

COVER PAGE

Document downloaded by @DAEL

Sat Nov 28 15:15:11 2026
For personal use

When automatic English translation is provided, only the original document is authentic.
The EAA cannot be held responsible of any translation error

Bibliographical reference

Effect of Familiar Masking Background on Target Speech Recognition, Zhang Bing, Long Zhang-Cai and Zhang Yang-Ping, Acta Acustica **vol. 98** (Number 2), 2012, pp. 328-333

DOI
https://doi.org/10.3813/AAA.918516

ACTA ACUSTICA UNITED WITH ACUSTICA
Vol. 98 (2012) 328 – 333
DOI 10.3813/AAA.918516

Effect of Familiar Masking Background on Target Speech Recognition

Zhang Bing[1)], Long Zhang-Cai[1)], Zhang Yang-Ping[2)]
[1)] Department of Physics, Huazhong University of Science and Technology, Wuhan, Luoyu Road 1037, 430074, P. R. China. longzc01@mails.tsinghua.edu.cn
[2)] Affiliated hospitals of Huazhong University of Science and Technology, Wuhan, Luoyu Road 1037, 430074, P. R. China

Summary
Real auditory environments are usually complex. The acoustic signals reaching our ears are mixtures of target and background masking sounds. It has been demonstrated before that the human auditory system can separate target speech from a masking background more easily when the listener have been cued to the target beforehand. However, less data are avaliable concerning the influence of masker cues on the ability to recognize the target speech. A psychoacoustic experiment conducted in this study indicates that cueing the listener to the masker for a short time actually impairs, rather than improves, performance. It suggests that the prior knowledge of the masker enhances the informational masking and makes the target recognition difficult. We conclude that the human auditory system is involved in a performance of only enhancing the familiar target, but not suppressing the familiar masker speech signals.

PACS no. 43.66.Dc, 43.71.Bp

1. Introduction

Real human auditory environments are usually complex. The acoustic signals reaching our ears are mixtures of sounds coming from different sources. Although the incoming speech signals are complicated, in most cases the human auditory system is powerful enough to extract and follow the component which the listener is interested in. This phenomenon is called the "cocktail party effect" [1]. In the past half century, this marvelous function of the human auditory system and its underlying mechanism have been arousing people's wide concern and study [2, 3, 4, 5].

Although the human auditory system is marvelous, the background masking sounds interfere with the target sound and reduce people's ability to recognize the target. Two factors arouse the interference: energetic masking and informational masking [6, 7]. Energetic masking occurs at the auditory periphery when target and masking sounds both contain energy in the same critical bands at the same time, so that the peripheral neural activity elicited by the target sound is overwhelmed by that elicited by the masking sounds. This leads to a degraded neural representation of the target signal, and makes it difficult for the subsequent higher level processes to extract the target. Informational masking is a kind of higher-level masking beyond the periphery. It occurs especially when the target and the masker are both speeches. In informational masking, the processing of the masking speech interferes with the processing of the target speech at cognitive levels, which makes the listener hard to extract the information of the target from the similar-sounding masking speech.

Understanding the mechanism of people's identification of target sound in a complicated listening environment has always been a hot topic. Bronkhorst gave a comprehensive summary of people's previous studies on the cocktail party phenomenon in 2000 [8]. He pointed out that the ability of the human auditory system to recognize a target sound in a complicated listening environment is influenced by the following factors: masker type, similarity of target and masking sounds, locations of target and masking sounds, reverberation of the listening environment, etc. When sound sources are at different locations, spatial cues, such as the interaural time difference, the interaural intensity difference and the interaural phase difference, help listeners to extract the target sound. The spatial separation of sound sources can greatly improve the target sound recognition performance [5, 9, 10, 11, 12]. However, when the target and the masking sound sources are overlapped, there is no any interaural spatial cue. In such circumstance, the human auditory system extracts the target sound from the masking background by monaural cues, such as vocal features of the target and the masking speakers (vocal tract size, fundamental frequency, accent, speaking style, etc.), differences in prosodic features and in intensities of the target and the masking sounds, etc. [13, 14].

In speech-on-speech masking, informational masking is more important than energetic masking. The amount of

Received 29 June 2011,
accepted 22 January 2012.

© S. Hirzel Verlag · EAA

informational masking highly depends on the similarity of target and masker voices [6]. It has ever been demonstrated by Freyman that two-talker masker is most effective in creating informational masking and the human auditory system can separate target speech from masking background more easily if the listener is cued to part of the target speech before the target is presented in masking [15]. Cued to the target makes the listener familiar with the voice characteristics of the target speaker. Hence when the mixed signals stimulate the listener's auditory system, it is easy for the listener to distinguish the familiar target from the background masker, and selectively attend to the target. This is called "priming effect". Two other kind of priming conditions were also studied by Freyman, in which the listener was primed by the target sentence spoken by another person, and primed by the print of the target sentence respectively. Interestingly, different kind of priming conditions improved performance relative to the no-prime condition by approximately the same amount. It shows that the prime needs not to be the exact target waveform. As long as the prime can help the listener attend to the target in the masking background, it is efficient, and attention plays important role in the target speech recognition. Chinese speech is quite different from English speech both in pronunciation and sentence structure, which may lead to a different informational masking effect. Zhigang Yang et al. did some experiments analogous to Freyman's in 2007. They found that there is also priming effect in Chinese speech-on-speech masking situation [7].

Work conducted by Freyman and Zhigang Yang both showed that the prior knowledge of the target cues helps the listener attend to the target and effectively improves the target recognition performance. The background masker is another important component in the mixed signal. We are curious about whether the background masker cues can also help the auditory system to rule the masker out from the mixture and extract the target effectively. In addition, an auditory enhancement effect put forward by Viemeister [16, 17] showed that the prior exposure to a harmonic complex with one pure tone component deleted can make the pure tone reinserted in the following harmonic complex perceptually "pop out". In this phenomenon, cueing to the masker (the harmonic complex without the pure tone component) improves the listener's ability to recognize the target signal (the pure tone component). Viemeister proposed that the decreased effectiveness of the masker was due to a simple adaption of its frequency components after exposure to the masker [17, 18]. Since there is auditory enhancement effect in the pure tone identification, can we find a similar enhancement effect in the speech-on-speech masking situation? Furthermore, as we know, if the characteristics of the target speech are well known, it is easy to separate the target from the masker with computer by some certain speech signal processing technologies. Similarly, if the masker characteristics are well known, we can also eliminate the masking signal in the mixture and improve the signal-to-noise ratio by means of spectral subtraction [19, 20, 21, 22, 23]. These make us to think that since machine can separate the target signal from a familiar masker, whether the human auditory system also have such advanced signal processing ability to extract the target from a familiar masking background. To answer all these questions, a psychoacoustic experiment involving target speech recognition in the brief familiar masking background and in the unfamiliar masking background was conducted. The listener's abilities to identify the target speech in these two different conditions were compared. Unexpectedly, we found that the brief familiar background masker can not improve but impair the listener's ability to recognize the target speech.

2. Methods

A psychoacoustic experiment was designed and conducted in this study to find the effect of the familiar masking background on the target speech recognition. Three factors were involved in the experiment: prime type (masker prime, no prime), masker type (two-female talker masker, two-male talker masker, one-female + one-male talker masker), and SNR (-8 dB, -4 dB, 0 dB, 4 dB). So there were $2\times3\times4 = 24$ different experiment conditions in total for each subject. In the masker-prime condition, the subject was primed to the background masker twice to be briefly familiar with the masker, and then listened to the target-masker mixture 2 s later. In the no-prime condition, the subject listened to the target-masker mixture directly without masker priming presented. The subject's task was always to repeat the entire target sentence he/she heard as well as possible. Percent-correct scores were recorded to measure the subject's target speech recognition ability. All the experiment conditions were tested in the order by Latin square design. The subject was guided to do the experiment in the designed order. There was a short-term training session for the subject to get familiar with the target speaker's voice and the experiment procedure before the experiment.

2.1. Stimuli

Target stimuli in the experiment were 480 short nonsense Chinese sentences with correct syntactic structure but incongruous semantics, e.g., "*Morning misses forest.*", for avoiding the subject guessing the rest after having heard the previous part of the target. Each target sentence contained three frequently used key words underlined as above for scoring purpose. Percent-correct scores were derived from the number of key words correctly identified by the subjects. All target stimuli were recorded by an adult Chinese female talker who could speak standard mandarin.

Masking stimuli were two-talker speech maskers. There were 960 masking utterances in all with both correct syntactic structure and semantics recorded by two adult Chinese males and two females respectively. The normal masking sentences were used to help the subjects more easily get familiar with the maskers. All the target and masking utterances were recorded digitally onto a notebook computer (Sony VAIO VGN-SZ562N) disk in

an anechoic chamber (inside dimension: $3.5\times2.4\times2.5\,m^3$, background noise $\geq$ 30 dB(A)) by the Cool Edit Pro 2.1 software, sampled at 22050 Hz and saved as 16-bit PCM wave files. Those evident pauses in each masking signal were removed so that the final waveform was a continuous stream without gap. The rms of all the targets and maskers were equated with one another.

Mixed stimuli were one target and one two-talker masker mixtures in which the masker led the target by 0.5 s while they ended simultaneously. The brief delay in target onset could help the listener attend to the target. The synthesis of the masking and the mixture stimuli was conducted by the Matlab 7.0 software. All the speech stimuli were stored in the notebook computer disk and presented through the internal loudspeakers of the computer.

2.2. Participants

24 university students (age range from 20 to 25, 10 males and 14 females) with self-reported normal hearing volunteered to participate in the psychoacoustic experiment. They had no any similar experiment experience before. They were instructed to do the experiment and were paid for their participation.

2.3. Design and Procedure

During the experiment, the subject sat at the center of the anechoic chamber. As the sound source of the target and the masker, the notebook computer was placed 1.2 m in front of the subject, and it nearly had same elevation as that of the two ears of the subject. During the experiment, the targets were presented at a fixed level of 50 dBA at the position where the subjects sat, while the maskers varied in level from block to block to create different signal-to-noise ratios.

There were three within-subject variables in the experiment:

1. prime type (masker prime, no prime),
2. masker type (two-female talker masker, two-male talker masker, one-female + one-male talker masker),
3. SNR (−8 dB, −4 dB, 0 dB, 4 dB).

Therefore, there were 2×3×4=24 different experiment conditions in total for each subject. The subjects did the experiment by a man-machine interface designed by matlab 7.0 software. There were 24 buttons on the interface, in which one button represented one experiment condition. All the experiment conditions were tested in the order by Latin square design to make every case presented with same probability of the sequence, and mitigate some possible influences (listener's increasing familiarity with the experiment as the experiment went on, the auditory fatigue, and listener's different perception ability to different sentences, etc.) on the accuracy of the experiment results. During the experiment, the subject was instructed to press the buttons in the designed order. There was a short-term training session before the experiment to familiarize the subject with the target speaker's voice and the experiment procedure. The utterances used in the training were different from those used in the formal testing.

In the masker-prime condition, after the button press, an individual trial consisted of a priming utterance (a two-talker masker) was presented in quiet twice, and then followed by a mixture utterance (a target utterance presented against the background masker) 2 seconds later. The twice masker prime was to make the listener get familiar with the background masker well, and the 2 s interval between the prime and the testing utterance was to make the listener get ready for the following recognition task and also mitigate the listener's attention on the priming masker. The listener's task was always to repeat the entire target speech as well as possible. One person sitting behind the listener recorded whether the key words were identified correctly. The no-prime condition was the same except the masker prime was not presented. Each subject did 10-time tests in every experiment condition. The target and masking sentences varied across each trial. A complete experiment lasted for about one hour and a half, and there were two intervals during the experiment to mitigate the auditory fatigue.

3. Results

Mean percent correct identification scores of key words in target sentences across 24 listeners as a function of signal-to-noise ratios for the three masking conditions and the two priming conditions are shown in Figure 1. Each data point in the Figure is based on 240 responses (24 subjects × 10 items per list). It shows that in all the three masking conditions there was no improvement in the target speech intelligibility for masker-prime as compared to no-prime conditions, but deterioration except for the performance at 4 dB in the female speech masking and the female + male speech masking conditions. A three-way, repeated measure ANOVA confirms that the main effect due to prime condition was significant, $F(1, 23) = 56.75$, $p < .0001$. Subjects performed better in no-prime than masker-prime conditions, $t(46) = 2.06$, $p < .05$. The other two main effects due to masker type ($F(2, 46) = 451.44$, $p < .0001$) and SNR ($F(3, 69) = 1173.50$, $p < .0001$) were also significant. The two-way interactions between prime condition and masker type was significant, $F(2, 46) = 56.27$, $p < .0001$. Subjects were easier to recognize the target speech in no-prime than masker-prime conditions in all the three different masker types (no-prime vs masker-prime for female masker, $t(46) = 4.15$, $p < .001$, for male masker, $t(46) = 5.73$, $p < .001$, and for female + male masker, $t(46) = 6.83$, $p < .001$). The two-way interactions between prime condition and SNR ($F(3, 69) = 117.49$, $p < .0001$), and between masker type and SNR ($F(6, 138) = 42.33$, $p < .00014$) were also significant. However, the three-way interaction among prime type, masker type and SNR was not significant, $F(6, 138) < 1$.

The signal-to-noise ratios required for a criterion performance of 50% correct were estimated through linear interpolation of the function in Figure 1. The derived data displayed in Table I and Figure 2 shows that in the three different masking conditions, the S-N ratios for 50%-correct

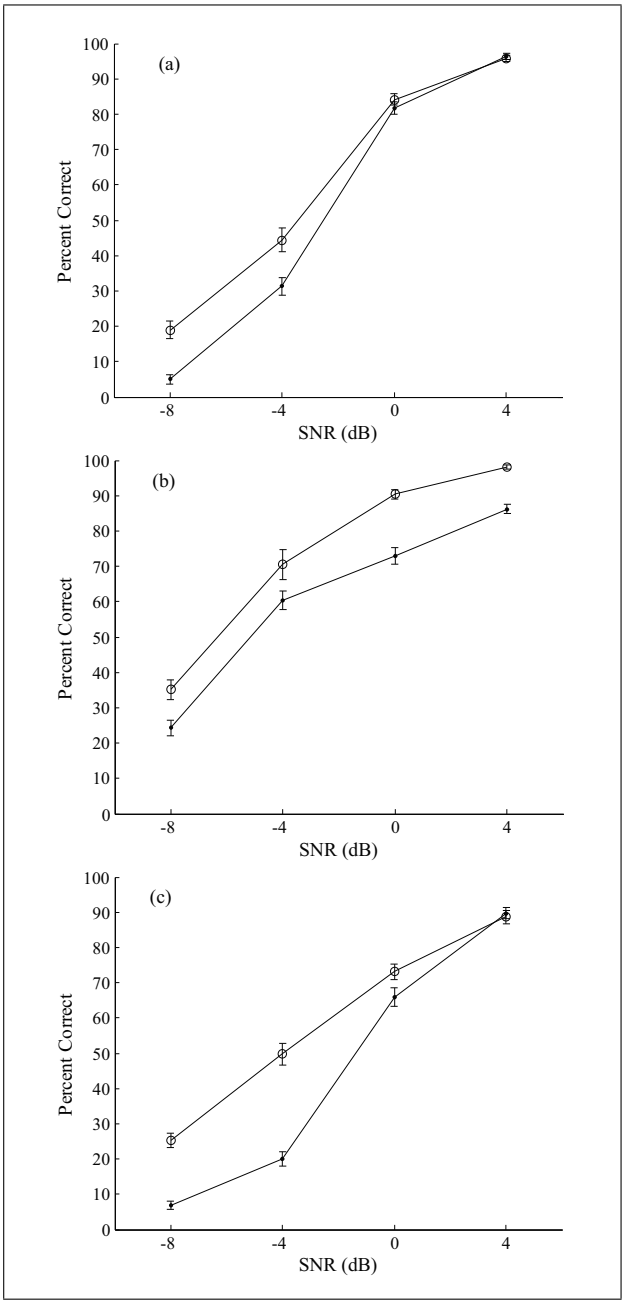

Figure 1. Mean percent correct scores of key words in the target sentences across 24 listeners as a function of signal-to-noise ratios for three masking conditions (a: female speech masking; b: male speech masking; c: male + female speech masking) and two prime conditions: (1) no prime (circles); (2) masker prime (dots). Error bars indicate the standard error of the mean.

responses in the masker-prime condition are all higher than that in the no-prime condition. It illustrates from another aspect that the familiar masking background might not lead to a better performance in the target speech recognition.

The three masking backgrounds in the experiment represent most masking situations in reality. The data of the no-prime situation in Table I shows that the masking effects of the three background maskers are different and the rank order from high to low is female-on-female speech masking, male + female-on-female speech masking, and male-on-female speech masking. It indicates that the degree of informational masking depends on the similarity of the target and masker's phonetic features, and the masking effect in the same-sex-talker condition is much greater than in the different-sex-talker condition. This is consistent with some previous studies [6, 24, 25]. The SNR differences of masker-prime vs no-prime conditions for 50%-correct response are also shown in Table I. The rank order from high to low is M+F-on-F masking, M-on-F masking, and F-on-F masking. This illustrates that, at the 50% target speech intelligibility, masker priming influences the masking effect most in the male + female masking condition.

Table I. SNR and SNR difference (dB) of masker-prime vs no-prime conditions derived by interpolation for 50% correct response under different masking conditions.

	F-on-F	M-on-F	M+F-on-F
No Prime	-3.4406	-6.3216	-3.9527
Masker Prime	-2.5193	-5.1539	-1.3897
ΔSNR	0.9213	1.1677	2.5630

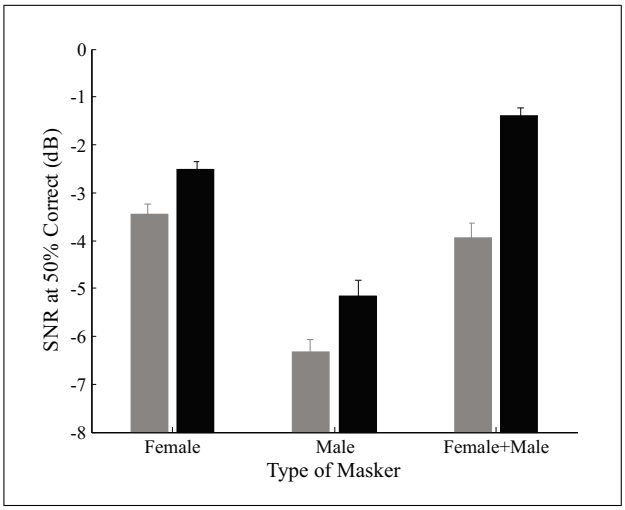

Figure 2. S-N ratios at 50%-correct performance as a function of the type of masker. The grey rectangles represent the data for the no-prime condition, and the black rectangles represent the data for the masker-prime condition. Error bars indicate the standard error of the mean.

4. Discussion

In a complicated listening environment, the masking sounds interfere with the target sound and make it difficult for listeners to extract the target. Some previous studies have demonstrated that the cues of the target speech can help the listener to trace and recognize the target [7, 15]. While there were very few studies about the effect of masking sound cues on the target speech identification. This paper is focusing on this problem and provides some cues for answering the question.

Auditory scene analysis contains bottom-up and top-down processes. They usually exist in the meantime, but

play different roles in different listening situations. The bottom-up is a lower level process conducted in the auditory periphery. It uses some cues of the incoming acoustic signals, such as frequency proximity, spectral similarity, common spatial location, common onset, rhythmic structure, etc., to decompose the incoming acoustic signals into a collection of sensory segments and group segments into different auditory streams. It is a simpler process fulfilled automatically without listener's prior knowledge and attention involved. While the top-down is a higher level process conducted in the auditory center. It is a controlled schema-driven process of more detailed analysis, in which previous knowledge and attention are involved [13, 26, 27]. In the top-down process, a comparison between the previous knowledge and the incoming signal is made to extract the meaning from the acoustic data [27]. Our study focuses on the speech-on-speech masking in which the top-down dominates over the bottom-up process. The experiment results provide some cues about the working modes of the auditory central system.

The worse target speech identification performance in the masker-prime condition than in the no-prime condition of this study indicates that the human auditory system has no such powerful processing ability to rule the familiar masker out from the mixture although the prior knowledge of the masker may help the auditory system to distinguish the target from the masker. However, some previous studies [7, 15] have demonstrated that if the recognition task is primed by some target speech cues, the performance can be effectively improved. All these results indicate that the auditory central system seems to be involved in a performance of only enhancing familiar target speech, not suppressing familiar masking speeches like signal processing technology in engineering. Identifying target speech from a familiar masker or a familiar target contained mixture is a higher-level process in human brain involving learning, knowledge storing and memory. Weinberger's studies have shown that when acoustic training stimuli are involved, associative plasticity develops in the primary auditory cortex (A1), in which the specific information of the acoustic stimuli is stored [28, 29, 30]. Lindsay also pointed out that the characteristic information of the acoustic stimuli can be stored in the brain as some patterns [31]. Based on their studies, we suggest that when the listener's auditory system was stimulated by the background masker several times, some characteristic information of the masker was stored in the brain. The stored information could not disappear immediately but stayed in the brain for a period of time due to the plasticity. When the mixed signals appeared several seconds later, the stored masker information was easily matched with the masker in the mixture automatically and the listener's attention was directed to the masker. Hence the listener couldn't attend to the target speech well and performed worse in the target identification task. However, in the target-prime situation, characteristic information of the cueing target speech was stored in the brain. When the mixed speech signals appeared, the stored target information was matched with the target signal in the mixture automatically, and a better target identification performance was presented.

The experiment results also indicate that the masker priming enhanced the speech-on-speech informational masking effect. Some previous studies have shown that the informational masking effect mainly depends on the similarity of the masker and the target voices. The more similar the masker and target are, the more significant the masking effect is [6, 7]. Our research reveals that the familiarity of the masker does also affect the informational masking effect, and under certain conditions (the time for the masker cueing is not too long to make the listener fatigue), a familiar masker enhances the informational masking effect. We propose that when the mixed signals reached the auditory system, higher level auditory processing of both the masker and the target started. Influenced by the prior knowledge of the background masker, the processing of the masker was greatly enhanced and dominated over the simultaneous target speech processing, which increased the informational masking and made the identification of the target difficult.

While maybe we all have such experience, after having been exposed to a noisy environment for a long time, we are sensitive to a new sound. In this case, it seems that the very familiar masking environment helps listeners to sense the target sound. This is a kind of adaptation or repetitive suppression of human central auditory system [32, 33]. The long-term exposure to the masking environment makes the listener to be very familiar with the background masker and easily to distinguish the masking sound from the target sound. However, in the meantime, it causes inhibition of the central auditory system and reduction of the neural activity. When a novel sound emerges, a sudden stimulation to the auditory nerve system arises and the listener's attention is easily drawn to the novel sound. Hence the listener can follow and extract the novel target sound in the masking background well. Some studies have shown that the neurons in primary auditory cortex and some neurons in the inferior colliculus respond strongly in the novel sound detection [33, 34].

According to above analyses, we find that the familiar masking background can give either positive or negative effect on the target speech identification. It depends on the specific situations, and the auditory processing mechanisms for different situations are different. In this study, the subjects were only primed to the masker for a short time and then did the identification task several seconds later. We find that the brief familiar masking background cannot improve but impair the target speech identification performance. We propose that our auditory centers only perform the familiar target speech enhancing process, do not perform the familiar masking speech eliminating process. The prior knowledge of the background masker enhances the informational masking and draws the listener's attention to the masker in the mixture. All these may help us to understand the higher level processing of human auditory center in the complicated listening environments.

Acknowledgement

This work was supported by Major Research Plan of National Natural Science foundation of China "Cognitive Computing of Visual and Auditory Information (Grant No.: 90820001)".

References

[1] E. C. Cherry: Some experiments on the recognition of speech, with one and with two ears. Journal of the Acoustical Society of America **25** (1953) 975–979.

[2] A. J. Duquesnoy: Effect of a single interfering noise and interfering speech upon the binaural sentence intelligibility of aged persons. Journal of the Acoustical Society of America **74** (1983) 739–743.

[3] W. A. Yost, R. H. Dye, S. Sheft: A simulated cocktail party with up to three sound sources. Perception & Psychophysics **58** (1996) 1026–1036.

[4] M. L. Hawley, R. Y. Litovsky, H. S. Colburn: Intelligibility and localization of speech signals in a multisource environment. Journal of the Acoustical Society of America **105** (1999) 3436–3448.

[5] T. L. Arbogast, C. R. Mason, G. Kidd, Jr.: The effect of spatial separation on informational and energetic masking of speech. Journal of the Acoustical Society of America **112** (2002) 2086–2098.

[6] D. Brungart: Informational and energetic masking effects in the perception of two simultaneous talkers. Journal of the Acoustical Society of America **109** (2001) 1101–1109.

[7] Z. G. Yang, J. Chen, Q. Huang, X. H. Wu: The effect of voice cuing on releasing Chinese speech from informational masking. Speech Communication **49** (2007) 892–904.

[8] A. Bronkhorst: The cocktail party phenomenon: A review of research on speech intelligibility in multiple-talker conditions. Acustica united with Acta Acustica **86** (2000) 117–128.

[9] G. Kidd, Jr., C. R. Mason, P. S. Deliwala, W. S. Woods, H. S. Colburn: Reducing informational masking by sound segregation. Journal of the Acoustical Society of America **95** (1994) 3475–3480.

[10] L. Freyman, R, U. Balakrishnan, K. S. Helfer: Spatial release from informational masking in speech recognition. Journal of the Acoustical Society of America **109** (2001) 2112–2122.

[11] R. L. Freyman, K. S. Helfer, U. Balakrishnan: Spatial and spectral factors in release from informational masking in speech recognition. Acta Acustica united with Acustica **91** (2005) 537–545.

[12] G. Kidd, Jr., C. R. Mason, V. Best, N. Marrone: Stimulus factors influencing spatial release from speech-on-speech masking. Journal of the Acoustical Society of America **128** (2010) 1965–1978.

[13] A. S. Bregman: Auditory scene analysis. MIT, Cambridge, 1994.

[14] C. Darwin, R. Hukin: Effectiveness of spatial cues, prosody and talker characteristics in selective attention. Journal of the Acoustical Society of America **107** (2000) 970–977.

[15] R. L. Freyman, U. Balakrishnan, K. S. Helfer: Effect of number of masking talkers and auditory priming on informational masking in speech recognition. Journal of the Acoustical Society of America **115** (2004) 2246–2256.

[16] N. F. Viemeister: Adaptation of masking. – In: Psychophysical, Physiological and Behavioural Studies in Hearing. G. van den Brink, F. A. Bilsen (eds.). Noordwijkerhout, The Netherlands, 1980, 190–199.

[17] N. F. Viemeister, S. P. Bacon: Forward masking by enhanced components in harmonic complexes. Journal of the Acoustical Society of America **71** (1982) 1502–1507.

[18] B. J. A., S. A. M., V. N. F.: The enhancement effect: Evidence for adaptation of inhibition using a binaural centering task. Journal of the Acoustical Society of America **129** (2011) 2088–209.

[19] R. Martin: Spectral subtraction based on minimum statistics. Proc. Eur. Signal Processing Conference, 1994, 1182–1185.

[20] Y. Ephraim, I. Cohen: Recent advancements in speech enhancement. the electrical engineering handbook. CRC Press, 2006.

[21] S. Rangachari, P. Loizou: A noise-estimation algorithm for highly non-stationary environments. Speech Communication **48** (2006) 220–231.

[22] R. Martin: Noise power spectral density estimation based on optimal smoothing and minimum statistics. IEEE Transaction on Speech and Audio Processing **9** (2001) 504–512.

[23] M. Cooke, D. P. W. Ellis: The auditory organization of speech and other sources in listeners and computational models. Speech Communication **35** (2001) 141–177.

[24] D. Brungart, B. Simpson: The effect of spatial separation in distance on the informational and energetic masking of a nearby speech signal. Journal of the Acoustical Society of America **112** (2002) 664–676.

[25] K. S. Helfer, R. L. Freyman: Aging and speech-on-speech masking. Ear Hearing **29** (2008) 87–98.

[26] W. M. Hartmann: Pitch perception and the segregation and integration of auditory entities. – In: Auditory function: Neurobiological bases of hearing. Wiley, New York, 1998, 623–645.

[27] C. Alain, S. R. Arnott, T. Picton: Bottom-up and top-down influences on auditory scene analysis: evidence from event-related brain potentials. Journal of Experimental Psychology: Human Perception and Performance **27** (2001) 1072–1089.

[28] N. M. Weinberger, D. M. Diamond: Physiological plasticity in auditory cortex: rapid induction by learning. Progress in Neurobiology **29** (1987) 1–55.

[29] N. M. Weinberger: Auditory associative memory and representational plasticity in the primary auditory cortex. Hearing Research **229** (2007) 54–68.

[30] N. M. Weinberger, J. S. Bakin: Learning-induced physiological memory in adult primary auditory cortex: receptive fields plasticity, model, and mechanisms. Audiology and Neurotology **3** (1998) 145–167.

[31] P. H. Lindsay, D. A. Norman: Human information processing. Academic Press, New York, 1977.

[32] N. Ulanovsky, L. Las, I. Nelken: Processing of low-probability sounds by cortical neurons. Nature Neuroscience **6** (2003) 391–398.

[33] K. Grill-Spector, R. Henson, A. Martin: Repetition and the brain: neural models of stimulus-specific effects. Trends in Cognitive Sciences **10** (2006) 14–23.

[34] D. Pérez-González, M. S. Malmierca, E. Covey: Novelty detector neurons in the mammalian auditory midbrain. European Journal of Neuroscience **22** (2005) 2879–2885.