A SYSTEM FOR MULTIPLE SOUND SOURCE LOCALIZATION

Jie Huang, Noboru Onishi, Noboru Sugie

Department of Electrical Engineering
Faculty of Engineering, Nagoya University
Furo-cho, Chikusa-ku, Nagoya 464-01, Japan

ABSTRACT

In the present article we present a new method for multiple sound source localization. The method relies on the phase differences among the corresponding frequency components derived from the signals picked up by a set of three microphones, where the sound sources are assumed to be coplanar with three microphones. In order to cope with echoes, the precedence effect discovered by Wallach is incorporated. As for the multiple sound source localization, the phase differences are measured around the transient portions of the signals and the candidates for the correspondences between the transient portions of the frequency components of the signals are extracted. A kind of histogram is used to determine the correct correspondences from the candidates, which results in multiple sound source localization.

1. Introduction

In robots as well as humans, the recognition of the surrounding environment is essential in order to behave adaptively. So far, robots are provided with vision[1] but not with audition for this purpose. It is well known that the human auditory system is capable of localizing sound sources. For a moble robot, too, it is very useful to be equipped with a system for sound source localization. The system will inform the robot of the existence of some possibly moving objects invisible from the robot. Then the robot can move towards the objects and see what they are. In the present article we present a new method for multiple sound source localization. The method relies on the phase differences among the frequency components derived from the signals picked up by a set of three microphones, where the sound sources are assumed to be coplanar with three microphones. In order to cope with echoes, the precedence effect[2] discovered by Wallach is incorporated. That is, the phase differences are measured around the transient onset portions of the frequency component. As for the multiple sound source localization, a kind of histogram is used to determine the correct correspondences between onset portions from a pair of microphones. The method has been implemented on a personal computer-based system and tested in both an anechoic chamber and an echoic chamber, where one or two sound sources are located. The results of the experiments have been quite satisfactory. The directions of sound sources have been determined within the accuracy of a few degrees.

2. Sound Source Localization System with Three Microphones

The system consists of four units as shown in Fig.1. That is, a set of three microphones as the input unit, three sets of bandpass filters as the pre-processing unit, a personal computer as the processing unit, and the data recorder unit used for offline processing.

We set up three microphones at the vertices of a regular triangle each side of which is 13.5cm in length. Note that the distance between the two human ears is about 18cm. The microphone set can be rotated around its center. Sound takes 396 microseconds to pass through the two microphones. The pre-processing unit consists of three sets of filters each triple of which have the same characteristics. As will be explained below, we use the interaural temporal difference as the cue. Thus if we use only two microphones, we can not resolve anterior-posterior ambiguity. This is the reason why we use three microphones.

3. Single Sound Source Localization

3.1 Detecting Onsets of Sound

The precedence effect[2)] in humans suggests an effective way for localization even in echoic environments. The effect means that the human auditory system gives priority to the preceding sound by masking the just succeeding one. Since the reflected sound passes a longer way than the direct sound before arriving microphones, so just after each onset there is no influence of reflected sound in the received signal. To incorporate the effect into the system, we must detect the onsets of sound. We set up a threshold level in the amplitude of sound and localize the sound source only during the short period of time after the onsets. The precedence effect is incorporated in the detection of onsets. The details of detection are described in section 4.

3.2 Measuring Interaural Temporal Difference

The frequency components of the sound obtained from the narrowly tuned bandpass filters are nearly pure tones, so we can use the zerocrossing method to detect the interaural(this term means between two microphones rather than two ears) temporal difference (ITD). We explain the method for measuring ITD in the following two cases: (1) there is at least one frequency component whose half wave length is longer than the intermicrophone distance (IMD); (2) for all frequency components their half wave lengths are shorter than intermicrophone distance. In the former case, we can determine the correspondence between waveforms from two microphones directly from the phase difference in the lowest frequency. In other word, we can determine ITD directly. in the latter case, we can not determine the correspondence between waveforms directly, since several waves appear during the time period of passage of sound through IMD. In this case, we must use multiple components of sound. For all of these components we can determine the candidates of ITD as shown in Fig.2. In this figure, the starting and ending points of arrows indicate the zerocrossing points of two microphones, respectively. The lengths of arrows indicate the values of ITD candidates, where left-ward arrows mean positive values and right-ward arrows mean negative values.

These ITD candidates, denoted as $\Delta tc^{(m)}$, have the following relation with real ITD, denoted as $\Delta t^{(m)}$.

$$\Delta t^{(m)}=\Delta tc^{(m)}+k_m/f_m \qquad (1)$$

Here, k_m is an integer, while m as the subscript and superscript corresponds to the component m the center frequency of which is f_m. For all components of sound, $\Delta t^{(m)}$ should indicate the same ITD, so we have the following relations.

$$\Delta t^{(1)}=\Delta t^{(2)}= . . . \qquad (2)$$

If the components are not related to one another as a fundamental frequency and its higher harmonics, then we can determine uniquely ITD from (2).

3.3 Determining Azimuth of Sound Source

The microphone set and the sound source are assumed to be coplaner as shown in Fig.3. The coordinates of the three microphones and the sound source are (x_1,y_1), (x_2,y_2), (x_3,y_3), $(d_s,0)$, respectively. If the d_s is much larger than IMD, ITD in microphones i,j, denoted as $\Delta t_{ij}'$, can be approximated as follows.

$$\Delta t_{ij}'=-(x_i-x_j)/C \qquad (3)$$

Here, i,j=1,2,3, while C is the velocity of sound. It is a function of sound source azimuth θ. Let the measured ITDs be Δt_{ij}. We use the following estimation formula. Note that $\Delta t_{ij}'$ depends on θ.

$$E(\theta)=\sum_{i,j}(\Delta t_{ij}-\Delta t_{ij}')^2 \qquad (4)$$

We can determine the sound source azimuth θ so that the estimation formula (4) is minimized.

4. Localization of Multiple Sound Sources

We decompose sound to several frequency components, detect the onsets of components and localize the sound source during the short period of time after onsets. As shown in Fig.4, there are many onsets in human voices.It is well accepted that completely continuous sound is hard to localize. We consider only such sound containing many onsets. So we can localize one sound source at one onset, and the other at another onset. It is natural to assume that the onsets of different sound sources rarely coincide. We introduce a method using histograms. We take a large number of ITD candidates to form histograms and from the peaks of the histograms, we can get the real ITDs. We make use of the following constraint as well to distinguish the ITDs of one sound source from others.

$$\Delta t_{12}+\Delta t_{23}+\Delta t_{31}=0 \qquad (5)$$

5. Experiments and Results

We conducted some experiments on multiple sound source localization in an anechoic chamber. This chamber has a dimension of $5.5*5.5*3.0m^3$. The cutoff frequency is 200Hz and the background noise is 20dB. The sound sources were, (1) the voices of a male announcer (s1), and (2) the conversation between a man and a woman (s2). We set up the equipments as shown in Fig.5. The difference in azimuth between s1 and s2 is about 38 degrees, while θ was changed systematically by rotating the microphone set.

To detect the onsets of sound sources, we set up the following two judging conditions:

(1) the increase in amplitude in a certain period is larger than k(k is a constant value larger than 1);

(2) the amplitude is larger than a certain threshold level.

We determined the ITD candidate at every onset and constructed a histogram as shown in Fig.6(a). By convolving with Gauss function it was smoothed as shown in Fig.6(b). While the histogram has some large peaks corresponding to ITD candidates, there are some spreads around the peaks. To make a comparison, we got a histogram for a single sound source. As shown in Fig.6(c), it also shows some spread as in multiple sound source localization. It suggests that the spreads are not due to the overlaps of onsets of s1 and s2.

We could get the ITDs from the peaks in the histogram. Then we could distinguish the ITDs of one sound source from the other employing the constraint (5). Thus finally we could localize the multiple sound sources by ITDs.

Fig.7 shows the results of multiple sound source localization. In this figure, the abscissa indicates the set up sound source azimuth, and the ordinate indicates the results of multiple sound source localization. The dashed lines show the ideal accurate localizations. From this figure, we could see that sound source azimuth was localized within the accuracy of a few degrees.

In the case of a single sound source, we conducted experiments in an echoic chamber. The results of localization were quite satisfactory, so we expect in the case of multiple sound source localization our system will work in echoic chambers.

6. Conclusion

In this paper, we proposed a method for multiple sound source localization. The method was implemented on a personal computer-based system and experiments were carried out successfully. In the present system, the processing is done offline due to the slow processing speed. But the recent rapid progress in digital signal processors will make the real time processing possible. Finally, if we use a set of four microphones constituting a regular pyramid, sound source localization in 3D will be possible.

Acknowledgement This work was supported in part by the Sound Technology Promotion Foundation. We gratefully acknowledge Nagoya Municipal Laboratory of Technology for offering us an anechoic chamber.

References

1) A.Pugh (ed.): "Robot Vision", Springer Verlag, Berlin, 1983

2) H.Wallach and others: "The Precedence effect in sound localization", Amer.J.Psych. Vol.LXII, No.3, pp.315-336 (1949)

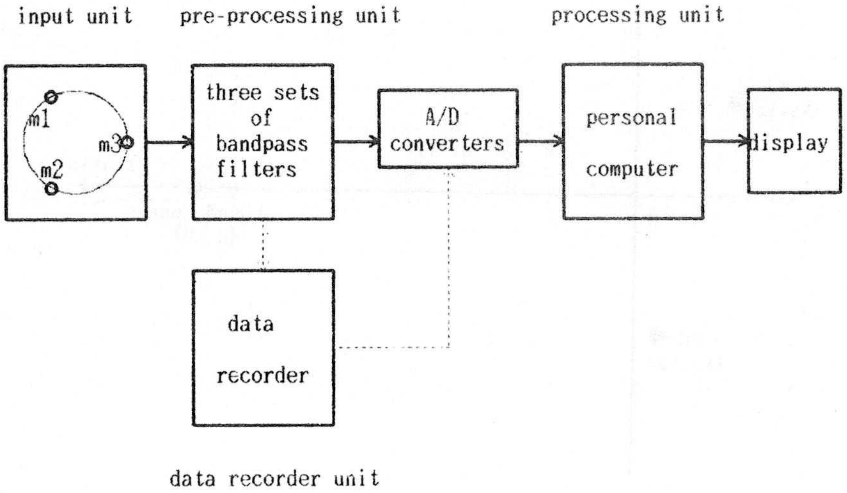

Fig.1 System organization.

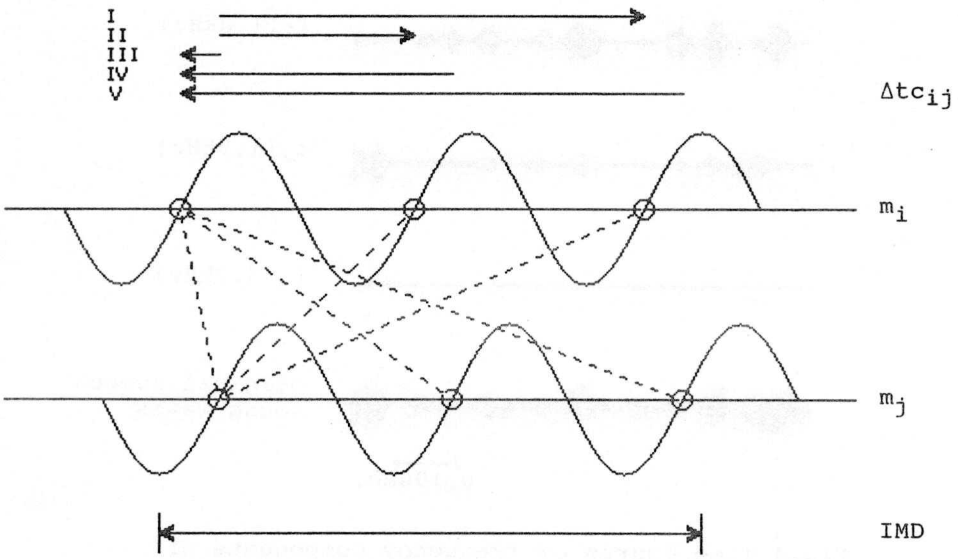

Fig.2 Correspondences of waveforms from two microphones.

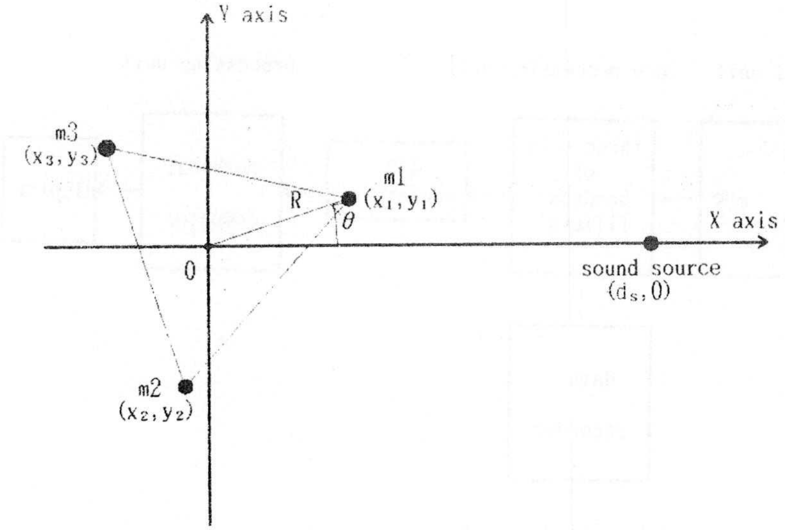

Fig.3 Geometry of sound source and microphones.

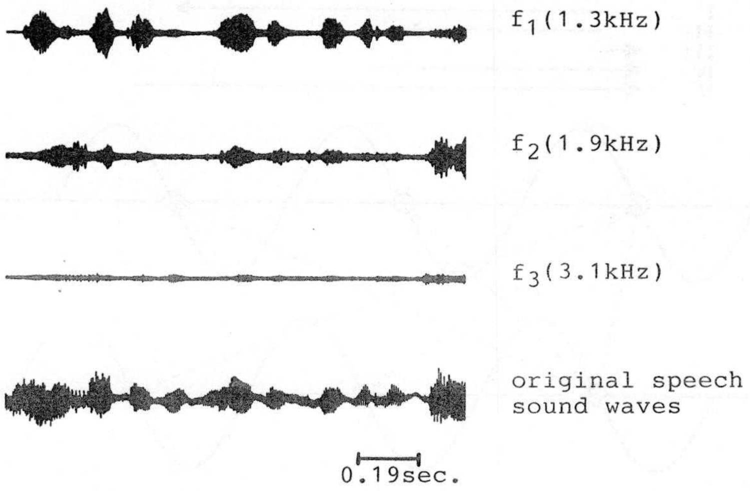

Fig.4 Time course of frequency components of human voice.The bottom trace shows original speech sound waves.

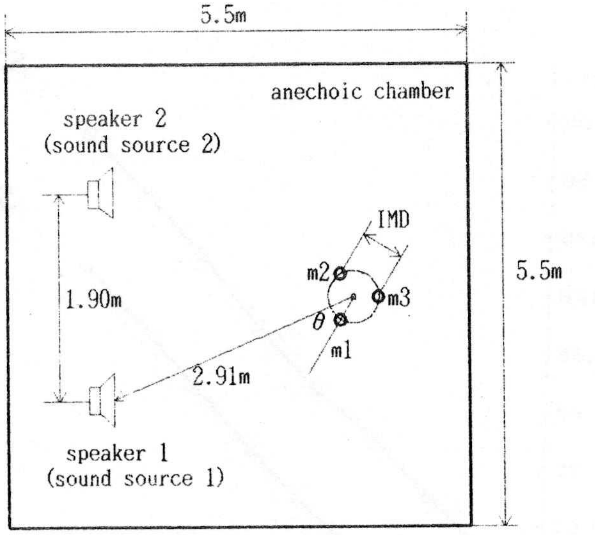

Fig.5 Experimental setup.

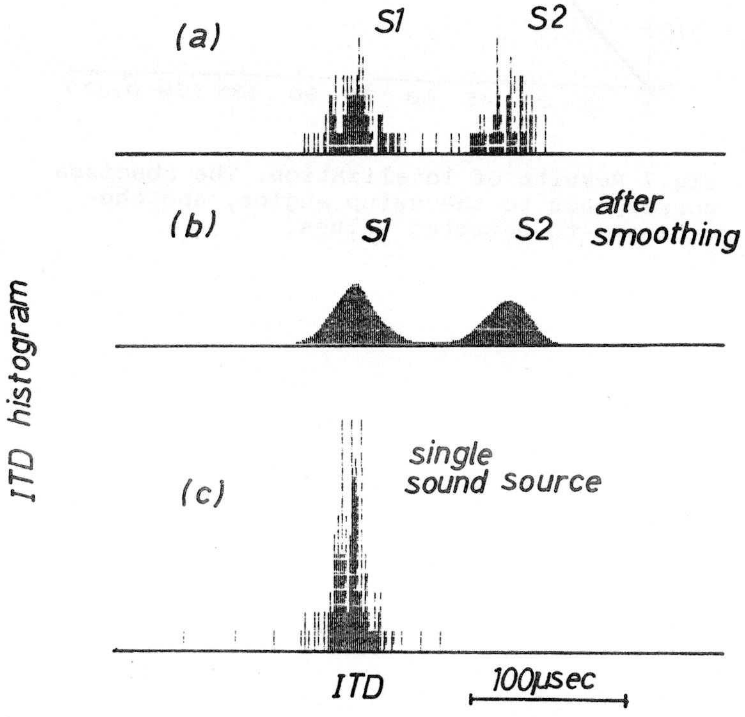

Fig.6 Histogram of ITD candidates. (a) Original histogram. (b)Smoothed histogram. (c)Histogram for a single sound source.

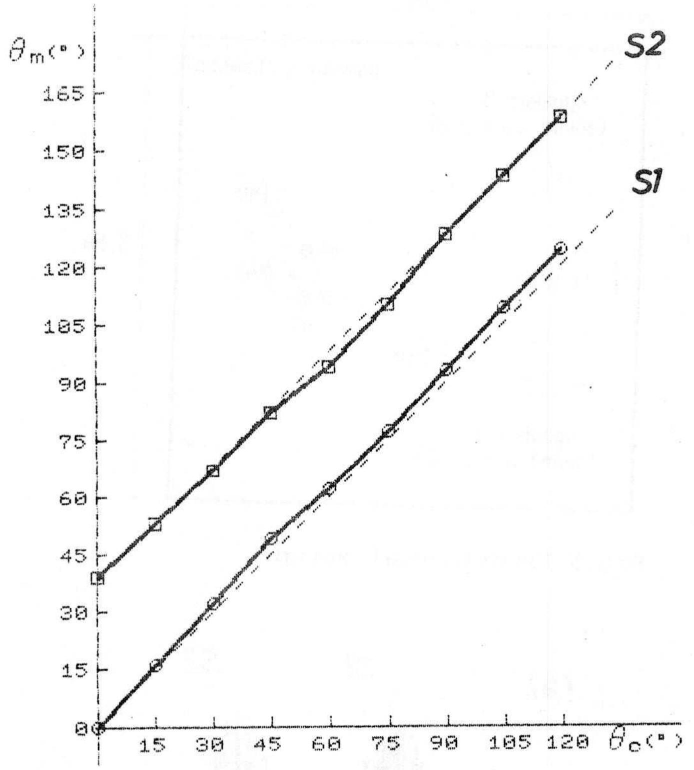

Fig.7 Results of localization. The abscissa corresponds to the setup angles, and the ordinate to detected values.