

SPEECH INTELLIGIBILITY IN VARIOUS SPATIAL CONFIGURATIONS OF BACKGROUND NOISE

J. KOCIŃSKI, A. SĘK

Adam Mickiewicz University
Institute of Acoustics
Umultowska 85, 61-614 Poznań, Poland
e-mail: jen@amu.edu.pl, oleksek@amu.edu.pl

(received May 5, 2004; accepted January 27, 2005)

This study is concerned with the influence of spatial separation of disturbing sources of noise on the speech intelligibility. Spatial separation of speech and disturbing sources without changing their acoustic power may contribute to the significant improvement in the speech intelligibility. This problem has been recently analysed in many papers [1–5]. These works have confirmed an important role of the spatial configuration of sources. However, there have been no work investigating this problem for nonsense words (logatoms) that may provide more rigorous tests of this phenomenon. Moreover, this problem has not been analysed for Polish speech. It is important to emphasize that the acoustic and phonetic properties of Polish speech are somewhat different from those of English one. Therefore, the attempt to investigate the influence of the spatial separation of sound sources was made in this study. In the situation with more than one spatially separated disturbances, there may occur a so-called spatial suppression phenomenon, that is “mutual suppression” of disturbing sounds in the auditory system that brings about an increase in the speech intelligibility. This phenomenon is also called the spatial unmasking of speech [5, 6].

The research consist in determination of the speech intelligibility in the presence of one or two statistically independent speech-shaped noise sources varying in configuration. Only two pairs of the spatial configurations were investigated.

Character of the dependences obtained in the study implies that the spatial suppression occurs in certain configurations of sources only. This effect brings about an increase in the speech intelligibility and can be explained on the basis of the binaural masking level difference (BMLD). It seems then that the BMLD may be a more general phenomenon and includes not only difference in the detection threshold of a pure tone masked by noise but also the improvement in the speech intelligibility, while speech is presented at the background of disturbing signals.

Key words: speech intelligibility, speech perception, spatial separation, spatial suppression.

1. Introduction

Natural environments typically contain sound sources other than the sources of interest (e.g. speech) that may interfere with the ability of the listeners to extract information about the target source. Therefore, a communication becomes often difficult, because the target sounds are masked by the coexisting disturbances. It is necessary to keep in mind that the auditory system is capable of separating out the incoming signals (coming from different directions) to some extent and to extract the desired information. This phenomenon is often called the “cocktail party effect”.

Many works were related to the influence of the binaural masking level difference (BMLD) on speech reception threshold (SRT) or to the spatial unmasking of speech [5, 7–10]. It was revealed that the SRT strongly depended on the mutual spatial configuration of the speech and the masker sources preserving the same SNR [2, 11–13]. The results of the works mentioned above proved that the spatial separation of sources has a very strong influence on the improvement of the speech intelligibility.

Furthermore, in the case where there exist more than one disturbing source, the so-called spatial suppression may occur, that is the “mutual suppression” of disturbances in the auditory system that brings about an increase in the speech intelligibility [1, 2, 5, 6]. This phenomenon is also called the spatial unmasking of speech or the release from masking. It should be emphasized that this effect is only a psychological phenomenon and has no connection with the physical suppression or mutual attenuation. It means that two (or more) spatially separated sources of noise are often less effective maskers than one noise at the same sound pressure level.

Beside the spatial suppression, the present study implies also the opposite effect: two spatially separated disturbances can sum up in the way that they become more effective maskers than one disturbance with the same level.

So far, most of the papers dealing with the spatial suppression phenomenon were based on the use of the word or short sentence tests. These types of tests are often used in the speech intelligibility analysis for normally hearing and hearing-impaired people [14] due to the redundancy, that characterizes these types of utterances. The acoustic waves of the speech signals conveying any semantic information, contain much more information than it is necessary for the utterances to be understood and repeated correctly. This “information overflow”, however, is effectively used by all the subjects.

As described by BOCCA and CALEARO [15], the subjects are able to perceive speech due, in part, to the extrinsic redundancy within the speech signal and the intrinsic redundancy within the auditory system. The extrinsic redundancy refers to the numerous overlapping cues within the speech itself; the intrinsic redundancy, on the other hand, refers to the multiple central auditory nervous system pathways and the sources of information that the human system possesses for the speech processing. It is because of these redundant cues that the message becomes somehow predictable, and it is this predictability which makes the transmission of information powerful.

Logatoms, i.e. nonsense syllables or words, are characterized by much lower redundancy, since in a given language, they have no meaning and they do not convey

any semantic information. Understanding and correct repetition of a logatom is possible only in the case when a subject can hear all successive phonemes occurring in the logatom. So the knowledge of the language rules and reach vocabulary or mental ability do not help markedly in performing the logatom intelligibility tests. Thus, it seems that the logatom tests provide the most robust and the most objective measure of the speech intelligibility and thus they are more often used than the other test types, especially when the speech transformed in some way is analyzed or presented at the background of any disturbing (masking) sounds.

Due to the above-mentioned reasons, in testing of the spatial suppression phenomenon we decided to use logatoms rather than short sentences or words. In the present study, a phonetically balanced Polish low-redundant logatom tests were used as the speech signals [16] to prevent the listeners from guessing or predicting linguistic units presented during auditory tests. It seems that the use of the logatoms provides a more robust, rigorous and objective test of the spatial suppression.

The spatial suppression effect is often explained on the basis of the binaural masking level difference (BMLD). Signal phase inversion in one ear may cause the decrease in the signal detection threshold. Furthermore, if a tone and a masker are delivered to one ear only, adding a noise to the opposite ear also decreases the detection threshold [17]. A general conclusion from the research on BMLD is that the interaural phase difference (IPD) is very significant for the speech intelligibility. LICKLIDER [7] revealed that in the case when the signal phase is inverted, the speech intelligibility increases by about 25% comparing to the situation with no phase shifts. DURLACH *et al.* [18] revealed that delivering uncorrelated maskers to each ear brings about an increase in the intelligibility in comparison to the situation when the same signal was delivered to both ears.

On the basis of the previous works on this topic [19, 20] it was implied that the speech understanding ability in the noisy environment is the lowest when the speech and the disturbance are coming from the same place (direction) in space and, generally, it increases with the increase in spatial separation of the sources. Therefore, the BMLD can be generalized. It is not only the problem of signal detection threshold changes but also the speech intelligibility changes in the case when phases or sound pressure levels of the masked or masking signal in both ears are different.

2. Experiment

2.1. Aim

Among all the works concerned with the analysis of the influence of the spatial separation of sources on the speech reception threshold or the signal-to-noise ratio enhancement, there is no work in which nonsense word tests were used. Moreover, there is no work concerned with this topic related to Polish speech. Polish language is somewhat different from the Western European ones and is characterized by a relatively substantial amount of noise sounds (fricatives) [21, 22]. Therefore, in the present study the attempt to investigate the influence of spatial separation of two sources on the intelligibility of

logatoms was proceeded, as it may provide much more robust test to the spatial suppression phenomenon. However, in the preliminary experiment we have also used the word tests. It was done to compare the changes in the intelligibility determined for different configurations of disturbing sound sources for both the word and logatom tests. We did not intend to formulate the universal conclusions concerning the spatial suppression. We have rather concentrated on showing that the spatial suppression is not attributed to the utterances conveying semantic information only, but it is a more general phenomenon that shows the ability of the auditory system to extract useful information from a meaningless acoustic waveform. Hence, only three normally hearing subjects participated in the experiments carried out, as it is very common in psychoacoustical research.

The main aim of this study was to investigate whether the spatial suppression occurred resulting in the increase in the speech intelligibility of nonsense words when two disturbing sources of noise were spatially separated. If the intelligibility of speech were improved when two spatially separated disturbing sources were used compared to the situation with one disturbing noise only, it would imply the existence of spatial unmasking of the speech that is spatial masking suppression. Comparison of the intelligibility for all the investigated cases (four spatial configurations) may enable to find the configuration with the highest intelligibility and refer the results to the spatial suppression and the BMLD assumptions.

2.2. Stimuli

The research consisted in recording of the test material in an anechoic chamber using a dummy head. Target speech stimuli were Polish numerical and one-syllable word tests [23] and nonsense word (logatom) tests [16]. All the tests were phonetically and structurally balanced. The interference sounds were either one or two sources of noise presented from the loudspeaker. Three subjects aged 20–25 with audiologically normal hearing were asked to listen to the tests and write down all words (or logatoms) that they heard out.

There were two stages of the research. In the first one (Experiment 1), the Polish numerical and one-syllable word tests were used as the target speech. For these tests the signal-to-noise ratio was adjusted to: 0, -3, -6, -9 and -12 dB. In the second one, Polish non-sense-word (logatom) tests were used and SNR was adjusted to: 3, 0, -3, -6, -9 dB, as there is a marked difference in the redundancy between numerical tests, one-syllable words and logatoms. Therefore different signal-to-noise ratios (SNR) were used.

There were ten one-syllable word tests with 20 words in each. There were also ten numerical tests, but with ten numerals only. Each of the logatom tests consisted of 300 elements.

Due to a marked difference in redundancy we have expected that the comparison of the intelligibility of one-syllable words and logatoms may reveal different factors that influence the spatial unmasking.

In natural environments the most common masker of the speech is another speech. Therefore, one or two sources of speech-shaped noises at the constant overall sound pressure level were chosen as disturbance sounds. The long-term spectrum of such a noise and a real speech spectrum are identical. The only difference is that noise does not convey any information. The speech-shaped noise spectrum is approximately constant up to the frequency of 500 Hz and above this frequency the Power Spectral Density (PSD) decreases by 12 dB per octave [24, 25].

It is important to keep in mind that sound pressure levels of one noise and that produced by two sources of noise in a checkpoint (just above the dummy head) were identical and were adjusted to 75 dB SPL. It implies that the sound pressure level of each of the two sources of noise appearing together was 3 dB lower than that in the configuration with one noise only.

The target speech was always presented from the loudspeaker placed directly in front of the dummy head (0°) because in real situations the interlocutor most frequently is situated in front of the listener. Such a configuration made the speech signals in both ears identical since there was no interaural phase difference nor interaural level difference.

2.3. The intelligibility determination

The number of correct answers to the number of all elements in the test ratio was taken as a measure of the intelligibility. The speech intelligibility was determined as the percentage of correct answers according to the equation (1):

$$I = \frac{p}{n} \cdot 100\%, \quad (1)$$

where: I – intelligibility, %; p – number of correct answers; n – number of all elements of a test.

Any mistake made by the subject caused that the subject's answer was treated as an incorrect one. Only the spelling mistakes (h-ch, u-ó, ź-rz) were not taken into consideration.

2.4. Apparatus

All recordings were carried out in the anechoic chamber using Tucker–Davies Technology (TDT), System 3 device at the sampling rate of 24414.0625 Hz and the resolution of 24 bits. One speech-shaped noise was generated in a real-time by the processor TDT-RP2 (channel 2), while the second one was played using a digital recorder Fostex D824 (in Experiment 1 another TDT-RP2 was used instead of Fostex D824). Both sources of noise were statistically independent. Next, the signals were amplified using PIONEER A-505R to the level of 75 dB SPL and delivered to the three-way loudspeakers ZG-60 placed in the anechoic chamber.

The target speech signals (numerical, one-syllable word and logatom tests) previously stored on the hard drive of the PC as a 24 bit binary files, were fed to the processor TDT-RP2 used as a D/A converter (channel 1). Next the speech signal was fed to the programmable attenuator, TDT-PA5, enabling the adjustment of SNR and amplified by the SONY STR-DE475 amplifier to the level of 78 dB SPL. Then the signal was delivered to the three-way loudspeaker ZG-60 placed in the anechoic chamber. The proper SNR was adjusted using programmable attenuator (TDT-PA5).

The generated signals were recorded using the dummy head, NEUMAN (separate channels for each ear) connected with two microphone amplifiers, TDT-MA2. Next both signals were fed to two inputs of the TDT-RP2 used as a A/D converter, delivered to the PC and saved on the hard drive as 24-bit binary files (one file for each ear).

During the listening sessions the signals recorded in the above described way were presented to the subjects in double-walled, acoustically isolated chambers. The signals stored on the hard drive were fed to the TDT-RP2 processor and then amplified in the headphone buffer, TDT-HB7, to the level of 75 dB SPL at the tympanic membrane. Next, the signals were delivered to the Sennheiser HDA580 headphones and presented binaurally to the subjects. All the recordings and presentations were carried out using MatLab 5.3 computing language.

The subject's task was to write down all the heard words, numbers or logatoms in a specially prepared form. Next, the intelligibility for all four spatial configurations was determined.

To determine whether the spatial suppression occurred, the speech intelligibility in two pairs of the spatial configurations were compared. If the intelligibility in the case with two sources of noise were higher than that with one noise only, it could imply the occurrence of the spatial suppression effect.

2.5. Spatial configuration of the sound sources in the anechoic chamber

In each of the spatial configurations of the sound sources used in the experiments, the speech (target) signal, S , was always presented from a loudspeaker placed directly in front of the dummy head (0°) and 3 m away from it. The dummy head was at a height of 1.5 m. The noise sources were also 3 m away from the dummy head, but they were varying in number (one or two) and in spatial configuration. The azimuth angle of the first noise source, $N^{(1)}$, varied (0° or -60° clockwise) while the azimuth of the second noise, $N^{(2)}$, was fixed at 45° (see Fig. 1 and 2).

The notation of the different configurations is as follows: the upper index stands for the number of the noise source, whereas the lower one stands for the azimuth of the noise source (e.g. $SN_{-60}^{(1)}$ stands for the target coming from the front, and one noise source at the azimuth of -60° , whereas $SN_0^{(1)}N_{45}^{(2)}$ stands for the signal coming from the front and two noise sources coming from 0° and 45° respectively).

Two pairs of spatial configurations of the source(s) of noise were used i.e. $\left\{ SN_0^{(1)}, SN_0^{(1)}N_{45}^{(2)} \right\}$ and $\left\{ SN_{-60}^{(1)}, SN_{-60}^{(1)}N_{45}^{(2)} \right\}$. These configurations are schematically depicted in Fig. 1 and 2.

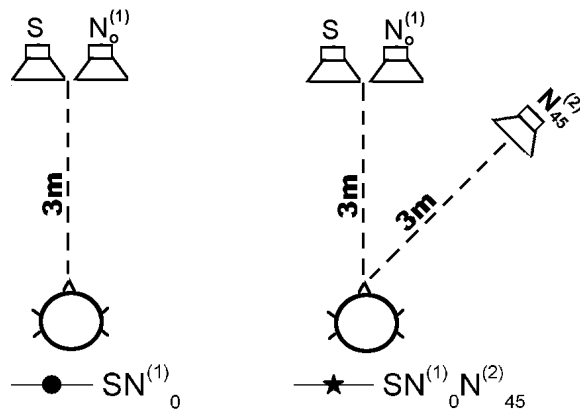


Fig. 1. Spatial configuration of the first pair of the compared situations. Filled circle and asterisk with solid lines refer to the notation in Fig. 3 and 4.

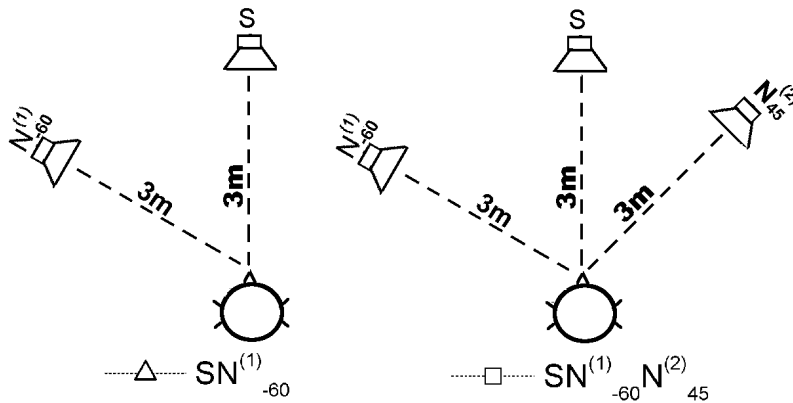


Fig. 2. Spatial configuration of the second pair of the compared situations. Empty triangle and square with dashed lines refer to the notation in Fig. 3 and 4.

2.6. Experiment 1: The numerical and the one-syllable word tests intelligibility masked by one or two disturbing sources of noise

In this experiment phonetically balanced one-syllable word tests and numerical tests were used [23]. It was a preliminary experiment in a sense and was carried out to investigate whether the spatial suppression effect is noticeable for spatially separated disturbing sources.

It must be emphasized that generally, the numerical tests intelligibility was much higher than one-syllable word tests. It is fully consistent with the results presented by other authors [26]. The intelligibility of the numerical tests was close to 100% and decreased to about 80% for the lowest SNRs only. For this reason the analysis of the intelligibility changes was carried out for one-syllable word tests only.

The results of this part of the study, i.e. the speech intelligibility as a function of the SNR for different spatial configurations of noise sources ($SN_0^{(1)}$, $SN_0^{(1)}N_{45}^{(2)}$, $SN_{-60}^{(1)}$ and $SN_{-60}^{(1)}N_{45}^{(2)}$), are depicted in Fig. 3. Successive panels of the figure show the data gathered for individual subjects and the last panel depicts the average for three subjects. All results for both pairs of the compared spatial configurations i.e. $SN_0^{(1)}$ with $SN_0^{(1)}N_{45}^{(2)}$ (circles and asterisks, respectively, connected with solid lines) and $SN_{-60}^{(1)}$ with $SN_{-60}^{(1)}N_{45}^{(2)}$ (triangles and squares, respectively, connected with dashed lines) are presented there. In spite of some intersubject scatter, the results for all subjects are qualitatively consistent with each other. The pattern of the dependences obtained in the study and mutual relations of the intelligibility for all spatial configurations are analogous. Generally, regardless of the spatial configuration, the significant increase in the intelligibility along the increase in the SNR is noticeable. This is fully consistent with the results presented by other authors (e.g. [7, 26]).

Let us analyse the situation in which the speech signal (always from the front) was masked by one noise only, i.e. the configurations $SN_0^{(1)}$ (circles) and $SN_{-60}^{(1)}$ (triangles). As it can be noticed in Fig. 3, the noise coming from the same direction as the speech signal is a very effective masker, since the speech intelligibility reaches the lowest values for this case. For the $SN_{-60}^{(1)}$ (the noise was placed at the azimuth of -60°), the speech intelligibility values are much higher. They were actually the highest values gathered in this experiment. Although the level of the masker in both configurations was the same, the masker placed at -60° was significantly less efficient. It is worth to emphasize that the difference in speech intelligibility for both configurations is not constant as a function of signal-to-noise ratio and reaches the largest values (up to 40%) for the lowest SNRs. These experimental findings seem to be fully consistent with the binaural masking level difference (BMLD) and are quite easy to explain on the basis of this phenomenon. In the $SN_0^{(1)}$ case there is no interaural phase nor level difference between the signals reaching the left and the right ear, so the intelligibility is the worst. However, in the $SN_{-60}^{(1)}$ case, the time courses of the noise in the left and in the right ear are different (delayed relatively to each other) and have different levels as a result of a significant acoustical shadow of the head, while the speech signals in both ears are identical. Thus, it is a typical situation in which the BMLD occurs contributing to the enhancement in SNR and as a consequence, in the speech intelligibility.

The data gathered in these two spatial configurations were subjected to a within-subject analysis of variance (ANOVA). The influence of the signal-to-noise ratio and the spatial configuration of the maskers on the speech intelligibility were tested. As it was expected, the influence of the SNR was proved to be statistically significant [$F(4, 8) = 163.22$, $p < 0.001$]. The spatial configuration (i.e. $SN_0^{(1)}$ or $SN_{-60}^{(1)}$) was also proved to be statistically significant: [$F(1, 2) = 25.44$, $p = 0.037$]. The interaction between these two factors was proved to be marginally statistically significant [$F(4, 8) = 3.19$, $p = 0.076$]. This implies that the difference in the intelligibility in both spatial configurations was only slightly dependent on SNR.

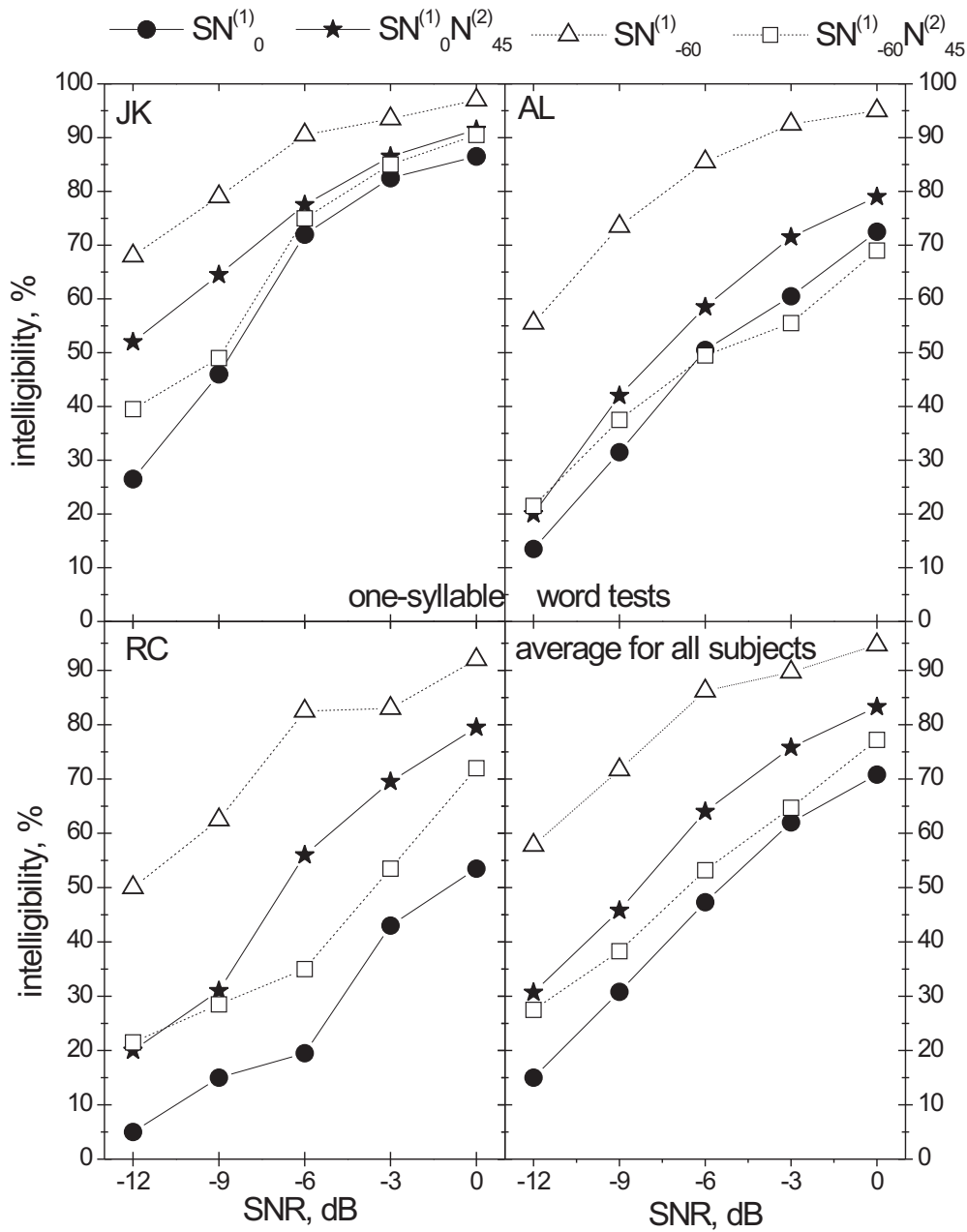


Fig. 3. One-syllable word tests intelligibility as a function of signal-to-noise ratio (SNR) for all four spatial configurations. Successive panels of the figure depict data gathered for individual subjects. $SN_0^{(1)}$ (circles connected with a solid line), $SN_0^{(1)}N_{45}^{(2)}$ (asterisks connected with a solid line), $SN_{-60}^{(1)}$ (triangles connected with a dashed line) and $SN_{-60}^{(1)}N_{45}^{(2)}$ (squares connected with a dashed line).

Next, the influence of the number of disturbing noise sources was analyzed. This analysis was divided into two parts, as two spatial configurations were compared, i.e. $SN_0^{(1)}$ with $SN_0^{(1)}N_{45}^{(2)}$ and $SN_{-60}^{(1)}$ with $SN_{-60}^{(1)}N_{45}^{(2)}$. For the first pair of configurations, e.g. $SN_0^{(1)}$ (circles) with $SN_0^{(1)}N_{45}^{(2)}$ (asterisks), for all subjects the addition of the second noise at the azimuth of 45° brought about a significant increase in the speech intelligibility by about a dozen percent. This increase can be also explained on the basis of the BMLD phenomenon. Addition of the second noise resulted in the fact that the time patterns of the noise (that is the superposition of two sources of noise $N_0^{(1)}$ and $N_{45}^{(2)}$) and their levels were not identical in the left and in the right ear while the speech signals in both ears had the same phase and level. When the second noise was added at 45° a significant increase in the speech intelligibility was noticed because in the lateralization process the “picture” of the noise was shifted away from the center of the head while the “picture” of the speech signal remained at the same position. The data gathered for these two spatial configurations were analysed using a within-subject ANOVA. The influence of the signal-to-noise ratio and the number of maskers on the speech intelligibility was examined. The SNR was proved to be statistically significant: [$F(4, 8) = 136.59$, $p < 0.001$]. The number of disturbing sources (i.e. $SN_0^{(1)}$ or $SN_0^{(1)}N_{45}^{(2)}$) was also proved to be statistically significant: [$F(1, 2) = 9.52$, $p = 0.049$]. However, the interaction between these two factors was proved to be statistically insignificant: [$F(4, 8) = 0.09$, $p = 0.982$]. That implies that the improvement in the speech intelligibility (i.e. the difference in the speech intelligibility for the two analyzed configurations) did not depend on SNR.

It seems that the comparison of these two configurations (i.e. $SN_0^{(1)}$ and $SN_0^{(1)}N_{45}^{(2)}$) implies the existence of a spatial suppression. Addition of the second noise, in the way that the overall level of the resulting masker was the same in both configurations, decreased the masking efficiency of the disturbances and increased the speech intelligibility. It seems to be reasonable to state that the spatial suppression phenomenon may be only a generalization of the BMLD phenomenon which, by definition, is mainly concerned with signal detection thresholds depending on the masking effect of noise. The data gathered so far have revealed that the phase and the level differences of noise between both ears lead to both lower signal-threshold values and to the increase in the speech intelligibility. Therefore, new definition of the BMLD phenomenon may be proposed not on the basis of the difference in the masking levels for a tone masked by a noise, but on the basis of the signal-to-noise ratio changes that leads to changes in the threshold values.

For the second pair of configurations, i.e. $SN_{-60}^{(1)}$ and $SN_{-60}^{(1)}N_{45}^{(2)}$, the masking enhancement was noticed instead of the spatial suppression effect. This fact is also fully consistent with the BMLD. The speech intelligibility was significantly better for one disturbing noise only (placed at -60°). In this case the difference in the overall levels of the noise in both ears was about 10 dB. This implies that the SNR in the ear that was farther from the noise source (a “better” ear), was much higher than that in the opposite

one. As a consequence, the intelligibility was much better for the configuration with one noise source only. The addition of the second source of noise at the side of the better ear decreased significantly the SNR in this ear. It is important to emphasize that the configuration $SN_{-60}^{(1)}N_{45}^{(2)}$ corresponds very well to the almost symmetrical configuration of the disturbing sources regarding the head: uncorrelated masking signals at the same level were coming to both ears. In this configuration any sound coming from the front was almost equally masked in both ears. Furthermore this spatial localization of the disturbing noise sources qualitatively corresponds to the $SN_0^{(1)}$ configuration, but in the latter the disturbances in both ears were fully correlated. This qualitative correspondence of both configurations has been proved by the results of the experiment. The comparison of the results for $SN_{-60}^{(1)}N_{45}^{(2)}$ and $SN_0^{(1)}$ presented in Fig. 3 reveal that they are almost identical. This was also proved by the within-subject analysis of variance: [$F(1, 2) = 2.82$, $p = 0.253$]. It must be reminded that the BMLD for the uncorrelated sources of noise reaches merely 4 dB [17]. This partly results from the lack of noticeable difference in speech intelligibility for $SN_{-60}^{(1)}N_{45}^{(2)}$ and $SN_0^{(1)}$ configurations.

For all four spatial configurations, the average (for three subjects) speech reception thresholds were determined using the intelligibility curves from Fig. 3. This kind of measure is consistent with other works on this topic [2, 13]. The values are expressed in terms of the signal-to-noise ratio at the SRT and for successive spatial configurations are shown in Table 1.

Table 1. The average speech reception thresholds (SRTs) expressed in terms of the signal-to-noise ratio for 4 spatial configurations of the sources investigated in Experiment 1.

Spatial configuration	$SN_0^{(1)}$	$SN_0^{(1)}N_{45}^{(2)}$	$SN_{-60}^{(1)}$	$SN_{-60}^{(1)}N_{45}^{(2)}$
SRT (dB)	-5.5	-8	-14	-6.5

As it can be seen, there is 1 dB difference in SRTs between the $SN_0^{(1)}$ and $SN_{-60}^{(1)}N_{45}^{(2)}$ configuration, while the $SN_{-60}^{(1)}$ is most convenient for speech understanding. However, PEISSIG and KOLLMEIER [2] reported that the SRT for symmetric and almost symmetric configurations ($SN_x^{(2)}N_{105}^{(1)}$, x range from 240° to 270°) was about 2 dB below the STR for the $SN_0^{(2)}N_{105}^{(1)}$ (non-symmetric) case. In the present study comparing analogous configurations i.e. $SN_{-60}^{(1)}N_{45}^{(2)}$ (almost symmetric), and $SN_0^{(1)}N_{45}^{(2)}$ (non-symmetric) the opposite effect can be noticed, that is the SRT for the symmetric configuration is about 1.5 dB above the SRT for the other one case.

2.7. Experiment 2: The logatom intelligibility masked by one or two sources of noise

In this experiment phonetically balanced logatom tests were used [16]. Speech intelligibility was measured for three subjects aged 18–24 with audiological normal hearing. One of subjects was different from those of Experiment 1. Due to the low

redundancy of the logatoms, it was necessary to increase the signal-to-noise ratio (3, 0, -3, -6, -9 dB) so that the intelligibility would be measurable. The same spatial configurations as those in Experiment 1 were used.

The results of this part of the study, i.e. the logatom intelligibility as a function of SNR for different spatial configurations of noise source(s) ($SN_0^{(1)}$, $SN_0^{(1)}N_{45}^{(2)}$, $SN_{-60}^{(1)}$ and $SN_{-60}^{(1)}N_{45}^{(2)}$), are depicted in Fig. 4. Successive panels of the figure depict the data gathered for individual subjects. All results for both pairs of the compared spatial configurations, i.e. $SN_0^{(1)}$ with $SN_0^{(1)}N_{45}^{(2)}$ (circles and asterisks connected with a solid line respectively) and $SN_{-60}^{(1)}$ with $SN_{-60}^{(1)}N_{45}^{(2)}$ (triangles and squares connected with a dashed line respectively) are presented in this figure.

The results for all subjects are qualitatively consistent with each other, however the intersubject scatter may be noticed. These results are also analogous to the results obtained in the Experiment 1. Generally it may be stated that the intelligibility is a monotonically increasing function of the SNR. However, the logatom intelligibility values are significantly below the one-syllable words intelligibility. This fact reflects the difference in the redundancy of these two tests.

Let us analyse the situation in which the speech signal was masked by one noise only, i.e. the configurations $SN_0^{(1)}$ (circles) and $SN_{-60}^{(1)}$ (triangles). As it can be noticed in Fig. 4, the noise coming from the same direction as the signal is, again, a very effective masker since the speech intelligibility reaches the lowest values. However, when the noise was coming from the azimuth of -60° (the $SN_{-60}^{(1)}$ configuration), the speech intelligibility was much higher, i.e. the highest values of the intelligibility were gathered for this configuration. Despite the fact of the same overall level of the masker in both configurations, the masker placed at -60° was significantly less efficient. The difference in the logatom intelligibility between these two configurations is approximately constant and does not depend on the SNR and reaches about 25%. This result is also consistent with the result of Experiment 1 and may be interpreted on the basis of the BMLD. In the $SN_0^{(1)}$ case there is no interaural phase nor level difference between the signals reaching the left and the right ear, so the intelligibility is the lowest. However, in the $SN_{-60}^{(1)}$ case the time patterns of the noise in the left and in the right ear are different while the time courses of the logatoms are still identical in both ears. Thus, it is a typical situation in which the BMLD occurs contributing to the enhancement of the speech intelligibility.

The intelligibility difference between these two spatial configurations (i.e. $SN_0^{(1)}$ and $SN_{-60}^{(1)}$) was subjected to the within-subject analysis of variance (ANOVA). The influence of the signal-to-noise ratio and the spatial configuration of the maskers on the speech intelligibility was examined. The effect of SNR was proved to be statistically significant [$F(8, 4) = 89.24, p < 0.001$]. The effect of spatial configuration was also proved to be statistically significant: [$F(1, 2) = 373.59, p = 0.003$]. The interaction between these two factors was proved to be marginally statistically significant

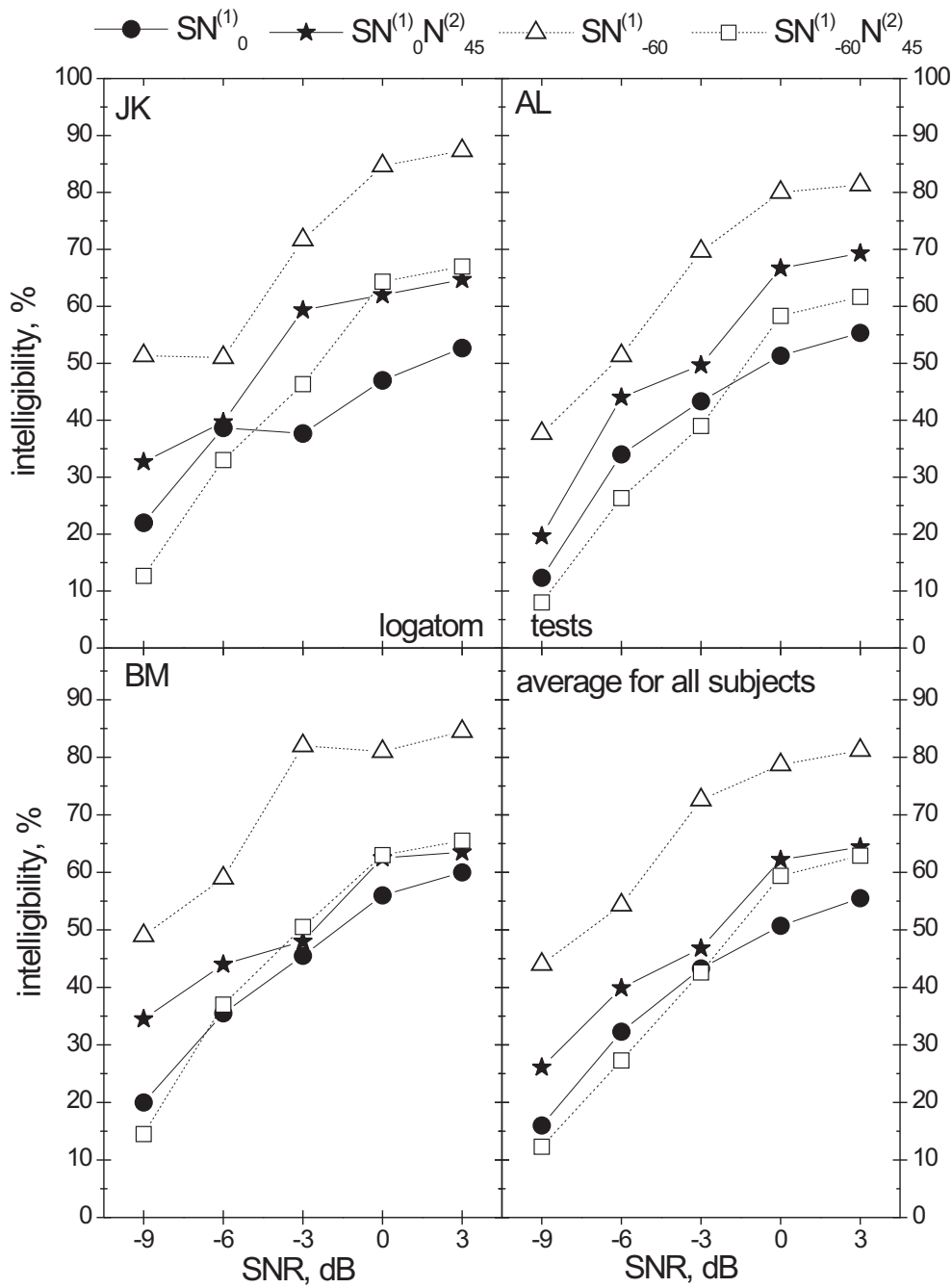


Fig. 4. Logatom tests intelligibility as a function of signal-to-noise ratio (SNR) for all four spatial configurations. Successive panels of the figure depict data gathered for individual subjects. $SN_0^{(1)}$ (circles connected with a solid line), $SN_0^{(1)}N_{45}^{(2)}$ (asterisks connected with a solid line), $SN_{-60}^{(1)}$ (triangles connected with a dashed line) and $SN_{-60}^{(1)}N_{45}^{(2)}$ (squares connected with a dashed line).

[$F(4, 8) = 3,60, p = 0.058$]. This implies that the difference in the intelligibility in both spatial configurations was only slightly dependent on SNR.

Addition of the second noise at the azimuth of 45° to the noise coming from 0° ($SN_0^{(1)}N_{45}^{(2)}$ configuration) contributed to a significant increase in the logatom intelligibility from a few to a dozen percent (compare $SN_0^{(1)}$ – circles and $SN_0^{(1)}N_{45}^{(2)}$ – asterisks in Fig. 4). This situation is analogous to that of the one-syllable word tests and may be explained in the same way. Addition of the second noise resulted in the fact that the time pattern and the level of the noise (that is superposition of two noises from two independent sources, $N_0^{(1)}$ and $N_{45}^{(2)}$) in the left and in the right ear were not identical while the speech signals in both ears had the same phase and level. Therefore, the addition of the second noise at 45° brought about a significant increase in the logatom intelligibility.

The analysis of variance in which the influence of the signal-to-noise ratio and the number of maskers on the logatom intelligibility was investigated revealed that the SNR was statistically significant: [$F(4, 8) = 57.05, p < 0.001$]. The number of disturbing sources (i. e. $SN_0^{(1)}$ or $SN_0^{(1)}N_{45}^{(2)}$) was also proved to be statistically significant: [$F(1, 2) = 45.36, p = 0.021$]. However, the interaction between these two factors was proved to be statistically insignificant: [$F(4, 8) = 0.34, p = 0.847$].

Comparison of the $SN_0^{(1)}$ and $SN_0^{(1)}N_{45}^{(2)}$ configurations and the statistically significant difference in the logatom intelligibility for these two cases implies the existence of the spatial suppression, since the addition of the second noise situated in the place different than the first one (in the way that the overall level of the masker was equal in both configurations) brought about the increase in the logatom intelligibility. These results are fully consistent with the data gathered by PEISSIG and KOLLMEIER [2]. They showed a significant improvement in the signal-to-noise ratio of short sentences when the second noise was added while the first one was fixed at 105° . The improvement, however, was much less prominent for azimuths from the range ($240^\circ, 270^\circ$), that is the symmetric or almost symmetric configuration, but still the SRT was about 2 dB below the SRT for $SN_0^{(2)}N_{105}^{(1)}$ configuration.

Similarly to the results of Experiment 1, for the second pair of configurations, i.e. $SN_{-60}^{(1)}$ and $SN_{-60}^{(1)}N_{45}^{(2)}$, the masking enhancement occurred, that is the opposite effect to the spatial suppression phenomenon. The logatom intelligibility was significantly better for one disturbing noise only, placed at -60° (see triangles in Fig. 4). This can be also explained in the same way as in Experiment 1. In this case the difference in the overall levels of noise in both ears was about 10 dB. This implies that the SNR in the “better” ear was much higher than in the opposite one. As a consequence, the intelligibility was much higher for the configuration with one noise source only. The addition of the second source of noise at the side of the “better” ear decreased significantly the SNR in this ear.

It is important to emphasize that the configuration $SN_{-60}^{(1)}N_{45}^{(2)}$ corresponds very well to a symmetrical configuration of the uncorrelated disturbing sources of speech-

shaped noise regarding to the head: uncorrelated masking signals at about the same level were coming to both ears. The BMLD for the uncorrelated sources of noise reaches merely 4 dB [17]. This partly corresponds to the lack of difference in the speech intelligibility for $SN_{-60}^{(1)}N_{45}^{(2)}$ and $SN_0^{(1)}$ configurations revealed by the ANOVA.

Again, the comparison of the data gathered in the experiments carried out for the $SN_{-60}^{(1)}$ and $SN_{-60}^{(1)}N_{45}^{(2)}$ conditions with the data presented by PEISSIG and KOLLMEIER [2] shows a qualitative agreement. However, it is difficult to state that this agreement is also quantitative. They showed that the addition of the second noise source generally brings about the increase in the speech reception threshold. For some of the positions of the second noise source, the second noise did not influence the SRTs whose values were very close to that measured with one noise source only. The data gathered in our experiments showed that the change in the speech intelligibility strongly depended on the spatial position of the first source of noise. In the PEISSIG and KOLLMEIER's [2] case, however, the position of the first source of noise was fixed at 105° while in our case the position of the first noise source was 45° . For this particular position of the first noise source, a marked decrease in the speech intelligibility was noticed after the second source of noise at the position of -60° was added. Thus, it seems that the existence of the spatial suppression is strongly influenced by mutual positions of the disturbing sound sources and the position of the speech source. In general it may be stated that when the positions of the disturbing sounds are approximately symmetrical, the spatial suppression may not occur at all. Moreover, the opposite effect may occur, that is the masking enhancement resulting in decrease of the speech intelligibility.

Similarly to the Experiment 1, the speech reception thresholds were determined using the intelligibility curves from Fig. 4. The values are shown in Table 2.

Table 2. The average speech reception thresholds (SRTs) expressed in terms of the signal-to-noise ratio for 4 spatial configurations of the sources investigated in Experiment 2.

Spatial configuration	$SN_0^{(1)}$	$SN_0^{(1)}N_{45}^{(2)}$	$SN_{-60}^{(1)}$	$SN_{-60}^{(1)}N_{45}^{(2)}$
SRT (dB)	0	-2.5	-7	-1.5

Again, there is 1.5 dB difference in SRTs between the $SN_0^{(1)}$ and $SN_{-60}^{(1)}N_{45}^{(2)}$ configuration while the $SN_{-60}^{(1)}$ is the most convenient for speech understanding and the $SN_0^{(1)}$ case is the most inconvenient one. Moreover, the difference in SRTs between the almost symmetric case ($SN_{-60}^{(1)}N_{45}^{(2)}$) and the non-symmetric ($SN_0^{(1)}N_{45}^{(2)}$) configuration observed in Experiment 1 was confirmed: the SRT for the symmetric configuration is about 1 dB above the SRT for the other one case while in data obtained by PEISSIG and KOLLMEIER, this dependence is reversed and reaches about 2 dB.

The data gathered in the first and the second experiment are in a good agreement. The only difference between these two experiments was the type of the test material used. Based on the gathered results it is possible to state that the spatial suppression

was noticed for both word tests and for the logatom tests. Also the effect opposite to the spatial unmasking was similar for these two types of speech sounds. Thus, it seems reasonable to state that the existence of the spatial suppression is not attributed to the utterances conveying any semantic information. It is rather a general ability of the auditory system to extract some information enclosed in acoustic waves. However, this ability is strongly influenced by exact position of the disturbing sources. Once the disturbing sources are more or less symmetrical, the spatial suppression does not occur.

As the spatial position of the disturbing sound sources is related to the interaural phase differences of sounds coming to the left and to the right ear, the results of our experiments imply an important role of phase preservation in the hearing aids. There should not be any phase shift between the input and the output signal. Otherwise the localization of the sound sources in the space as well as the speech intelligibility may be decreased as the spatial suppression may not occur at all. Furthermore, in some cases the masking enhancement may be noticed.

3. Conclusion

The carried out experiments enable us to formulate the following conclusions:

- All results are fully consistent with the binaural masking level difference and can be explained on the basis of this phenomenon. Therefore, it can be stated that the spatial suppression (in the first pair of spatial configuration) and masking enhancement (in the second pair) is only a generalized case of the BMLD.
- The decrease in speech intelligibility is consistent with the data on the decrease in the signal-to-noise ratio obtained by other authors.
- The results gathered in the present study revealed a significant difference between the speech reception thresholds determined by the authors and PEISSIG and KOLLMEIER [2] in analogous spatial configurations. However, in the present study more rigorous and objective tests were used for the speech intelligibility determination.
- The present study implies that the spatial suppression occurs only in certain cases (i.e. $SN_0^{(1)}$ and $SN_0^{(1)}N_{45}^{(2)}$). In some other spatial configurations the masking enhancement may occur (e.g. $SN_{-60}^{(1)}$ and $SN_{-60}^{(1)}N_{45}^{(2)}$).
- The logatom tests were proved to be more rigorous tests compared to the one-syllable word tests what can be seen from the comparison of the speech reception thresholds for those tests. However, the results obtained for both types of test are fully consistent with each other and across all the subjects. Even though the speech intelligibility differs for those two kinds of tests, the character of the dependences are similar and the spatial suppression and the masking enhancement are noticed to the same extent. It implies that the spatial suppression is not attributed to the utterances conveying semantic information only, but it is more general phenomenon that shows an ability of the auditory system of extracting useful information from a meaningless signal.

- In the $SN_{-60}^{(1)}$ case there was a significant difference in the overall levels of noise in both ears leading to a much better signal-to-noise ratio in one ear. Therefore, this configuration was most convenient for speech understanding.
- Data for both $SN_{-60}^{(1)}N_{45}^{(2)}$ and $SN_0^{(1)}$ cases are almost identical. This fact can be also explained on the basis of BMLD. The $SN_{-60}^{(1)}N_{45}^{(2)}$ case qualitatively corresponds to the $SN_0^{(1)}$ configuration, but in the latter the maskers in both ears were fully correlated whereas in the first one it could be assumed that the uncorrelated masking signals at the same level were coming to both ears. According to [17] when uncorrelated rather than correlated maskers of the same level are coming to separate ears, the BMLD value reaches merely 3–4 dB. This small value explains the lack of significant statistical difference in the intelligibility between spatial configurations $SN_{-60}^{(1)}N_{45}^{(2)}$ and $SN_0^{(1)}$.
- Since in the presented experiments Polish speech was used, it is possible to state that the spatial suppression was similar to that observed for the other languages. It suggests, then, that spectral structure of the speech signals is not an important factor for the spatial unmasking occurrence. Spatial position of the disturbing sound sources (which is related to the interaural phase and level differences) plays a primary role in the spatial suppression occurrence that may improve the speech intelligibility.

Acknowledgements

The authors would like to thank professor Andrzej Rakowski and an anonymous reviewer for the very helpful comments on an earlier version of this paper.

This work was supported by The State Committee for Scientific Research, project number 4T11E 01425 and 1H01F 08127.

References

- [1] BRONKHURST A.W., PLOMP R., *The effect of head-induced interaural time and level differences on speech intelligibility in noise*, Journal of the Acoustical Society of America, **83**, 4, 1508–1516 (1988).
- [2] PEISSIG J., KOLLMEIER J., *Directivity of binaural noise reduction in spatial multiple noise-source arrangements for normal and impaired listeners*, Journal of the Acoustical Society of America, **101**, 1660–1670 (1997).
- [3] HAWLEY M.L., LITOVSKY R.Y., COLBURN H.S., *Speech intelligibility and localization in a multi-source environment*, Journal of the Acoustical Society of America, **105**, 3436–3448 (1999).
- [4] SHINN-CUNNINGHAM B.G. KOPČO N., *Tori of confusion: Binaural localization cues for sources within reach of a listener*, Journal of the Acoustical Society of America, **107**, 1627–1636 (2000).

- [5] SHINN-CUNINGHAM, B.B., SCHICKLER J., KOPČO N., LITOVSKY R., *Spatial unmasking of nearby speech sources in a simulated anechoic environment*, Journal of the Acoustical Society of America, **110**, 1118–1129 (2001).
- [6] FREYMAN R.L., BALAKRISHNAN U., HELFER K.S., *Spatial release from informational masking in speech recognition*, Journal of the Acoustical Society of America, **109** (5 Pt 1) 2112–2122 (2001).
- [7] LICKLIDER J.C.R., *The influence of interaural phase relations upon the masking of speech by white noise*, Journal of the Acoustical Society of America, **20**, 150–159 (1948).
- [8] DRIKS D.D., WILSON R.H., *The effect of spatially separated sound sources on speech intelligibility*, J. Speech Hear. Res., **12** (5–38) (1969).
- [9] PLOMP R., MIMPEN A.M., *Effect of the orientation of the speaker's head and the azimuth of a noise source on the speech reception thresholds for sentences*, Acoustica, **48**, 325–328 (1981).
- [10] ZWICKER E., FASTL H., *Psychoacoustics – facts and models, second edition*, Berlin: Springer-Verlag, 1999.
- [11] HIRSH I.J., *The relation between localization and intelligibility*, Journal of the Acoustical Society of America, **22**, 196–200 (1950).
- [12] BRONKHORST A.W., PLOMP R., *The effect of head-induced interaural time and level differences on speech intelligibility in noise*, Journal of the Acoustical Society of America, **83**, 4, 1508–1516 (1988).
- [13] HAWLEY M.L., LITOVSKY R.Y., CULLING J.F., *The benefit of binaural hearing in a cocktail party: Effect of location and type of interferer*, Journal of the Acoustical Society of America, **115**, 2, 833–843 (2004).
- [14] BAER T., MOORE B.C.J., GATEHOUSE S., *Spectral contrast enhancement of speech in noise for listeners with sensorineural hearing impairment: effects on intelligibility, quality and response times*, Journal of Rehabilitation Research and Development, **30**, 49–72 (1993).
- [15] BOCCA E., C. CALEARO, *General hearing processes*, [in:] *Modern developments in Audiology*, J. J., Editor. 648–662, Academic Press, New York 1963.
- [16] BRACHMAŃSKI S., STARONIEWICZ P., *Phonetic structure of test material used in subjective measurement of speech quality* [in Polish], [in:] *Speech and Language Technology*, 71–80, Poznań 1999.
- [17] MOORE B.C.J., *An introduction to the psychology of hearing*, 5-th Ed. Academic Press, London 2003.
- [18] DURLACH N.I., GABRIEL K.J., COLBURN H.S., TRAHOTIS C., *Interaural correlation discrimination: II. Relation to binaural unmasking*, Journal of the Acoustical Society of America, **79**, 1548–1557 (1986).
- [19] CHERRY E.C., *Some experiments on the recognition of speech, with one and with two ears*, Journal of the Acoustical Society of America, **25**, 975–979 (1953).
- [20] POLLACK I., PICKETT J.M., *Stereophonic listening and speech intelligibility against voice babbles*, Journal of the Acoustical Society of America, **30**, 131–133 (1958).
- [21] JASSEM W., *Noise spectra of fricative consonants*, IV International Congress on Acoustics: Copenhagen 1962.

-
- [22] JASSEM W., *The formants of fricative consonants*, LS, **8**, 1, 1–16 (1965).
- [23] PRUSZEWICZ A., SURMANOWICZ–DEMENKO G., WIKI T., RICHTER L., *Polish numerical and word test for hearing examination and tests for hearing training* [in Polish], Marke-Med, Poznań 1999.
- [24] DUQUESNOY A.J., PLOMP R., *The effect of a hearing-aid on the speech-reception threshold of hearing-impaired listeners in quiet and in noise*, Journal of the Acoustical Society of America, **73**, 2166–2173, 1983.
- [25] CULLING J.F., COLBURN H.S., *Binaural sluggishness in the perception of tone sequences and speech in noise*, Journal of the Acoustical Society of America, **107**, 1, 517–527 (2000).
- [26] SĘK A., SKRODZKA E., OZIMEK E., WICHER A., *Intelligibility of speech processed by a spectral contrast enhancement procedure and a binaural procedure*, Archives of Acoustics, **29**, 1, 25–34 (2004).