



COVER PAGE

Document downloaded by @DAEL

Wed Oct 28 23:03:50 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

Spatial Separation Benefit for Speech Detection in Multi-Talker Babble-Noise with Different Egocentric Distances, I. G. Andreeva, M. Dymnikowa, A. P. Gvozdeva, E. A. Ogorodnikova and S. P. Pak, *Acta Acustica* **vol. 105** (Number 3), 2019, pp. 484-491

DOI

<https://doi.org/10.3813/AAA.919330>

Spatial Separation Benefit for Speech Detection in Multi-Talker Babble-Noise with Different Egocentric Distances

I. G. Andreeva^{1,2)}, M. Dymnikowa^{4,3)}, A. P. Gvozdeva¹⁾, E. A. Ogorodnikova³⁾, S. P. Pak³⁾

¹⁾ I. M. Sechenov Institute of Evolutionary Physiology and Biochemistry, Russian Academy of Sciences, Toreza pr. 44, 194223 Saint-Petersburg

²⁾ Speech Technology Center-innovations, Krasutskogo st. 4, 196084 Saint-Petersburg

³⁾ I. P. Pavlov Institute of Physiology, Russian Academy of Sciences, Makarova nab. 6, 199034 Saint-Petersburg

⁴⁾ Association of Musical Psychologists and Psychotherapists, Akademika Anohina st. 38 k.3, 119602 Moscow. dmwl@bk.ru

Summary

Experiments performed in free field showed that a task of speech stimuli detection is solved more successfully when a speech signal and a speech-like noise source are spatially separated in distance. The spatial separation of the sources of the target speech signal and the speech-like masker improved detection of the target speech signal in noise at a low signal-to-noise ratio of -14 dB. This indicated the release from masking when the sources of the target signal and the masker were separated by a distance of 3 m, the value of which amounted to 2–3 dB. Release from masking at equal ratios of the intensity level of the target signal and the masker did not depend on which source was closer to the listener. This result provides a possibility of adjusting the auditory system to a specified azimuth angle, at which the sound wave arrives at the ears of the person.

PACS no. 43.60.-c

1. Introduction

In everyday conditions successful communication is possible when the target speech signal can be detected, and the level of speech intelligibility in the case of multiple noise and speech sources is sufficient (cocktail-party problem). The source of the target speech signal, as a rule, is spatially separated from masking noise and/or speech sources. Therefore, the task of detection of the target signal in a noisy environment is greatly aided by the effects of spatial release from masking, among which the most studied example is the binaural release from masking. It occurs during dichotic listening to non-speech and speech signals [1, 2]. This phenomenon is based on the ability of the auditory system to analyze separately the signals arriving at the right and left ears of participants with an inter-aural time difference [3, 4, 5]. It is shown that inter-aural time differences are effective for frequencies below 1000 Hz, with the maximum amount of release from masking reaching 15–16 dB at frequencies of 100–300 Hz, which corresponds to the fundamental frequency range for male and female voices [6]. At frequencies above 1000 Hz, the release from masking when signals are played back through

headphones is determined by binaural decorrelation of the auditory analysis and amounts to only 2–3 dB.

The separation of sources of the target speech signal and the masker along the azimuthal coordinate can lead to a release from masking up to 12 dB [7, 8, 9]. Appearing in the presence of stimuli in free field, inter-aural intensity differences determined by the acoustic effects of the shadow of the head and the outer ear demonstrate a complex dependence on the frequency of sound signals. Inter-aural intensity differences are close to zero in the median (or sagittal) plane of the head and maximal when the source is shifted to the left or to the right. Inter-aural intensity differences are frequency-dependent and reach 20 dB at frequencies above 4 kHz.

When the target speech signal and the masker are spatially separated, a part of the binaural advantage occurs in one ear in which the signal-to-noise ratio rises due to the shadow effect of the head for high frequencies. In other words, the head creates an acoustic shadow that gives the advantage to one of the ears, depending on the location of the sources of the speech signals [10].

In the last two decades, the mechanisms of binaural release from masking and their influence on the intelligibility of speech in noise have been intensively studied in the context of hearing aid technology improvement [11]. These studies allow the creation of a new generation of hearing aid technology focused on the mechanisms of bin-

Received 24 September 2018,
accepted 24 April 2019.

aural interaction. The aim of the technology is to ensure successful communication under conditions of constantly increasing real-life noise. Such technology is in demand in patients with chronic sensorineural hearing loss and age-related impairment of auditory function i.e. presbycusis, concerning hearing loss with gradually progressive inability to hear, especially for high frequency sounds [12, 13].

Target speech signals and maskers can be separated not only in azimuth, but also by distance. When the sound sources are located at different distances from a listener, the time lag of the signal from a more remote source occurs. It means that spatial separation by distance leads to temporal separation of the target signal and the masker. For comfortable communication conditions that do not require voice enhancement, the distance between talking people is from about 0.5 m to 4–5 m and the time lag can reach 15 ms. This delay is enough to perceive short signals separately. In case of longer signals the time effects are complex and could be accompanied by their mutual overlapping or echo-effects [14, 15].

At distances to point sources of sound comparable to the inter-aural distance, along with the time lag, a difference arises in the azimuthal angles by which the sound wave arrives at the right and left ears. Earlier it was shown that when the distance is varied within 1 m to the listener, inter-aural differences in the signal level change noticeably [16], which can contribute into spatial release from masking. The study of speech intelligibility thresholds under anechoic conditions was performed at distances from the listener to the sources of the speech signal and the masker is lower than 1 m [17, 18, 19]. A significant distance effect was found for speech intelligibility, especially when the masker characteristics were close to the ones of the target signal.

The perception of speech stimuli at distances of more than 1 m in free field i.e. in the absence of reflecting surfaces has not been investigated. It is believed that inter-aural differences will be too small to form a noticeable effect. At the same time, it has long been known that there is a so-called acoustic horizon - a distance of 4–5 m from the listener, after which pronounced underestimation of distance is observed compared to its overestimation at closer distances [20, 21, 22, 23]. Such differences in the auditory estimation of distance indicate the shutdown of some vital mechanism. This mechanism is involved in auditory processing of information on egocentric distance. According to our hypotheses, this is a binaural mechanism that takes part in the analysis of auditory information in spatial configurations typical for speech communication. When assessing the azimuth angle formed by a point source and two ears of the participant, it turns out that its threshold value (about 1 degree) is reached at distances that coincide with the acoustic horizon. It is possible that up to the distances on which an acoustic horizon is formed, the binaural mechanism participates in the evaluation of distance and affects the spatial release from masking. Thus, at least two factors can cause the thresholds of speech signal detection to be reduced and speech intelligibility to

be increased when the sources of the target signal and the masker are separated by distance: delay of signals in time and change of inter-aural differences by distance. The effect of spatial separation by distance up to 3 m on detection thresholds of speech signals on the background of speech-like masker has not been studied yet.

The purpose of this study was to estimate the amount of spatial release from masking when distances from the listener to the target signal sources and to the masker were in the range of 1–4 m. To create a typical communication situation, speech-like noise was used as a noise masker [2], which was produced by mixing speech of several speakers - multi-talker babble noise. Such noise created an illusion of being in a place where a lot of people speak simultaneously, so as to reproduce a typical situation for a communicative contact.

2. Method

2.1. Participants

Twenty-four native Russian-speaking participants aged 18–40 (mean age 24) were involved in the study. Most of them (21 participants) were naive listeners recruited from students of psychology courses who participated in fulfillment of a course requirement. Three participants were members of our laboratory and had experience of participation in psychoacoustic experiments. Experiments were conducted in accordance with the International Ethical Guidelines for Biomedical Research Involving Human Subjects. Hearing status of the participants was assessed using a pure tone audiometry (Biomedilen AA-02 audiometer) and a random gap-detection test [24, 25]. Successful passing of both tests was taken to indicate the absence of peripheral and central hearing impairments.

2.2. Stimuli

The stimuli were disyllabic words of the Russian language, pronounced by four native speakers whose fundamental frequencies were 117, 139, 208 and 234 Hz. Words (“pochva”, “stroka”, “rubezh”, “plata”, “vypusk”, “kredit”, “ruchka”, “nabor”) had low frequency of use in spoken language according to the new national corpus of Russian language [26]. Monophonic audio recording of words was carried out in a quiet room. The background noise level was 30 dB (A) SPL. Dynamic vocal microphone Sennheiser E845 with cardioid directional pattern and an external USB audio interface Creative E-MU 0202 connected to a desktop PC with Adobe Audition 1.5 software were used. In this way, 32 different speech files were recorded (4 speakers $\times$ 8 words).

As a speech-like masker, the result of mixing 32 audio files with duration of 10 s was applied, in each of which the same speech stimulus repeated several times successively without pauses. When listening to this speech-like masker, the participants had an auditory sensation similar to the feeling of being in a place where many people speak simultaneously. The technique for creating such babble-noise is described in detail in [2]. From the obtained multi-talker babble-noise audio file, ten random fragments with

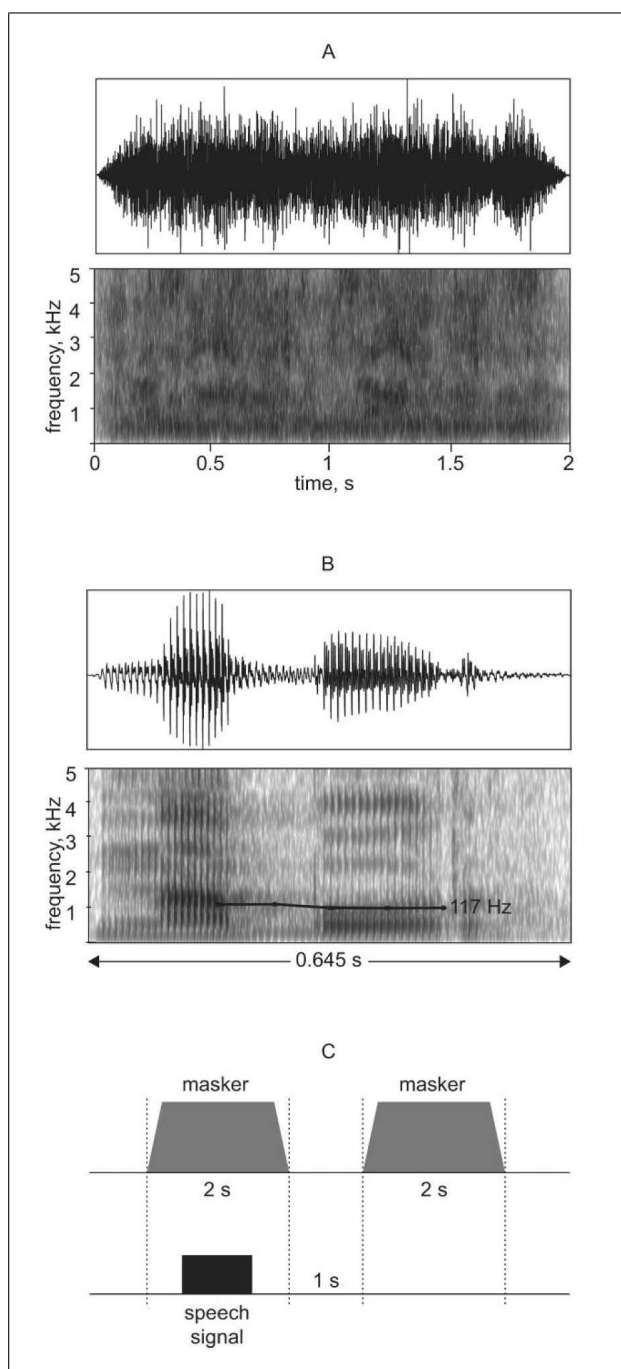


Figure 1. Examples of speech-like masking and target speech signals and a diagram of their presentation to the participant. (a) waveform (top) and spectrogram (bottom) of a speech-like noise signal used in experiments as a masker. The abscissa indicates the duration of the signal, s; the ordinate is the frequency, kHz. (b) waveform (top) and spectrogram (bottom) of the Russian word “NABOR” (in English: meaning *§. SETĬ*), pronounced by a speaker with average fundamental frequency of voice 117 Hz. (c) scheme of presentation of speech and masking speech-like stimuli in the implementation of the procedure of two-interval two-alternative forced choice. Example of location of target speech signal in the first time interval.

duration of 2 seconds were cut out. Each of the fragments had a linear rise and fall time of 100 ms. These fragments were used in experiments as masking signals. The duration

Table I. Spatial configurations of the sources of the target speech stimulus (speech) and the speech-like masker, used in different experiments.

Number of Series	Egocentric distance, m		
	1	2	4
Experiment 1 Series 1	speech	–	masker
Experiment 1 Series 2	–	speech	masker
Experiment 1 Series 3	–	–	speech masker
Experiment 2	masker	–	speech

of 32 speech stimuli was from 515 to 935 ms, so the duration of the masker was more than twice as long as their duration. Figures 1A,B show waveforms and spectrograms of a masking speech-like signal that contains fragments with a formant structure characteristic for speech and a speech signal - the word “nabor”, pronounced by a male speaker with a pitch frequency of 117 Hz.

2.3. Apparatus

Acoustic measurements and experiments to evaluate the spatial release from masking were carried out under conditions of a soundproof anechoic chamber with a volume of 62.5 m³ (5 × 5 × 2.5 m). The attenuation of external noise in the chamber was = 40 dB in the frequency range 0.5–16 kHz. Acoustic measurements were carried out using 4145 microphone, 2639 preamplifier and 2606 amplifier of Bruel and Kjaer Company in a mode of averaging the mean square power over a time interval of 1000 ms.

The sound stimuli were played back on a MicroXperts computer with an external USB audio interface Creative E-MU 0202, analog signals from which were fed to a two-channel ONKYO power amplifier, and then to a pair of SONY XSF1720 loudspeakers, located facing the subject at the level of his or her ears (0° azimuth, 0° elevation) at different distances.

2.4. Speech detection assessment

Spatial release from masking was evaluated by measuring thresholds of detecting the speech signal in babble-noise for various spatial configurations (Table I). The thresholds were estimated in the two-interval forced choice paradigm by the method of constant stimuli (Figure 1C). The threshold level was equal to 75% of correct judgments. Masking signals were presented by pairs - two intervals with 0.5 s pause. Randomly selected target speech stimulus had been appearing on the background of one of the two intervals equally likely. The level of speech stimuli was 50 dB (A) SPL and was the same in all experimental series. The level of the speech-like masker was changed to 3 dB (A) increment and was equal to 55, 58, 61 or 64 dB (A) SPL at the listening position. In other words, the signal-to-noise

ratios (SNRs) were -5 , -8 , -11 and -14 dB. Each series consisted of 40 pairs of masking signals with different SPLs. The pause between the pairs amounted 5 s.

2.5. Experiment I

The influence of speech source remoteness from the listener on signal detection in speech-like noise was studied in experiment I. Sixteen participants listened to three series of stimuli with different positions of two loudspeakers (Table I). The distance from the listener to the one loudspeaker, which was the source of the speech-like masker, was constant and equal to 4 m. In Series 1 and 2 the speech signals were emitted by the other loudspeaker placed at 1 or 2 m from the listener. In Series 3 the only loudspeaker was used for playback of both the masker and speech signals, while the other loudspeaker was removed. Therefore, the distance between sources of speech-like noise and of the speech signal, or spatial separation, was 3, 2 or 0 m (Table I). The spatial configuration of loudspeakers was changed by an experimenter manually. The order of serial presentation was random for each of the participants.

During listening to each of the experimental series, the participants were seated in an armchair so that the loudspeakers were directly in front of them at the height of their ears. Participants were not blindfolded and could see the loudspeakers. Participants listened to all the experimental series once. After listening to a pair of masking signals, the participant had to respond to the question: "What interval in the pair, first or second, contains speech signal?" by pushing one of two buttons on a control panel. Listeners' answers were recorded automatically. Series had approximately 10 minutes duration and after each of them the participant was provided with 5 minutes of rest.

2.6. Experiment II

The effect of mutual placement of the masker and target speech signal sources on detection of the latter under condition of their maximal spatial separation was investigated in Experiment II. Eight participants took part in the experiment. Speech stimuli were presented to the participants at a distance of 4 m, and speech-like noise - from a distance of 1 m (Table I). The experimental procedure, temporal structure of the experimental series and the levels of masking signals and speech signals were the same as in Experiment I. Percentages of correct judgments at different SNRs in Experiment II were compared to the data from Experiment I series 1.

3. Results

3.1. Experiment I

The percentages of correct judgments were evaluated for four SNRs (-5 , -8 , -11 , -14 dB) and for three distances between sources of the speech signal and the masking signal - 3, 2 and 0 m. The data are presented as histograms showing the distribution of participants' number by these percentages for different SNRs and distances (Figure 2). These histograms demonstrate that with decrease of SNR

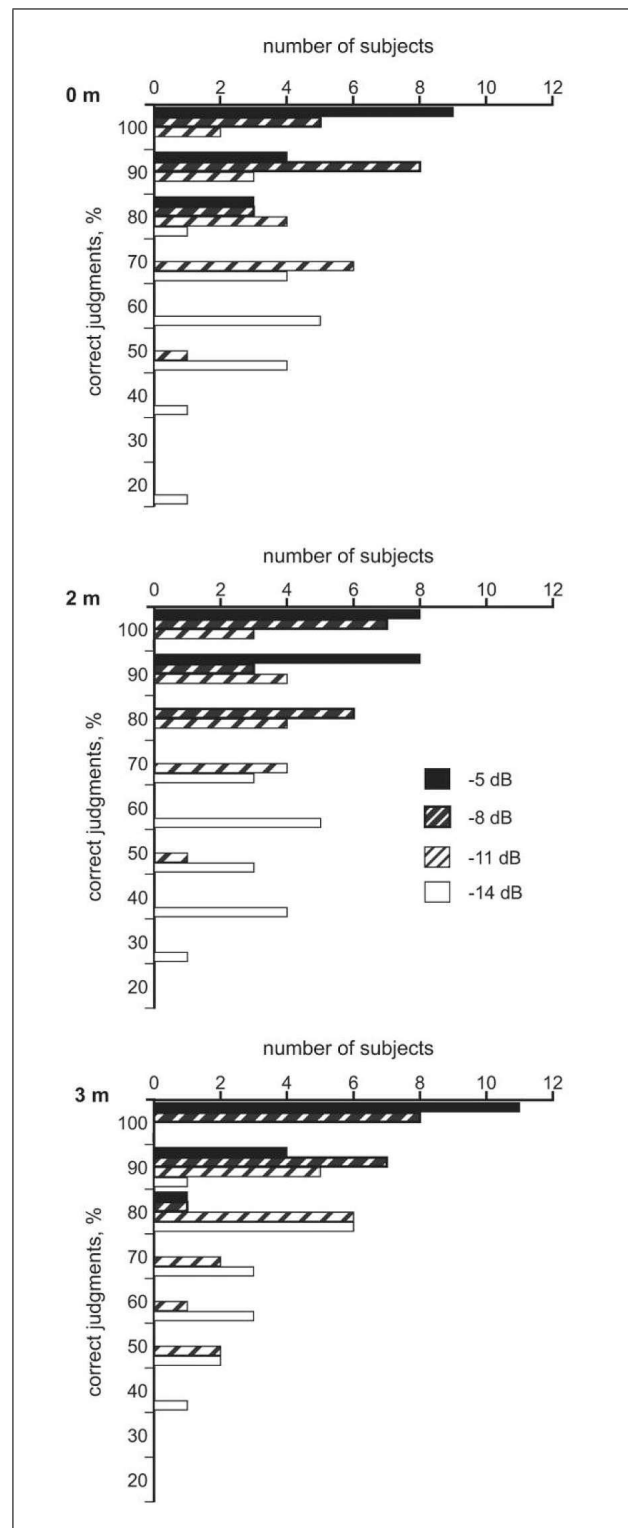


Figure 2. Distribution of the participants in the percentage of correct judgments in task of speech detection in the Experiment I. On the abscissa axis - the number of participants with a given percentages of correct judgments, along the ordinate axis - percentage of correct judgments, %. Symbols above indicate the distance between sources of speech and speech-like noise, m. The numbers on the left denote the signal-to-noise ratio, dB.

the peaks of the distribution shift towards lower percentages and the range of percentages become wider. For

−5 dB SNR the peak was observed at 100% of correct judgments for all three distances. In case of the lowest SNR (−14 dB), the peak was at 60% of correct judgments for distances of 0 and 2 m, and 80% for 3 m distances. At high SNRs (−5 and −8 dB), regardless of spatial separation, all participants detected speech signals in 80–100% of cases, which means that the detection was higher than the 75% threshold level. With SNR decreasing to −11 dB, variability of percentages became wider (50–100%), and in some participants the percentage became less than 75%. Moreover, the number of such participants varied depending on the spatial separation: it was seven listeners when the speech signal and the masker were presented from the same loudspeaker (Series 3), and five listeners when spatial separation of loudspeakers was 2 (Series 2) and 3 m (Series 1). For −14 dB SNR in cases of co-location of sources and 2 m spatial separation between them only one participant had above threshold percentage of correct judgments - 80%. On contrary, with 3 m distance between sources seven of 16 participants could detect speech stimuli in 80% or more cases. Thus, for the lowest SNR the distribution of the percentage was the narrowest for spatial separation of 3 m.

Data averaged by the group are presented by psychometric curves in Figure 3. These curves demonstrate more clearly the same changes in percentage of correct judgments with SNR as the distribution in histograms described above. Step-by-step lowering of the SNR led to a decrease in percentage of correct judgments for all three spatial configurations of sources.

When there was no spatial separation of sources of speech signals and the masker average percentages of correct answers were 94, 91, 79 and 58% for −5, −8, −11 and −14 dB SNR, respectively. No statistically significant difference in percentages of correct judgments was revealed between −8 and −5 dB cases. The differences in correct judgments between −8 and −11 dB, and −11 and −14 dB were significant ($N\text{-ties}=3$, $p=0.004$; $N\text{-ties}=1$, $p=0.001$, both one-tailed). The percentage of correct judgments was significantly higher than the threshold level of 75% for −5 and −8 dB SNR ($p<0.001$, one-sided binomial criterion), did not differ from the threshold level for −11 dB SNR ($p=0.2$, one-sided binomial criterion), and was significantly lower for −14 dB SNR ($p<0.001$, one-sided binomial criterion), i.e. with co-location of sources the threshold amounted −11 dB SNR.

For 2 m spatial separation, average percentages were 95, 91, 82 and 53% for −5, −8, −11 and −14 dB SNR, respectively. No significant difference of percentages was observed between −5 and −8 dB SNR. Percentages corresponding to the following SNRs was significantly different from each other: −8 and −11 dB ($N\text{-ties}=8$, $p=0.044$, one-tailed), and −11 and −14 dB ($N\text{-ties}=1$, $p=0.0007$, one-tailed). Percentages were significantly higher than the threshold level in case of −5 and −8 dB SNR ($p<0.001$, one-sided binomial criterion), did not differ from the threshold level for −11 dB SNR ($p=0.064$, one-sided binomial criterion) and were significantly lower for −14 dB

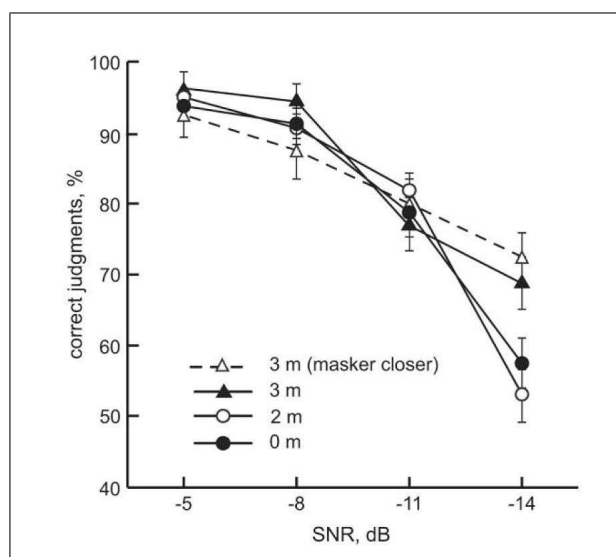


Figure 3. The total percentages of correct judgements in task of speech stimuli detection on the background of a speech-like masker for groups of participants at different signal-to-noise ratios. The abscissa is the signal-to-noise ratio, dB, and the ordinate is the percentage of correct judgements,%. The legend indicates data from Experiment I (solid lines, $N=160$) and Experiment II (dashed line, $N=80$).

SNR ($p<0.001$, one-sided binomial criterion). Thus, the threshold in the case of 2 m spatial separation was the same as for co-location and amounted −11 dB SNR.

In the case of 3 m spatial separation average percentages of correct judgments amounted 96, 94, 77 and 69% for −5, −8, −11 and −14 dB SNR, correspondingly. The difference between percentages for −8 and −11 dB were significant (Wilcoxon matched pairs test, $N\text{-ties}=4$, $p=0.002$, one-tailed), while for −5 and −8 dB, or −11 and −14 dB pairs, no significant difference was revealed. The percentages for −5 and −8 dB SNRs were significantly greater than threshold level ($p<0.001$, one-sided binomial criterion). The percentage did not differ from the threshold level when SNRs amounted −11 and −14 dB ($p=0.34$ and $p=0.12$, one-sided binomial criterion), i.e. the threshold for 3 m spatial separation was between −11 and −14 dB SNR. Thus, step-by-step lowering of SNR resulted in a steady decrease of percentage of correct judgments in all three spatial configuration. The percentage reliably dropped to sub threshold level for −14 dB SNR only in cases of 0 and 2 m spatial separation of sources. The percentage still did not differ from the threshold level when separation amounted 3 m.

The following comparison of data obtained for different spatial configurations with the same SNR makes the observed differences more noticeable. In case of −14 dB SNR the percentage of correct judgments was significantly higher for 3 m separation than for 2 and 0 m (3 and 2 m: $N\text{-ties}=3$, $p=0.007$; 3 and 0 m: $N\text{-ties}=5$, $p=0.016$, both one-tailed). The improvement of speech signal detection for −14 dB SNR amounted to 11% when distance between speech and noise sources increased from 0 to 3 m. Im-

provement of speech signal detection observed between conditions at 2 and 3 m separation amounted to 15%. There was no statistically significant difference in speech detection between conditions of 0 and 2 m separation of masker and speech signal sources ($N\text{-ties}=3$, $p=0.5$, one-tailed). No differences were found between percentages in the case of employed spatial separations for -11 , -8 or -5 dB SNRs.

To assess the role of spatial separation and the signal-to-noise ratio in detection of speech stimuli against a speech-like noise background, we used non-parametric Friedman 4×3 ANOVA for comparing multiple dependent samples. In this case, the percentage of correct judgments was used as a dependent variable, and SNR (-5 , -8 , -11 , and -14 dB) and spatial separation (0, 2 and 3 m) were utilized as independent variables. The Friedman test showed that the SNR dramatically influenced speech detection in the entire range of employed distances between the sources of speech and masking signals, 0.2 (3, $N=48$, $p<0.001$). Mean speech detection percentages were 95, 92, 79 and 60% for SNRs of -5 , -8 , -11 and -14 dB, respectively. Thus, in case of -14 dB SNR the detection percentage was lower than 75% threshold. The distance between sources of speech and noise signals was also found to be significant for detection percentage of the former, 0.2 (2, $N=64$, $p<0.05$).

Mean detection percentage amounted 80% for 0 m, and 90% for 2 and 3 m separations of the sources, i.e. distance parameter significantly affected speech signal detection. The data indicate that spatial separation of sources of the masker and the speech signal by distance improves speech detection when SNRs are low. For -11 dB SNR percentage did not differ from the threshold level in all three spatial configurations. In case of -14 dB differences between detection percentage and threshold level were not found only for 3 m separation. It means that in the latter case the threshold was between -11 and -14 dB, i.e. spatial release from masking was less than 3 dB SNR. A second experiment was performed to test the hypothesis that the observed improvement of speech detection under condition of near threshold values of SNR was due to the spatial separation of their sources, but not due to their mutual location relative to the listener.

3.2. Experiment II

In Experiment II the distance from the listener to the source of the target speech signal was 4 m, and the source of the speech-like noise was placed at 1 m away from the listener, while using the same set of SNRs at the place of listening that was used in Experiment I. The histograms, which show the distribution of participants' number by percentages of correct judgments for different SNRs, are shown in Figure 4. The peaks of distributions shifted with a decrease of SNR towards lower percentages, and the range of percentages became wider, as in Experiment I (Figure 2). The distribution peaks for the case of Experiment II (spatial configuration) were slightly moved towards lower percentages, in comparison with the corre-

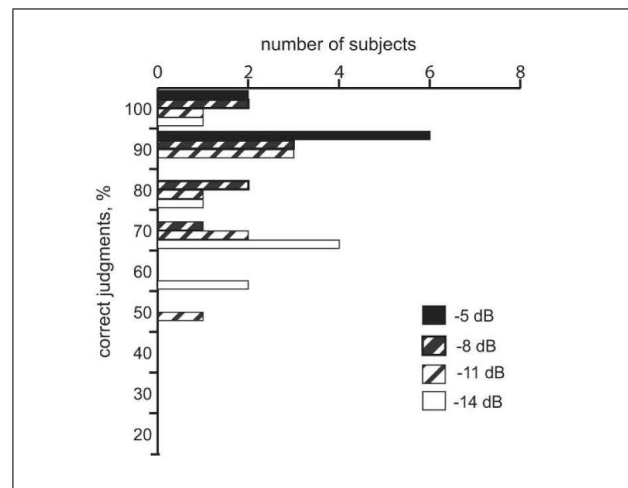


Figure 4. Distribution of the participants by the percentage of correct judgments in task of speech detection in the Experiment II. On the abscissa axis - the number of participants with a given percentages of correct judgments, along the ordinate axis - the percentage of correct judgments, %. The numbers on the right indicate the signal-to-noise ratio, dB.

sponding peaks for the case of Experiment I, series 1. For SNRs of -5 and -8 dB, all participants except one detected speech signals with above-threshold levels, one participant showed 70% of correct judgments in the case of -8 dB SNR. For -11 dB SNR five of eight participants had above-threshold percentage of correct judgments. At a near threshold SNR (-14 dB), two participants detected target signals at 80% or higher, half of the participants showed 70% detection level, and two demonstrated 60% detection level.

The data averaged by group are presented in Figure 3. The percentage of correct judgments decreased with decreasing SNR: 93, 88, 80 and 73% for -5 , -8 , -11 and -14 dB SNRs, respectively. A statistically significant decrease of detection percentage was found only when comparing percentages of correct judgments at -8 and -14 dB SNR ($N\text{-ties}=2$, $p=0.027$, one-tailed), and at -5 and -14 dB SNRs ($N\text{-ties}=1$, $p=0.017$, one-tailed). In cases of -5 and -8 dB SNRs the percentages were above the threshold level ($p<0.01$ and $p<0.05$, respectively, one-sided binomial criterion), while for -11 and -14 dB SNRs the percentage did not differ from the threshold level ($p=0.22$ and $p=0.39$, one-sided binomial criterion). Thus, speech detection threshold was between -11 and -14 dB SNRs similar to results of Experiment I, series 1. This means that in both cases the release from masking in comparison with the situation of co-location of sources was observed and its value was less than 3 dB SNR.

Differences of percentages were evaluated for the cases of maximal spatial separation and both spatial configurations of speech and speech-like noise sources (Experiment II and Experiment I, series 1). No differences were observed, at any SNR, between the two spatial configurations (Mann-Whitney test for independent samples). As shown above, in both configurations the percentages did not differ

significantly from 75% threshold value in case of -11 dB SNR and -14 dB SNRs, and were higher than the threshold level for -5 dB SNR and -8 dB SNR, i.e. percentages of speech signals detection on the background of a speech-like masker did not depend on which source was near to the listener.

4. Discussion

Spatial separation of the sources of the target speech signal and the speech-like masker increased the detection percentage at the lowest signal-to-noise ratio. This indicated the existence of release from masking when the sources of the target signal and the masker are separated by a distance when they are more than 1 m from the listener. The release from masking did not depend on which source was closer to the listener, at equal ratios of the intensity levels of the target signal and the masker. This symmetrical situation, when sources are placed directly in front of a subject at different distances, allows one to assume the possibility of auditory spatial tuning to egocentric distance characterized by a certain opening angle, under which the sound wave arrives to the left and right ears of the participant.

It is known that for azimuth angles even a separation of 4–10 degrees is enough to achieve significant improvement of performance in detection and speech discrimination tasks [4]. In case of maximal spatial separation by distance used in our study (3 m) the opening angle difference for the speech signal and the masking signal sources was about 8 degrees. Decrease of distance between the sources led to a reduced difference in their opening angles. This could be the reason why we did not observe any significant improvement of speech signals detection in the case of 2 m separation of sources compared to the situation of their co-location.

Thus, the experiments showed the efficiency of separating sources (at 0 degrees azimuth) of speech-like noise and speech by distance for the task of detecting speech stimuli. In the scientific literature, we have not found any work in which the thresholds of detection of speech signals at different distances between the target speech signal and the masker were performed. Therefore, a comparison can be made for the parameters of release from masking in the case of speech discrimination tasks. Release from masking at distances greater than 1 m was observed when simulating the frontal location of speech sources and the masker in reverberant spaces where the ratio of the direct wave energy and reverberant energy is an important cue for distance estimation [27]. This research was carried out using virtual reality technology based on binaural room impulse responses. Reproduction of the location of sound sources in acoustic virtual reality technology was achieved by interpolating a set of head-related transfer functions that were previously measured for a set of source positions [28]. Release from masking was estimated for 0.5 m distance to the target speech source, with the co-location of the masker and with a fourfold increase of the masker distance – up to 2 m.

Release from masking in these conditions amounted 3 dB on an average data [27]. It should be noted that with a further increase in the distance to the masker (5 and 10 m), no additional release from masking was observed. The authors also report an increase in the individual scatter of data with increasing distance, which may be caused by the acoustic horizon effect described in the introduction. Thus, despite the significant differences in the method of presentation of stimuli and in the evaluation of thresholds for detecting and intelligibility of the speech signal, the data of the abovementioned work confirm our result on the presence of release from masking when the sources are separated by distance within the limits of acoustic horizon.

The selection of stimuli used in our study was determined by the need to create a Russian version of the Coordinate Response Method (CRM), which is widely used for investigations on the release from masking of target speech signals [29, 30]. The present experimental work is the first step to creation of a Russian-language version of the CRM that will allow us to compare the release from masking effects in subsequent experiments to study the effect of separation over the distance of the target speech source and the masker on speech intelligibility. Since the speech signal has information redundancy, it can be assumed that in the problem of speech intelligibility, the separation efficiency by distance will be higher than that found in this study.

Acknowledgement

The work is supported by means of the state budget on the state assignment for 2013–2020 years (themes no. AAAA-A18-118013090245-6 in the Sechenov Institute of evolutionary physiology and biochemistry of Russian Academy of Sciences and no. 0134-2014-0002 in the Pavlov Institute of physiology of Russian Academy of Sciences).

References

- [1] J. C. R. Licklider: The influence of interaural phase relations upon the masking of speech by white noise. *J Acoust Soc Am* **20** (1948) 150–159.
- [2] V. Best, E.R. Thompson, C.R. Mason, G. J. Kidd: An energetic limit on spatial release from masking. *J Assoc Res Otolaryngol* **14** (2013) 603–610.
- [3] P. Bremen, J. C. Middlebrooks: Weighting of Spatial and Spectro-Temporal Cues for Auditory Scene Analysis by Human Listeners. *PLoS ONE* **8**(3) (2013) e59815.
- [4] A. W. Bronkhorst: The cocktail-party problem revisited: early processing and selection of multi-talker speech. *Atten Percept Psychophys* **77**(5) (2015) 1465–1487.
- [5] A. W. Bronkhorst: The cocktail party phenomenon: A review of research on speech intelligibility in multiple-talker conditions. *Acta Acust united Ac* **86** (2000) 117–128.
- [6] P. M. Zurek: Binaural advantages and directional effects in speech intelligibility. *Acoustical Factors Affecting Hearing Aid Performance*. Allyn and Bacon, Boston, 1993.
- [7] J. Blauert: *Spatial Hearing: The Psychophysics of Human Sound Localization*. MIT Press, Cambridge, 1997.
- [8] M. L. Hawley, R. Y. Litovsky, J. F. Culling: The benefit of binaural hearing in a cocktail party: Effect of location and type of interferer. *J Acoust Soc Am* **115** (2004) 833–843.

- [9] R. Y. Litovsky: Spatial release from masking. *Acoustics today* **8** (2012) 18–24.
- [10] A. W. Bronkhorst, R. Plomp: The effect of head-induced inter-aural time and level differences on speech intelligibility in noise. *J Acoust Soc Am* **83**(4) (1988) 1508–1516.
- [11] N. Marrone, C. R. Mason, Jr. G. Kidd: Evaluating the Benefit of Hearing Aids in Solving the Cocktail Party Problem. *Trends Amplif* **12** (2008) 300–315.
- [12] M. Yu. Boboshko, L. E. Golovanova, E. V. Zhilinskaya, E.A. Ogorodnikova: Speech intelligibility in elderly hearing impairment people. *Adv Gerontol* **29** (4) (2016) 663–669.
- [13] M. Yu. Boboshko, L. E. Golovanova, E. V. Zhilinskaya, E. A. Ogorodnikova: The effectiveness of hearing aids in elderly people. *Adv Gerontol* **30**(1) (2017) 114–120.
- [14] M. Iu. Agaeva: Echo thresholds of the precedence effect in the vertical plane. *Zh Vyssh Nerv Deiat Im. I. P. Pavlova*. **58**(4) (2008) 443–448.
- [15] M. Iu. Agaeva: Buildup of echo suppression for the high frequency signal // *Zh Vyssh Nerv Deiat Im. I. P. Pavlova*. **60** (4) (2010) 303–308.
- [16] D. S. Brungart, W. M. Rabinowitz: Auditory localization of nearby sources. Head-related transfer functions. *J Acoust Soc Am* **106**(3) 1999 1465–1479.
- [17] B. G. Shinn-Cunningham, J. Schickler, N. Kopco, R. Litovsky: Spatial unmasking of nearby speech sources in a simulated anechoic environment. *J Acoust Soc Am* **110** (2) (2001) 1118–1129.
- [18] J. F. Culling, M. L. Hawley, R. Y. Litovsky: The role of head-induced inter-aural time and level differences in the speech reception threshold for multiple interfering sound sources. *J Acoust Soc Am* **116**(2) (2004) 1057–1065.
- [19] D. S. Brungart, B. D. Simpson: Effect of target-masker similarity on across-ear interference in a dichotic cocktail-party listening task. *J Acoust Soc Am* **122**(3) (2007) 1724–1734.
- [20] M. B. Gardner: Distance estimation of 0° or apparent 0°-oriented speech signals in anechoic space. *J Acoust Soc Am* **45**(1) (1969) 47–53.
- [21] W. Simpson, L. D. Stanton: Head movement does not facilitate perception of the distance of a source of sound. *Am J Psychol* **86** (1) (1973) 151–159.
- [22] G. Kearney, M. Gorzel, H. Rice, F. Boland: Distance perception in interactive virtual acoustic environments using first and higher order ambisonic sound fields. *Acta Acust united Ac* **98**(1) (2012) 61–71.
- [23] G. Parsehian, C. Jouffrais, B. F. Katz: Reaching nearby sources: Comparison between real and virtual sound and visual targets. *Front Neurosci.* **8** (2014) 269.
- [24] R. W. Keith: Random Gap Detection Test. Auditec of St. Louis, St. Louis, 2000.
- [25] F. E. Musiek, G. D. Chermak: Handbook of central auditory processing disorder .Vol.1: Auditory neuroscience and diagnosis. 2nd edition. Plural Publishing, San Diego, 2014.
- [26] O. N. Lyashevskaya, S. A. Sharov: Chastotnyy slovar' sovremennogo russkogo yazyka (na materialakh natsional'nogo korpusa russkogo yazyka). Azbukovnik, Moscow, 2009. (in Russian)
- [27] A. Westermann, J. M. Buchholz: The influence of informational masking in reverberant, multi-talker environment. *J Acoust Soc Am* **138**(2) 2015 584–593.
- [28] F. Wightman, D. Kistler: Measurement and validation of human HRTFs for use in hearing research. *Acta Acust united Ac.* **91**(3) (2005) 429–439.
- [29] N. K. Srinivasan, K. M. Jakien, F. J. Gallun: Release from masking for small spatial separations: Effects of age and hearing loss. *J Acoust Soc Am* **140**(1) (2016). EL73–EL78.
- [30] K. M. Jakien, S. D. Kampel, M. M. Stansell, F. J. Gallun: Validating a Rapid, Automated Test of Spatial Release From Masking. *Am J Audiol* **21** (2017) 1–12.