed in the SOPA data are one stream of a monaural signal and spatial information of sine waves that compose the monaural signal. In SOPA version 1 and 2, spatial information consisted of only the propagation directions of the sine waves. In SOPA version 3, it consists of not only the propagation direction but also the propagation speed of each sine wave. In the SOPA technology, panoramic sounds can be generated by using a monaural audio data stream and directional information conveyed in SOPA data.

1.2 Direction of propagation

Propagation of a sine wave is shown in Fig. 1. The left figure shows a sine wave generated by a single sound source. A sine wave generated by 2 sources is shown in the right figure. The triangles on the top indicate the sound source positions. The arrows in the figures represent the directions of propagation. The observation point is represented by the ‘X’ in each figure.

The arrows in Fig. 2 represent the sound image directions. In the SOPA format, the direction from which the sine wave is propagating is defined as the ‘Sound Image Direction.’ It represents the opposite of the propagation direction. In the case of a sine wave generated by a single source, the sound
image direction corresponds to the wave normal direction and it also corresponds to the sound source direction.

When the sine wave is generated by more than 2 sources, a composite wave of the signals from the sources is observed. The composite wave has only one direction of propagation at the observation point and it does not correspond to the sound source directions. Please note that there is only one sound image direction for each sine wave no matter how many sound sources exist.

1.3 Speed of propagation

Since the directions of propagation in the left and right figures of Fig. 2 are the same, it should be difficult to discriminate these signals at the observation point if the propagation directions are the only cues available. To discriminate them, the speed of propagation has to be considered.

The arrows in Fig. 3 represent the wave length of the sine wave at the observation point. It is longer in the right figure than in the left indicating that the speed of propagation in the right figure is faster than that in the left.

The sound speed or the wave length affects reflection and diffraction of the sound. Consequently, it affects the interaural phase difference and interaural level difference in the actual life.

It is important, therefore, to make each sine wave propagate not only from the proper direction but also at the proper speed. Both of the direction
and speed of propagation can be obtained by a multi-point measurement. In the 3-dimensional space, for example, they can be obtained by a 4-point measurement. A miniature head simulator system that has a tetrahedral microphone system can be used for this purpose [3].

In SOPA version 3, the propagation speed of the sine wave is implemented by assuming 2 imaginary sound sources for each frequency bin. Let us call A and B the imaginary sound sources.

As illustrated in Fig. 4, the propagation speed of the sine wave can be controlled by the angle between A and B. The ‘X’ in each figure represents the reproduction point and A and B are the imaginary sources. The directions of these imaginary sources are determined so that the sound image direction locates at the middle of them.

The directions of the imaginary sources are supposed to be represented as 3D vectors. In the SOPA version 3 file, however, they are represented by 8-bit numbers instead of 3D vectors to save space.

Conversion from 3D vectors to 8-bit numbers is accomplished by the following procedure [3].

• Assume an imaginary sphere around the reproduction point (Fig. 5)

• Divide the surface of the sphere into 254 sectors (Fig. 6)
Fig. 4: Imaginary sound sources and the propagation speed

\( \theta \) is the angle between the imaginary sources and \( c \) is the normal sound speed that is approximately 340 m/s. When \( \theta \) is 0 (left figure), the sine wave propagates at the normal sound speed. By modifying the angle \( \theta \), the propagation speed can be controlled.

Fig. 5: Imaginary sphere around the reproduction point

Fig. 6: Sectors on the surface of the sphere

- Numbers between 0 and 253, inclusive, are assigned, one by one, to the sectors (Fig. 7)
- The propagation direction can be represented by the number between 0 and 253

Fig. 5 shows an imaginary sphere around the reproduction point. The surface of the sphere is divided into 254 sectors as illustrated in Fig. 6 and the numbers between 0 and 253, inclusive, are attached, one by one, to the sectors. The numbered sectors are shown in Fig. 7. The front face is the up direction in Fig. 7. By using these numbers, directions from the reproduction point can be encoded as 8-bit numbers instead of 3D vectors.

By assuming 2 imaginary sources for each frequency bin, in SOPA version
Integers between 0 and 253 inclusive are assigned, one by one, to the sectors on the surface of the sphere.

3, it is possible to make each sine wave propagate from the intended direction and at the intended speed. Since accuracy of spatial information improved, directionality of sounds can be programatically manipulated in SOPA version 3.
2 Data configuration

At the top of a SOPA file, there is a header that is followed by a data chunk. In SOPA version 3, directions of 2 imaginary sound sources, instead of the sound image direction, are specified for each frequency bin. Fig. 8 shows the data configuration and an example of the header is shown in Fig. 9. Each frame (the frame No. is on the bottom of Fig. 8) has \( N \) samples and the frames overlap one another by 50%. In SOPA version 3, frames always overlap by 50%.

Fig. 8: Data stream of SOPA version 3

The data stream of the SOPA version 3 consists of the directional data and the monaural audio data stream.

Each audio sample is a 16-bit number while each imaginary source direction is an 8-bit number.

The yellow cells and green cells in Fig 8 contain the directional data of the imaginary source A and B, respectively.

In the data chunk, the imaginary source directions for the frequency bins are represented by 8-bit unsigned integers. In the case of the \( f_0 \) (DC) component, however, the direction of propagation has no importance. The addresses corresponding to the \( f_0 \) Hz component (area with red colored text in Fig. 8) can contain any integer, and we put 255s there.

‘FF’ in hex, ‘255’ in decimal can be seen at the address corresponding to \( f_0 \) Hz in Fig. 9 (colored by red). Since it is not used for any other frequencies, ‘255’ appears only at addresses corresponding to \( f_0 \) Hz and can be used as the marker that indicates the beginning of the frame. By counting the
In the previous version of the format, the overlap factor that is colored by green in Fig. 9 was used to specify whether the frames overlap by 50% or by 75%. It is no longer used since the frames always overlap by 50% in the SOPA version 3.

Panoramic sound can be synthesized either by presenting each sine wave from both of its imaginary source directions or by superimposing the mean of the HRTFs of these directions to the amplitude and phase of each sine wave.

Fig. 9: Header of SOPA version 3

<table>
<thead>
<tr>
<th>Offset</th>
<th>0</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>A</th>
<th>B</th>
<th>C</th>
<th>D</th>
<th>E</th>
<th>F</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ox00</td>
<td>5249</td>
<td>4648</td>
<td>Ri</td>
<td>FF</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>E47B</td>
<td>2D00</td>
<td>2980836</td>
<td>(Chunk size)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>534F</td>
<td>5041</td>
<td>SQ</td>
<td>PA</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>fmt</td>
</tr>
<tr>
<td>Ox10</td>
<td>1000</td>
<td>0000</td>
<td>16</td>
<td>0</td>
<td>0100</td>
<td>0400</td>
<td>1</td>
<td>4</td>
<td>(PCM)</td>
<td>Not used</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ox20</td>
<td>0400</td>
<td>1000</td>
<td>4</td>
<td>16</td>
<td>0000</td>
<td>0003</td>
<td>3.000</td>
<td>(SOPA file version)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ox30</td>
<td>7E7E</td>
<td>0500</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>126</td>
<td>126</td>
<td>5</td>
<td>0</td>
<td>111</td>
<td>126</td>
<td>3</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(f2, f2)</td>
<td>(S2)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

References