Acoust. Sci. & Tech. **24**, 5 (2003)

INVITED REVIEW

Spatial unmasking and attention related to the cocktail party problem

Masanao Ebata*

Kumamoto National College of Technology,
2659–2 Suya, Nishigoshi, Kumamoto, 861–1102 Japan

(*Received 4 February 2003, Accepted for publication 28 May 2003*)

Abstract: Although there are several factors causing "cocktail party effect" after more than half a century of research, the major one is considered to be the spatial separation of the target signal and the interferer. This paper will overview developments of the improvement of performance resulting from the directional separation of the target signal from interferers when listening in a field or through headphones. The basic assumption concerning the cocktail party effect is that there are one or more interfering sound sources in addition to the target signal source. In this situation it is important to remember the selective attention effect, which attenuates the interfering sound by concentrating the attention on a specific signal. Pitch of sound is the simplest cue for selective attention; however, spatial information can also be one. The latter half of this review discusses the effect of spatial filtering and an attention filter on the frequency domain.

Keywords: Spatial unmasking, Cocktail party effect, Interaural difference, Selective attention, Attention filter

PACS number: 43.66.Pn, 43.66.Qp, 43.66.Rq, 43.66.Dc, 43.71.Gv [DOI: 10.1250/ast.24.208]

1. INTRODUCTION

It has already been 50 years since Cherry [1] published his paper on the cocktail party problem. Since then several review papers have been published. Yost [2] reviewed both selective attention and masking level difference in relation to binaural hearing and the cocktail party problem. In 2000 Bronkhorst [3] reviewed the effect of the number of interference noises on speech intelligibility under binaural hearing conditions. Despite the wealth of studies conducted over the last five decades, the five factors for understanding the cocktail party problem listed by Cherry are still thought to be suitable because of the adequateness of their description.

(1) Information on the direction of the sound source.
(2) Visual information such as gestures and lip movement.
(3) Tonal quality of the voice, speed of speech, average pitch and information concerning the speech sound itself such as the sex of the speaker.
(4) Information on accents.
(5) Information on transition-probabilities.

In observations of the cocktail party phenomenon all five factors are thought to be utilized in an attempt to increase the intelligibility of speech under severe Signal to Noise Ratio (SNR) conditions. In the special case concerning the directional information of sources Cherry made the following experiment: the subject heard separate speeches in each ear and was asked to repeat the sentence heard in one of the ears in order to have the subject focus on the speech presented to that ear. He then observed the degree of processing, perception and cognition of the speech given to the other ear. Since Cherry's paper, many papers have mentioned the cocktail party problem, particularly in relation to the improvement of hearing performance by the directional separation of sound sources. Some papers have explored source separation under actual acoustic environments, but most have dealt with the improvement of performance by the use of artificial binaural information: for example, Interaural Time Difference (ITD), Interaural Intensity Difference (IID), or Interaural Phase Difference (IPD) using headphones. These experiments including Cherry's do not, however, give us a direct key to the cocktail party problem in a real-world environment. In the 1940's Licklider [4] and Hirsh [5] showed that the threshold of pure tone is decreased and speech intelligibility is increased by the interaurally out-of-phase signal against a diotic masker. Although this work maintains its attraction even now, the experiments are also far from an actual acoustic environment. Following their papers, many studies were done on the Masking Level

*e-mail: ebata@knct.ac.jp

Difference (MLD). Those papers conclude that from 15 dB to 18 dB of threshold decrease is observed when utilizing an out-of-phase signal or noise. Although it is not possible to reproduce this situation in the real world, the improvement of performance under binaural hearing conditions is one of the most interesting phenomena in understanding the functioning of the auditory system and in determining its limits for segregating a signal from interference. Due to its interesting nature, this theme was the focus of research for a long period in the 1950's and 60's. On the other hand, in order to investigate the cocktail party problem, research testing the effect of spatial separation continues in experiments under the free field. There are, however, alternatives. In the first alternative, the recorded signals obtained by an artificial ear or dummy head in the free field are used for binaural hearing in order to investigate the effect of source separation on performance. Also the Head Related Transfer Function (HRTF) obtained a priori is convolved with the signal and noise in order to identify the stimuli. In the present paper, release from masking due to spatial separation of sources in a free field is reviewed first, and then masking release using an artificial head, and MLD due to interaural differences, are overviewed in section 2.

Under the cocktail party problem we may assume that the information from the signal sound source is segregated from amongst others. That means selective attention is always utilized in this situation. For example, when the subject is asked to focus on a pure tone, there is an attentional filter on the frequency axis, and there is a spatial filter when the subject is asked to focus on a specific direction. In section 3, the methods of controlling attention and the attentional filter obtained by those methods are reviewed.

When considering the cocktail party problem in a more realistic situation we need to think about the mechanism for following a continuous dialog. Of course the obvious cue to following a dialog is the direction of the speaker. But as Cherry's third point suggests, there are other cues such as the quality of the voice, average pitch, etc. Research related to these topics has advanced in the last decade, but due to limitations of space these subjects are omitted from this review.

2. RELEASE FROM MASKING DUE TO SEPARATION OF THE SOUND SOURCES

2.1. Release from Masking due to Spatial Separation of the Sources in a Free Field

Cherry's research into the cocktail party problem showed that we can separate two signals and attend to one of them. Since then the direction of the source has long been known as a cue to signal separation. Along with the development of the stereophonic LP record, the two-track recording technique became popular, and research on the improvement of perceptual performance based on source separation became active. However, research into directional separation in the sound field did not progress as remarkably as that using headphones under dichotic hearing conditions. Research into the free field was resurrected in the 1990's after a period of inactivity. On the contrary, during the inactive period, studies using headphones were particularly active, most of them aimed at investigating effective physical parameters as cues of binaural release from masking.

In this subsection the effects of directional separation in a sound field will be oveviewed from the aspect of parameter extraction. In the early stages of signal separation research based on directional information, Spieth *et al.* [6] showed that the percentage of correct answers to the messages increases more than 20% at 10 to 20° of source separation. Although they made experiments for the range of 90 to 180°, they did not vary the direction of the source systematically. In an anechoic room, Motor [7] measured detection thresholds of 500 Hz pure tone within a masking noise when the direction of the source was varied systematically. The thresholds of pure tone located in front of the subjects were measured when the source of the broadband masking noise was located at every 15°. According to their results, the threshold of detection decreases by 7 dB when the separation of directions is 90°. This decrease is the masking level difference due to the directional separation of sources. Ebata *et al.* [8] performed similar experiments. They measured the threshold of 1,000 Hz pure tone which was presented by a loudspeaker located at one of 0, 10, 20, 30, 45, 60, 90 and 135° while the masker was presented from 0°, or in front of the subject. MLD due to the directional separation reaches its saturated level, 10 dB, when the directional difference is 60°. Ebata *et al.* [8] also carried out the experiments on release from masking due to the directional separation using monosyllabic words. The results showed that the maximum improvement of intelligibility due to directional separation amounts to 10% when SNR is −5 dB and it corresponds to 3.5 dB in SNR. If the directivity characteristics of the ears are taken into account, the improvement decreases to 2.7 dB, which is significantly less than the MLD obtained by pure tone (10 dB). The amount of release from masking decreases further if SNR is positive. Levitt and Rabinaer [9] obtained the similar result that a binaural release from masking by phase inversion gives a 13 dB gain for the detection, but only about 6 dB for intelligibility. In addition they showed that the binaural release from masking by interaural time difference reaches 13 dB for the detection task but only 3 dB for intelligibility. The reason for these differences between detection threshold and intelligibility has not been clearly explained. Ebata *et al.* [10] performed experiments in which masked loudness was measured using loudness

balance as a function of Sensation Level (SL). Using the same experimental setup as the signal detection task, not only the threshold but also the loudness of pure tone, whose frequency was 500 Hz, 1,000 Hz, or 2,000 Hz with an SL between 5 dB and 35 dB, were measured. The results showed that the increase of masked loudness becomes smaller as SL increases, and it becomes almost 0 dB at 25 dBSL. This means that the above mentioned difference is not due to the characteristics of speech, but that masking release is at its maximum at 0 dBSL and decreases when the SL increases. On this issue, there have been similar studies which show the maximum improvement of hearing ability around threshold and a decrease in the amount of improvement as an increase of SL. Licklider [4] explored the relation between SNR and release from masking. For example, he showed that when SNR $= -8$ dB, interaural phase inversion of the signal gives about 20% improvement of intelligibility but becomes 2 to 3% when SNR $= 0$ dB. Later, Carhart *et al.* [11,12] performed similar experiments and showed that a larger improvement can be obtained due to the phase inversion of a signal when SNR is low. Also Townsend and Goldstein [13] measured the loudness increase due to phase inversion at various SLs using 250 Hz and 500 Hz pure tone. Their results show that the effect reaches 11 dB around the threshold level, but 8 dB at 5 dBSL, 4 dB at 10 dBSL, and only 1 dB at 20 dBSL. Henning [14] also showed that the increase in correct response due to phase inversion of a pure tone is greater at low SNR than at high SNR in frequency discrimination.

Research on the improvement of performance due to directional separation in a sound field continued. Plomp [15] measured the masking release of intelligibility due to the separation of the source direction when the source of the target signal was located in front and the noise source was located at every 45°. He used connected discourses as the representative of everyday situations and measured the threshold in which the discourse was just intelligible. The results showed that the gain due to directional separation is about 5 dB. Plomp also discussed the effect of reverberation in order to further understand the cocktail party effect in daily life conditions. According to his results, the gain due to directional separation with a reverberation time of 1.4 s decreases to 2 dB and that with 2.0 s reverberation almost disappears. Koehnke and Besing [16] discussed the effect of reverberation on the release from masking due to directional separation using monosyllabic words. They used an earphone to present stimuli made by the convolution technique using a source-to-eardrum transfer function obtained from an artificial head, *a priori*. They obtained the results that an improvement in SNR in an anechoic chamber is 13 dB for 90° directional separation but only 4 dB in simulated sound field environment with a reverberation time of 0.25 s to 0.4 s. On the other hand, Freyman *et al.* [17,18] investigated the effect of single reflection on masking release due to directional separation. The subjects were asked to figure out the key word included in a nonsense sentence. The interference sound was either noise with a speech-like spectrum shape or nonsense sentences uttered by different speakers. Although reflections were added both to target speech and interference, sound localization was determined by the direct sound due to the precedence effect because the delay between direct and reflected sounds was only 4 ms. The results obtained using speech-shaped noise interference showed that SNR improvement is 8 dB without reflection when the directional separation is 60°, but it reduces to 1 dB with reflection. On the other hand, when nonsense sentences were used as the interferer, the advantage due to directional separation reaches 14 dB, and it remains 6 to 10 dB even if a reflection is added. The reason for the difference of advantages due to the type of interference was thought to be as follows: The interference is due mainly to energetic masking when noise is used as the interference sound, but it is also due to informational masking when speech interference is used, so that the effect of the directional separation becomes greater.

The increase of performance due to spatial separation remained several dB for the monosyllables and phonetically balanced words (PB words). Since the 1980's, however, the Speech Reception Threshold (SRT) has been used as a measure of speech recognition, and an improvement in SNR of more than 10 dB has been reported. The SRT measure is obtained subjectively as the level of speech corresponding to 50% intelligibility of the sentence. This method was introduced by Hawkins and Stevens [19] as the threshold of intelligibility and has been popular after being reevaluated by Plomp [20].

Plomp and Mimpen [21] measured the release from masking by directional separation in the SRT using a speech signal. They measured the SRT for speech located in front of the subject while interference noise whose spectrum shape was the same as the long-term average of speech was presented every 22.5° on the semicircle. Their results showed about 10 dB masking release when the noise source was located at the side of the subject, 90 to 125° from the front, compared with the SRT when both the speech and noise sources were located in front of the subject. Duquensnoy [22] also obtained the effect of directional separation using a measure of SRT. His results show that the masking release is 9.6 dB for the stationary noise when the direction of arrival is separated by 90°, but decreases to 6.7 dB when speech is used as an interference. The reason why the improvement in the speech interferer is not so great is thought to be that the overall intelligibility under interfering speech is originally high due to silent durations between utterances even if the directions of

signal and interference are the same. Bronkhorst and Plomp [23] measured the masking release due to directional separation using SRT. In their experiments, speech spectrum noise was used. They obtained about a 6 dB improvement in SRT when the directions of the signal and the noise were separated by 90°.

All of the above mentioned researches discuss the masking release due to directional separation in the horizontal plane. On the other hand, Saberi *et al.* [24] discussed one in the vertical median plane. Because there are no interaural differences for sound presented in the vertical median plane, it is thought to be less advantageous in release from masking due to spatial separation. Therefore, the amount of research on the vertical separation of signal and masker is much less than that on the horizontal. Saberi *et al.* measured the MLD of directional separation for a short train of 50 μs clicks within broadband noise. Although maximum MLD reached 15 dB in the horizontal plane, only 8 dB was obtained at 60° in the vertical median plane. One reason for the advantage due to directional separation in vertical median plane might be the ability of the subjects to utilize direction-dependent peaks and troughs in the power spectrum resulting from the interactions of the stimulus from the convolutions of the pinna. Gilkey and Good [25] also measured the MLD in the vertical median plane using a filtered pulse train as the signal. A signal train of 20 μs pulses was filtered by a low, middle, or high frequency band pass filter, and the band width of the masker was set wide enough so that it sufficiently covered signal band width. When the masker was located in front of the subject, a threshold decrease of about 11 dB was observed for low and high pass filtered signals in the horizontal plane. In the middle frequency range it reached about 5 dB. In the vertical median plane, however, MLD reached 10 dB for a high frequency range signal, but there was almost no MLD for low frequency range and only 2 dB for middle frequency range. The results suggest that the MLD in the vertical median plane comes from direction-dependent peaks and troughs in the spectral cue as Saberi *et al.* mentioned.

2.2. Spatial Release from Masking Using an Artificial Head

In order to evaluate the effect of directional separation in an actual sound field, stimuli recorded by a dummy head microphone system in a field can be reproduced by headphones. For example, by comparing the performance obtained by binaural and monaural reproductions, the cues for release from masking in an actual sound field are examined. This area of research started before 1950 and is very active even now.

At first, the effect of directional separation on performance for a signal obtained by a human head in an actual sound field is greater than for a signal obtained by an artificial ear. The difference is a 3 to 5 dB decrease for the threshold and a 20 to 30% increase for speech intelligibility. Koch [26] measured the performance when the speech source was located in front of the subject and the noise source was at the back, both in free field and in binaural reproduction conditions. Results showed that the threshold of intelligibility obtained in a free field condition is 3 to 4 dB lower than that in a binaural one. Hirsh [27] also measured the effect of directional separation under the conditions of a free field and headphone reproduction of stimuli recorded by a spherical dummy head. Locations of target and interference sound sources were varied: 180° by front-back or left-right, and 90° by front-left/right or back-left/right. The results show that the threshold of intelligibility obtained in the free field is lower, by as much as 5.5 dB in the 90° case. In recent research, an accurate artificial head was introduced, and a comparison was made between a human head and the dummy head on the effect of source separation. Some reports [18,28] show that the performance by a human head is 20 to 30% better than that of the dummy head. One report [27] shows no difference between them. There was no complete agreement between the results. This difference in performance between free field listening and binaural listening using a signal obtained by a dummy head seems to be due to the inaccurate modelling of the dummy head on the subject's own head. There are many reports using an accurate Head Related Transfer Function (HRTF) for a subject. Details are reported in the last part of this section; first the research on the differences in performance between binaural and monaural hearing conditions using a dummy head are overviewed. In the beginning, a spherical shaped dummy head equipped with two microphones was used. Generally there is a 4 to 7 dB advantage for binaural over monaural hearing conditions. However, as shown by Bronkhorst and Plomp [30], the threshold for a monaural hearing condition can be lower than that of the binaural condition because the effect of the head shadow increases the SNR of the ear on the opposite side to the interference noise source. Hirsh [27] obtained a 4.5 dB advantage for the binaural hearing condition over the monaural one when the directional difference between the target and interference sources was 90° using an artificial head. Dirks and Wilson [31] compared the hearing conditions with 90° separation of source direction and obtained about a 7 dB improvement for the binaural condition. They also reported that the advantage due to 90° separation in binaural hearing was 6 to 9 dB.

Recently artificial heads are being used in many studies. In order to make a dummy head similar to a human head, the KEMAR mannequin was developed by Burkhard and Sachs [32] in 1975. Since then, it has been

used in many research papers [18,28–30,33,34]. These papers reported a 9 dB decrease in threshold and a 30 to 40% improvement of intelligibility due to spatial separation. Bronkhost and Plomp [30] showed that SRT is improved by 9.4 dB due to a 90° spatial separation of speech and speech-like noise, which has the same long-term average spectrum as the sentences. Under the monaural hearing conditions by means of the better SNR ear, the threshold increased 2.5 dB. However, this threshold is still about 7 dB lower than one obtained by binaural hearing without spatial separation, i.e. when the noise and masker are located in the same direction. Because they thought that the lower threshold obtained in monaural hearing condition was due to SNR improvement resulting from the head shadow effect, they measured the performance for stimuli which had only either IID or ITD. For the stimuli with IID alone and ITD alone, the thresholds decrease 6.5 dB and 4.7 dB, respectively, and neither of these can explain the spatial unmasking of 9.4 dB. Yost *et al.* [28] also compared the binaural and monaural hearing conditions using the KEMAR dummy head, and showed that the performance on the identification of 26 letters and 9 numbers is 30% higher for the binaural hearing condition than for the monaural one. Hawley *et al.* [29] measured the increases in intelligibility of key words under spatial separation conditions using the KEMAR dummy head. They showed that the error rate deceased 50% due to spatial separation and that the difference in the error rate between binaural hearing and monaural hearing by the better ear reached 30%. Freyman *et al.* [18] showed that correct responses under binaural hearing conditions for a key word identification task are 35% higher than under monaural hearing conditions.

Recent developments in information processing technologies let us perform experiments using stimuli made virtually by the convolution of signal and HRTF measured *a priori*. Based on this technique, Peissing and Kollmeier [35] measured SRT by means of simulated experiments under the condition where the interference moves around the subject while the target is fixed in front. When a continuous, speech-spectrum-shaped noise is introduced as an interferer, about 10 dB SNR improvement is obtained when the direction difference between the target and interferer is 105°. Furthermore, when two interferers are fixed at 105° and 225° and a third interferer is added, SNR improvement remains only 2 dB regardless of the third interferer's location. Drullman and Bronkhorst [36] also used speech, both words and sentences, as target and interferer, and compared the intelligibility obtained under (1) a simulated condition based on HRTF, (2) a dichotic hearing condition where the different combinations of target and interferers were fed into each ear, and (3) the monaural hearing condition. For the single interferer case, the performance obtained in the monaural condition was 40 to 50% lower than those in other conditions. However, in the case of two interferers, the performance degraded for all conditions. The performance obtained in the binaural condition was 20 to 30% lower than that in the simulated condition based on HRTF, while in the monaural condition was 20 to 30% lower than in the binaural condition. When the number of interferers increased to three, the worst performance, almost 0% intelligibility, was obtained for the monaural hearing condition, and the worst performance for all of three conditions was reached in the case of four interferers. Note that they reported no differences in performance between a subject's own HRTF and another's. This result can be understood because the stimuli were low-pass filtered at 4k Hz. Shinn-Cunningham *et al.* [37] obtained the spherical-head HRTFs for all the positions from which sources were to be simulated and evaluated the effect of directional separation assessing the SNR of each frequency band. According to their results, the SNR was improved considerably in the wide frequency range.

All the studies mentioned above discuss only the directional separation of the target and the interferer. They do not discuss the effect of the distance between the listener and the sound sources. As mentioned by Brungart and Rabinowitz [38] there is almost no effect of distance on the HRTF when the distance from the sound source is more than 1 m, but the characteristics of the HRTF change for distances less than 1 m. In conditions when the cocktail party effect occurs, the distance of sources can often be within 1 m. Shinn-Cunninghum *et al.* [37] investigated the effect of the separation in distance on performance while the sound sources were located at 15 cm and 1m from the subject. Results demonstrated that the advantage due to 90° separation from noise was only 5 dB when sources were far from the head, but it was about 20 dB when the target source was near. Brungart and Simpson [39] obtained a 4 to 5 dB improvement in the SNR by the separation in distance.

There were huge variations in both the type and number of interferers in the above-mentioned investigations. However, in general the tendencies were (1) a steady noise shows a larger masking effect than speech, and release from masking due to the separation of sound sources is 3 to 6 dB greater in a noise interfering condition than in a speech one; (2) meaningful speech may interfere much more than noise because of the informational masking effect. Furthermore, Brungart [40] showed in the monaural hearing condition that speech is more disturbing than random noise, that the speech of someone of the same sex is more disturbing than that of someone of the opposite sex, and that the same talker's speech (the subject's own speech) is the most disturbing. To investigate the effects of energetic masking and informational masking separately,

instead of a simple noise, random noise with a long-term averaged spectrum of words or sentences was used by many researchers including Duquensnoy [22], Bronkhorst and Plomp [30,33], Koehnke and Besing [16], Ericson and McKinly [34], and Peissing and Kollmeier [35]. Random noise modulated by a speech envelope was also used by Bronkhorst and Plomp [33] to simulate the temporal pattern of speech. In addition, in order to examine the effect of informational masking, many studies are measuring the masking effects for both actual speech and noise; for example, Plomp [15], Ericson and McKinley [34], Peissing and Kollmeier [35], and Freyman *et al.* [17]. On the other hand, Yost *et al.* [28], Pessing and Kollmeir [35], Hawley *et al.* [29], Drullman and Bronkhorst [36] and Freyman *et al.* [18] used speech both as target and interferer.

Speech intelligibility using monosylabic speech and PB words was frequently used as a measure. In this case, however, a large number of presentations of stimuli is required to obtain accuracy. To shorten the experimental time, the SRT has been widely used by Plomp and Mimpen [21], Duquesnov [22], Bronkhorst and Plomp [23,30,33], Pessing and Kollmeir [35], and Freyman and Helfer [17]. Using this measure, the time requirement for an experiment is reduced by means of method of adjustment or adaptive method, but the load for subjects becomes large.

2.3. Speech Intelligibility Difference due to ITD and IPD

The MLD has been well known since the 1960's as the threshold shift around 15 dB for the detection of pure tone whose frequency is 500 Hz or less when the pure tone signal is out-of-phase under a dichotic condition. This effect can be understood to be due to the blurring of the intracranial image over the whole head of the subject resulting from phase inversion, while the image of in-phase noise can be located at the center of the subject's head. It is understood that these image differences lead to partial improvement of the SNR in the head even if this phenomenon is completely different from spatial separation. This effect of phase inversion is measured not only for the threshold of pure tone but also for speech intelligibility.

Many studies on the MLD on speech intelligibility utilize a speech signal whose phase is inverted in its lower frequency range, because the MLD on pure tone is observed only at a frequency lower than 1,000 Hz, especially less than 500 Hz. Schubert and Schultz [41] divided the speech signal into three frequency ranges as low (200 to 1,660 Hz), middle (880 to 2,200 Hz) and high (1,600 to 6,100 Hz), and measured the effect of interaural polarity reversal on the improvement of intelligibility. The results show a 25% improvement of intelligibility in the low frequency range, and 10% in the middle frequency range, but no improvement in the high frequency range, when the interferer is broadband noise at SNR = 0 dB. On the other hand, when speech is introduced as the interferer, the maximum MLD of phase inversion reaches 9 dB when the speech stimuli by the same talker are used for the target and the interferer, but a very small MLD on speech intelligibility is obtained when the speech stimuli uttered by different talkers are used. Levitt and Rabiner [9] also measured MLD for speech whose phase in the lower or higher frequency bands was inverted. The intelligibility increases when the phase of frequency band, including components lower than 500 Hz, is inverted. Note that this improvement in MLD is the same amount observed by full range phase inversion. Mosko and House [42] showed the effect of phase inversion on the detection of single vowels /*a*/, /i/, and /u/ and obtained 12 dB of MLD as the average of these three vowels. Flanagan and Watson [43] measured MLD by phase inversion of the pulse train instead of speech. MLD changes according to the frequency of the pulse train and reaches the maximum of 10 to 15 dB when the repetition is 100 to 150 pps. The release from masking is found to relate primarily to the fundamental component, so that the elimination of a fundamental component substantially reduces the MLD.

Besides research on intelligibility improvement due to phase inversion in a dichotic listening condition, there are many reports on the effect of intelligibility by controlling the interaural time difference. Obviously phase inversion cannot explain the phenomenon of the cocktail party effect, but the interaural time difference is understood to be one of the major physical cues for directional separation in the sound field, and its effects have been researched over several decades. According to the results, the improvement of intelligibility is only 3 to 4 dB by controlling the interaural time difference, while it is 5 to 9 dB by phase inversion, and the improvement obtained by phase inversion is always larger than that obtained by the interaural time difference. Schubert [44] compared the intelligibility improvement obtained by interaural delay and phase inversion. He obtained a 3 dB improvement due to 470 μs of ITD, less than the 5 dB resulting from phase inversion. Similar results are obtained by the treatment of the interferer instead of the target speech. Schubert and Schultz [41] used speech as an interferer and measured the intelligibility improvement for 0.5 ms ITD. They obtained a maximum of 4 dB improvement for various maskers under these conditions, while a 9 dB improvement was obtained with phase inversion. Levitt and Rabiner [9] compared the effect of phase inversion and interaural delay. They found a 6 dB improvement of speech intelligibility due to phase inversion, but only about 3 dB improvement due to 0.5 to 10 ms of ITD.

Carhart *et al.* [12] researched the effect of phase

inversion and interaural delay on speech intelligibility under various conditions. They examined the effect of phase inversion on speech intelligibility under continuous noise, and obtained a 4 to 5 dB improvement. They also obtained about a 3 dB improvement by stimuli which had a 0.8 ms interaural delay. The amount of improvement was decreased under modulated noise. Although the amount of release from masking due to interaural delay is usually not larger than that obtained by phase inversion as mentioned above, it reaches almost the same amount when the target and the interferer have 0.8 ms delays but the location of the sound image of both target and noise is perceived to be on the opposite side. Furthermore, Carhart *et al.* [45] compared the amount of release from masking due to phase inversion and interaural delay using modulated noise and connected speech as the interferers. Discrimination of monosyllabic speech improves up to 5 dB due to phase inversion and only 2.5 dB due to interaural delay. On the other hand, the change in the threshold level for spondee words reaches 4.9 dB for phase inverted stimuli but only 1.5 dB for interaurally delayed stimuli. Carhart *et al.* [46] reported on masked threshold under 37 binaural listening conditions using four kinds of interferers, i.e. two kinds of modulated noise and two kinds of spoken sentences uttered by male speakers. According to their results, the largest MLD (4.4 to 6.8 dB) is obtained under the antiphasic condition, and the MLD under the delay condition is always 1 dB less, in average, than that under the antiphasic condition. They also reported that MLD under opposed time-delay conditions (0.8ms delay given to all portions of the masker complex but not all masker signals delayed to the same ear) was reduced another 0.6 dB from the antiphasic condition, on average, ranging from 2.6 to 6.4 dB. Carhart *et al.* [46] showed that (1) the masking level decreases by 1 dB due to an increase of the duty cycle of modulated noise from 50 to 75%, (2) there is a 3.2 dB excess of masking by means of mixing speech with interfering noise, and (3) the excess masking reaches 6.6 dB when two kinds of speech are included in the interfering noise. Recently, Darwin and Hukin [48] obtained an improvement of performance in the identification of key words in a sentence separating the direction of the signal speech from the interfering speech due to ITD. They showed a 10% improvement in identification due to 91 μs of interaurally different ITD in the signal and the interferer. They concluded that the subjects can easily concentrate on the signal and suppress the interferer because the images of the signal and the interfering speech are separated during utterance of the key words.

3. ATTENTIONAL FILTER IN FREQUENCY DOMAIN

When experiencing the cocktail party effect, it is necessary to pay attention to a single speech. Research on the physiological mechanisms of this selective attention is still in the preliminary stage. The studies done by Puel *et al.* [49], LePage [50], Teder and Nattanen [51], Scharf *et al.* [52], and Scharf *et al.* [53], are examples. In this section, the psychophysical research on selective attention related to the cocktail party effect is reviewed.

Over the years there have been many researches into the strategy for dividing attention and its modelling in the case of uncertain frequency stimuli. It was in the late 1960's that a very important method for researching selective attention was introduced. Greenberg and Larkin [54] used the two alternative forced choice procedure, and measured the performance of the probe tone detection task under the condition where 1,100 Hz primary tones were presented to 70% of the whole trials with different frequency tones accounting for the remaining 30%. With this method the subjects can focus their attention on a specified frequency of a primary tone. They called this method the "probe-signal method." Before their work, Ebata *et al.* [8] examined the role of attention using a similar method. They observed selective attention due not only to frequency but also to the direction of the arriving signal. In their experiments, eight loudspeakers were located at points 45° apart around the subject. In 90% of the whole trials, 1,100 Hz pure tone lasting 1.2 s was presented by the loudspeaker located in front of the subject, and in the remaining 10% of the trials, tones either 220 Hz or 550 Hz were presented by one of seven loudspeakers other than the one in front. In quiet conditions, the threshold of a 220 Hz or 550 Hz tone increases by 6 to 7 dB compared with that in control condition for all directions. By contrast, Greenberg and Larkin [54] measured the threshold change of a test tone whose frequency was close to that of a primary tone. They used a 1,000 Hz or 1,100 Hz tone as the primary tone and measured the threshold for test tones whose frequency varied in the range of 700 Hz to 1,300 Hz. The results showed that detection performance falls nearly to the level of chance when the frequency of the test tone differs 100 Hz or more from that of the primary tone. As shown in their results, by focusing attention on the primary tone, the subject seems to construct a "filter" to decrease sensitivity to frequency ranges other than the attended one. Later, this filter is called as the "attention filter," and the frequency range of the filter is called the "attention band" [56,59].

Greenberg and Larkin obtained an attention band with nearly the same width as the critical band. After that, many other studies also obtained band widths close to the critical band as an attention band. Examples are Scharf *et al.* [55], Scharf [56], Schlauch and Hafter [57], and Hafter *et al.* [58]. The shape and bandwidth of the attention filter were compared with those of the auditory filter, and their

similarities were presented by Dai *et al.* [59] Schlauch and Hafter [57], Hafter *et al.* [58], Wright and Dai [60,61], and Botte [62]. Many other researchers obtained attention band widths significantly narrower than that of the critical band as shown by Penner [63], Macmillan and Schwartz [64], Yama and Robinson [65], Dai *et al.* [59], and Ison *et al.* [66]. However, results of recent research have shown that the band width of the attention filter depends on experimental conditions. In particular, it depends on the subject's strategy for focussing their attention on signal detection. Penner [63] measured the detection performance weighted by payoff on the probe tone in order to control the subject's intention and showed that the width of the attention band is broader when compared with that obtained by the typical probe signal method. Macmillan and Schwartz [64] also showed that the band width can be broadened by the use of two primary tones instead of one, and Dai *et al.* [59] and Wright and Dai [60,61] showed that the band width obtained by a gated masker is broader than that obtained by a continuous masker. Wright and Dai [60,61] also showed that band width introduced by a short duration tone becomes broader.

Although the characteristics of the attention filter obtained by the probe signal method show frequency characteristics similar to that of a band pass filter, the dynamic range of response is significantly smaller. In 1968, Ebata *et al.* [8] already obtained a threshold shift of 7 dB for a frequency range substantially lower than the frequency of the primary tone, and in 1991 Dai *et al.* [59] also obtained 7 dB of threshold shift. However, at frequencies close to the edge of the attention band, the shift of psychometric function is 4 dB at most as mentioned by Ebata *et al.* [67] and Dai *et al.* [59]. The dynamic range obtained through the conversion of d' based on the correct response rate into dB attenuation is in the range of 3 dB to 10 dB as shown by Yama and Robinson [65], Dai *et al.* [59], Schlauch and Hafter [57], Hafter *et al.* [58], Wright and Dai [60], and Botte [62], but the reliability of converted data is not high because the large threshold shift of a few dB was only obtained under conditions near the chance level.

The original probe signal method controlled the attention of the subject only by frequent presentation of the primary tone. Then a revised method was introduced using a cue tone inserted as the first presentation in each trial. Although the original probe signal method was able to use only one or two probe tones within a single session, the multiple probe technique, in which multiple probe tones are used in a session, makes it possible to focus attention more effectively on the primary tone. As the number of multiple cue tones increases, the rate of correct responses decreases and the shape of the attention filter flattens and broadens. The musical 5th and missing fundamental are also useful to construct the attention filter as shown by Hafter *et al.* [58], Hafter and Sbaeri [68], and Ebata *et al.* [69].

While the above-mentioned researches looked into the effect of the attention filter on frequency domain, there have been some studies about selective attention on directivity, in other words, a spatial filter focused in a specific direction. Ebata *et al.* measured the effect of spatial filtering based on the primary tone, which was presented in front of the subject at a ratio of 90% out of whole trials, and the probe tone, which was presented at 45, 90, or 135° from the front. According to the results of this research, the effect is quite small. Only a 1 dB increase of threshold against the probe tone, except from in front of subject, was obtained when the frequencies of target and primary tones were the same. However, if a masking noise was introduced from either in front or in back of the subject, the effect of filtering increased by 2 to 4 dB. This means that the attention on a specific tone configures a spatial filter even if the dynamic range of the filter is not large. As the effect on performance due to spatial selective attention was not so obvious, some studies were concerned with the effect on the Response Time (RT). The larger the angle between the direction of a cue tone and the direction of a target tone, the longer the latency of response. And this latency increases linearly until an angle of 90° (Rhodes [70] and Mondor and Zatorre [71]). In these experiments, the direction (right or left) was inquired when the cue tone was attended both under "valid" and "invalid conditions." The valid condition represents the condition that the direction of the target tone is the same as that of the cue tone, while the invalid condition represents the condition that the direction of the target tone is opposite of that of the cue tone. The RT for the valid condition is always shorter than that for the invalid one. The decreasing rate of RT depends on the ratio of the number of trials on the valid condition against the whole number of trials. At higher ratios such as 70 to 80%, the difference becomes significant and the RT of the valid condition is 10 to 20% smaller than that of the invalid condition (Spence and Driver [72], and Quinlan and Bailey [73]).

Although the change in RT and in performance due to spatial attention is very small, as mentioned above, the effect of spatial separation on performance becomes apparent when complex tasks are introduced into the experiment. According to Arbogast and Kidd [74], the performance of discriminating between an upward or downward frequency change on a series of short tones is degraded when the direction of frequency change on attended tones is not the same as that on the test tones. When six types of frequency change are tested, the effect of directional separation becomes larger, and the maximum effect corresponds to 10 to 20 dB in SNR. As Kidd *et al.* [75] suggest, this effect may not be due to the simple

masking release, but to the release from informational masking.

4. CONCLUSION

In this review paper, the author has tried to overview 50 years of history on the performance improvement due to spatial separation as the one of the major cues of the cocktail party effect. A variety of presentation methods has been used, such as hearing in a sound field, headphone reproduction using an artificial ear, and binaural hearing with a special signal processing either on the target signal or the interferer. The performance improvement due to spatial information has been reported in a range from several dB to more than 10 dB. Scattering of the performance improvement due to spatial separation is observed, and it increases especially during severe hearing conditions such as low SNR or more than one speech interferer. It has been confirmed by many studies that spatial separation of the speech signal from an interferer improves the performance of speech reception and contributes to the cocktail party effect as a primary factor. Table 1 shows the advantage due to spatial separation of sound sources or separation of sound images by interaural differences as reported by many researchers. These results can be utilized in the development of a more effective speech acquisition system by treating it as a human interface.

REFERENCES

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

[2] W. A. Yost, "The cocktail party problem: Forty years later," in *Binaural and Spatial Hearing in Real and Virtual Environments*, R. H. Gilkey, T. R. Anderson, Eds. (Lawrence Erlbaum Associates, New Jersey, 1997), Chap. 17, pp. 329–347.

[3] A. W. Bronkhorst, "The cocktail party phenomenon: A review of research on speech intelligibility in multiple-talker conditions," *Acustica*, **86**, 117–128 (2000).

[4] J. C. R. Licklider, "The influence of interaural phase relations upon the masking of speech by white noise," *J. Acoust. Soc. Am.*, **20**, 150–159 (1948).

[5] I. J. Hirsh, "The influence of interaural phase on interaural summation and inhibition," *J. Acoust. Soc. Am.*, **20**, 536–544 (1948).

[6] W. Spieth, J. F. Curtis and J. C. Webster, "Responding to one of two simultaneous messages," *J. Acoust. Soc. Am.*, **26**, 391–396 (1954).

[7] J. T. Moter, "Binaural detection of signal with angular dispersion of masking noise," *BS thesis, Dept. of Electrical Engineering, MIT, Cambridge, Massachusetts* (1964).

[8] M. Ebata, T. Sone and T. Nimura, "Improvement of hearing ability by directional information," *J. Acoust. Soc. Am.*, **43**, 289–297 (1968).

[9] H. Levitt and L. R. Rabiner, "Binaural release from masking for speech and gain in intelligibility," *J. Acoust. Soc. Am.*, **42**, 601–608 (1967).

[10] M. Ebata, T. Sone and T. Nimura, "Improvement of detectability of pure tone and intelligibility of speech in noise by directional hearing," *Sci. Rep. Res. Inst. Tohoku Univ. B Electr. Commun.*, **17**, 1–20 (1965).

[11] R. Carhart, T. W. Tillman and K. R. Johnson, "Binaural masking of speech by periodically modulated noise," *J. Acoust. Soc. Am.*, **39**, 1037–1050 (1966).

[12] R. Carhart, T. W. Tillman and K. R. Johnson, "Release of masking for speech through interaural time delay," *J. Acoust. Soc. Am.*, **42**, 124–138 (1967).

[13] T. H. Townsend and D. P. Goldstein, "Suprathreshold binaural unmasking," *J. Acoust. Soc. Am.*, **51**, 621–624 (1972).

[14] G. B. Henning, "Effect of interaural phase on frequency and amplitude discrimination," *J. Acoust. Soc. Am.*, **54**, 1160–1178 (1973).

[15] R. Plomp, "Binaural and monaural speech intelligibility of connected discourse in reverberation as a function of azimuth of a single competing sound source (Speech or Noise)," *Acustica*, **34**, 200–211 (1976).

[16] J. Koehnke and J. M. Besing, "A procedure for testing speech intelligibility in a virtual listening environment," *Ear Hear.*, **17**, 211–217 (1996).

[17] R. L. Freyman, K. S. Helfer, D. D. McCall and R. K. Clifton, "The role of perceived spatial separation in the unmasking of speech," *J. Acoust. Soc. Am.*, **106**, 3578–3588 (1999).

[18] R. L. Freyman, U. Balakrishnan and K. S. Helfer, "Spatial release from informational masking in speech recognition," *J. Acoust. Soc. Am.*, **109**, 2112–2122 (2001).

[19] J. E. Hawkins and S. S. Stevens, "The masking of pure tones and of speech by white noise," *J. Acoust. Soc. Am.*, **22**, 6–13 (1950).

[20] R. Plomp, "Auditory handicap of hearing impairment and the limited benefit of hearing aids," *J. Acoust. Soc. Am.*, **63**, 533–549 (1978).

[21] R. Plomp and A. M. Mimpen, "Effect of the orientation of the speaker's head and the azimuth of a noise source on the speech-reception threshold for sentences," *Acustica*, **48**, 325–328 (1981).

[22] A. J. Duquesnoy, "Effect of a single interfering noise or speech source upon the binaural sentence intelligibility of aged persons," *J. Acoust. Soc. Am.*, **74**, 739–743 (1983).

[23] A. W. Bronkhorst and R. Plomp, "A clinical test for the assessment of binaural speech perception in noise," *Audiology*, **29**, 275–285 (1990).

[24] K. Saberi, L. Dostal, T. Sadralodabai, V. Bull and D. R. Perrott, "Free-field release from masking," *J. Acoust. Soc. Am.*, **90**, 1355–1370 (1991) .

[25] R. H. Gilkey and M. D. Good, "Effects of frequency on free-field masking," *Hum. Factors*, **37**, 835–843 (1995).

[26] W. E. Kock, "Binaural localization and masking," *J. Acoust. Soc. Am.*, **22**, 801–804 (1950).

[27] I. J. Hirsh, "The relation between localization and intelligibility," *J. Acoust. Soc. Am.*, **22**, 196–200 (1950).

[28] W. A. Yost, R. H. Dye Jr. and S. Sheft, "A simulated "cocktail party" with up to three sound sources," *Percept. Psychophys.*, **58**, 1026–1036 (1996).

[29] M. L. Hawley, R. Y. Litovsky and H. S. Colburn, "Speech intelligibility and localization in a multi-source environment," *J. Acoust. Soc. Am.*, **105**, 3436–3448 (1999).

[30] A. W. Bronkhorst and R. Plomp, "The effect of head-induced interaural time and level differences on speech intelligibility in noise," *J. Acoust. Soc. Am.*, **83**, 1508–1516 (1988).

[31] D. D. Dirks and R. H. Wilson, "The effect of spatially separated sound sources on speech intelligibility," *J. Speech Hear. Res.*, **12**, 5–38 (1969).

[32] M. D. Burkhard and R. M. Sachs, "Anthropometric manikin for acoustic research." *J. Acoust. Soc. Am.*, **58**, 214–222

Table 1 Masking release due to spatial information.

Researcher (reference No.)	Stimulus (Signal)	Interferer (Masker)	Advantage	Measure	Other condition
(1) Masking release due to source separation in free field					
Hirsh (1950)[27]	2 syllables	broad band noise	5 dB	threshold of intelligibility	90° separation
Moter (1964)[7]	500 Hz tone	broad band noise	7 dB	threshold	90° separation
Ebata *et al.*	pure tones	broad band noise	10 dB	threshold	90° separation
(1965,1968)[8,10]	monosyllabic words (male talker)	broad band noise	2.7 dB	SNR	90° separation
Plomp (1976)[15]	connected discourse	speech spectrum noise	5 dB	threshold of intelligibility	135° separation
	connected discourse	connected discourse	5 dB	threshold of intelligibility	135° separation
Plomp and Mimpen (1981)[21]	sentences	speech spectrum noise	10 dB	SRT	90° 125° separation
Duquensnog (1983)[22]	sentences	speech spectrum noise	9.6 dB	50% intelligibility	90° separation
	sentences (female talker)	speech (male talker)	6.7 dB	50% intelligibility	90° separation
Brankhourst and Plomp (1990)[23]	short sentences	speech spectrum noise	6 dB	SRT	90° separation
Yost *et al.* (1996)[28]	words (male talker)	words (male talker)	25%	identification	90° separation
Hawley *et al.* (1999)[29]	key words (male talker)	sentences (male talker)	50%	error rate	30° separation
Freyman *et al.* (1999,2000)[17,18]	key words	speech spectrum noise	8 dB	SNR	60° separation
	key words (female talker)	sentences (female talker)	14 dB	SNR	60° separation
(2) Masking release using artificial head					
Hirsh (1950)[27]	2 syllables	broad band noise	2.5 dB	threshold of intelligibility	90° separation
	2 syllables	broad band noise	4.5 dB	threshold of intelligibility	90° separation (2 mike vs monaural)
Dirks and Wilson (1969)[31]	spondee words	broad band noise	7 dB	intelligibility	90° separation (2 mike vs monaural)
	spondee words	broad band noise	9 dB	intelligibility	90° separation (dummy head)
Bronkhorst and Plomp (1988)[30]	sentence (female talker)	speech spectrum noise	9.4 dB	SRT	90° separation (KEMAR)
Koehnke and Besing (1996)[16]	monosyllabic words	speech spectrum noise	13 dB	50% intelligibility	90° separation (KEMAR)
Yost *et al.* (1996)[28]	key words (male talker)	words (male talker)	30%	identification	binaural vs monaural
Peissig and Kollmeier (1997)[35]	speech (male talker)	speech spectrum noise or multi-talker speech	10 dB	SRT	105° separation (HRTF)
Hawley *et al.* (1999)[29]	key words (male talker)	competing sentence	30%	error rate	30° separation(KEMAR)
Drullman and Bronkhorst (2000)[36]	words and sentence (male talker)	words (male and female talker)	40∼50%	intelligibility	binaural vs monaural
Freyman *et al.* (2001)[18]	key words (female talker)	sentence(female talker)	35%	word score	60° separation (KEMAR)
Shinn-Cunningham *et al.* (2001)[37]	key words(male talker)	speech spectrum noise	6 dB	SRT	45° separation (HRTF)
(3) Masking release due to interaural time difference					
Schubert (1956)[44]	words	noise	3 dB	intelligibility	$\Delta t = 470\ \mu$s
Schubert and Schultz (1962)[41]	words	sentences	4 dB	intelligibility	$\Delta t = 0.5$ ms
Levitt and Rabiner (1967)[9]	words	broad band noise	3 dB	50% intelligibility	$\Delta t = 0.5 \sim 10$ ms
	speech	broad band noise	13 dB	threshold	$\Delta t = 0.5 \sim 10$ ms
	speech	broad band noise	3 dB	intelligibility	$\Delta t = 0.5 \sim 10$ ms
Carhart *et al.* (1967)[12]	spondee words	white noise	3 dB	intelligibility	$\Delta t = 0.8$ ms
	spondee words	broad band noise	6 dB	threshold	$\Delta t = 0.8$ ms
	spondee words	broad band noise	4 dB	intelligibility	$\Delta t = 0.8$ ms
Carhart *et al.* (1968)[45]	spondee words	broad band noise	1.5 dB	intelligibility	$\Delta t = 0.8$ ms
	monosyllable	broad band noise	2.5 dB	intelligibility	$\Delta t = 0.8$ ms
Carhart *et al.* (1969)[46,47]	spondee words	modulated noise or speech	4.5 dB	intelligibility	$\Delta t = 0.8$ ms
Bronkhorst and Plomp (1988)[30]	sentence	speech spectrum noise	4.7 dB	SRT	$\Delta t = 0.7$ ms
Darwin and Hukin (2000)[48]	words	sentences	10%	intelligibility	$\Delta t = \pm 90\ \mu$s
(4) Masking release due to phase reverse					
Licklider (1948)[4]	speech	broad band noise	3 dB	intelligibility	Sπ
Carhart *et al.* (1996)[11]	spondee words	broad band noise	6∼7 dB	threshold	Sπ
(1997)[12]	spondee words	broad band noise	3∼6 dB	intelligibility	Sπ
Levitt and Rabiner (1967)[9]	speech	broad band noise	13 dB	threshold	Sπ or Nπ
	speech	broad band noise	6 dB	intelligibility	Sπ or Nπ

(1975).
[33] A. W. Bronkhorst and R. Plomp, "Effect of multiple speech-like maskers on binaural speech recognition in normal and impaired hearing," *J. Acoust. Soc. Am.*, **92**, 3132–3139 (1992).
[34] M. A. Ericson and R. L. McKinley, "The intelligibility of multiple talkers separated spatially in noise," in *Binaural and Spatial Hearing in Real and Virtual Environments*, R. H. Gilkey, T. R. Anderson, Eds. (Lawrence Erlbaum Associates, New Jersey, 1997), Chap. 32, pp. 701–724.
[35] J. Peissig and B. Kollmeier, "Directivity of binaural noise reduction in spatial multiple noise-source arrangements for normal and impaired listeners," *J. Acoust. Soc. Am.*, **101**, 1660–1670 (1997).
[36] R. Drullman and A. W. Bronkhorst, "Multichannel speech intelligibility and talker recognition using monaural, binaural and three-dimensional auditory presentation," *J. Acoust. Soc. Am.*, **107**, 2224–2235 (2000).
[37] B. G. Shinn-Cunningham, J. Schickler, N. Kopco and R. Litovsky, "Spatial unmasking of nearby speech sources in a simulated anechoic environment," *J. Acoust. Soc. Am.*, **110**, 1118–1129 (2001).
[38] D. S. Brungart and W. M. Rabinowitz, "Auditory localization of nearby sources. Head-related transfer functions," *J. Acoust. Soc. Am.*, **106**, 1465–1479 (1999).
[39] D. S. Brungart and B. D. Simpson, "The effects of spatial separation in distance on the informational and energetic masking of a nearby speech signal," *J. Acoust. Soc. Am.*, **112**, 664–676 (2002).
[40] D. S. Brungart, "Informational and energetic masking effects in the perception of two simultaneous talkers," *J. Acoust. Soc. Am.*, **109**, 1101–1109 (2001).
[41] E. D. Schubert and M. C. Schultz, "Some aspects of binaural signal selection," *J. Acoust. Soc. Am.*, **34**, 844–849 (1962).
[42] J. D. Mosko and A. S. House, "Binaural unmasking of vocalic signals," *J. Acoust. Soc. Am.*, **49**, 1203–1212 (1971).
[43] J. L. Flanagan and B. J. Watson, "Binaural unmasking of complex signals," *J. Acoust. Soc. Am.*, **40**, 456–468 (1966).
[44] E. D. Schubert, "Some preliminary experiments on binaural time delay and intelligibility," *J. Acoust. Soc. Am.*, **28**, 895–901 (1956).
[45] R. Carhart, T. W. Tillman and K. R. Johnson, "Effects of interaural time delays on masking by two competing signals," *J. Acoust. Soc. Am.*, **43**, 1223–1230 (1968).
[46] R. Carhart, T. W. Tillman and E. S. Greetis, "Release from multiple maskers: Effects of interaural time disparities," *J. Acoust. Soc. Am.*, **45**, 411–418 (1969).
[47] R. Carhart, T. W. Tillman and E. S. Greetis, "Perceptual masking in multiple sound backgrounds," *J. Acoust. Soc. Am.*, **45**, 694–703 (1969).
[48] C. J. Darwin and R. W. Hukin, "Effectiveness of spatial cues, prosody, and talker characteristics in selective attention," *J. Acoust. Soc. Am.*, **107**, 970–977 (2000).
[49] J. L. Puel, P. Bonfils and R. Pujol, "Selective attention modifies the active micromechnical properties of the cochlea," *Brain Res.*, **447**, 380–383 (1988).
[50] E. L. LePage, "Functional role of the olivo-cochlear bundle: A motor unit control system in the mammalian cochlea," *Hear. Res.*, **38**, 177–188 (1989) .
[51] W. Teder and R. Naatanen, "Event-related potentials demonstrate a narrow focus of auditory spatial attention," *Cognit. Neurosci. Neuropsychol.*, **5**, 709–711 (1994).
[52] B. Scharf, J. Magnan, L. Collet, E. Ulmer and A. Chays, "On the role of the loivocochlear bundle in hearing: A case study," *Hear. Res.*, **75**, 11–26 (1994).
[53] B. Scharf, J. Magnan and A. Chays, "On the role of the loivocochlear bundle in hearing: 16 case studies," *Hear. Res.*, **103**, 101–122 (1997).
[54] G. Z. Greenberg and W. D. Larkin, "Frequency-response characteristic of auditory observers detecting signals of a single frequency in noise: The probe-signal method," *J. Acoust. Soc. Am.*, **44**, 1513–1523 (1968).
[55] B. Scharf, S. Quigley, C. Aoki, N. Peachey and A. Reeves, "Focused auditory attention and frequency selectivity," *Percept. Psychophys.*, **42**, 215–223 (1987).
[56] B. Scharf, "Spectral specificity in auditory detection: The effect of listening on hearing," *J. Acoust. Soc. Jpn.* (*E*), **10**, 309–316 (1989).
[57] R. S. Schlauch and E. R. Hafter, "Listening bandwidths and frequency uncertainty in pure-tone signal detection," *J. Acoust. Soc. Am.*, **90**, 1332–1339 (1991).
[58] E. R. Hafter, R. S. Schlauch and J. Tang, "Attending to auditory filters that were not stimulated directly," *J. Acoust. Soc. Am.*, **94**, 743–747 (1993).
[59] H. Dai, B. Scharf and S. Buus, "Effective attenuation of signals in noise under focused attention," *J. Acoust. Soc. Am.*, **89**, 2837–2842 (1991).
[60] B. A. Wright and H. Dai, "Detection of unexpected tones with short and long durations," *J. Acoust. Soc. Am.*, **95**, 931–938 (1994).
[61] B. A. Wright and H. Dai, "Detection of unexpected tones in gated and continuous maskers," *J. Acoust. Soc. Am.*, **95**, 939–948 (1994).
[62] M.-C. Botte, "Auditory attentional bandwidth: Effect of level and frequency range," *J. Acoust. Soc. Am.*, **98**, 2475–2485 (1995).
[63] M. J. Penner, "The effect of payoffs and cue tones on detection of sinusoids of uncertain frequency," *Percept. Psychophys.*, **11**, 198–202 (1972).
[64] N. A. Macmillan and M. Schwartz, "A probe-signal investigation of uncertain-frequency detection," *J. Acoust. Soc. Am.*, **58**, 1051–1058 (1975).
[65] M. F. Yama and D. E. Robinson, "Comparison of frequency selectivity for the monaural and binaural hearing system: Evidence from a probe-frequency procedure," *J. Acoust. Soc. Am.*, **71**, 694–700 (1982).
[66] J. R. Ison, T. M. Virag and P. D. Allen, "The attention filter for tones in noise has the same shape and effective bandwidth in the elderly as it has in young listeners," *J. Acoust. Soc. Am.*, **112**, 238–246 (2002).
[67] M. Ebata, H. Miyazono, S. Suzuki, T. Usagawa and B. Scharf, "Auditory detection of multiple targets," *J. Acoust. Soc. Jpn.* (*E*), **18**, 163–171 (1997).
[68] E. R. Hafter and K. Saberi, "A level of stimulus representation model for auditory detection and attention," *J. Acoust. Soc. Am.*, **110**, 1489–1497 (2001).
[69] M. Ebata, H. Miyazono, K. Kumamaru, Y. Chisaki and T. Usagawa, "The formation of attention filters for a missing-fundamental complex tone and frequency-gliding tones," *Acoust. Sci. & Tech.*, **22**, 401–406 (2001).
[70] G. Rhodes, "Auditory attention and the representation of spatial information," *Percept. Psychophys.*, **42**, 1–14 (1987).
[71] T. A. Mondor and R. J. Zatorre, "Shifting and focusing auditory spatial attention," *J. Exp. Psychol. Hum. Percept. Perform.*, **21**, 387–409 (1995).
[72] C. J. Spence and J. Driver, "Covert spatial orienting in audition: Exogenous and endogenous mechanisms," *J. Exp. Psychol. Hum. Percept. Perform.*, **20**, 555–574 (1994).
[73] P. T. Quinlan and P. J. Bailey, "An examination of attentional control in the auditory modality: Further evidence for auditory orienting," *Percept. Psychophys.*, **57**, 614–628 (1995).

[74] T. L. Arbogast and G. Kidd Jr., "Evidence for spatial tuning in informational masking using the probe-signal method," *J. Acoust. Soc. Am.*, **108**, 1803–1810 (2000).
[75] G. Kidd Jr., C. R. Mason, T. L. Rohtla and P. S. Deliwala, "Release from masking due to spatial separation of sources in the identification of nonspeech auditory patterns," *J. Acoust. Soc. Am.*, **104**, 422–431 (1998).

Masanao Ebata was born in Fukui, Japan in 1940. He received the B.E., the M.E., and the Dr. Eng. Degree from Tohoku University, Sendai, Japan, in 1963, 1965 and 1968, respectively. In 1968, he jointed the Faculty of Engineering, Tohoku University. From 1983 to 2001, he was the professor in Kumamoto University and now he is a president of Kumamoto National College of Technology. He has been engaged in researches on psychological acoustics, auditory information processing, noise evaluation, noise control, acoustic measurement and so on. Dr. Ebata is a member of ASJ, ASA, IPSJ, INCE Japan and IEICE.