25

Sound Localization for Robot Navigation

Jie Huang
School of computer science and engineering The University of Aizu
j-huang@u-aizu.ac.jp

1. Introduction

For mobile robots, multiple modalities are important to recognize the environment. Visual sensor is the most popular sensor used today for mobile robots Medioni and Kang (2005). The visual sensor can be used to detect a target and identify its position Huang et al. (2006). However, since a robot generally looks at the external world from a camera, difficulties will occur when a object does not exist in the visual field of the camera or when the lighting is poor. Vision-based robots cannot detect a non-visual event that in many cases with sound emissions. In these situations, the most useful information is provided by auditory sensor. Audition is one of the most important senses used by humans and animals to recognize their environments Heffner and Heffner (1992). Sound localization ability is particularly important. Biological research has revealed that the evolution of the mammalian audible frequency range is related to the need to localize sound, and the evolution of the localization acuity appears to be related to the size of the field of best vision (the central field of vision with high acuity) Heffner and Heffner (1992). Sound localization enables a mammal to direct its field of best vision to a sound source. This ability is important for robots as well. Any robot designed to move around our living space and communicate with humans must also be equipped with an auditory system capable of sound localization.

For mobile robots, auditory systems also can complement and cooperate with vision systems. For example, sound localization can enable the robot to direct its camera to a sound source and integrate the information with vision Huang et al. (1997a); Aarabi and Zaky (2000); Okuno et al. (2001).

Although the spatial accuracy of audition is relatively low compared with that of vision, audition has several unique properties. Audition is omnidirectional. When a sound source emits energy, the sound fills the air, and a sound receiver (microphone) receives the sound energy from all directions. Audition mixes the signals into a one-dimensional time series, making it easier to detect any urgent or emergency signal. Some specialized cameras can also receive an image from all directions, but still have to scan the total area to locate a specific object Nishizawa et al. (1993). Audition requires no illumination, enabling a robot to work in darkness or poor lighting condition. Audition also is less effected by obstacles, so a robot can perceive the auditory information from a source behind obstacles. Even when a sound source is outside a room or behind a corner, the robot can first localize the sound source in the position of the door or corner then travel to that point and listen again, until it finally localize the sound source.

Many robotic auditory systems, similar to the human auditory system, are equipped with two microphones Bekey (2005); Hornstein et al. (2006). In the human auditory system, the sound localization cues consist of interaural time difference (ITD), interaural intensity difference (IID) and spectral change. Among them, ITD and IID cues are more precise than the spectral cue which, caused by the head and outer ears (pinnas), is ambiguous and largely individual dependent Blauert (1997). While the ATD and IID cues are mainly used for azimuth localization, the spectral cue is used for front-back judgment and elevation localization of spatial sound sources. The localization accuracy of elevation is far lower than azimuth.

In section 5, we describe a four-microphone robotic auditory system, three around the spherical head at the same horizontal plane with the head center and one at the top, for localization of spatial sound sources Huang et al. (1997b). The arrival time differences (ATD) to the four microphones were used to localize the sound sources. A model of the precedence effect was used to reduce the influence of echoes and reverberation. Azimuth or azimuth-elevation histograms were introduced by integrating the ATD histograms with the restrictions between different ATDs. From the histograms, the possibilities of sound sources can be obtained from the position of histogram peaks.

Comparing with other microphone array based methods Johnson and Dudgeon (1993); Valin et al. (2007), the array based methods use more microphones and need more computation power. While the array based methods are basically armed to obtain the best average of cross-correlation between different microphones, our method is based on the ATDs for different frequency components and different time frames. Since cross-correlation based methods calculate the cross-correlation function for all of the frequency components, when a sound source has high intensity than others, the low intensity sound sources will be easily masked. However, as we experience by our auditory system, we can usually localize sound sources with an intensity difference concurrently. It is because the different sound sources have different frequency components and appear in different time frames. By this mean, the histogram based method can have the advantage to distinguish sound sources with an intensity difference.

2. Bio-mimetic approach for Sound localization

When we design a sound localization system for a robot used in real environments, the tasks with first priority will be the robustness, high efficiency and less computation. It will be largely benefited by learning from the bio mechanisms.

High efficiency

Different animals will have different approaches to localize sound sources. For example, the humans use both ITD and IID cues for horizontal sound localization, and left the ambiguous spectral cue for elevation sound localization. It is because horizontal localization is the most important task the humans in daily
life.

The barn owls since need to localize both azimuth and elevation exactly in the darkness to hunt mice, they take a strategy to use ITD for azimuth localization and IID for elevation localization. Their right ear is directed slightly upward and the left ear directed downward Knudsen (1981). This up-down disparity provides the barn owls IID for elevation sound

localization. Insects like the bush crickets use their legs as acoustic tracheal system to extend the time and/or intensity differences for sound localization Shen (1993).
For a robot, it is possible to choose more than two microphones for sound localization. Since the ITD cue is the most precise one, if we use four microphones and arrange them at different directions covering the ITD cue for all of the directions, it will be highly efficient for the robot to localize sound sources in both azimuth and elevation.

Available for multiple sound sources
When there are multiple sound sources concurrently, we need some methods to distinguish them each other. The cross correlation method is a most popular method for time difference calculation. It gives the similarities between two signals for different time disparities, and we can find the most similar point by its peak position. By this method, we usually need to analyze signals for a relitave long time period to identify multiple sound sources. Moreover, since the cross correlation method is a averaging method for all of the frequency components, if there are two signals with different intensities, the smaller signal will be masked by the louder signal.
In the human auditory system, the time difference is calculated by local information, limited in both time and frequency range. The local time difference information is then integrated by an histogram-like method, with the weights not only depends on its intensity but also effected by the precedence effect, to find out the correct ITDs (characteristic delays) for different sound sources. This method can distinguish multiple sound sources even with intensity disparities.

Robustness
Since many robots are to be used in the human daily environments, the robustness for echoes and reverberation is very important for robotic auditory systems. As in the human auditory system, the robotic auditory systems also need to incorporate the precedence effect. By the EA model of the precedence effect, we can calculate the weights depend on the estimated sound-to-echo ratios to give more priority to echo free onsets to reduce the influence of echoes. The precedence effect not only reduce the influence of echoes, but also can increase the separation rate for multiple sound sources. It is because the echo free onsets are usually the parts of the beginning of a sound, or the parts where the sound level increases sharply. In all cases, the sound intensity is contributed by a single dominant sound source.

3. Sound localization cues in the human auditory system

3.1 Cues for azimuth localization

Sound localization is to identify the direction, including azimuth and elevation, of sound sources by the information obtained from the sound signals. For horizontal sound sources, it is well known that the interaural time differences (ITD) and interaural intensity differences (IID) are the most important localization cues Blauert (1997).
The ITD cue is caused by the arrival distance disparity from sound source to the two ears. When the distance of sound source is far away from the head $(D \gg d)$, the incidence of sound is parallel and the arrival time difference can be approximated as

$$\Delta t = \frac{d}{2}(\theta + sin\theta) \ -\frac{\pi}{2} \leq \theta \leq \frac{\pi}{2} \tag{1}$$

where d is the diameter of head and θ is the azimuth of sound source.
In the human auditory system, the time disparities of a sound signal is coded by the neural phase-lock mechanism, i.e., auditory neurons fire at a particular phase angle of a tonal stimulus up to about 5 kHz Gelfand (1998). For signals have frequency components of more than about 1.5kHz, where the wavelength becomes shorter than the distance between the two ears, the time difference information can not be recovered from the phase difference uniquely because of the phase wrapping phenomenon.
Suppose the phase difference of frequency component f is $\Delta\psi_f(-\pi < \Delta\psi_f \leq \pi)$, the possible real phase difference $\Delta\hat{\psi}_f$ can be

$$\Delta\hat{\psi}_f(n) = \Delta\psi_f + 2\pi n_f \tag{2}$$

where $n_f$is an integer depends on each frequency.
Biological studies about owl's auditory system revealed that the phase difference is detected by a neural coincident detector Konishi (1986), and the reduction of redundancy is done by finding out the characteristic delay (CD) Takahashi and Konishi (1986), the common time difference among multiple different frequency components.
For sound signals containing multiple frequency components, the task is to find an integer n_f for each frequency component *f*, so that the time difference $t_{CD}$is the same for all frequency components

$$\Delta t_{CD} = \frac{1}{2\pi f}\Delta\hat{\psi}_f = \frac{1}{2\pi f}(\Delta\psi_f + 2\pi n_f) \tag{3}$$

On the other hand, the IID cue, caused by the shadow effect of the head and outer ears, is significant for high frequency sounds but becomes weak as the frequency decreases. It is large when the sound comes from side directions (left and right) and small when the sound is from front and back. It is more complex to formulate the intensity difference compared to the time difference. Addition to the interaural cues, the spectral cue is used to disambiguate frontback confusion of the ITD and IID cues Blauert (1997).

3.2 Cues for azimuth and elevation localization

For sound sources in the 3D space, interaural cues are not enough for both azimuth and elevation localization. For example, the possible source positions which create the same ITD to the two ears will be approximately a locus of conical shell which known as the cone-of-confusion in psychoacoustic studies Blauert (1997). The changes of spectral characteristics are important for elevation localization. For example, sound sources in the median plane will not create any interaural difference (assume the auditory system is left-right symmetry), sound localization is mainly due to the spectral cues.
The directional spectral changes can be represented by the transfer function from the sound source to the two ears, the so-called head-related transfer functions (HRTFs). The frequency characteristics of HRTFs, influenced by the head, ears, shoulder and body, are variant with azimuth θ and elevation ρ. By the spectral changes together with ITD and IID cues, the

auditory system can identify the azimuth and elevation of a sound source concurrently. However, compared to the interaural cues, the spectral cues are weaker, individual dependent, and easy to be confused.

In general, the HRTFs contain not only spectral cues, but also interau-ral cues. Representing the HRTFs for left and right ears as $Hi(9,ip,f)$ and $H_r\ (9,ip,f)$, the cross interaural transfer function can be defined as

$$H_x(\theta,\rho,f) = \frac{|H_l(\theta,\rho,f)|}{|H_r(\theta,\rho,f)|} e^{jp(\theta,\rho,f)} \tag{4}$$

or the opposite $(H_x = H_r/H_l)$, where p is the phase difference. The amplitude part of H_x provides the ITD, and the group delay of the phase part provides the IID information.

4. The Echo-Avoidance Model of the Precedence Effect

4.1 Introduction

When a sound is presented in a reverberant environment, listeners usually can localize the sound at the correct source position, being unaware and little influenced by the surrounding reflections. This phenomenon, including its different aspects, is referred to by different names, Haas effect Haas (1951), Franssen effect Franssen (1959), the first front effect Blauert (1997) and the precedence effect Wallach et al. (1949).

The precedence effect has been a topic of continuous theoretical interest in the field of psychoacoustics for more than half a century Gardner (1968). Evident from developmental psychological studies suggest it is a learned effect in the human auditory system to cope with echoes in ordinary reverberant environments Clifton et al. (1984). Because humans spend much time indoors in a typical reverberant environment, the needs for a human to localize sound may cause the human auditory system to adapt to reverberant environments.

Recent studies also show that the precedence effect is active and dynamic. Blauert and Col Blauert and Col (1989, 1992), Clifton and Freyman Clifton (1987); Clifton and Freyman (1989) and Duda Duda (1996) reported that the precedence effect can break down and become re-established in an anechoic chamber or hearing test by headphones.

The precedence effect is an important factor for acoustical architecture design Haas (1951) and stereo sound reproduction Snow (1953); Parkin and Humphreys (1958). It is also important for computational sound localization in reverberant environments Huang et al. (1995, 1997b). Moreover, since the precedence effect influences the spatial cues in reverberant environments, it is also important for the perceptual segregation and integration of sound mixtures, the so-called cocktail-party effect Blauert (1997); Bodden (1993); Cherry (1953), or auditory scene analysis Bregman (1990) and its computational modeling Cooke (1993); Ellis (1994); Huang et al. (1997a); Lehn (1997).

Despite the large number of psychological studies on the precedence effect, there have been few computational modeling studies. Some abstract models, e.g. funneling models von Bekesy (1960); Thurlow et al. (1965), inhibition models Haas (1951); Harris et al. (1963); Zurek (1980, 1987); Martin (1997) and others McFadden (1973); Lindemann (1986a,b); Litovsky and Macmillan (1994) have been proposed for the precedence effect. Basically, while the funneling models proposed that the localization of succeeding sounds is biased toward the direction which has been established by the first-arriving sound, the inhibition

models argued that the onset of a sound may trigger a delayed reaction which inhibits the contribution of succeeding sounds to localization.
In all those models, the precedence effect is considered to be triggered by an "onset". Zurek argued that the onset should be a very "rapid" one, but no quantitative criterion was given for a rapid onset. Furthermore, neither funnel-ing nor inhibition models provide a consistent explanation for psychoacoustic experiments with different types of sound sources.
According to the Zurek model, the inhibition signal takes effect after a delay of about 800μs, and lasts for a few milliseconds. The inhibition interval was determined based on the just-noticeable difference (JND) tests of interaural delay and intensity judgment which showed that the JND level increases in the interval range from about 800μs to 5 ms. However, the psychological experiments conducted by Franssen indicated that the sound image of constant level pure tone was localized by the transient onset and could be maintained for a time interval of seconds or longer Franssen (1959); Hartmann and Rakerd (1989); Blauert (1997). Other psychological experiments, e.g. those by Haas (1951) using speech and filtered continuous noise, have shown that the inhibition occurs after a time delay of about 1 ms to about 50 ms according to the type of sound source used in the tests.
The Zurek model cannot distinguish the different phenomena caused by different types of stimuli. One more point to be noted is that the inhibition in the Zurek model was absolute, i.e., a very small onset can inhibit any high-intensity succeeding sound. This obviously conflicts with the fact that the precedence effect can be canceled by a higher-intensity succeeding sound.
A computational model on the precedence effect must give a systematic interpretation of the results of psychological tests and provide a theoretical explanation for the phenomenon.
Because of the needs for human to localize sounds in reverberant environments, it is our opinion that there should be a mechanism which can estimate the level of reflected sounds and emit an inhibition signal to the sound localization mechanism, so that the neural pathway from low to high level of localization processing can be controlled to avoid the influence of reflections. Such a mechanism is possibly located in the cochlear nucleus Oertel and Wickesberg (1996). From this point of view, the precedence effect can be interpreted as an "echo-avoidance" effect. Here as well as later, the term "echo" is used with the wide meaning of all reflected sounds by the surrounding.
In this section we will propose a new computational model of the precedence effect, the Echo-Avoidance (EA) model (Section 1.4.2), with an echo estimation mechanism. We will show that the EA model of the precedence effect can be used to detect available onsets which are relatively less influenced by echoes. The model can explain why the precedence effect occurs in transient onsets and can interpret the data obtained by several psychological experiments consistently
(Section 1.4.3).

4.2 The Computational Echo-Avoidance (EA) Model

The EA model of the precedence effect, similar to the Zurek model, consists of two paths, one for localization cue processing and one for inhibition signal generation as shown in Fig. 1.1. We assume that the echo estimation and inhibition mechanism is independent for different frequencies Hafter et al. (1988). Both binaural and monaural localization processes are effected by the precedence effect Rakerd and Hartmann (1992). In the integration process, averaging for different localization cues in all of frequency subbands will take place.

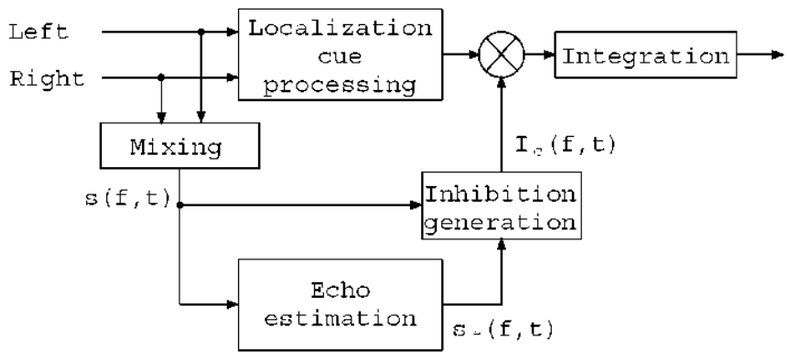


Fig. 1. EA model of the precedence effect.

When an impulsive sound is presented in a reverberant environment, the resulting signals arrive our ears are first the direct sound and then followed by a series of reflections. The sequence of reflections depends on the shape and reflection rates of the surfaces in the environment, and the position of sound source and observation points.
It is impossible for the auditory system to distinguish all of the reflections one-by-one. The sound image we perceive is a series of sound impulses whose amplitudes decay over time in an minus exponential manner approximately. Thus, the auditory system may learn about two features, decay and delay, related to the reflections.
These two features can provide a prospective pattern of echoes:

$$h_e(t) = \{ \begin{cases} 0 & ,0\tau_0 \\ ke^{-(t-\tau_0)/\tau_d} & ,t \geq \tau_0 \end{cases} \tag{5}$$

where k, τ_0 and τ_d are learned parameters, correspond to the strength and delay time of the first reflections, and the time constant of decay respectively.
Denote the sound level of a particular frequency by $s(f,t)$. Thus, the possible echo can be estimated as:

$$s_e(f,t) = \underset{t}{Max}\{s(f,t-t')h_e(t')\}, \quad \tau_0 < t' < \infty \tag{6}$$

Here, the operator Max, instead of sum, is to take the maximum value for all of t', since $h_e(t)$ is not a real impulse response but an approximation pattern.
An abstract illustration of the relation between received sound and estimated echoes is given in Fig. 2(a). The signal is represented in discrete time with an interval t_s which can be the sampling interval or the length of a time frame. By the exponential decay feature in the echo estimation mechanism, the algorithm

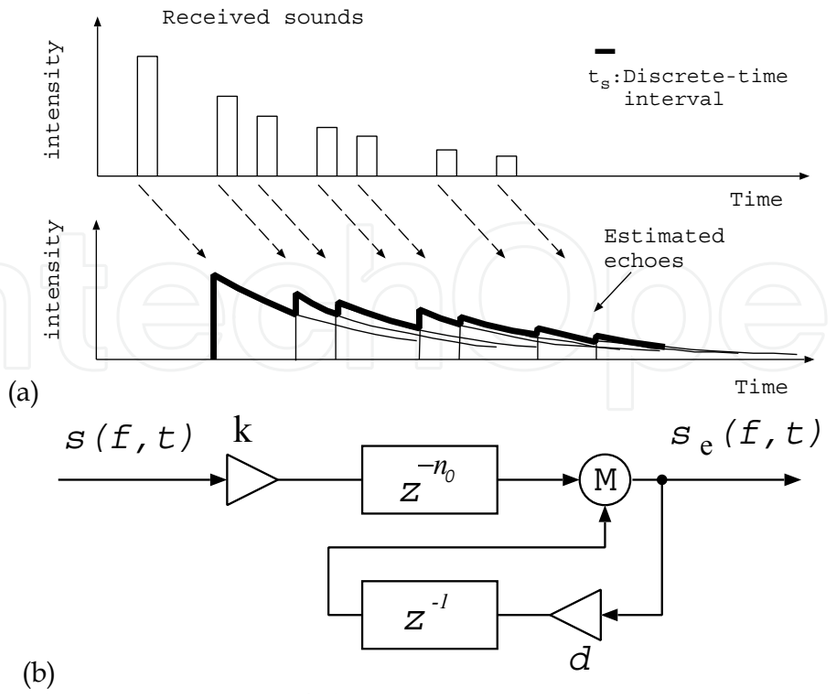


Fig. 2. (a) An abstract illustration of the relation between received sound and estimated echoes. (b) A feedback algorithm for echo estimation.

can be implemented using a feed-back algorithm as shown in Fig. 2(b), where $n_0 = \tau_0/t_s$ and $d = e^{-t_s/\tau_d}$.

An example of echo estimation and echo-free onset (see Section 1.4.3) detection is shown in Fig. 1.3. In this figure, the last plot shows the positions of echo-free onsets given as $s(f,t)/s_e(f,t) \geq 3$ (The value 'three' of sound-to-echo ratio is not critical. From our experience, a weight of 0.1 to 1 corresponding to sound-to-echo ratios 2.1 to 3.0 were used for averaging in sound localization algorithm.) It is clear that each echo-free onset corresponds to a sharp increase point in amplitude, and the sound level before the echo-free onset is relatively low. Another clear fact is that the echo-free onset does not depend on the absolute strength of the sound.

Fig. 1.4 shows the influence of echoes on arrival time differences for a time period around echo-free onsets detected by the EA model. In this figure, the horizontal axis is time with the origin about 0.03 second before each detected onset. The first plot is the average amplitude of sound signals arround the echofree onsets. The second plot is the average magnitude of estimated sound-to-echo ratios. The third plot shows the distribution of time differences between a microphone pair for every echo-free onset. The last (fourth) plot is the standard divergence of the time differences (the dashed-line is the real time difference for reference.) From the results, we can see that the standard divergence of time differences is remarkably smaller near the onsets than in other portions. This demonstrated that the onsets detected by the EA model are indeed less.

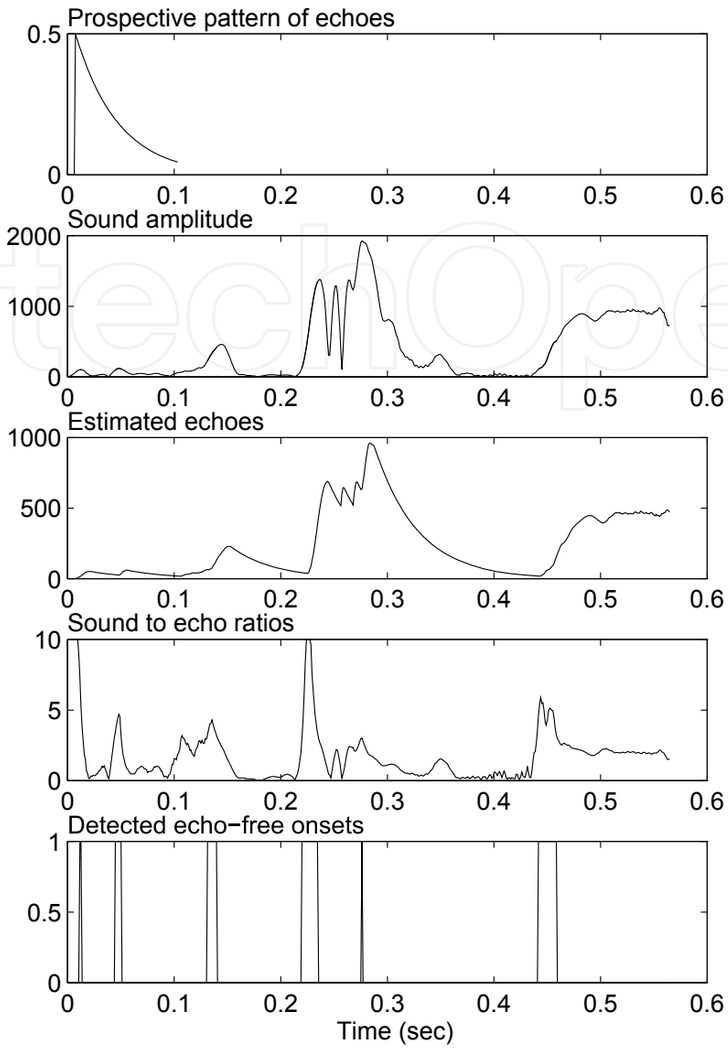


Fig. 3. An echo-free onset detection test by the model based algorithm. The first plot is the prospective pattern of echoes $h_e(t)$. The second plot shows the amplitude envelope of a test sound $s(f; t)$ (filtered speech signal, central frequency 255 Hz and bandwidth 30 Hz). The third and fourth plots show the estimated echoes $s_e(f; t)$ and the sound-to-echo ratios $s(f; t)=s_e(f; t)$. The last plot shows the positions of echo-free onsets as $s(f; t)=s_e(f; t) \geq 3$.

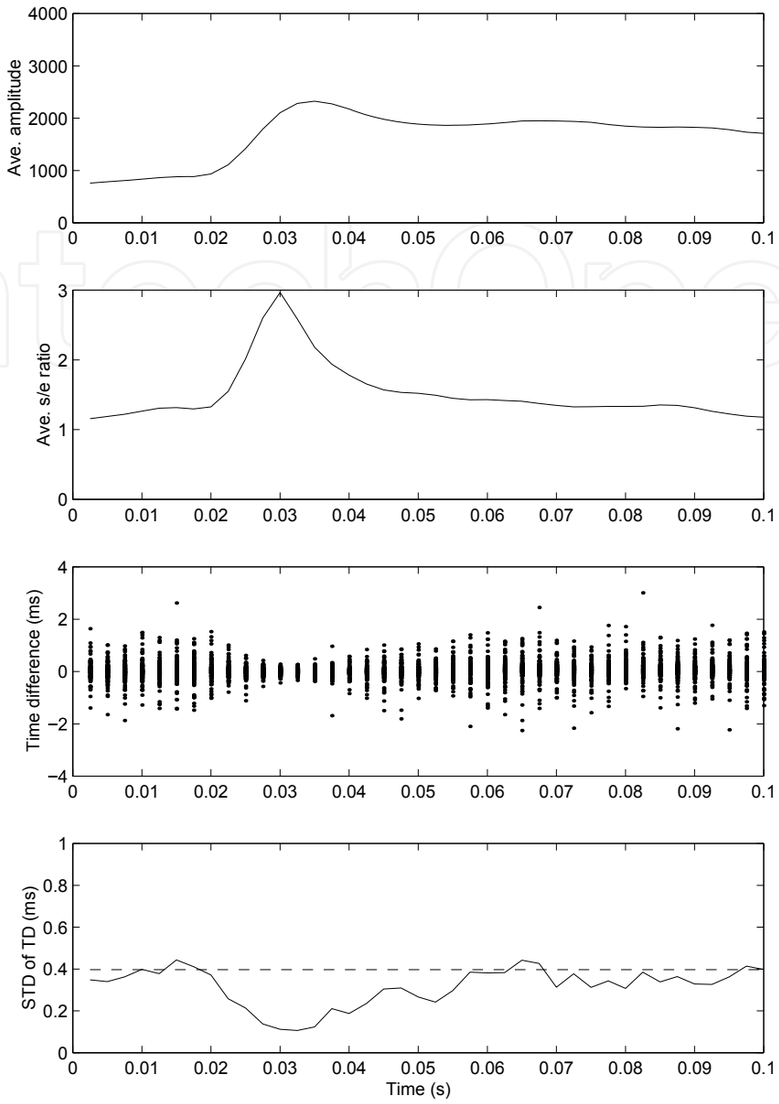


Fig. 4. The time variance of phase difference around onsets (speech source, ordi-nary room, 5 s length, 137 detected onsets).

4.3 Psychological Experiments and Their Explanations Increase of just-noticeable difference

Zurek Zurek (1980) reported that the just-noticeable difference (JND) of interaural delay and intensity difference, tested by headphones, increased temporarily following a preceding burst sound. The time period of the increase ranges from 800 / s to 5 ms after the preceding burst. Saberi and Perrott Saberi and Perrott (1990) reported similar results by a different method. They showed the perceptual thresholds of interaural time difference increased after the preceding click sound in a time range from 500 / s to 5 ms. The results of Saberi and Perrott' experiments are shown in Fig. 1.5 (reproduced from FIG.3 in Saberi and Perrott (1990)). In this figure, the 'x' marks show the increase of the perceptual threshold of interaural time difference. We can see that the increase pattern of the perceptual threshold of

the interaural time difference also has similar features (delay and decay) as we mentioned in the prospective pattern of echoes.

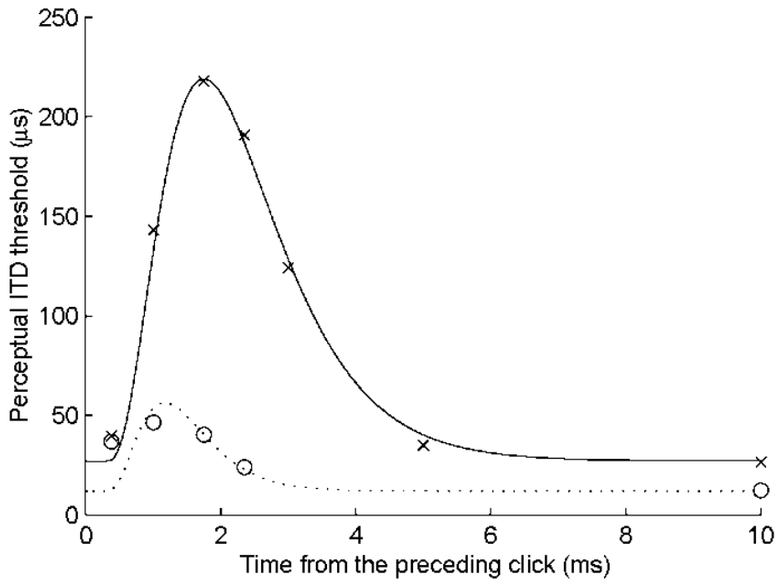


Fig. 5. The increase in the interaural time difference perceptual threshold after a preceding impulsive sound. Saberi and Perrott's experiment results are indicated by 'x' marks (experiment I) and 'o' marks (experiment II, same subjects after several extended practice sessions). The solid line is the calculated approximation of Saveri and Perrott's experiment I, and the dashed line for experiment II.

The perceptual threshold increase in the interaural time difference is quite like the post-synaptic or membrane potential of a neuron stimulated by an impulsive signal Kuffler et al. (1984). We can approximate the pattern by an exponential function $a\ h_i(t)\ + b$ (solid line in the figure), where

$$h_i(t) = \begin{cases} 0 & ,0 \leq t < \tau_o \\ k_i(e^{(t-\tau_o)/\tau_a} - 1)^3 e^{-(t-\tau_0)/\tau_b} & ,t \geq \tau_o \end{cases} \quad (7)$$

with $\tau_a = 1.50ms, \tau_b = 0.31ms, \tau_o = 0.3ms, k_i = 12.5$(a normalization value calculated from the peak position $t_p = -\tau_a ln(1 - 3\tau_b/)$) , $a = 192\mu s$ and $b = 27\mu s$.

This pattern is possibly a kind of impulsive aftereffect of an echo-estimation neural mechanism. However, there is a marked difference between the prospective pattern of echoes and the impulsive aftereffect. While the time extension of the prospective pattern of echoes is usually from several hundreds of milliseconds to more than 1 s depending on the reverberation time of the environment, the duration of the impulsive aftereffect is merely 5 ms.

To bridge the gap between the prospective pattern of echoes and the impulsive aftereffect derived from psychological tests, let us consider an impulsive sound and a series of its reflections as shown in Fig. 1.2. We note that the impulsive sound and its reflections will

trigger a chain aftereffect so that the duration of the total aftereffect will be extended. It is because, in fact, the received sound itself includes echoes and thus can be represented as the convolution of the original sound and the impulse response of the environment. Therefore, the decay feature of the impulse response is already included in the received sound, i.e., the decay parameter in the EA model is not critical for a normal reverberant environment.

Time variance of echo estimation

Recent studies have shown that the precedence effect is active and dynamic. Blauert and Col Blauert and Col (1989, 1992), Clifton and Freyman Clifton (1987); Clifton and Freyman (1989) and Duda Duda (1996) reported that the precedence effect can break down and become re-established in an anechoic chamber or hearing test by headphones.

The echo threshold of the precedence effect changed when the subjects were presented with a series of "conditioning" clicks before tests Freyman et al. (1991) and also depends on what the subjects expect to hear Clifton et al. (1994).

The above results suggest that the precedence effect is time variant and adaptive to different environments and different sound sources. The time variance of the precedence effect is also shown in the results of Saberi and Perrott's experiments. As shown in Fig. 1.5 ('o' marks), the temporal increase of perceptual thresholds of the interaural time difference almost disappeared after some dense practice sessions by headphones. The reduced increase of the perceptual threshold can be approximated by the same exponential function Eq (7) as mentioned above, by just modifying two parameters ($\tau_b = 0.22ms, b = 12\mu s$, dashed line in Fig. 5). The decrease of the offset *b* may be due to the perceptual improvement by repeating the tests, and the change of time constant τ_b means a change in the condition of echo-estimation mechanism. However, the dynamics of the precedence effect are complex and more computational studies are required in the future to solve the problem completely.

Criteria for a transient onset

As mentioned above, the inhibition control I_c is assumed to depend on the relative strength of observed sounds and estimated echoes. We use a sigmoid like function for inhibition generation as shown in Fig. 6, where $\rho(t) = s(t)/s_e(t)$**.** The inhibition control will take value 1 when the sound-to-echo ratio exceeds than an open (or free-pass) threshold ρ_o**,** and in the opposite the inhibition control will take the value 0, when *rho*(*t*) is less than a closed (or inhibition) threshold ρ_c**.**

Since the echo-inhibition is depending on the relative strength of the sound-to-echo ratio, the model can explain the results of the experiments by Haas

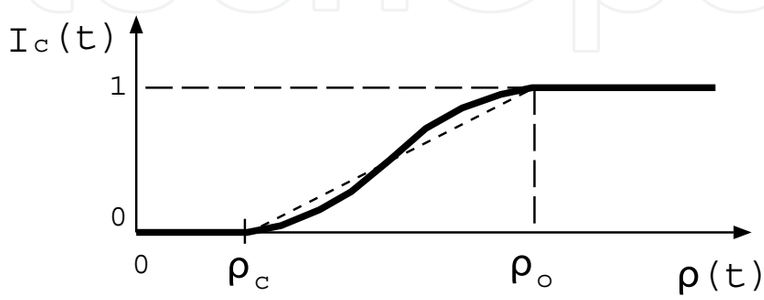


Fig. 6. The inhibition control as function of sound-to-echo ratio $\rho(t)$.

(1951) and Snow (1953) shown that the precedence effect can be canceled by a higher intensity succeeding sound.
By the EA model, we can check the availability of an onset for sound localization. We refer to a sound portion where $\rho \geq \rho_o$as an "echo-free" onset and that where $\rho \geq \rho_c$ as an available onset.
Assume that the sound level is a constant value s_0 before t_0, and after t_0 the sound increased a level Δs for time period τ_0 (the dacay parameter in the EA model). The sound-to-echo ratio at $t = t_0 + \tau_0$ can be estimated as

$$\rho(t)\,|_{t=t_0+\tau_0} = \frac{s(t)}{\underset{t'\leq t_0+\tau_0}{Max}\ \{s(t-t')s_e(t')\}} = \frac{s(t_0)+\Delta s}{ks(t_0)} = \frac{1}{k}(1+\frac{\Delta s}{s_0}). \tag{8}$$

It shows that the sound-to-echo ratio depends on the relative increase of the sound level. This explains why a transient onset is needed for the precedence effect Rakerd and Hartmann (1985), and is also consistent with the results obtained by Franssen's experiments Franssen (1959) that a constant-level continuous sound contributions little to sound localization.

Explanation for the Haas effect
To predict the inhibition related to the results of Haas's tests, we need to analyze the "continuity" of the stimuli. Here, the continuity is evaluated by the average length of the components in individual frequencies. Since the precedence effect is considered as independent for different frequencies, a continuous articulation with a change of frequency is also considered as a discontinuity. From this meaning, we consider that the average length of the phonemes is useful for roughly estimating the continuity for speech sound.
Statistical investigations of the Japanese phonemes Kimura (1988) have shown that the five vowels in Japanese speech have an average duration of about 100 ms and the consonants have an average duration of 14 ms to over 100 ms (e.g. 't': 15 ms, 'r': 14 ms, 'sh': 146 ms and 's': 129 ms). Thus, we can roughly conclude that the continuity of Japanese speech is more than 15 ms. Regardless of the difference for different languages, we assume the speech stimuli used in Haas's tests have the same continuity.
With this knowledge in mind, we can easily understand that the onsets in the delayed speech will overlap with the continuous portions of non-delayed speech. The contribution of delayed speech to localization will thus be inhibited, as illustrated in Fig. 7(a). The overlap probability is high when the time delay is small. When the time delay becomes longer than about 14 ms (**t_2** in Fig. 7(a)), the onsets of the delayed speech begin to neet with the next quiet portions of non-delayed speech. The inhibition will thus turn back to low according to the average ratio of silence-to-continuity portion of the speech. We can therefore expect that the inhibition of the delayed speech will have the curve as shown in Fig. 7(b). For noise stimuli, the overlap probability will decrease more quickly and remain at a relatively high level because the rate of the quiet portion is less and the average length of the continuous portion becomes shorter than speech stimuli. This tendency is more conspicuous for high frequency noise.

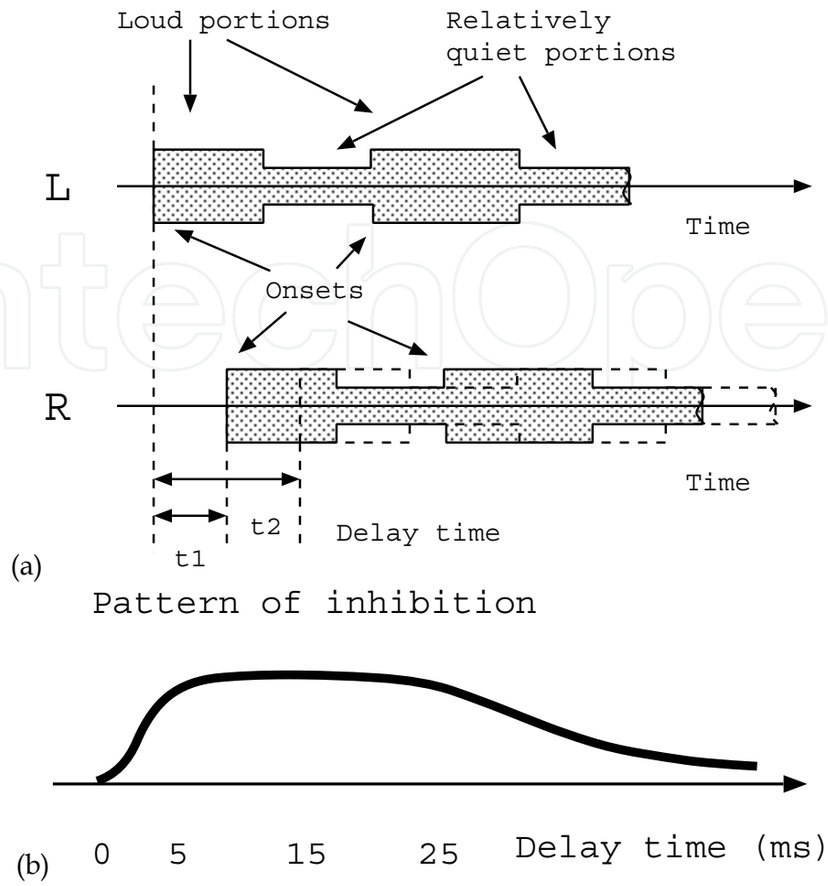


Fig. 7. (a) The \continuity" of a speech sound; (b) the pattern of inhibition.

5. Localization of multiple sound sources

We can construct a robot auditory system with four microphones arranged in the surface of a spherical head. Three of the microphones at the level of the center and one on the top of the spherical head. By the four microphones, we can solve the front-back and up-down ambiguity. To reduce the number of time difference candidates estimated from the phase differences, restrictions between the time differences of different microphone pairs were introduced. To cope with the echoes and reverberation caused by walls, ceiling, floor and other objects the EA model of the precedence effect are to calculate the weights for the ATDs. The time difference histograms are then mapped to a 2-D azimuth-elevation histogram where we can identify sound sources by its peaks Li et al. (2007).

5.1 ATD as the function of azimuth and elevation

As shown in Fig. 8, M_1, M_2, M_3 and M_4 are the four microphones. We define the positions by the spherical polar coordinates with the center of the robot head as the origin, (r (cm), θ (deg), ρ (deg)). Than, the locations of the four microphones can be described by M_1(15, 0, 90), M_2(15, 180, 0), M_3(15, 60, 0)and M_4(15, 300, 0), where the radius of the robot head is equal to 15 cm.

Consider a plane contains sound source *S*, robot head center *O* and microphone M_i as shown in Fig 9., the distance between the sound source and each microphone can be obtained as

$$d_i = \begin{cases} |\vec{OS} - \vec{OM_i}| & (SP \geq SM_i) \\ SP + arc(PM_i) & (SP < SM_i) \end{cases} \tag{9}$$

Denoting the azimuth of sound source as θ and the elevation as ρ, and D as the distance from the robot head center to the sound source, we have

$$\vec{OS} = D(cos\rho cos\theta, sin\theta, cos\rho sin\rho) \tag{10}$$

$$\vec{OM_i} = R(cos\rho M_i cos\theta M_i, sin\rho M_i, cos\rho M_i sin\theta M_i) \tag{11}$$

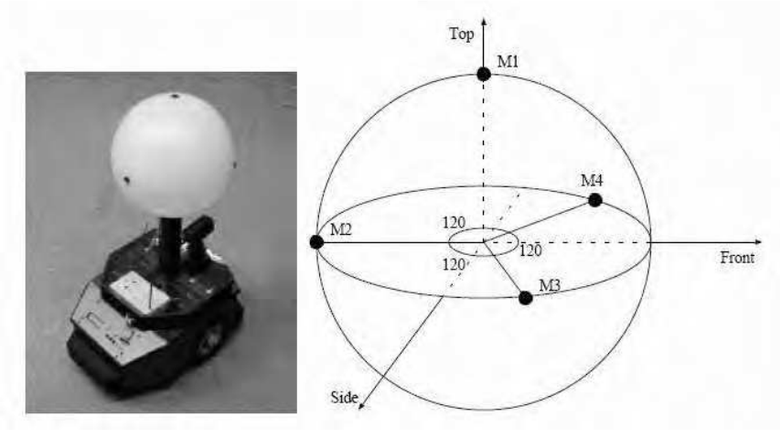


Fig. 8. The robot system and a spherical head with a four-microphone set.

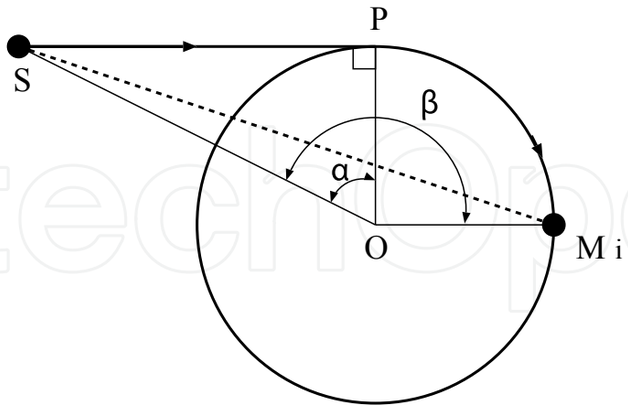


Fig. 9. The distance of sound path from the source to each microphone.

where ***R*** is radius of the robot head, and ρ_{M_i} are the azimuth and elevation of microphone *i*. The ATD between two microphones mainly depends on the azimuth and elevation of sound source. When D is very larger than $R(D \gg R)$, the influence caused by the difference of D can be ignored. Thus, the arrival time differences can be denoted as the function of θ and ρ

as

$$\Delta t_{ij}(\theta\rho) = \frac{d_i(\theta,\rho) - d_j(\theta,\rho)}{v} \tag{12}$$

where v is the sound velocity and $i, j = 1, 2, 3, 4\ i \neq j$.
Fig 10 shows the calculation results of ATDs between microphone M_1 and M_2.

5.2 Restrictions between ATD candidates

Because of the phase wrapping of high frequency components, the ATDs can not be determined uniquely. We obtain the ATD candidates from the phase

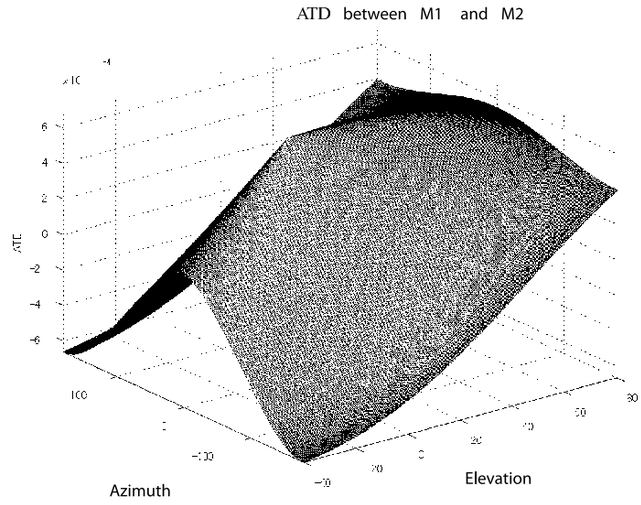


Fig. 10. ATDs between SM_1 and SM_2
differences of each frequency band for each microphone pair,

$$\Delta z_{ij}^{(k)} = \frac{\Delta\psi_{ij} + 2\pi n}{2\pi f_k} \tag{13}$$

where $\Delta\psi_{ij}$ is the measured phase difference between a microphone pair i and j in the fcth frequency band with center frequency f_k, and n is an arbitrary integer limited by $|\Delta t_{ij}^{(k)}| \leq \Delta T = \frac{2}{3}\pi R/C$.
To reduce the number of ATD candidates, we consider the sum ATDs which forms a circulation start from one microphone (ex, microphone 1), through the other three microphones and finally finish at the same microphone. The sum of ATDs in such a circulation should be zero. We have six different such paths, and if we ignore the difference in direction we can have the following three restrictions.

Zero summation restrictions:

$$\begin{aligned}\Delta t_{12}^{(k)} + \Delta t_{23}^{(k)} + \Delta t_{34}^{(k)} + \Delta t_{41}^{(k)} = 0 \\ \Delta t_{13}^{(k)} + \Delta t_{34}^{(k)} + \Delta t_{42}^{(k)} + \Delta t_{21}^{(k)} = 0 \\ \Delta t_{14}^{(k)} + \Delta t_{42}^{(k)} + \Delta t_{23}^{(k)} + \Delta t_{31}^{(k)} = 0\end{aligned} \tag{14}$$

Similarly, we can have the restrictions for the absolute sum between the circulation ATDs

$$\begin{aligned}T_{min} \leq |\Delta t_{12}^{(k)}| + |\Delta t_{23}^{(k)}| + |\Delta t_{34}^{(k)}| + |\Delta t_{41}^{(k)}| \leq T_{max} \\ T_{min} \leq |\Delta t_{13}^{(k)}| + |\Delta t_{34}^{(k)}| + |\Delta t_{42}^{(k)}| + |\Delta t_{21}^{(k)}| \leq T_{max} \\ T_{min} \leq |\Delta t_{14}^{(k)}| + |\Delta t_{42}^{(k)}| + |\Delta t_{23}^{(k)}| + |\Delta t_{31}^{(k)}| \leq T_{max}\end{aligned} \tag{15}$$

where $T_{min} = 2R/C$ for sound from top direction with 90 degrees elevation, and $T_{max} = (\frac{4}{3}\pi + \frac{3}{2})R/C$ for sound from the back direction with 180 degrees azimuth and 0 degree elevation.
For sound sources in the horizontal plane, the restrictions are reduced to the following forms,

$$\Delta t_{12}^{(k)} + \Delta t_{23}^{(k)} + \Delta t_{31}^{(k)} = 0 \tag{16}$$

and

$$T'_{min} \leq |\Delta t_{12}^{(k)}| + |\Delta t_{23}^{(k)}| + |\Delta t_{31}^{(k)}| \leq T'_{max}$$

where $T'_{min} = (2 + \frac{2}{3}\pi)R$ and $T'_{max} = \frac{4}{3}\pi R$.

5.3 Integration and mapping of ATD histograms

By summarizing all of the ATD candidates of different frequency bands for microphone pair *i* and *j* with the weights calculated by the EA model, we can form an histogram $h_{ij}(\Delta t_{ij})$. Since $\Delta t_{ij}$is a function of θ and ρ, we can denote the histogram as $h_{ij}(\Delta t_{ij}(\theta, \rho))$. This histogram, after a smoothing operation, can be a continuous function for Δt_{ij} and thus for θ and ρ. We now consider a new function $p_{ij}(\theta, \rho)$ of θ and ρ, and let

$$p_{ij}(\theta, \rho) = h_{ij}(\Delta t_{ij}(\theta, \rho)) \tag{17}$$

Here, $p_{ij}(\theta, \rho)$ is a two dimensional histogram, differing with h_{ij} which is an one dimensional histogram of Δt_{ij}. We call this transformation as a 'mapping' from ATD histogram to azimuth-elevation histogram, a one-to-many mapping.
The azimuth-elevation histogram p_{ij} of a single microphone pair is not enough to determine the θ and ρ of sound sources. The peaks in the azimuth-elevation histogram provide a necessary condition only but not sufficient. The sufficient condition can be integrated by taking the geometric average for different microphone pairs,

$$P(\theta, \rho) = \left[\prod_{i \neq j} p_{ij}(\theta, \rho)\right]^{1/6} \tag{18}$$

An horizontal version of this mapping is shown in Figure 11.
The final histogram $P(\theta, \rho)$, after normalization, can be consider as a possibility function for spatial distribution of sound source in all of the directions (θ, ρ).

5.4 Experiments and Results

The experiments were carried out in both an anechoic chamber and an ordinary room. The anechoic chamber has a size of 5.5 $\times$ 5.5 $\times$ 5.5 **m³,** and the ordinary room is of size 4 $\times$ 6 $\times$ 3 **m³,** without any acoustic treatment for its floor, ceiling and walls. The robot head was set on a height, 1 m above the floor.

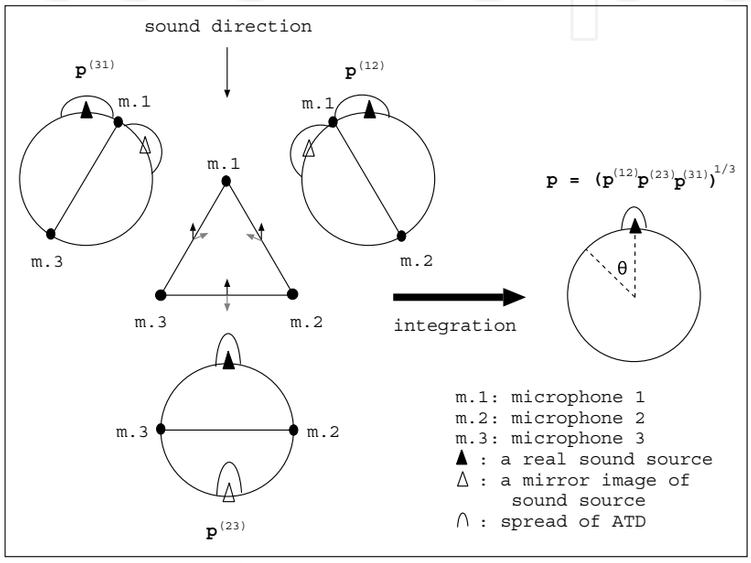


Fig. 11. Mapping and integration of ATDs to the azimuth histogram.

Localization of horizontal sound sources

Sound source (1) was a recording of a radio weather forecast presented by a male announcer (speaker 1 at 0 degrees of azimuth), and sound source (2) was a recording of a fast radio talk show with one male and one female host (speaker 2 at 38 degrees azimuth). The overlap portion of Sound source (1) and sound source (2) is about 16 seconds.
The resulting azimuth histograms are shown in Figure 12. The time segment for each azimuth histogram is 0.5 s.
The peaks appear clearly around 0 degree and 38 degrees over all time segments in the histograms of the anechoic chamber. The positions of major peaks are in the regions of [0,4] and [35,39] degrees, *i.e.,* the sound source 1 was localized in 2(±2) degrees and sound source 2 in 37(±2) degrees. The maximum absolute error is 4 degrees (regardless of the setup errors).
The histograms of the normal room, however, show more disorder comparing to the anechoic chamber. The scores are smaller and the size of the peaks is not consistent. We smoothed the histograms by a two-dimension gaussian function ($\sigma_t = 1$ second and $\sigma_\theta = 2$degrees) as shown in Figure 12.(c). In the smoothed histogram, the time resolution decreased to about 2 seconds, but the peak positions became more consistent.
The positions of major peaks are in the regions of [-2,2] and [33,37] degrees, *i.e.,* the sound

source 1 was localized in 0(±2) degrees and sound source 2 in 35(±2) degrees.

Localization of spatial sound sources
We used two loudspeakers to generate two different sound sources concurrently at different locations with a intensity difference.

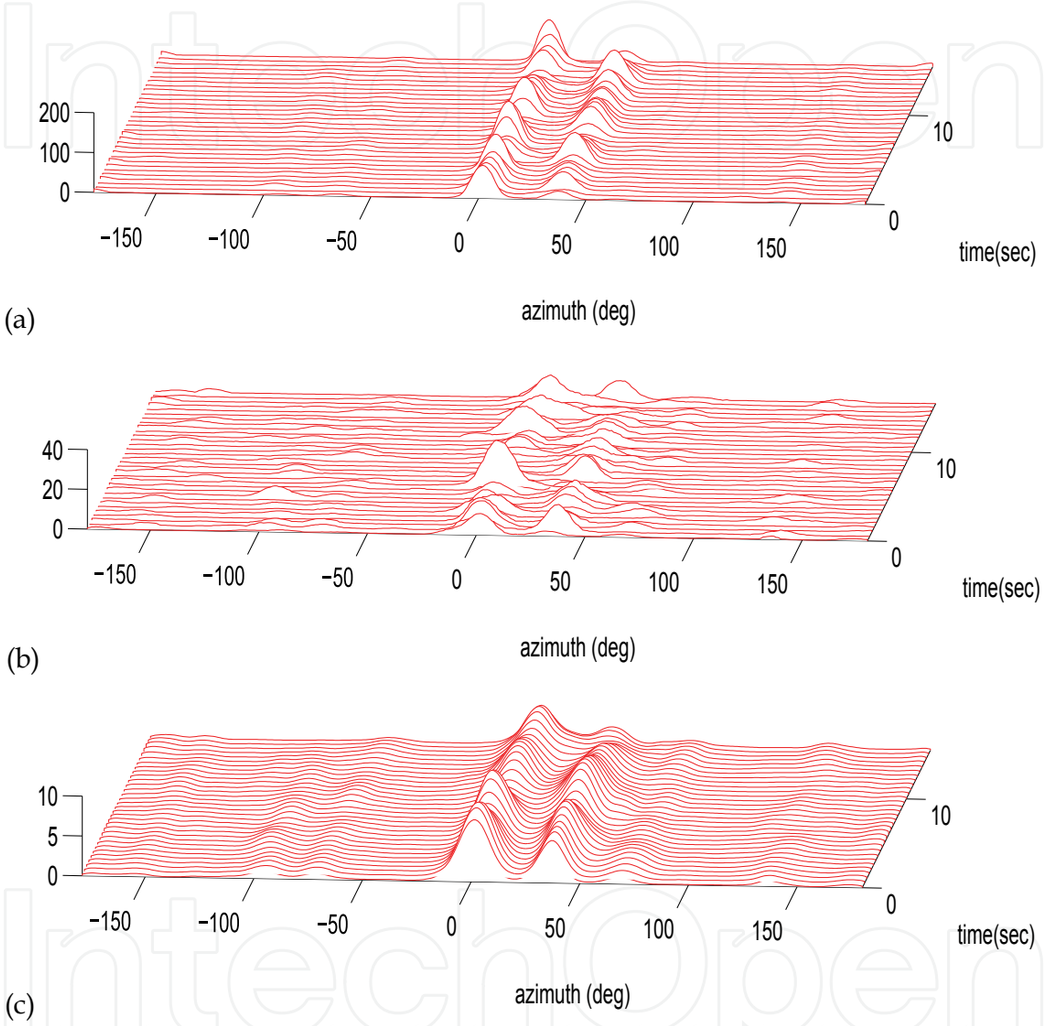


Fig. 12. Azimuth histograms obtained in the anechoic chamber (a) and the normal room (b), with time segments of 0.5 second and total length of 16 seconds. Figure (c) is the smoothed azimuth histograms in the normal room, by two-dimension gaussian function ($\sigma_t = 1$ second and $\sigma_\theta = 2$ degrees).

The signals received by the four microphones were divided into frames of 250 ms length. The azimuth-elevation histogram was calculated for each time frame. To reduce noise and increase accuracy, histograms of every 10 frames were summed and applied with a threshold.
Fig. 13 shows a resulting azimuth-elevation histogram for sound sources at S_1 (0°, 0°) and $S_2(-90°, 0°)$ in the anechoic chamber.

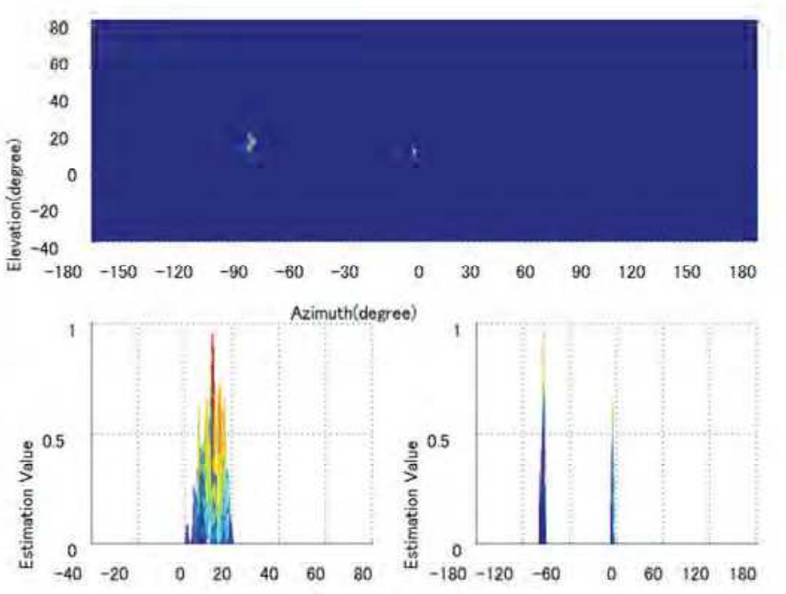


Fig. 13. A resulting average azimuth-elevation histogram for 10 time frames for sound source S_1(0°, 0°) and S_2(−90°, 0°) in an anechoic chamber.

Fig. 14 shows the resulting azimuth-elevation histogram for sound sources in the ordinary room. From the figure, we can identify the two sound sources by the peaks clearly.
The average localization errors by the peak positions of multiple sound sources at different locations in the ordinary room are shown in Table. 1.1.

Sound sources direction	S_1(0°, 0°) S_2(-90°, 0°)	S_1(0°, 15°) S_2(90°, 15°)	S_1(0°, 15°) S_2(90°,3 0°)
Average peaks direction errors	(7°, 8°) (4°, 12°)	(4°, 2°) (4°, 2°)	(2°, 6°) (6°, 8°)

Table 1. An average azimuth-elevation histogram for 10 time frames of two sound sources in an ordinary.

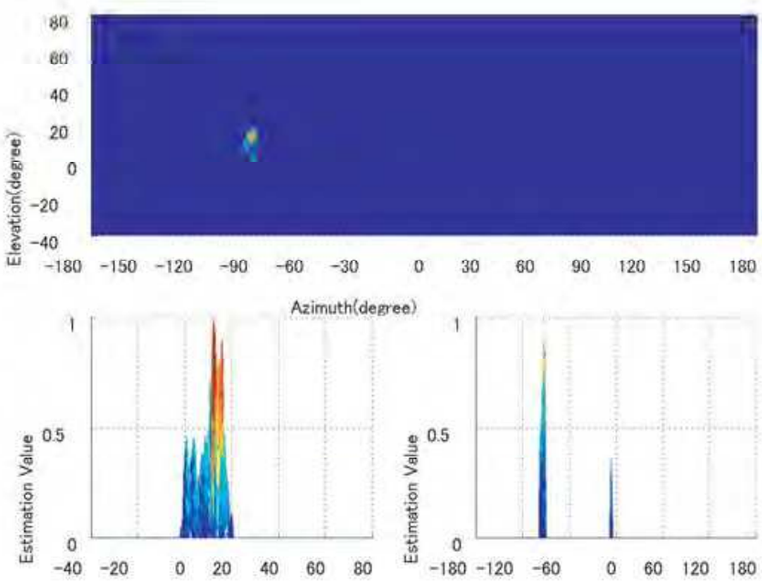


Fig. 14. A resulting average azimuth-elevation histogram for 10 time frames for sound source $S_1(0°, 0°)$ and $S_2(-90°, 0°)$ in an ordinary room.

6 Navigating a robot to a sound source

6.1 System

In this paper, we describe a mobile robot equipped with a simplified real time sound localization system which involves the function to cope with echoes and reverberations. By audition, a robot can find an object behind obstacles and even outside of a room by tracing back the sound path from the source Huang et al. (1999, 1997a). Experiments of robot navigation by the sound localization system have been conducted in various setups and conditions to demonstrate the effectiveness of the system.

As shown in Figure 15, the three omnidirectional microphones are placed on the circumference of a circle in a manner that the three microphones form a regular triangle on a horizontal plane, where the radius of the circle was variable. The mobile robot (RWI B-12) is wheel-based and equipped with a sonar system for obstacle detection and a CCD camera system reserved for visual processing.

To make it easy to implement the method in a mobile robot, the system uses only a single frequency band (central frequency 1kHz, band width 600 Hz). The frequency band and its width is chosen so that the maximum phase difference between two microphones will be less than n. Therefore the time differences can be uniquely determined from the zero-crossing points of wave forms.

The steering unit of the system has four modes,

• Localization mode: localizing sound sources

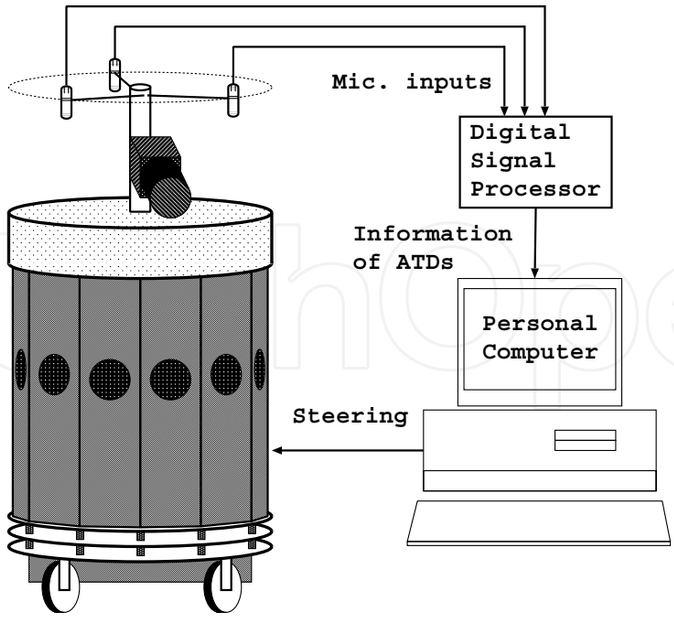


Fig. 15. A robot equipped with ears (microphones)

- Forward movement mode: going straightforward
- Avoidance mode: avoiding obstacles
- Stop mode: waiting for instructions.

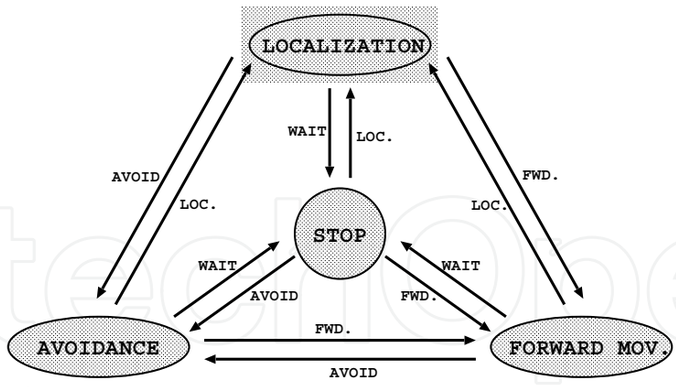


Fig. 16. Modes for robot steering

Transfers among the four modes are controlled by the commands from the personal computer: "localize sound", "avoid obstacle", "go forward" and "wait", as shown in Figure 16. The method of obstacle avoidance is simply implemented by turning the direction of the robot clockwise whenever an obstacle is detected in front of it by the sonar system.

6.2 Experiments

The experiments were conducted in an ordinary room, where background noise of up to 40 dB is generated by computer fans, an air conditioner and other environment noise coming in

through the windows. The sound sources used were a 1 kHz brief sinusoidal sound presented by a loudspeaker and hand-clapping sounds.

Obstacle Avoidance

In this experiment, an obstacle was located between the initial position of the robot and the sound source of goal (a speaker emitting intermittent 1kHz sounds) as shown in Figure 1.17. The obstacle was about 0.47m high, lower than the mount position of the microphone set, so that the direct sound path from the sound source to the robot was not obstructed. The result shows that the mobile robot could correctly localize the sound source and move toward the sound source position avoiding the obstacle in between. Here, the source direction was provided by the sound localization unit and the obstacle avoidance was achieved by the sonar system.

If the obstacle is big enough and higher than the position of microphones and sound source, there will be no direct path from sound source to microphones. In this case, the sound localization unit will localize the sound source by the minimum path of sound-to-microphone (see next section). In generally, it means the system will localize the sound source at the edge of the obstacle (the turning point of the minimum path). The edge position of the obstacle is helpful, since after the robot approached the edge the robot can localize the sound source again and then finally approach the source.

Approaching an Invisible Sound Source

In this experiment, the sound source was invisible to the robot and the direct path from the sound source to the robot was blocked. The robot was initially positioned outside a room, while the sound source was positioned inside (Figure 18). In this situation, instead of the direct sound, the sound from the minimum path plays an important role. The sound direction perceived by the robot will depend on the last turning position of the minimum path from the sound source to the robot.

The results of estimation in positions (X1 - X6) are shown in Table 2. From the results, the robot did tend to localize the sound at the last turning position of the minimum path, the edge of the door. The localization accuracy was lower than what was obtained with a direct path, especially when the robot was far from the sound source. This is because the narrow corridor can act as a large guiding tube, where reflections will be funneled in the direction of the corridor to form a mixed sound with expanded arrival directions. However, the localized sound direction always pointed to the same side of the corridor, and after some approaches and corrections the robot was able to find the turning position and finally localize the sound source inside the room.

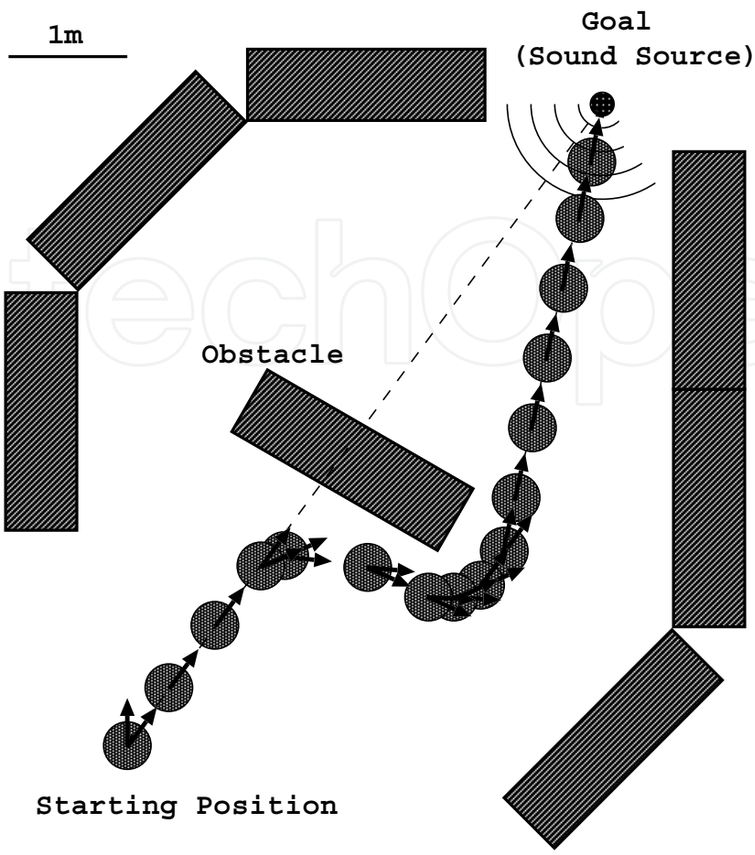


Fig. 17. Obstacle avoidance

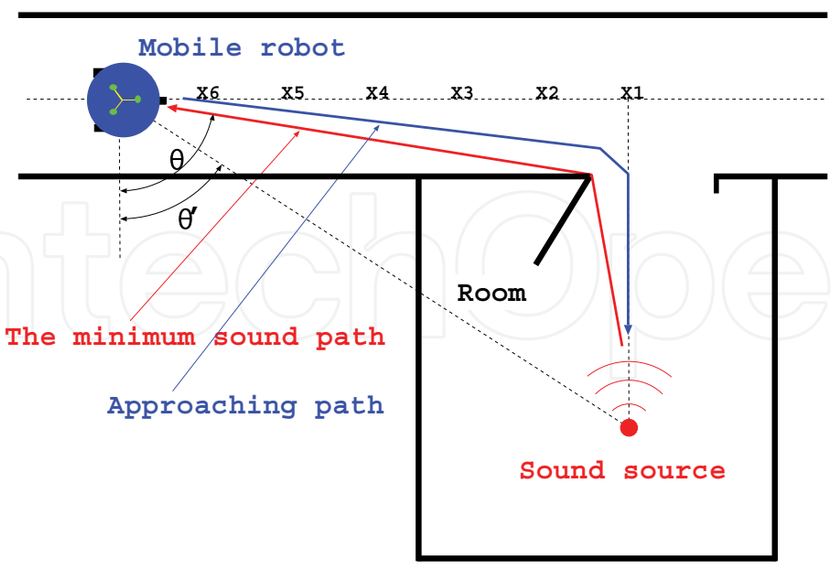


Fig. 18. Approaching a invisible sound source

measurement	X1	X2	X3	X4	X5	X6
1	0	49	60	75	*	62
2	1	41	*	79	76	81
3	0	46	64	76	80	87
4	1	39	65	75	*	*
5	0	43	67	81	85	*
6	1	44	69	71	83	68
7	1	44	67	82	81	50
8	1	45	66	79	*	81
9	0	45	55	76	78	75
10	1	46	61	78	81	47
average	0.6	44.2	63.8	77.2	80.6	68.9
θ'	0	23.1	40.5	52.0	59.7	64.9
θ	0	44.5	66.7	74.7	78.8	81.0

*: failure of localization

Table 2. Azimuth estimation of a invisible sound source

7. References

Aarabi, P. and Zaky, S. (2000). Integrated vision and sound localization. In *Proc. 3rd Int. Conf. Information Fusion,* Paris.

Bekey, G. A. (2005). *Audonomous Robots: From Biological Inspiration to Implementation and Control (Intellegent Robotics and Autonomous Agents).* MIT Pr.

Blauert, J. (1997). *Spatial hearing: the psychophysics of human sound localization.* The MIT Press, London, revised edition.

Blauert, J. and Col, J. P. (1989). Etude de quelques aspects temporels de l'audetion spatiale (A study of certain temporal effects of spatial hearing). Note-laboratoire LMA 118, CRNS, Marseille.

Blauert, J. and Col, J. P. (1992). A study of temporal effect in spatial hearing. In Cazals, Demany, and Horner, editors, *Auditory Psychology and Perception,* pages 531-538. Pergamon Press.

Bodden, M. (1993). Modeling human sound-source localization and the cocktail-party-effect. *Acta Acustica,* 1:43-55.

Bregman, A. S. (1990). *Auditory Scene Analysis: The Perceptual Organization ofSound.* The MIT Press, London.

Cherry, E. C. (1953). Some experiments on the recognition of speech with one and with two ears. *J. Acoust. Soc. Am.,* 25:975-979.

Clifton, R. K. (1987). Breakdown of echo suppression in the precedence effect. *J. Acoust. Soc. Am.,* 82:1834-1835.

Clifton, R. K. and Freyman, R. L. (1989). Effect of click rate and delay and breakdown of the precedence effect. *Percept. Psychophys.,* 46:139-145.

Clifton, R. K., Freyman, R. L., Litovsky, R. Y., and McCall, D. (1994). Listern-ers' expectations about echoes can raise or lower echo threshold. *J. Acoust. Soc. Am.,* 95:1525-1533.

Clifton, R. K., Morrongiello, B. A., and Dowd, J. M. (1984). A developmental look at an auditory illusion: The precedence effect. *Dev. Psychobiol.,* 17:519536.

Cooke, M. (1993). *Modeling Auditory Processing and Organisation.* Cambridge University Press, Cambridge.

Duda, R. O. (1996). Auditory localization demonstrations. *Acustica,* 82:346355.

Ellis, D. P. W. (1994). A computer model of psychoacoustic grouping rules. In *Proc. 12th Int. Conf. on Pattern Recognition.*

Franssen, N. V. (1959). Eigenschaften des naturlichen Richtungshorens und ihre Anwendung auf die Stereophonie (The properties of natural directional hearing and their application to stereophony). In *Proc. 3rd Int. Congr. Acoustics,* volume 1, pages 787-790.

Freyman, R. L., Clifton, R. K., and Litovsky, R. Y. (1991). Dynamic processes in the precedence effect. *J. Acoust. Soc. Am.,* 90:874-884.

Gardner, M. B. (1968). Historical background of the Haas and/or precedence effect. *J. Acoust. Soc. Am.,* 43:1243-1248.

Gelfand, S. A. (1998). *Hearing – An Introduction to Psychological and Physiological Acoustics – (Third Edition, Revised, and Expanded,).* Marcel Dekker, Inc., New York.

Haas, H. (1951). Uber den eingluss eines einfachechos auf die horsamkeit von sprache. *Acustica,* 1:49-58. English translation in: "The influence of a single echo on the audibility of speech", *J. Audio Eng. Soc.,* Vol. 20, pp. 146-159, (1972).

Hafter, E. R., Buell, T. N., and Richards, V. M. (1988). Onset-coding in later-alization: Its form, site, and function. In Edelman, G. M., Gall, W. E., and Cowan, W. M., editors, *Auditory Function: Neurobiological Bases of Hearing,* pages 647-676. Wiley and Sons, New York.

Harris, G. G., Flanagan, J. L., and Watson, B. J. (1963). Binaural interaction of a click with a click pair. *J. Acoust. Soc. Am.,* 35:672-678.

Hartmann, W. M. and Rakerd, B. (1989). Localization of sound in room IV: The Franssen effect. *J. Acoust. Soc. Am.,* 86:1366-1373.

Heffner, R. S. and Heffner, H. E. (1992). Evolution of sound localization in mammals. In D. B. Webster, R. R. F. and Popper, A. N., editors, *The evolutionary biology ofhearing,* pages 691-715. Springer-Verlag, New York.

Hornstein, J., Lopes, M., Statos-Victor, J., and Lacerda, F. (2006). Sound localization for humanoid robot - building audio-motor maps based on the hrtf. In *Proc. IEEE/RSJ Int. Conf. Intelligent Robots and Systems.* IEEE/RSJ.

Huang, J., Ohnishi, N., and Sugie, N. (1995). A biomimetic system for localization and separation of multiple sound sources. *IEEE Trans. Instrum. and Meas.,* 44(3):733-738.

Huang, J., Ohnishi, N., and Sugie, N. (1997a). Building ears for robots: Sound localization and separation. *Artificial Life and Robotics,* 1(4):157-163.

Huang, J., Ohnishi, N., and Sugie, N. (1997b). Sound localization in reverberant environment based on the model of the precedence effect. *IEEE Trans. Instrum. and Meas.,* 46(4):842-846.

Huang, J., Supaongprapa, T., Terakura, I., Wang, F., Ohnishi, N., and Sugie, N. (1999). A model based sound localization system and its application to robot navigation. *Robotics and Autonomous Systems,* 27(4):199-209.

Huang, J., Zhao, C., Ohtake, Y., Li, H., and Zhao, Q. (2006). Robot position identification using specially designed landmarks. In *Proc. Instrum. Meas. Technol. Conf.,* Sorrento. IEEE.

Johnson, D. H. and Dudgeon, D. E. (1993). *Array Signal Processing: Concepts and Techniques.* PTR Prentice-Hall, NJ.

Kimura, S. (1988). Investigation on variations of acoustic segment duration. No. 1-2-14, Proc. Spring Meet. Acoust. Soc. Jpn. (Japanese).

Knudsen, E. I. (1981). The hearing of the barn owl. *Sci. Am.,* 245(6):82-91.

Konishi, M. (1986). Centrally synthesized maps of sensory space. *Trends in Neuroscience,* 9(4):163-168.

Kuffler, S. W., Nicholls, J. G., and Martin, A. R. (1984). *From Neuron to Brain: A Cellular Approach to the Function of the Nervous System.* Sinauer Associates Inc., Sunderland, MA, second edition.

Lehn, K. H. (1997). Modeling binaural auditory scene analysis by a temporal fuzzy cluster analysis approach. In *Proc. IEEE Workshop WASPAA'97.*

Li, H., Yoshiara, T., Zhao, Q., Watanabe, T., and Huang, J. (2007). A spatial sound localization system for mobile robots. In *Proc. Instrum. Meas. Technol. Conf.,* Warsaw. IEEE.

Lindemann, W. (1986a). Extension of a binaural cross-correlation model by contralateral inhibition. i. simulation of lateralization for stationary signals. *J. Acoust. Soc. Am.,* 80:1608-1622.

Lindemann, W. (1986b). Extension of a binaural cross-correlation model by contralateral inhibition. ii. the low of the first wavefront. *J. Acoust. Soc. Am.,* 80:1623-1630.

Litovsky, R. Y. and Macmillan, N. A. (1994). Sound localization precision under conditions of the precedence effect: Effects of azimuth and standard stimuli. *J. Acoust. Soc. Am.,* 96:752-758.

Martin, K. D. (1997). Echo suppression in a computational model of the precedence effect. In *Proc. IEEE Workshop WASPAA'97.*

McFadden, D. (1973). Precedence effects and auditory cells with long characteristic delays. *J. Acoust. Soc. Am.,* 54:528-530.

Medioni, G. and Kang, S. B., editors (2005). *Emerging Topics in Computer Vision.* PTR Prentice-Hall.

Nishizawa, Y., Yagi, Y., and Yachida, M. (1993). Generation of environmental map and estimation of free space for a mobile robot using omnidirectional image sensor COPIS. *J. Robotics Soc. Japan,* 11(6):868-874. (Japanese).

Oertel, D. and Wickesberg, R. E. (1996). A mechanism for the suppression of echoes in the cocklear nuclei. In Ainsworth, W. A., editor, *Advances in Speech, Hearing and Language Processing,* volume 3 (Part B), pages 293-321. JAI Press Inc., London.

Okuno, H. G., Nakadai, K., Hidai, K., Mizoguchi, H., and Kitano, H. (2001). Human-robot interaction through real-time auditory and visual multiple-talker tracking. In *Proc. IEEE/RSJ Int. Conf. Intelligent Robots and Systems,*

pages 1402-1409. IEEE/RSJ.

Parkin, P. H. and Humphreys, H. R. (1958). *Acoustics, Noise and Buildings.* Faber & Faber, London.

Rakerd, B. and Hartmann, W. M. (1985). Localization of sound in rooms, II: The effects of a single reflecting surface. *J. Acoust. Soc. Am.,* 78:524-533.

Rakerd, B. and Hartmann, W. M. (1992). Precedence effect with and without interaural differences — Sound localization in three planes. *J. Acoust. Soc.*

Am., 92:2296(A).

Saberi, K. and Perrott, D. R. (1990). Lateralization thresholds obtained under conditions in which the precedence effect is assumed to operate. *J. Acoust. Soc. Am.,* 87:1732-1737.

Shen, J. X. (1993). A peripheral mechanism for auditory directionality in the bushcricket gampsocleis gratiosa: Acoustic tracheal system. *J. Acoust. Soc. Am.,* 94:1211-1217.

Snow, W. B. (1953). Basic principles of stereophonic sound. *J. Soc. Motion Pict. Telev. Eng.,* 61:567-587.

Takahashi, T. and Konishi, M. (1986). Selectivity for interaural time difference in the owl's midbrain. *J. Neuroscience,* 6(12):3413-3422.

Thurlow, W. R., Marten, A. E., and Bhatt, B. J. (1965). Localization aftereffects with pulse-tone and pulse-pulse stimuli. *J. Acoust. Soc. Am.* , 37:837-842.

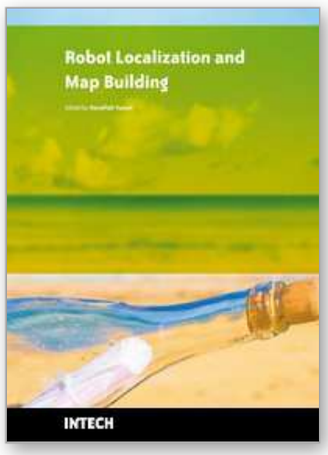
Valin, J. M., Michaud, F., and Rouat, J. (2007). Robust localization and tracking of simultaneous moving sound sources using beamforming and particle filtering. *Robotics and Autonomous Systems,* 55(3):216-228.

von Bekesy, G. (1960). *Experiments in Hearing (pp.288-301,609-634).* McGraw Hill, New York. Translated from: Zur Theorie des Horens: Uber das Richtung-shoren bei einer Zeitdifferenz oder Lautstarkenungleichheit der beiderseitigen Schalleinwirkungen, *Physik. Z.,* **31,** pp.824-835, 857-868, (1930).

Wallach, H., Newman, E. B., and Rosenzweig, M. R. (1949). The precedence effect in sound localization. *J. Psychol. Am.,* 62(3):315-336.

Zurek, P. M. (1980). The precedence effect and its possible role in the avoidance of interaural ambiguities. *J. Acoust. Soc. Am.* , 67:952-964.

Zurek, P. M. (1987). The precedence effect. In Yost, W. A. and Gourevitch, G., editors, *Directional hearing,* pages 85-105. Springer-Verlag, New York.



Robot Localization and Map Building

Edited by Hanafiah Yussof

ISBN 978-953-7619-83-1
Hard cover, 578 pages
Publisher InTech
Published online 01, March, 2010
Published in print edition March, 2010

Localization and mapping are the essence of successful navigation in mobile platform technology. Localization is a fundamental task in order to achieve high levels of autonomy in robot navigation and robustness in vehicle positioning. Robot localization and mapping is commonly related to cartography, combining science, technique and computation to build a trajectory map that reality can be modelled in ways that communicate spatial information effectively. This book describes comprehensive introduction, theories and applications related to localization, positioning and map building in mobile robot and autonomous vehicle platforms. It is organized in twenty seven chapters. Each chapter is rich with different degrees of details and approaches, supported by unique and actual resources that make it possible for readers to explore and learn the up to date knowledge in robot navigation technology. Understanding the theory and principles described in this book requires a multidisciplinary background of robotics, nonlinear system, sensor network, network engineering, computer science, physics, etc.

InTech Europe
University Campus STeP Ri
Slavka Krautzeka 83/A
51000 Rijeka, Croatia
Phone: +385 (51) 770 447
Fax: +385 (51) 686 166
www.intechopen.com

InTech China
Unit 405, Office Block, Hotel Equatorial Shanghai
No.65, Yan An Road (West), Shanghai, 200040, China
中国上海市延安西路65号上海国际贵都大饭店办公楼405单元
Phone: +86-21-62489820
Fax: +86-21-62489821

IntechOpen