PAN ARCHIVES OF ACOUSTICS
Vol. **43**, No. 4, pp. 581–592 (2018)

DOI: 10.24425/aoa.2018.125152

Research Papers

Detection and Recognition of Environmental Sounds by Musicians and Non-Musicians[1]

Andrzej MIŚKIEWICZ(1), Teresa ROŚCISZEWSKA(1), Jan ŻERA(2)
Jacek MAJER(1), Barbara OKOŃ-MAKOWSKA(1)

(1) *Chair of Musical Acoustics*
Department of Sound Engineering
The Fryderyk Chopin University of Music
Okólnik 2, 00-368 Warszawa, Poland

(2) *Institute of Radioelectronics and Multimedia Technology*
Faculty of Electronics and Information Technology
Warsaw University of Technology
Nowowiejska 15/19, 00-665 Warszawa, Poland; e-mail: j.zera@ire.pw.edu.pl

(*received November 8, 2017; accepted July 5, 2018*)

The article reports three experiments conducted to determine whether musicians possess better ability of recognising the sources of natural sounds than non-musicians. The study was inspired by reports which indicate that musical training develops not only musical hearing, but also enhances various non-musical auditory capabilities. Recognition and detection thresholds were measured for recordings of environmental sounds presented in quiet (Experiment 1) and in the background of a noise masker (Experiment 2). The listener's ability of sound source recognition was inferred from the recognition-detection threshold gap (RDTG) defined as the difference in signal level between the thresholds of sound recognition and sound detection. Contrary to what was expected from reports of enhanced auditory abilities of musicians, the RDTGs were not smaller for musicians than for non-musicians. In Experiment 3, detection thresholds were measured with an adaptive procedure comprising three interleaved stimulus tracks with different sounds. It was found that the threshold elevation caused by stimulus interleaving was similar for musicians and non-musicians. The lack of superiority of musicians over non-musicians in the auditory tasks explored in this study is explained in terms of a listening strategy known as casual listening mode, which is a basis for auditory orientation in the environment.

Keywords: environmental sounds; detection threshold; recognition threshold.

1. Introduction

Over the last decades there has been an increasing interest in the study of human abilities of acquiring and processing acoustic information from the environment. In addition to providing basic scientific knowledge on the processes of auditory perception in the environment, psychoacoustic studies concerning environmental sounds also have an applied aspect, as those kind of sounds are used as test and communication signals in medicine and in various branches of engineering.

The term "environmental sounds" refers in the literature to sounds naturally occurring in the environment, other than speech and music (Gygi, 2001; Gygi *et al.*, 2007). This paper reports a study in which recordings of environmental sounds were used as target signals for the measurement of recognition thresholds and detection thresholds. The purpose of the study was twofold: (1) to determine whether musicians possess better sound recognition abilities than non-musicians and are able to recognise the sources of environmental sounds at signal levels closer to the detection threshold,

[1]Portions of the data reported in this paper were presented at the 63rd Open Seminar on Acoustics in Białowieża (Rościszewska *et al.*, 2016) and at the Acoustics 2017 ASA-EAA Meeting in Boston (USA).

(2) to determine whether there exists any systematic relationship between the acoustic characteristics of environmental sounds and the minimum signal level increment above detection threshold required for a sound source to be correctly recognised by a listener.

The sound level difference between the thresholds of sound recognition and sound detection is termed as the recognition-detection threshold gap, RDTG, in the literature (Abouchacra *et al.*, 2007). The size of RDTG is equivalent to the minimum sensation level, SL, at which a sound source is correctly recognised by the listener.

Comparison of RDTGs measured for recordings of environmental sounds may provide considerable insight into various aspects of perception of this kind of sounds: (1) estimation of the RDTG may be used for assessment of the difficulty with which environmental sound sources are recognised by listeners and for prediction of utility of particular sounds as communication signals in various applications; (2) the size of RDTG may also serve as an indicator of the listener's sound recognition ability.

Abouchacra *et al.* (2007) measured detection and recognition thresholds for a set of 30 recordings of environmental sounds played back to the listeners in quiet and in the presence of multitalker masking noise. The results demonstrated that, both in quiet and in noise, some sound sources were recognised by the listeners at very low levels, almost as soon as the sound was detected, while other became recognisable only when the signal level was set well above the detection threshold. Depending on the sound, the RDTG ranged from 2 to nearly 13 dB in quiet and from 0.2 to nearly 12 dB in masking noise (Abouchacra *et al.*, 2007).

Andringa and Pals (2009) determined the RDTGs for various acoustic categories of environmental sounds presented in the background of a pink noise masker. The mean RDTG ranged from 7 dB for tonal sounds to 11 dB for noise-like sounds in their experiment. Large variability of RDTGs, spanning a range from 2 to 22 dB in quiet, was also observed by Myers *et al.* (1996) in an experiment conducted with the use of filtered sound recordings, spectrally limited to one-octave bands.

In the present study we sought to verify a working hypothesis assuming that RDTGs, determined for environmental sounds, might be smaller for musicians than for non-musicians. This hypothesis was derived from reports which indicate that training in music develops not only musical hearing, but also enhances various auditory abilities not related to music. Musicians, compared to non-musicians, demonstrate enhanced ability of understanding speech in noise (Parbery-Clark *et al.*, 2009), better pitch discrimination and timbre discrimination (Bogusz-Witczak *et al.*, 2015), better discrimination of voice timbre in speech (Chartrand, Belin, 2006), better ability of hearing out individual components in tone complexes (Fine, Moore, 1993), enhanced auditory working memory (Chan *et al.*, 1998), superior auditory attention selectivity (Lee *et al.*, 2007), better abilities in identification of the temporal order of sounds (Jakobson *et al.*, 2003), and faster reaction to sounds (Strait *et al.*, 2010). Musicians are less susceptible than non-musicians to backward masking (Strait *et al.*, 2010) and informational masking (Oxenham *et al.*, 2003). The auditory advantages of musicians observed in behavioral experiments have also been supported by evidence from cognitive neuroscience studies of the brain processes of sound perception (e.g., Musacchia *et al.*, 2007; Pantev *et al.*, 2007; Herholz *et al.*, 2011).

At signal levels close to detection threshold not all acoustic signatures of a sound are clearly audible, so only a limited set of perceptual cues can be used for the recognition of the sound source. One may therefore expect that listeners with highly refined auditory skills would be able to obtain more information from very weak auditory cues and recognise the sound source more readily than those who possess only average auditory skills.

The experiments reported in this paper were conducted using a set of stimuli belonging to various categories of environmental sounds distinguished on the basis of their acoustic and perceptual characteristics (Gygi *et al.*, 2007). The measurements enabled therefore to examine whether there is any systematic relationship in the size of RDTG among the typological categories of environmental sounds.

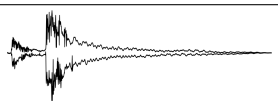
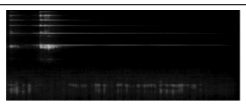
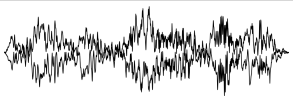
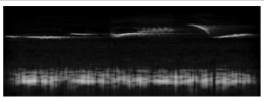
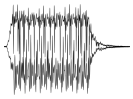
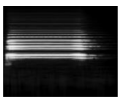
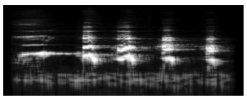
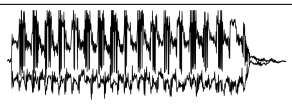
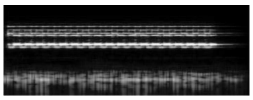
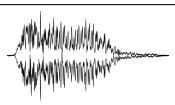
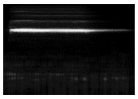
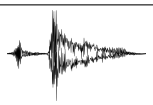
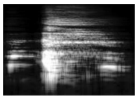
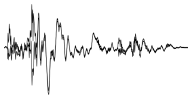
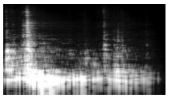
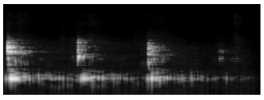
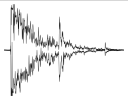
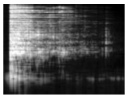
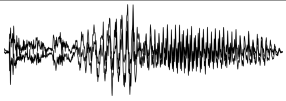
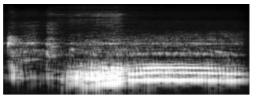
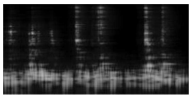
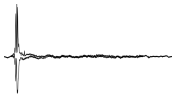
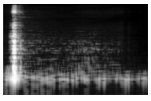
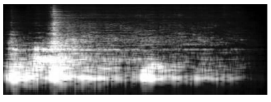
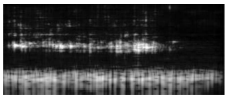
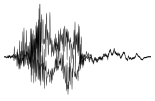
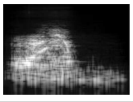
The present study comprised three experiments conducted on groups of musicians and non-musicians. In Experiments 1 and 2, the RDTGs were determined for recordings of environmental sounds presented, respectively, in quiet and in the background of a multitalker noise masker. In Exp. 3 only detection thresholds were measured, for selected sounds, with the use of a single-track adaptive procedure and a procedure with three interleaved stimulus tracks.

2. Method

2.1. Target sounds

The target sounds were binaural recordings of 16 environmental sounds commonly encountered in everyday life and had either a form of a single homogenous sound or consisted of a series of brief sounds forming an acoustic event. The sounds were exemplars of three categories distinguished by Gygi *et al.* (2007) in a typology of environmental sounds based on the similarity of their acoustic and perceptual characteristics: (1) harmonic sounds, (2) impulsive sounds, (3) non-harmonic sounds. Table 1 lists the sound sources or sound events that produced the target sounds and

Table 1. Duration, waveforms, and spectra of the recordings of environmental sounds used in Exps 1 and 2. The acronyms in the first column denote the acoustic category of the sounds: H – harmonic sound, I – impulsive sound, NH – non-harmonic sound. The abscissa on the spectrograms is sound duration represented on a linear scale and the ordinate shows the frequency on a logarithmic axis, within a 20–20 000 Hz range.

Sound source/event	Duration [ms]	Waveform	Spectrogram
Bicycle bell ringing (H)	1296		
Bird calling (H)	1442		
Car honking (H)	637		
Laughter (H)	1350		
Telephone ringing (H)	1376		
Whistle blowing (H)	738		
Coughing (I)	724		
Door handle pressing (I)	906		
Footsteps (I)	1445		
Glass breaking (I)	673		
Car starting (NH)	1368		
Typing on keyboard (NH)	1049		
Lighting a match (NH)	809		
Toilet flushing (NH)	1457		
Water pouring (NH)	1245		
Zipper (NH)	702		

shows the durations, the waveforms, and rough amplitude spectra of the sounds. The waveforms and spectra shown in Table 1 were measured for the sum of signals in the left and in the right channel.

For the determination of RDTGs (Exps 1 and 2) a full set of 16 target sounds was used. The sounds were played back in quiet in Exp. 1 and in the background of a continuous noise masker in Exp. 2. The masker added to the target sounds in Exp. 2 was a binaural recording of 25-voice multitalker noise played back at a sound level of 65 dB(A). The sound spectrum of the masker is shown in Fig. 1. The measurements of detection thresholds in Exp. 3 were conducted with the use of three target sounds presented in the background of the multitalker noise masker used in Exp. 2.

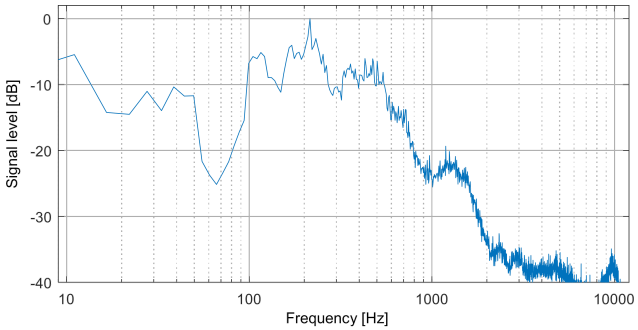


Fig. 1. Long-term average spectrum of the multitalker noise masker used in Exps 2 and 3.

2.2. Apparatus

Both the environmental sounds and the mutitalker noise masker were recorded with a dummy head (Neumann KU 100) and digitally stored on a Tascam DR120 recorder in two-channel (44 100 Hz/16 bit) wave format. The threshold measurements were conducted in a sound-attenuating booth. A PC-compatible computer controlled the presentation of the sounds, executed the psychophysical procedures and stored the listeners' responses. The wave files, played back from hard disk, were D/A converted (Focusrite Scarlett 2i2 audio interface), led through two programmable attenuators (Tucker-Davis Technologies, PA4) and a headphone buffer (Tucker-Davis Technologies, HB6), and reproduced dichotically through a Beyerdynamic DT 990 headset. An artificial ear (Brüel & Kjær, type 4143) coupled with a 1/4-inch microphone (Brüel & Kjær, type 4134) and a spectrum analyser (Brüel & Kjær, type 4144) were used for the measurement of the sound pressure levels reproduced from the earphones and for spectral analysis of the sounds.

All experiments were conducted in a soundproof booth. Depending on the task performed during a listening session, the listeners entered their responses using a computer terminal or a button box (Tucker-Davis Technologies, RBOX) (see Subsec. 2.4).

2.3. Listeners

All experiments were conducted on a group of 10 musicians and 10 non-musicians. The groups of listeners were different in Exps 1, 2, and 3. None of the listeners had any history of hearing difficulties and all of them had pure-tone audiometric thresholds at 15 dB HL or less, at octave frequencies between 0.25 and 8 kHz. The musicians were students at the Fryderyk Chopin University of Music in Warsaw and the non-musicians were students from various non-musical academic schools. None of the non-musician listeners had any experience in amateur musical activity.

2.4. Procedure

2.4.1. Experiments 1 and 2

In Exps 1 and 2 the listeners took part in a recognition task and in a detection task. The recognition task was conducted first and the detection task began after all the sound recognition tests had been completed in a given experiment.

(A) *Recognition task:* Recognition thresholds were measured using a one-interval, 16-alternative, forced choice (16-AFC) procedure. In each trial a target sound was presented and the listener had to identify the sound source by selecting one of the 16 available responses on the computer screen. After the listener had entered the answer, a next trial began. When the listener was unable to recognise the sound source, he/she gave a guessed answer.

Each target sound was presented at seven signal levels in each presentation of a block of trials. The signal levels were set individually for each sound such as to span, in six steps, a range from the expected detection threshold to a level of 21 dB above threshold. Five signal level steps were 3 dB in size and the size of the highest step was 6 dB. The signal levels corresponding to the expected detection threshold, used as reference for setting the signal levels in the recognition task, were roughly estimated by three members of the laboratory staff in a pilot test, by adjusting the signal level with a manual attenuator and then verified in a 16-AFC recognition task, in a practice run completed by selected listeners.

The set of stimuli comprised 112 test items (16 target sounds × 7 signal levels) presented in random order, different in each presentation of the set. The observation interval within which the target sound was played back was marked by a visual signal on the screen. To prevent the listeners from using the visual marker's duration as a cue for selecting the response the marker had the same duration of 1.5 s in all trials, regardless of the duration of the target sound in a given trial. When the listener was unable to identify the sound source or the sound was inaudible, he/she selected a response by guessing. To reduce the possibility of constant er-

rors in guessed responses the 16 sound sources were displayed in different order on the computer screen in each trial.

Each listener completed 12 series of judgments of the set of 112 test items. The responses obtained from 10 listeners were then used to determine a psychometric function for recognition showing the percentage of correct recognitions obtained for a given sound at each of the seven signal levels, in the group of musicians and in the group of non-musicians. The recognition threshold was estimated as the signal level corresponding to 53.125% correct recognitions, a value midway between random recognition in a 16-AFC task (6.25%) and 100% correct recognition.

(B) *Detection task:* Detection thresholds were measured using an adaptive, two-interval, two-alternative up-down forced-choice procedure (2I, 2-AFC) with feedback (Levitt, 1971). The observation intervals and correct-answer feedback were indicated in each trial by lights on the listener's response box. The signal level was varied according to a two-down/one-up decision rule that estimated the 70.7% correct point on the psychometric function. Each adaptive run started with a signal level approximately 15 dB above the listener's threshold and terminated after 50 trials. The initial step size of 5 dB was reduced to 2 dB after the fourth reversal of signal level. The threshold was estimated as the average signal level at the reversals, following the fourth reversal. Each listener completed three adaptive runs for each sound. The detection threshold determined for a listener was taken as the mean of thresholds determined in three runs conducted with the use of a given sound.

2.4.2. Experiment 3

A characteristic feature of a single track adaptive procedure, such as that used for the measurement of detection thresholds in Exps 1 and 2, is that the listener knows what sound will be presented in each trial and may focus attention on that sound. An alternative way of stimulus presentation is the interleaving of multiple adaptive tracks in a block of trials (Leek, Watson, 1994). When adaptive tracks are interleaved the stimulus to be presented in a given trial is chosen at random, out of the set of tracks currently used in a run, and the listener cannot focus attention on an expected sound, as is the case of a single-track procedure.

The purpose of Exp. 3 was to examine the effect of track interleaving on the detection thresholds measured for musicians and non-musicians in the presence of a continuous multitalker noise masker. The masker was the same as that used in Exp. 2. The thresholds were measured in two stimulus conditions: (1) separately for each target sound in single-track runs, (2) with interleaved adaptive tracks run at the same time for three target sounds. In both conditions, the procedures of signal level setting and calculation of the detection thresholds in each of the interleaved tracks were the same as in the single track procedure in Exp. 2 (see Subec. 2.4.1).

3. Results and discussion

3.1. Recognition-detection gaps (Experiments 1 and 2)

Figure 2 shows, for the groups of musicians (circles) and non-musicians (squares), the mean detection threshold and the standard deviation around the mean for 16 target sounds played back in quiet (Exp. 1, open symbols) and in the presence of a continuous multitalker noise masker (Exp. 2, filled symbols). The thresholds are expressed in terms of unweighted sound exposure level, *SEL*, according to the following formula:

$$SEL = 10\log\left(\frac{1}{T_0}\int_{t_1}^{t_2}\frac{p^2(t)}{p_0^2}\right)\mathrm{d}t, \qquad (1)$$

where T_0 is the reference duration of 1 s, $p(t)$ is the sound pressure, p_0 is the reference sound pressure of 20 μPa, t_1 and t_2 are the starting and ending times of the measurement. The sounds are grouped along the abscissa axis by the following categories: (1) harmonic sounds, (2) impulsive sounds, (3) non-harmonic sounds.

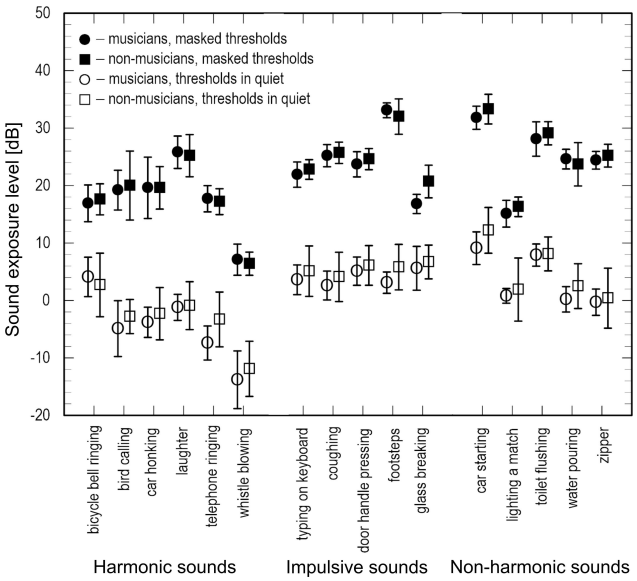


Fig. 2. Detection thresholds measured in Exps 1 and 2 for the recordings of 16 target sounds. Group means across 10 musicians (circles) and 10 non-musicians (squares). The open symbols show the thresholds in quiet (Exp. 1) and the filled symbols show the thresholds measured in the presence of a multitalker noise masker (Exp. 2). Error bars indicate the standard deviation of threshold values around the means.

The open symbols in Fig. 2 show that the group mean detection thresholds measured for individual target sounds in quiet ranged from −13.8 dB (whistle blowing) to 9.1 dB (car starting) in the group of musicians and from −11.9 dB (whistle blowing) to 12.2 dB (car starting) in the group of non-musicians. The masked thresholds (filled symbols) spanned a range from 7.1 dB (whistle blowing) to 33.1 dB (footsteps) in the group of musicians and from 6.4 dB (whistle blowing) to 33.3 dB (car starting) in the group of non-musicians. The relatively large overall ranges of detection thresholds obtained for both groups of listeners, in quiet and in masked conditions, was an expected effect as the target sounds used in the experiment considerably differed in their spectral and temporal characteristics. In general, the thresholds were lower for harmonic sounds than for the two other acoustic categories. The lower detection thresholds resulted from the spectral characteristics of harmonic sounds: those sounds were high pitched and their spectra contained strong formants at frequencies between about 2 and 5 kHz, in the range of maximum hearing sensitivity.

The detection thresholds determined for individual target sounds were similar for the groups of musicians and non-musicians and did not differ by more than 2 dB in the case of most sounds. Somewhat larger differences, amounting to 4 dB between the two groups of listeners, were observed for the sound of telephone ringing in quiet and the masked sound of glass breaking. However, a two-tailed t-test for independent pairs of values has indicated that those differences were statistically non-significant (telephone ringing in quiet: $t = 0.734$; $p = 0.469$; masked sound of glass breaking: $t = 1.176$; $p = 0.249$).

Figure 3 presents the mean sound recognition scores obtained from musicians (circles) and non-musicians (squares) in quiet (Exp. 1, open symbols) and in the presence of a continuous, multitalker noise masker (Exp. 2, filled symbols). The results for each target sound are plotted in a separate panel. Each data point is based on 120 responses (10 listeners × 12 repetitions) and shows the percentage of correct recognitions at the signal level indicated on the abscissa. The horizontal, broken line marks the value of 53.125% correct responses taken as the recognition threshold.

When threshold signal levels are derived from empirical psychometric functions a key issue is to decide what percentage of correct responses should be taken as the threshold level. The point of 53.125% correct responses shown by the broken line in the panels in Fig. 3 lies midway between the level of guessing (6.25%) and 100% recognition in a 16-AFC task. Such a rule of estimating thresholds from psychometric functions has been commonly used in classical psychophysics (Fechner, 1860). However, in modern studies, based on the signal detection theory, psychophysical thresholds are specified in terms of the index of detectability, d' (Green, Swets, 1966). The d' values given for m-alternative forced choice tasks in the literature (Macmillan, Creelman, 2005) refer to a model based on an assumption that all observation intervals in a trial contain random noise and the target signal is added to the noise in one of the intervals (Green, Swets, 1966). The application of such a model might be, however, problematic in the present experiment as the decision process in the recognition task was more complex than assumed in the model of an m-alternative forced choice task. The target sounds presented in the 16-AFC recognition task differed in the degree of their mutual, perceptual similarity. It is therefore very likely that the listeners responded by choosing an alternative not from the full set of 16 target sounds, but from a smaller number of sounds which seemed to be reasonably matched with the auditory cues perceived in a given trial. A straightforward application of the theory underlying the m-alternative forced choice technique would not therefore correspond to the actual process of sound recognition performed by the listeners in Exps 1 and 2.

It should also be noted that the choice of the point on the psychometric function taken as recognition threshold was of no critical importance in this study. A comparison of auditory abilities of musicians and non-musicians in sound recognition at near-threshold signal levels could be made in a reliable way with reference to any, arbitrarily chosen point of the psychometric function, located within the range from above chance recognition to below 100% correct recognition.

As seen in Fig. 3, the set of group mean percentages of correct recognitions calculated for the signal levels used in the experiment did not yield, in the case of almost all sounds, a target value of 53.125%, so the recognition threshold was calculated by interpolation of the two adjacent data points surrounding the target percentage level.

The data plotted in individual panels in Fig. 3 show that the psychometric functions for recognition had for most sounds a typical shape of an ogive curve extending from chance level to 100% correct recognition. Such a pattern of data is apparent in Fig. 3 in both conditions – in quiet and in noise, for musicians as well as for non-musicians. In the case of some sounds (e.g., door handle pressing and toilet flushing in quiet) the highest signal levels used in Exp. 1 were still too low to yield a level of 100% correct recognition scores. It also should be noted that the mean percentage of correct recognitions obtained for the sound of toilet flushing in quiet, for the group of non-musicians in Exp. 1, did not at all reach the target level of 53.125%.

The signal levels corresponding to the recognition threshold were, for most target sounds, almost identical for the groups of musicians and non-musicians. This convergence of the data between the two groups of lis-

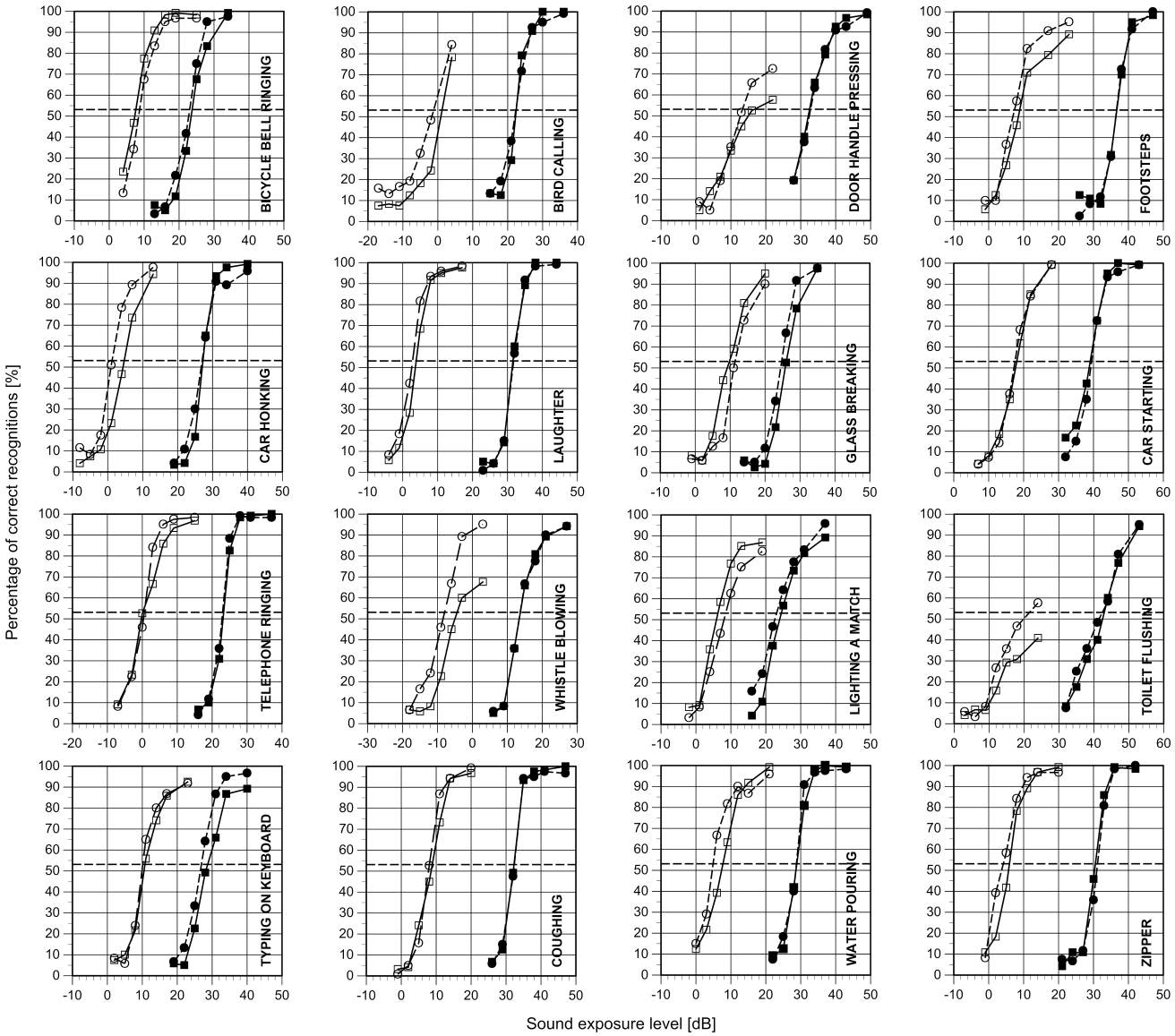


Fig. 3. Percentage of correct recognitions of environmental sound sources as a function of unweighted sound exposure level. Means across 10 musicians (circles) and 10 non-musicians (squares). Open symbols show the results obtained in quiet (Exp. 1) and the filled symbols show those obtained in the presence of a multitalker noise masker (Exp. 2). The horizontal broken line shows the target level of 53.125% correct recognitions used for the determination of the recognition threshold.

teners was strongly apparent in Exp. 2, where most of the recognition psychometric functions, determined for musicians and non-musicians, practically overlapped (Fig. 3).

Figure 4 shows the RDTGs determined for each of the 16 sounds and two groups of listeners – musicians and non-musicians, in quiet (Exp. 1) and in noise (Exp. 2). The RDTG was calculated for each sound as the difference in decibels between the recognition threshold and the detection threshold measured for a given group of listeners. The unmasked RDTG could not be calculated for the sound of toilet flushing in the group of non-musicians as the percentage of correct recognitions did not reach the target level of 53.125% in that case (see Fig. 3b).

As seen in Fig. 4, the RDTGs varied, depending on the sound, and ranged in quiet (Exp. 1) from 2.8 dB (bird calling) to 13.4 dB (toilet flushing) in the group of musicians and from 3.1 dB (footsteps) to 10.4 dB (door handle pressing) in the group of non-musicians. The RDTGs measured in the presence of the multitalker noise masker were within the range from 3.3 (bird calling) to 14.2 dB (toilet flushing) in the group of musicians and from 2.5 dB (bird calling) to 14.0 dB (toilet flushing) in the group of non-musicians. The RDTG ranges determined in the present study agree

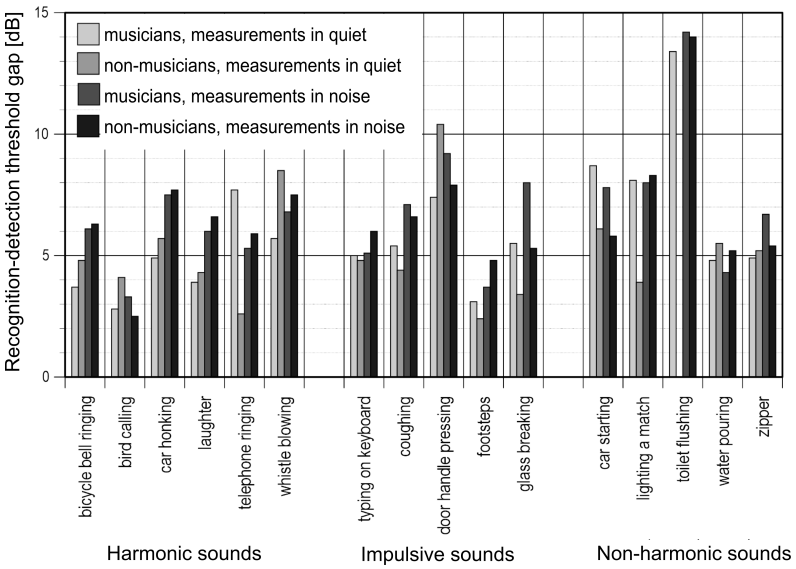


Fig. 4. Recognition-detection threshold gaps for 16 environmental target sounds. Group means for musicians and non-musicians in Exp. 1 (in quiet) and Exp. 2 (in noise).

fairly well with those reported by Abouchacra *et al.* (2007). The RDTGs ranged in their study 3.0–12.6 dB in quiet, 3.9–11.6 dB in noise, except for one outlying sound in noise, which was recognised practically at the detection threshold (Abouchacra *et al.*, 2007).

A noteworthy finding becomes evident when the RDTG values shown in Fig. 4 are examined with reference to the psychometric functions plotted for individual sounds in Fig. 3. It is clearly apparent that the psychometric functions for recognition are shallower for the sounds with the largest RDTGs (toilet flushing, door handle pressing) than for all the other sounds. This finding indicates that the difficulty with which a sound is recognised at low signal levels, manifested by a larger RDTG, is also reflected by a shallower psychometric function for sound recognition.

As the present study and the literature reports (Abouchacra *et al.*, 2007) show that RDTGs considerably vary for different environmental sounds an important question arises as to whether there are any specific acoustic cues that facilitate sound recognition at very low signal levels? A comparison of the RDTGs plotted in Fig. 4 with the sound waveforms and spectrograms shown in Table 1 suggests that the easiest ones to recognise are the sounds consisting of repeated, regular or irregular brief acoustic events (bird calling, laughter, footsteps, typing on keyboard) and sounds with abrupt onset and/or offset (car honking, glass breaking, coughing, lighting a match). The hardest ones to recognise are sounds with a more gradual onset and offset and a broadband spectrum (toilet flushing, door handle pressing, car starting).

To verify the main hypothesis put forward in the present study, assuming that musicians outperform non-musicians in the recognition of environmental sound sources at low signal levels, it is essential to determine whether the RDTGs measured in Exps 1 and 2 were smaller for musicians than for non-musicians. To facilitate such a comparison of the two groups of listeners, the RDTGs shown in Fig. 4 were replotted in Fig. 5 in such a way that the coordinates of each data point show the RDTGs obtained for a given sound for the group of musicians (abscissa) and for the group of non-musicians (ordinate). The open symbols indi-

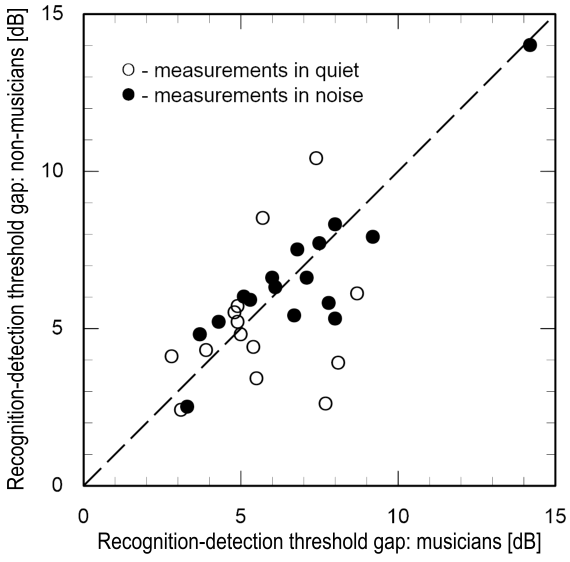


Fig. 5. A comparison of RDTGs determined for a group of 10 musicians and a group of 10 non-musicians, for 16 target sounds presented in quiet (Exp. 1, open symbols) and in the background of a multitalker noise masker (Exp. 2, closed symbols). The abscissa of each point is the RDTG determined for a given sound for musicians; the ordinate shows the RDTG for the same sound, for non-musicians.

cate the RDTGs obtained in quiet and the filled ones show the RDTGs measured in noise.

If musicians had better ability of sound source recognition the RDTGs should be larger for non-musicians and the data points would fall into the area above the diagonal line in Fig. 5. It is, however, apparent in Fig. 5 that most of the data points are clustered in close vicinity of the diagonal line, above and below the line, which means that the differences in the size of RDTG were small between the groups of listeners and did not exhibit any systematic pattern that would indicate a pronounced advantage of one group over the other one in sound recognition.

It should also be noted that some of the data points seen in Fig. 5 are located at a somewhat larger distance from the diagonal line than the rest of the points. Those points, plotted by open symbols, represent the RDTGs determined in quiet, and are distributed both above and below the diagonal line. At the present stage it is difficult to determine what exactly has caused a larger difference between the group RDTGs in the case of some individual target sounds presented in quiet. One possible explanation is that the measurements of sound detection and sound recognition were to a larger degree affected by the fluctuations of physiological noise in quiet (Exp. 1) than in the presence of the noise masker (Exp. 2).

3.2. The effect of stimulus interleaving on the assessment of detection thresholds

Figure 6 shows the results of Exp. 3 in which the detection thresholds were measured in the presence of a continuous multitalker noise masker, with the use of a single track procedure and with a procedure with three interleaved stimulus tracks, each containing a different target sound. The data plotted in Fig. 6 are group mean detection thresholds obtained for 10 musicians (circles) and 10 non-musicians (squares). Open symbols show the thresholds measured in single-track runs and filled symbols are the data obtained with stimulus track interleaving. Each listener completed three runs in each condition and the data points in Fig. 6 indicate the group means. The error bars show the standard error of individual listeners' thresholds around the means.

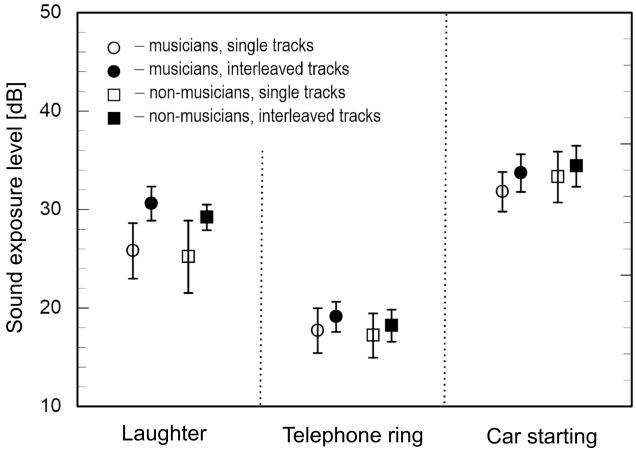


Fig. 6. Masked detection thresholds determined for recordings of three environmental sounds with the use of a single-track procedure (open symbols) and a procedure with three interleaved stimulus tracks (closed symbols). Group means for 10 musicians (circles) and 10 non-musicians (squares).

The data in Fig. 6 show that interleaving of stimulus tracks caused a slight elevation of the detection threshold both in the group of musicians and in the group of non-musicians. The magnitude of threshold elevation depended on the sound and ranged from about 1 dB (telephone ringing) to nearly 5 dB (laughter). A noteworthy observation is that, for individual target sounds, the threshold elevation caused by interleaving of stimulus tracks was similar for musicians and non-musicians.

4. General discussion

The lack of evidence for the assumed superiority of musicians over non-musicians in the ability of recognising the sources of environmental sounds needs a more thorough comment, in light of the body of published reports that have demonstrated the effect of "musician hearing advantage", in a variety of non-musical auditory tasks presented in the Introduction. The present results pose a question of why the enhanced auditory skills of musicians, manifested in a variety of listening tasks in psychoacoustical experiments, could not be observed when the listeners had to recognise the sources of environmental sounds?

The answer to this question should be sought in the specific nature of the perception of environmental sounds. Sound perception may be based upon different listening strategies, termed the *modes of listening* in the literature (Gaver, 1993a; 1993b; Chion, 1994). The listening mode upon which the listener's response is based in a given auditory task depends on the type of sounds being listened to and the purpose of listening.

A number of classifications of listening modes, based on various criteria, were proposed in the literature (Schaeffer, 1966; Gaver, 1993a; 1993b; Chion, 1994; Preis, Klawiter, 2005; Tuuri *et al.*, 2007). The most relevant to the present study is a classification presented by Chion (1994) in which he distinguished three basic modes of listening: causal listening, semantic listening, and reduced listening. *Causal listening*, also termed *everyday listening* by Gaver (1993a; 1993b), is focused on the identification of the sound sources and gathering information about the events in the environment that are reflected by the sounds. The focus of *semantic listening* is to extract the meanings conveyed by the sounds by means of

a certain code or language. The most common example of this mode of listening is the reception of spoken language, but the concept of semantic listening may also refer to various other situations in which a meaning is inferred by a sound associated with a certain type of formal or habitual code. The aim of *reduced listening* is to perceive the inherent sonic characteristics of sound with no connotations to any sound sources or events that might produce the sounds. The term *reduced listening* was coined by Schaeffer (1966) in a treatise on the phenomenology of musical and auditory objects. Reduced listening has also been called *musical listening* (Gaver, 1993a; 1993b). The classification of listening modes proposed by Chion (1994), as well as other similar classifications (e.g., Gaver, 1993a; 1993b), are very general in their character. One should therefore keep in mind that in real-life situations of sound perception different modes of listening may operate concurrently and complement each other in the accomplishment of the goal of listening (Tuuri *et al.*, 2007).

The task of recognition of the sources of environmental sounds performed by the listeners in the present study belongs to the class of the causal listening mode whereas the findings of enhanced auditory abilities of musicians, reported in the literature, were obtained in experimental tasks based on semantic listening, conducted with the use of speech sounds (e.g., Parbery-Clark *et al.*, 2009), and on reduced listening, in the case of artificial test signals presented in basic research studies in psychoacoustics (e.g., Oxenham *et al.*, 2003; Strait *et al.*, 2010). This difference may explain, at least in part, why the auditory skills developed by musical training and resulting in better performance of musicians in a variety of non-musical auditory tasks, were not manifested by improved recognition of the sources of environmental sounds in the present study.

Each mode of listening engages a different set of auditory perceptual mechanisms. Gygi (2001) noted an important difference between listening to environmental sounds, and to speech and music. When listening is aimed at auditory orientation in the environment the listener is focused on short-term spectral and temporal properties of the acoustic signals, as such a strategy facilitates quick recognition of the sound sources. In listening to speech and music the listener's attention is spread over a longer time interval which is needed to extract the semantic or artistic messages conveyed by the sounds.

The present data, interpreted in terms of Chion's (1994) classification of listening modes, indicate that the auditory listening abilities developed by musical education and musical professional experience are much more closely related to semantic listening and reduced listening than to the casual listening mode which is the basis of sound source recognition and auditory orientation in the environment. Although this study has shown that musical training does not improve the ability of recognising the sources of environmental sounds, it should be emphasised that such an ability may be improved by specialised, but not necessarily musical training, focused on the perception of specific classes of sounds. A variety of specialised training courses aimed at the development of auditory skills needed for the identification of specific classes of sound sources and sound events in the environment have been developed and successfully implemented in various branches of science and technology, for example, in the automotive industry (e.g., Miśkiewicz, Letowski, 2014) and in the military (Fluitt *et al.*, 2010; Scharine *et al.*, 2010).

Another finding of the present study that adds new insight to the reports of the "musician hearing advantage" effect is that interleaving of stimulus tracks containing different sounds has a similar effect on the assessment of sound detection thresholds in musicians and in non-musicians. Studies of auditory attention selectivity (see (Scharf, 1998) for a review) have demonstrated that spectral cues have a strong effect on signal detection. When the listener's attention is focused on a particular frequency band the detection threshold is typically lower than in conditions when the signal frequency is not known in advance to the listener. In the present study the use of track interleaving resulted in an elevation of detection threshold by the same level in the groups of musicians and non-musicians. This finding suggests that auditory attention, involved in sound detection, has a similar spectral selectivity in musicians and non-musicians.

Although the main objective of this study was to compare the abilities of musicians and non-musicians in the recognition of environmental sound sources, the measurement conducted in the experiments also gave an opportunity to determine the physical properties of sounds that facilitate sound source recognition at low signal levels. The finding that no clear-cut relations between the physical characteristics of sound and the size of RDTG were apparent in the experiments adds to the reports which indicate that recognition of the sources of environmental sounds is a complex process based on multiple acoustic cues (Pastore *et al.*, 2008). Humans exhibit extensive abilities of extracting environmental information from sounds and can recognise the sources of the sounds as well as their various physical properties communicated in form of acoustic information. For example, listeners are able to estimate by sound the size and shape of impacted bars (Lakatos *et al.*, 1997), the length of dropped rods (Carello *et al.*, 1998) and the posture of walkers (Pastore *et al.*, 2008). The acoustic cues upon which such sophisticated auditory judgments are made and the interrelations between those cues are, however, very complex and still poorly understood.

5. Conclusions

The main conclusions of the present study are as follows.

- Contrary to what might be inferred from the reports of the so called "musician hearing advantage" effect, musicians did not demonstrate better ability of recognising the sources of environmental sounds than non-musicians in the present study.
- The lack of auditory superiority of musicians over non-musicians in the ability of recognising environmental sound sources may be explained in terms of a listening strategy, known as the casual listening mode, which is a basis for auditory orientation and recognition of the sound sources in the environment. The auditory advantages of musicians over non-musicians, reported in a number of studies, were observed in non-musical experimental tasks based not on the casual listening mode, but on different cognitive auditory processes of reduced listening and semantic listening.
- The recognition thresholds measured for individual sounds exceeded the detection thresholds by 3–14 dB and had a similar range in quiet and in noise. The difficulty with which a sound is recognised by the listener, manifested by a larger RDTG, seems to be also reflected by the steepness of the psychometric function for recognition. The psychometric functions determined for sounds with the largest RDTGs were shallower than those obtained for all the other sounds in the present study.
- The elevation of detection threshold caused by interleaving of different sounds during the presentation of stimuli is similar for musicians and non-musicians. This finding shows that the frequency selectivity of auditory attention which causes such a threshold elevation does not improve by musical training and practice in music performing.

Acknowledgment

This work was supported by the Polish National Center for Scientific Research, Grant UMO-2013/11/B/HS6/01252, *Recognition of environmental sounds by musicians and non-musicians.*

References

1. Abouchacra K., Łętowski T., Gothie J. (2007), *Detection and recognition of natural sounds*, Archives of Acoustics, **32**, 3, 603–616.

2. Andringa T., Pals C. (2009), *Detection and recognition threshold of sound sources in noise*, Proceedings of the 31st Annual Conference of the Cognitive Science Society, CogSci09, pp. 1798–1803.

3. Bogusz-Witczak E., Skrodzka E., Turkowska H. (2015), *Influence of musical experience of blind and visually impaired young persons on performance in selected auditory tasks*, Archives of Acoustics, **40**, 3, 337–349.

4. Carello C., Anderson K.L., Kinkler-Peck A.J. (1998), *Perception of object length by sound*, Psychological Science, **9**, 211–214.

5. Chan A.S., Ho Y.C., Cheung M.C. (1998), *Music training improves verbal memory*, Nature, **396**, 128.

6. Chartrand J.-P., Belin P. (2006), *Superior voice timbre processing in musicians*, Neuroscience Letters, **405**, 164–167.

7. Chion M. (1994), *Audio-vision: Sound on screen*, Columbia University Press, New York.

8. Fechner G.T. (1860), *Elemente der Psychophysik*, Breitkopf & Härterl, Leipzig.

9. Fine P.A., Moore B.C.J. (1993), *Frequency analysis and musical ability*, Music Perception, **11**, 39–53.

10. Fluitt K., Gaston J., Karna V., Letowski T. (2010), *Feasibility of audio training for identification of auditory signatures of small arms fire*, Report ARL-TR-5413, Army Research Laboratory, Aberdeen, MD.

11. Gaver W.W. (1993a), *What in the worlds do we hear? An ecological approach to auditory event perception*, Ecological Psychology, **5**, 1–29.

12. Gaver W.W. (1993b), *How do we hear in the world? Explorations in ecological acoustics*, Ecological Psychology, **5**, 285–313.

13. Green D.M., Swets J.A. (1966), *Signal detection theory and psychophysics*, New York: Wiley.

14. Gygi B. (2001), *Factors in the identification of environmental sounds*, Doctoral dissertation, Department of Psychology, Indiana University.

15. Gygi B., Kidd G.R., Watson C.S. (2007), *Similarity and categorization of environmental sounds*, Perception and Psychophysics, **69**, 839–855.

16. Herholz S.C., Boh B., Pantev C. (2011), *Musical training modulates encoding of higher-order regularities in the auditory cortex*, European Journal of Neuroscience, **34**, 524–529.

17. Jakobson L., Cuddy L., Kilgour A. (2003), *Time tagging: A key to musicians' superior memory*, Music Perception, **20**, 307–213.

18. Lakatos S., McAdams S., Caussé R. (1997), *The representation of auditory source characteristics: Simple geometric form*, Perception and Psychophysics, **59**, 1180–1191.

19. Lee Y., Lu M., Ko H. (2007), *Effects of skill training on working memory capacity*, Learning and Instruction, **17**, 336–344.

20. Leek M.R., Watson C.S. (1984), *Learning to detect auditory pattern components*, Journal of the Acoustical Society of America, **76**, 1037–1044.

21. Levitt H. (1971), *Transformed up-down methods in psychoacoustics*, Journal of the Acoustical Society of America, **49**, 467–477.

22. Macmillan N.A., Creelman C.D. (2005), *Detection theory: a user's guide*, 2nd Ed., Lawrence Erlbaum Associates, Mahwah, New Jersey.

23. Miśkiewicz A., Letowski T. (2014), *Timbre Solfege training in automotive industry*, Proceedings of the 7th Forum Acusticum, Kraków.

24. Musacchia G., Sams M., Skoe E., Kraus N. (2007), *Musicians have enhanced subcortical auditory and audiovisual processing of speech and music*, Proceedings of the National Academy of Sciences, **104**, 15894–15898.

25. Myers L.L., Letowski T.R., Abouchacra K.S., Kalb J.T., Haas E.C. (1996), *Detection and recognition of octave-band sound effects*, Journal of the American Academy of Audiology, **7**, 346–357.

26. Oxenham A.J., Fligor B.J., Mason C.R., Kidd G., Jr. (2003), *Informational masking and musical training*, Journal of the Acoustical Society of America, **114**, 1543–1549.

27. Pantev C., Ross B., Fujioka, T., Trainor L.J., Schutte M., Schulz M. (2007), *Music and learning-induced cortical plasticity*, Annals of the New York Academy of Sciences, **999**, 438–450.

28. Parbery-Clark A., Skoe E., Lam C., Kraus N. (2009), *Musician enhancement for speech in noise*, Ear and Hearing, **30**, 653–661.

29. Pastore R.E., Flint J.D., Gaston J.R., Solomon M.J. (2008), *Auditory event perception: the source – perception loop for posture in human gait*, Perception and Psychophysics, **70**, 13–29.

30. Preis A., Klawiter A. (2005), *The audition of natural sounds – its levels and relevant experiments*, Proceedings of Forum Acusticum, Kraków, pp. 1595–1599.

31. Rościszewska T., Miśkiewicz A., Żera J., Majer J., Okoń-Makowska B. (2016), *Detection and recognition thresholds of environmental sounds*, [in:] *Advances in Acoustics*, M. Meissner [Ed.], Polish Acoustical Society, Warsaw Division, Warszawa, pp. 257–268.

32. Schaeffer P. (1966), *Traité des objets musicaux*, Editions du Seuil, Paris.

33. Scharf B. (1998), *Auditory attention: The psychoacoustical approach*, [in:] *Attention*, H. Pashler [Ed.], Psychology Press, Hove, pp. 75–118.

34. Scharine A., Letowski T., Mermagen T., Henry P. (2010), *Learning to detect and identify acoustic environments from reflected sound*, Military Psychology, **22**, 24–40.

35. Strait D.L., Kraus N., Parbery-Clark A., Ashley R. (2010), *Musical experience shapes top-down auditory mechanisms: Evidence from masking and auditory attention performance*, Hearing Research, **261**, 22–29.

36. Tuuri K., Mustonen M.-S., Pirhonen A. (2007), *Same sound – different meanings: A novel scheme for modes of listening*, Proceedings of Audio Mostly, Fraunhofer Institute for Digital Media Technology IDMT, pp. 13–18.