www.czasopisma.pan.pl PAN www.journals.pan.pl

ARCHIVES OF ACOUSTICS
Vol. **41**, No. 4, pp. 637–648 (2016)

DOI: 10.1515/aoa-2016-0061

Research Papers

Loudness Scaling Test Based on Categorical Perception

Bożena KOSTEK[(1)], Piotr ODYA[(2)], Piotr SUCHOMSKI[(2)]

[(1)] *Audio Acoustics Laboratory*
Faculty of Electronics, Telecommunications and Informatics
Gdańsk University of Technology
Narutowicza 11/12, 80-233 Gdańsk, Poland; e-mail: bokostek@audioacoustics.org

[(2)] *Multimedia Systems Department*
Faculty of Electronics, Telecommunications and Informatics
Gdańsk University of Technology
Narutowicza 11/12, 80-233 Gdańsk, Poland; e-mail: {piotrod, pietka}@sound.eti.pg.gda.pl

(*received March 2, 2016; accepted May 5, 2016*)

The main goal of this research study is focused on creating a method for loudness scaling based on categorical perception. Its main features, such as: way of testing, calibration procedure for securing reliable results, employing natural test stimuli, etc., are described in the paper and assessed against a procedure that uses 1/2-octave bands of noise (LGOB) for the loudness growth estimation. The Mann-Whitney U-test is employed to check whether the proposed method is statistically equivalent to LGOB. It is shown that loudness functions obtained in both methods are similar in the statistical context. Moreover, the band-filtered musical instrument signals are experienced as more pleasant than the narrow-band noise stimuli and the proposed test is performed in a shorter time. The method proposed may be incorporated into fitting hearing strategies or used for checking individual loudness growth functions and adapting them to the comfort level settings while listening to music.

Keywords: loudness scaling; categorical perception; musical instrument signals stimuli.

1. Introduction

Research studies on categorical perception are carried out in multiple fields, such as loudness measurements and scaling, diagnosing hearing impairments, measuring recruitment in the inner ear hearing loss, determining parameters of dynamics compression customized for hearing aids users, etc. It may be said that categorical loudness scaling by definition is a psychoacoustic measurement procedure which registers individual subjective loudness perception (HoerTech). It should be noted that determining loudness perception for hearing impaired persons is one of the most important areas. The number of people having hearing loss problems has increased significantly over the recent years. Audiometric tests may not be sufficient to fully assess hearing, especially in the context of loudness recruitment in persons with sensorineural hearing loss (SNHL) (Moore, 2007). Although importance of loudness scaling tests is well-documented in hearing aid fitting (Pettit, Keefer, 2011; Keefer, 2012), the same procedure is often applied to discover loudness recruitment (Brand, 2007). Categorical loudness scaling (CLS) procedure was introduced by Heller (1985), it used a verbal scale with seven categories, and a subscale with 10 fine subdivisions, that's why is time-consuming. However, since more than 1000 subjects were involved in tests, it is referred as the gold standard in loudness scaling because of its precision in an individual hearing loudness growth function determination. The WHS (Würzburger Hörfeld) procedure, proposed by Hellbrück and Moser (1985), was based on Heller's method, but implemented in its simpler and shorten version. One of the loudness scaling tests, i.e. LGOB (Loudness Growth in 1/2-octave bands) procedure (Allen *et al.*, 1990) used an even simpler scale consisting of seven response alternatives. Another test, named ACALOS (Adaptive Categorical Loudness Scaling) (Brand, 2000; 2007; Brand, Hohmann, 2001; 2002) was introduced in 2000 and served later as a basis for the new standard ISO 16832 on CLS, i.e. "Acoustics – Loudness scaling by means of categories", that was released in 2006 (ISO 16832, 2006).

Recently, a new set of publications appeared on this subject. One of the studies (Al-Salim *et al.*, 2010) was performed to estimate reliability of loudness scaling results. This involved both normal hearing and hearing-impaired subjects that were tested in two separate sessions, separated by varying time intervals (one week to six months). The authors of this research study found that mean stimulus-level difference between visits ranged from 6.6 to 7.8 dB, depending on frequency. They have reported that CLS measurements were reliable within-subject across sessions both for individual loudness categories and for slope of the CLS functions. They concluded that their work supports the assumption that audiometric threshold and response growth (loudness) are both determined by the same underlying cochlear mechanisms. Contrarily, Marozeau and Florentine (2007) demonstrate that loudness scaling for hearing-impaired individuals show large individual differences. Moreover, in the case of hearing impairment individual loudness-growth functions encompass a wide range of shapes. This may indicate a problem of loudness category estimation. Their study confirm the criticism conveyed earlier by Elberling (1999), who formulated a notion that different methods produce different loudness functions that cannot be compared. Other researchers continued studies on CLS in various contexts. Oetting *et al.* (2014) introduced a categorical loudness scaling procedure that tries to avoid a problem with uncomfortable loudness level (UCL) estimation, especially if responses are not present in the upper loudness range. Rasetshwane *et al.* (2015) proposed a method to measure impact of hearing loss on the estimates of loudness. Their study is especially important when loudness recruitment and reduced cochlear compression occurred in the presence of hearing loss. Moreover, their work describes constructing equal-loudness contours (ELCs) in phons from categorical loudness scaling (CLS), as in their opinion representing CLS data in phons may lead to wider acceptance of CLS measurements.

Even though LGOB method is not often (or no longer) used in hearing evaluation nowadays, the assumptions that underlie this approach make it is easy to implement on computers. Therefore, principles of this method are presented in Sec. 3. However, LGOB is still a method that takes a considerable amount of time and needs a good attention span from a tested person. On the other hand, artificial test signals (narrowband noise) may be replaced by more naturally sounding signals. Therefore, the aim of this paper is to propose a method which allows for scaling loudness sensation that works similarly to the LGOB test but takes less time, and test stimuli are more easily acceptable. For this purpose a computer-based application is constructed that uses only chosen categories from those prescribed in the LGOB test, which are most often rated by tested persons with normal hearing. The method proposed is described in Sec. 3. Tests are performed with students of the Multimedia Systems Department (MSD), Gdańsk University of Technology (GUT), as well as with a group of hearing impaired persons. Results obtained in LGOB and the proposed method are compared, and then statistically analyzed in Sec. 4. Examples of pure tone audiometry are presented against the loudness scaling results.

Another problem with categorical loudness scaling tests is that results obtained in those methods may differ as much as 30 dB despite using the same loudness scale (Oetting *et al.*, 2014). As mentioned before, different methods are comparable to some extent. The problem lies in differences between the loudness categories and stimuli parameters (Keefer, 2012; Oetting *et al.*, 2014). This may create problems for hearing aid fitting process, but it should still be possible to detect problem of hearing impairment. Therefore the additional goal was to investigate whether the proposed methodology loudness scaling makes possible to determine loudness growth function in young people, active listeners of music. It was observed that one of the noise-induced adverse effects might be associated with loudness recruitment, especially in young persons when their normal hearing seems to be still preserved. Research in this area proved that noise causes hearing loss. Worse, the harmful impact on hearing is not limited only to noise defined as unwanted sound. Music can even be more dangerous (Bulla, Hall, 1996; Dibble, 1995; Kozłowski, Młyński, 2014), as in most cases it is treated as desirable sound. Musicians often suffer from too loud music and are indicated as a group with potential severe hearing losses (Jaroszewski *et al.*, 1998; Pawlaczyk-Łuszczyńska *et al.*, 2013). The number of people who are threatened by listening to music is growing rapidly over years. It is caused by popularity of portable audio players. They are used on the way to work and school, during relaxation and sport activities. People try to avoid noise, but they do not perceive music as being dangerous to their hearing. Another problem is related to the sound level set by the users. They want to listen to music as loud as possible (Vogel *et al.*, 2008). As a result, sound from the headphones (usually In-The-Ear technology, ITE) may exceed 100–110 dB SPL (Sound Pressure Level). Of course, there exist standards concerning the output level from the portable audio players, e.g. in Poland this regards the BS EN 50332-1 standard. Unfortunately, many manufacturers simply ignore all recommendations and standards. Others give an opportunity to set a limit over sound level, but it requires the deliberate attention of the user. One has to find the appropriate option in the device menu and turn it on. It can be assumed that many people do not even know or care about the existence of such an option to decrease the output level.

This paper is an extended and revised version of the paper entitled: "Loudness Scaling Tests in Hearing Problems Detection", presented at the 58th AES Conference in Aalborg (Kostek *et al.*, 2015). The study involved 100 young people with normal hearing and the results obtained were used to explore which test signals were most frequently rated within a particular loudness category in the loudness scaling test. This issue is to be explained in Sec. 4.

2. Loudness growth

Diagnostic standard pure-tone threshold audiometry does not always result in an overall evaluation of hearing. A shift in the hearing threshold is not the only symptom of hearing loss. Incorrect perception of loud and very loud sounds can also be an important indicator of deterioration in hearing, i.e. loudness recruitment typically occur in the case of sensorineural hearing impairment (Moore, 2007; Keefer, 2012; Hood, 1977). This phenomenon is related to incorrect growth of loudness sensation, i.e. perception of loud sounds is too intense in relation to perception of such sounds in the case of normal hearing.

A typical test used to detect and measure the abnormal growth of loudness sensation is the binaural alternate loudness balance test (Allen *et al.*, 1990; Hood, 1977). This test assumes that the person tested has one ear healthy (with normal hearing). There is no such requirement in the case of the loudness scaling tests. Such tests consist in an assessment of the loudness sensation caused by test signals, which most often take the form of narrow-band noise with a different center frequency and different levels of sound. For the loudness assessment, the scale of loudness categories is defined according to the particular method. As mentioned before, the LGOB test is an example of such tests (Allen *et al.*, 1990). It is a relatively effective method, that may be easily implemented as a software application. In the LGOB procedure test signals are in the form of half octave narrow-band noise with the following center frequencies: 500 Hz, 1000 Hz, 2000 Hz and 4000 Hz. The sound level is changed gradually in steps of 5 dB in the range from 20 to 120 dB SPL. The test signal consists of three 0.5 s noise bursts separated by a 0.5 s silence gap (BSA, 2011). A waveform of the LGOB test signal is presented in Fig. 1.

The average RMS amplitude of the reference LGOB test signals is between −50 and −58 dBFS (i.e. dB Full Scale; 0 dBFS represents the highest possible level in digital system), loudness of these audio files varies from −7 to −10 LUFS (LUFS = Loudness Units relative to Full Scale). It should be reminded that LUFS is EBU loudness unit, related to full scale, described in EBU R 128 and ITU-R BS.1770 standards (EBU, 2010; Lund, 2007; ITU-R BS.1770). Silence gaps significantly influence these results. It is to observe that the average RMS amplitude for a single noise burst is between −6 and −9 dBFS and loudness measured varies from −5 to −8 LUFS. Overall, loudness normalization is concerned with balancing audio signal according to the actually perceived loudness.

The loudness sensation within the LGOB test is assessed using a 7-point loudness category scale, i.e.: I CAN'T HEAR (0), VERY SOFT (1), SOFT (2), COMFORTABLE (3), LOUD (4), VERY LOUD (5), TOO LOUD (6). The test results are compared with the reference loudness scaling functions for people with normal hearing.

LGOB test results allow for assessing perception of loudness sensation for the threshold levels, but also for comfort hearing levels and uncomfortable levels (Brand, 2000, 2007; Jaroszewski *et al.*, 1998). In addition, the use of individual loudness scaling turns out to be advantageous for the fitting of hearing aids (Brand, 2007). Moreover, as mentioned before, the comparison of the obtained loudness scaling function and the reference function can be used to detect the loudness recruitment problem. If the obtained function increases faster than the reference function, the loudness recruitment problem is observed. That refers to the sensorineural hearing impairment and the loudness recruitment effect (see examples in Fig. 2).

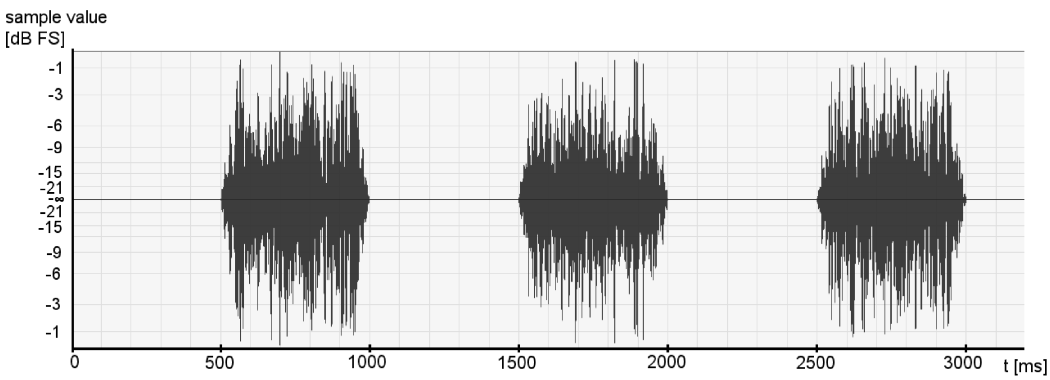


Fig. 1. An example of the LGOB test signal (filter center frequency: 1000 Hz).

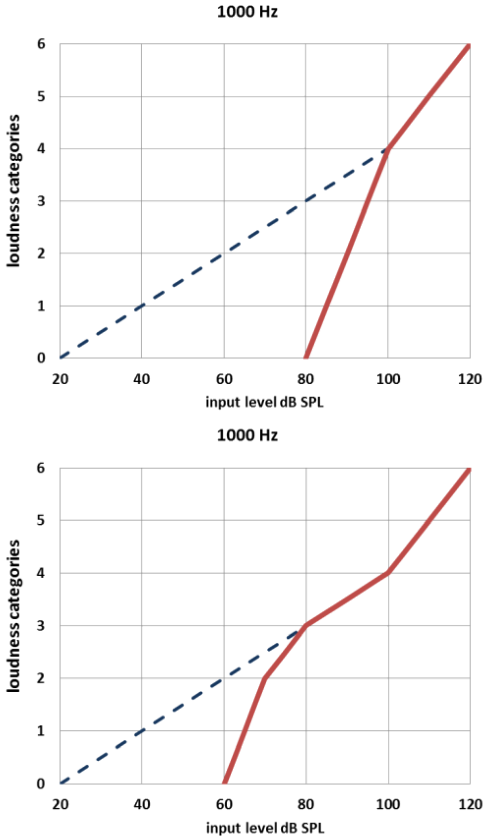


Fig. 2. Examples of the loudness recruitment effect (dashed line – the reference function, y-axis corresponds to the loudness sensation categories: I CAN'T HEAR (0), VERY SOFT (1), SOFT (2), COMFORTABLE (3), LOUD (4), VERY LOUD (5), TOO LOUD (6).

It should be noted that a normal ear is one not only with good hearing but also with a full dynamic range for different intensities of sound. Contrarily, a recruiting ear is one in which the dynamic range is narrowed or contracted. This loss of compression is often stated to be consistent with the presence of recruitment in listeners with cochlear hearing losses (Moore *et al.*, 1999). Some researchers also argue that the amount of loudness recruitment caused by cochlear hearing loss is closely related to the amount of hearing loss (Moore *et al.*, 1999; Miskolczy-Fodor, 1960).

However, Buus' and Florentine's findings indicate that it is likely to be more appropriate to define recruitment as an abnormally large loudness at an elevated threshold instead of as an abnormally rapid growth of loudness above an elevated threshold (BUUS, Florentine, 2001).

3. Method proposed

Digital technology, especially the omnipresent multimedia technologies allow for the implementation of accessible tools for audiological tests. Usually, such tools are used only for screening, although quality of the audio path of modern multimedia devices is sufficiently high to perform audiological measurements. As a result, it is possible to create an application to conduct tests that can provide reliable results. For the purpose of the carried out investigations the LGOB and the proposed test were implemented on the PC platform. High quality of sound is provided by an external sound card with a 24-bit converter and clinical audiometric headphones (Tonsil SD-307) integrated with the sound level calibrator. The calibrator is a kind of a sound analyzer that measures the RMS of the acoustical signal. It is equipped with diodes that light up when a certain reference level of a signal is achieved. This device was set up together with the audiometric headphones and the artificial ear for reference (Czyżewski *et al.*, 2000; Suchomski *et al.*, 2008). This configuration enables to obtain sound levels from approx. 30 dB SPL to 110 dB SPL. The calibration procedure is further presented in Subsec. 4.2.

Before the proposed method is presented, some additional issues should be considered at the stage of assumptions for building a loudness scaling test. Loudness is a subjective measure related to the properties of human hearing, as well as individual evaluation of pitch, volume, tone, and duration of sounds (Krumhansl, Iverson, 1992; Moore, 2012). Thus for loudness level assessment we use phon, a loudness unit related to sound levels in dB SPL with regard to psychophysical properties of the human ear, or sone to describe perception of loudness. The sound level in dB SPL is the objectively measured quantity to be measured and analyzed. Also, when comparing CLS methods such issues as stimulus parameters, randomization of stimulus levels and frequencies, application of loudness categories, instructions to the test subjects, etc. should be taken into account (Oetting *et al.*, 2014). Overall, the accuracy of loudness scaling tests is problematic as the scale used for assessment of the loudness sensation depends on subjective perception of a person. A lot of examined persons have problems with a correct interpretation of loudness categories (Brand, 2007). Therefore, the results obtained in two CLS methods are comparable only to some extent.

As mentioned before, the main purpose of using loudness scaling tests is to determine the loudness growth function which enables to optimally set the parameters of dynamic range compression in hearing aids (Pettit, Keefer, 2011; Keefer, 2012). "Classical" loudness scaling tests have several disadvantages (Allen *et al.*, 1990) of which the long time required to obtain the results is the most important one. In addition, test signals (mostly in a form of narrowband noise) are considered as unpleasant. Furthermore, these tests require attention span from an examined person. A typical LGOB test lasts for about 10 minutes (for one ear). During this time, an examined person has to focus on the task: the loudness sensation evaluation by selecting the appropriate loudness category from the proposed scale. Taking all the above mentioned factors, it is desirable to enhance the performance of such a testing. It also important to reduce time required to complete the test. Artificial test signals (filtered noise samples) may be replaced by more naturally sounding signals, e.g. recordings of musical instruments can be used, as such sounds are easily accepted. Furthermore, signal parameters, i.e. frequency parameters (e.g. a desired bandwidth) and amplitude, can be controlled, however as stated by Klonari *et al.* (2011) the problem of loudness estimation and the control of musical sounds has long been an issue in various applications of audio engineering and technology.

To reduce time needed for completing the test one can simply increase the step size from 5 dB to 10 or 15 dB, but this is a trivial operation. A more sensible way to reduce test time is to analyze results obtained in tests with normal hearing people and selecting only these test signals of particular levels, which are most frequently rated within a given loudness category. This may result in reducing test time even up to three times. For reducing the number of tests signals a group of 100 persons with normal hearing was tested in an earlier stage of this research study, and the main outcome was an indication of levels that were most frequently rated by the subjects within a given loudness category (Kostek *et al.*, 2015). Figure 3 shows an example of assigning the levels of test signals derived from the loudness scaling results for subjects with normal hearing. The dashed vertical lines show the selected levels of test signals for a given frequency band. Analyzing results of loudness scaling tests for each loudness category can help to select one level of the test signal which correlates loudness sensation evaluated within a particular loudness category with indications of normal hearing subjects. In this way five levels (one per each loudness category) can be determined.

In the LGOB test, one of the problems is related to the interpretation of loudness categories, such as: VERY SOFT and SOFT and also LOUD, and VERY LOUD (Brand, Hohmann, 2001). Therefore, for simplicity and for ambiguity avoidance, loudness category scale can be reduced to five labels: I CAN'T HEAR (0), SOFT (1), COMFORTABLE (2), LOUD (3), TOO LOUD (4). The test proposed by GUT contains test signals in the form of musical instrument sounds with bandwidth limited to one octave to preserve their naturalness. The preliminary tests showed that a narrow bandwidth (e.g. 1/2 octave bands such as in the LGOB test) adversely affects sound naturalness. Frequency bands of test signals have been limited using band-pass filtering. The filter slope is not greater than 24 dB/oct. to preserve the naturalness of sound. The length of each test signal is equal to 3 s. An example of the waveform of the test signal (piano) is presented in Fig. 4.

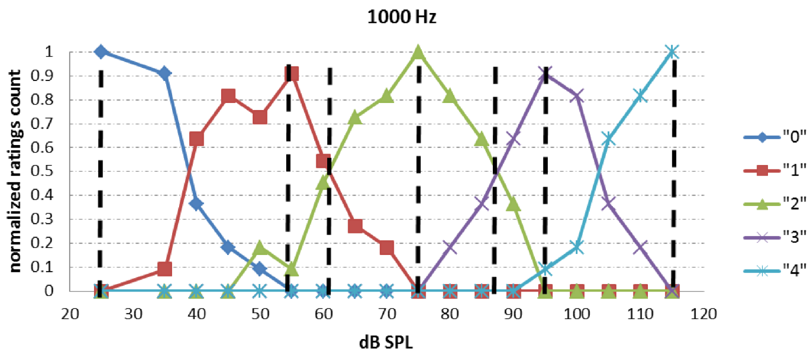


Fig. 3. An example of the loudness function for loudness categories obtained for 1000 Hz, y-axis corresponds to the ratio of the given category ratings for the test signal to the number of all evaluations carried out for this category; dashed vertical lines show the selected levels of test signals for a given frequency band.

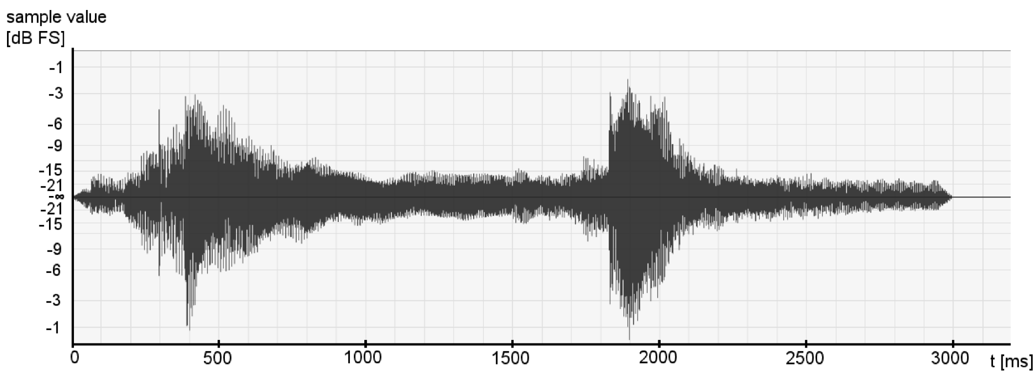


Fig. 4. An example of the test signal (piano, 1000 Hz).

The average RMS amplitude of the reference signals was equal to −17 dBFS, loudness of these audio files varied from −10 to −8 LUFS. The reference signals correspond to 110 dB SPL. For each frequency band, seven test signals were prepared by attenuating the reference signals. The silence signal was the eighth signal. They were as follows: drums (center frequency 500 Hz; −10.0 LUFS), piano (center freq. 1000 Hz; −10.01 LUFS), electric guitar (center freq. 2000 Hz; −8.90 LUFS) and violin (center freq. 4000 Hz; −8.62 LUFS). Differences in values result from the fact that it was not possible to increase the level of musical instrument sounds without dynamic compression. Besides, additional processing could affect the naturalness of the sounds prepared.

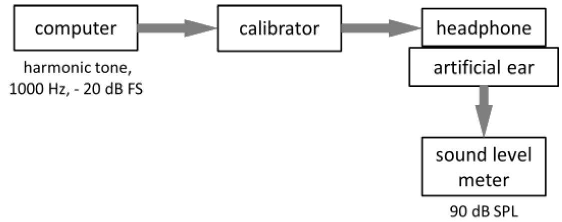
4. Loudness scaling tests

4.1. Subjects

In order to verify the test developed, first a hearing examination was performed. Two groups of subjects were chosen: without and with hearing impairments. The first group consisted of 15 GUT students of age of 21–25 years, including five women. All subjects had hearing thresholds ≤ 20 dB HL in the frequency range from 0.5 to 8 kHz, based on standard pure-tone audiometry. Thresholds were measured in 5-dB steps, following the standard clinical audiometric procedures. The second group comprised five people with diagnosed a hearing impairment, aged between 43–87 years (four men, one woman). Both audiometric and LGOB-based tests have been carried out in two silent, acoustically treated rooms, with a background noise level below 30 dB SPL (100–10000 Hz). Due to time constraints, only one ear was examined for each individual.

4.2. Calibration procedure

Tests were performed in a laboratory with a partly noise-insulated computer workstation. The workstation was equipped with an external sound card and clinical audiometric headphones (Tonsil SD-307) integrated with the sound level calibrator. The calibration procedure consists of two stages (Suchomski *et al.*, 2008). During the initial calibration, the headphones (integrated with the sound level calibrator) are coupled with an artificial ear and acoustic analyzer (Fig. 5). Through the analyzer, measurements of the acoustic pressure level are made. The calibration test signal (a tone at a frequency of 1000 Hz with level −20 dB FS (Full Scale)) is generated in the computer. The level of acoustic pressure is adjusted using the internal potentiometer of the calibrator to obtain value of 90 dB SPL. In this way, it is possible to generate signals with the maximum value of 110 dB SPL. Additional potentiometer is used to set the green LED on the calibrator housing lightning. The LED indicates that the required level is achieved.

a)


b)


Fig. 5. Calibration procedure (a), view of headphones (Tonsil SD 307) and calibration (b).

On the user's side no costly artificial ear or sound level meter are required. After the application is activated, it requires playing back the calibration signal (the same as in the previous stage). The operator/user has to adjust the software fader until the LED on the calibrator lights up. Then the system is ready to carry out the tests (Suchomski *et al.*, 2008).

4.3. Measurement procedure

Each subject performed two tests: LGOB and the proposed one. The order of tests was determined randomly for each subject to avoid any dependencies between tests. In addition, tests were repeated after seven days to verify consistency and repeatability of gathered data. During the second examination, the test order was reversed for each subject. Stimuli were presented monaurally. Only one ear (right or left) was examined.

In the experiment, the subjects were asked to wear headphones and to assess loudness of the presented sound samples according to the given categories. The subject had to choose the category by clicking on an icon related to a particular category. The developed application was used during the experiment. All subjects' answers were stored in text files. The average time to complete the LGOB test for one ear was 10 minutes whereas for the developed test was equal to about 4 minutes. Moreover, subjects found the new test signals more pleasant than the noise signals.

4.4. Reference functions

Based on the results of people with normal hearing, reference functions of loudness scaling were obtained and are shown in Fig. 6. The reference function was calculated as follows:

- each loudness scaling result consisted of three parameters: selected loudness category, frequency band and the sound level of the test signal,
- for each frequency band loudness scaling results obtained for normal hearing people were gathered,
- next, for each sound level used in the test a median value of the results obtained for this sound level was calculated,
- the last step involved connecting the calculated results points on a graph, including standard deviation.

During the verification process of the developed test, results for the proposed method were compared with results obtained in the LGOB test. Examples of results of correct (normal hearing) loudness scaling are presented in Fig. 7.

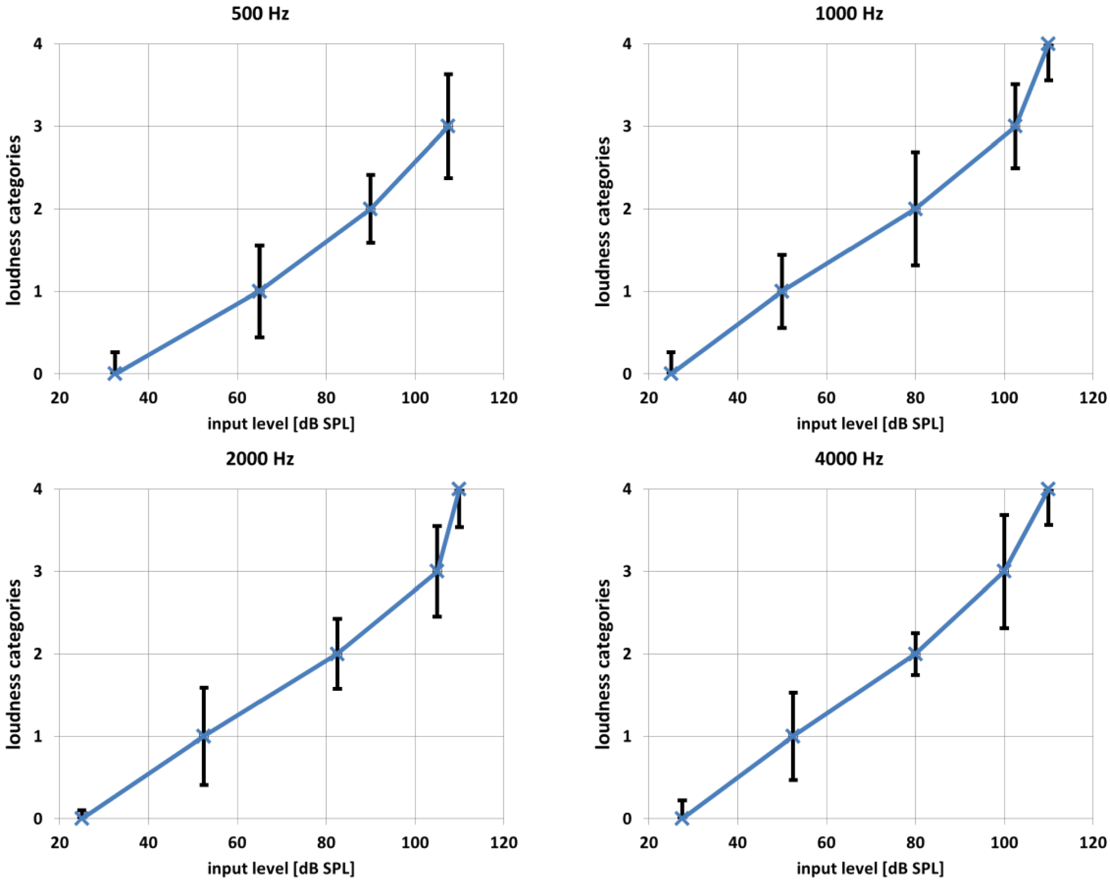


Fig. 6. Reference functions obtained using the proposed test procedure (it should be noted that the vertical axis is different from the LGOB test, i.e. 0 corresponds to I CAN'T HEAR, 1 – SOFT, 2 – COMFORTABLE, 3 – LOUD, 4 – TOO LOUD).

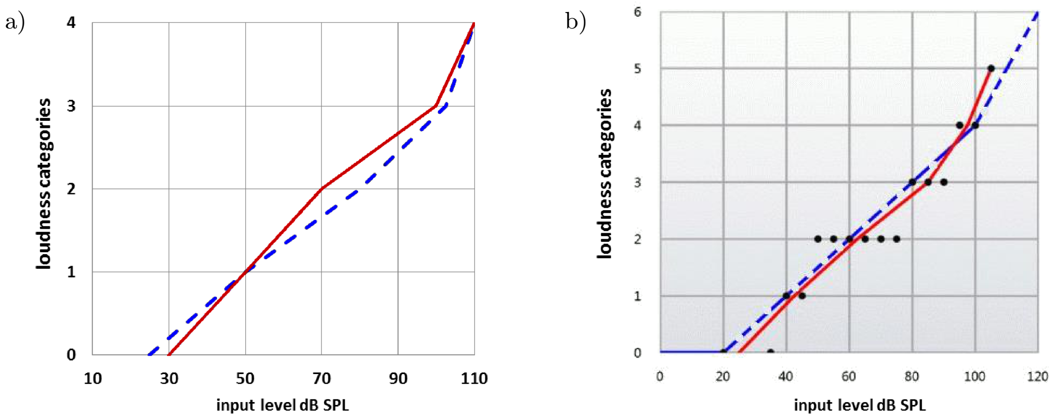


Fig. 7. Examples of loudness scaling results: 1000 Hz, normal hearing, a) obtained in the developed test – vertical scale as in Fig. 6; b) obtained in the LGOB test – vertical scale as denoted in Fig. 2; the dashed line refers to the LGOB reference function.

4.5. Results

As mentioned before, a characteristic reflecting loudness recruitment is characterized by the steep slope of the function obtained (there is also a raised threshold of hearing). This indicates a distorted (compressed) loudness perception. Such a phenomenon is typically observed in the case of sensorineural hearing loss as presented in Fig. 8. However, an increased loudness sensation may also occur for subjects with normal hearing. It can be related to hyperacusis or resulted from excessive music-listening. An example of hearing test results of a person with such a problem is shown in Fig. 9. In such as case, for finding the cause of the loudness growth shape, additional diagnostic tests should be performed, such as for example UCL audiometric measurement and otoacoustic emission (OAE) to determine whether outer hair cells damage is present or absent.

In Figs. 8 and 9 audiogram charts including air and bone conduction measurement data are presented first, then data for two LGOB-based tests are included for a subject with hearing loss and a normal hearing person. Due to time constraints audiometric measurements for

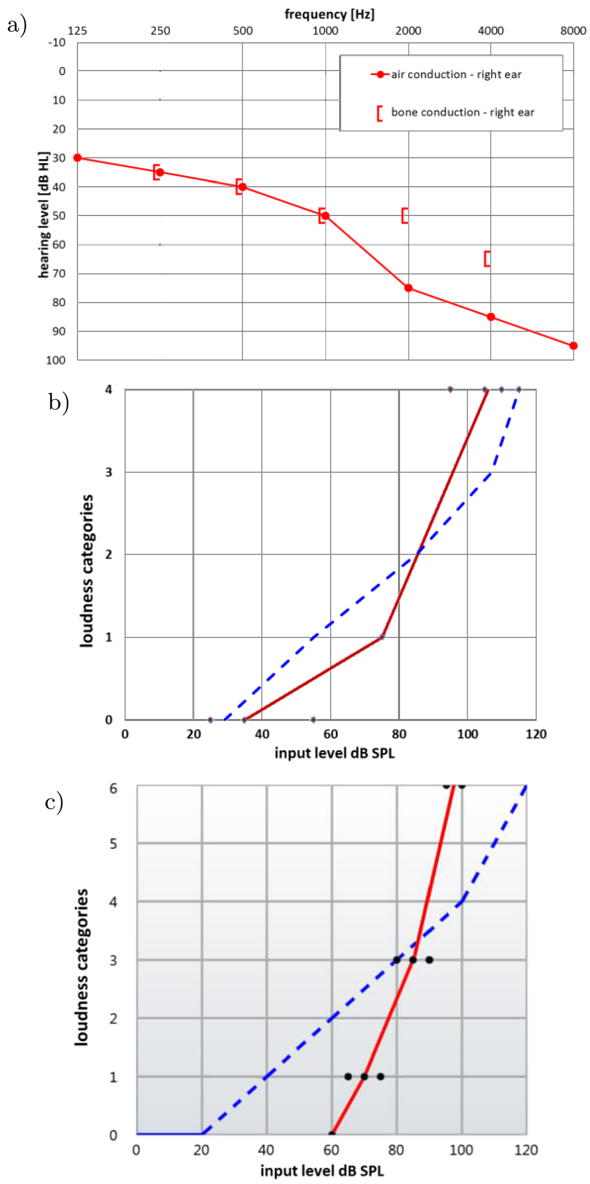


Fig. 8. An example of loudness recruitment effect for the person with sensorineural hearing loss: a) air and bone conduction hearing test results for the right ear; b) loudness function (2000 Hz) resulted from the proposed procedure, denotations as in Fig. 6, c) data obtained in the LGOB test (2000 Hz), denotations as in Fig. 2; the dashed line refers to the LGOB reference.

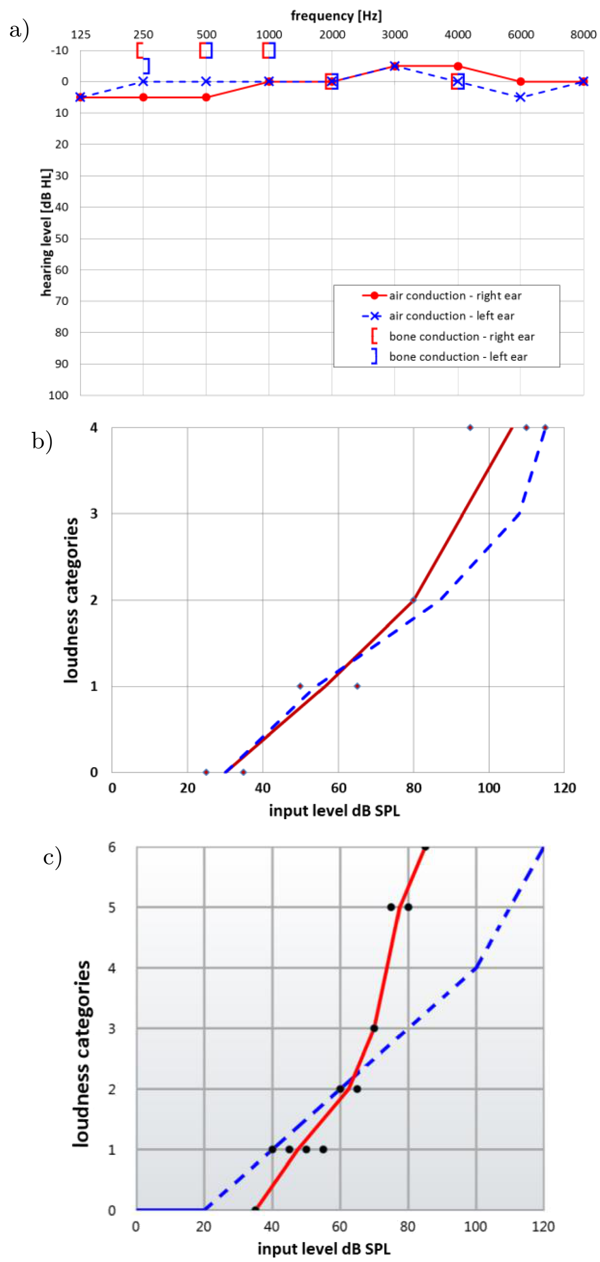


Fig. 9. An example of loudness function for the subject with normal hearing: a) results of the audiometric test, b) loudness scaling results (2000 Hz) obtained in the developed test, vertical scale as in Fig. 6, c) loudness scaling results (2000 Hz) obtained in the LGOB test; the dashed line refers to the LGOB reference.

the person with hearing loss were performed for right ear only.

To recall some background notions a short description of audiometric measurements is provided. In audiometric examination pure tones of varying intensities are delivered via air conduction, using the audiometric headphones or bone conduction by placing a bone conductor behind the ear, which stimulates the cochlear directly without going through the middle ear (Hearing Loss Examples; Franks, 2015; ISO8253-1, 2010). The pure-tone audiogram allows determination of thresholds at the given test frequencies. An audiogram shows frequency in Hz increasing from left to right as a logarithmic scale, so that there is equal distance between the octaves. Thresholds have units dB HL (hearing level). Thresholds for the right ear are drawn as circles, in red, while thresholds for the left ear are drawn as x's. The average "normal" threshold is represented as a horizontal line at the top of the plot, and the degree of the hearing loss is indicated by how much the threshold falls below this line (ISO 8253-1, 2010). It is important to note that the audiometric measurements and the loudness scaling tests use different types of signals (pure tones vs. narrow-band signals). Therefore, obtained results cannot directly be compared.

4.6. Statistical analysis

An accurate comparison of the two loudness scaling tests is to some degree difficult, because different loudness scales and different test signals are involved. It should also be remembered that in subjective listening, as is in case of the LGOB test, the results may vary each time they are repeated within a given category. However, it can be expected that the shape of the obtained characteristics for each frequency band in both tests should be comparable. Also, a range of sound levels and frequencies bands are similar.

To perform a statistical analysis of gathered data, the Mann-Whitney U-test was used. There are some basic assumptions before running the Mann-Whitney U-test, such as: one dependent variable is measured at the ordinal level, one independent variable consists of two categorical, independent groups and there is no relationship between the observations in each group of the independent variable or between the groups themselves, the distribution of scores for both groups for the independent variable has the same shape (or a different shape), i.e. the null hypothesis is the distributions of the two groups are equal.

In order to employ the Mann-Whitney U-test, the following assumptions were made before running a test:

- the number of loudness categories in the LGOB was reduced to five – according to the loudness scale used in the proposed method. It means that VERY SOFT and SOFT categories were assigned to one category (SOFT). Similarly, categories labelled as LOUD and VERY LOUD were reduced to category LOUD,
- only sound levels occurring in both methods are used in the analysis. The number of sound levels was equal to eight for each frequency band, but the particular values differed,
- only results from the second session were used. During the second session subjects were familiar with the test procedure, hence the number of potential mistakes was lower.

The results for people with normal hearing are shown in Table 1, for the group with hearing loss – in Table 2.

Table 1. The Mann-Whitney U-test results for people with normal hearing.

frequency: 500 Hz	level [dB SPL]	25	50	65	75	90	100	110	115
	p	0.3506	0.1577	0.5373	0.7376	**0.0341**	0.0654	**0.0278**	0.0626
frequency: 1000 Hz	level [dB SPL]	25	35	55	75	95	105	110	115
	p	0.1580	0.4620	0.0790	0.7350	0.0770	0.1890	0.7280	0.4330
frequency: 2000 Hz	level [dB SPL]	25	35	50	65	80	95	110	115
	p	0.5765	0.1570	0.3506	0.1577	0.0786	0.2913	0.2923	0.3856
frequency: 4000 Hz	level [dB SPL]	25	40	50	65	85	100	110	115
	p	0.3506	0.2132	0.3506	0.6920	0.7011	0.1146	**0.0277**	0.1950

Table 2. The Mann-Whitney U-test results for people with hearing loss.

frequency: 500 Hz	level[dB SPL]	25	50	65	75	90	100	110	115
	p	1	1	1	1	1	0.6825	1	1
frequency: 1000 Hz	level [dB SPL]	25	35	55	75	95	105	110	115
	p	1	1	1	1	0.4760	0.7620	1	1
frequency: 2000 Hz	level [dB SPL]	25	35	50	65	80	95	110	115
	p	1	1	1	1	0.5556	0.6349	0.4048	0.1667
frequency: 4000 Hz	level [dB SPL]	25	40	50	65	85	100	110	115
	p	1	1	1	0.5238	0.4444	1	0.6429	0.8730

Only in three cases (cells indicated with a grey background) the differences between results are statistically significant ($p < 0.05$). It means that the data derived from tests provide little or no evidence that the null hypothesis (both methods give identical or similar results) is false.

5. Conclusions

The main purpose of this paper was to design a novel way to test loudness sensation in a reduced time compared to the standard LGOB test, and at the same time to make it more acceptable with regard to stimuli utilized. The procedure proposed uses band filtered music excerpts instead of narrow-band noise.

Within the carried out investigations both tests, i.e. LGOB and the test developed were implemented on the PC platform. Both applications were calibrated to obtain comparable results using audiometric headphones integrated with the sound level calibrator and coupled with an artificial ear and an acoustic analyzer. Two groups of subjects participated in the carried out study, young persons with normal hearing and individuals with a hearing loss.

The data obtained for both groups were analyzed statistically employing the Mann-Whitney U-test. This was used to verify whether there are no significant differences between obtained results. The data analyzed provide little or no evidence that the null hypothesis, i.e. both methods give similar results, is false.

The results show that the effect of the loudness recruitment is visible for hearing impaired persons in the test results obtained both in LGOB and the developed test. Furthermore, the proposed test allows for observing atypical loudness sensation function for a person with normal hearing.

In addition to the main objective of this paper, it was observed that the designed application may be useful for checking an individual loudness growth function and setting the comfort level of listening to music. It should be remembered that audiological tests require calibration of sound level. Also, reference characteristics should be available. This is due to the fact that the obtained results are usually compared with reference results. That's why an implementation of such tests on any audio device may be difficult. This problem can however be solved either by calibrating an audio device at the production stage or equipping appropriately calibrated headphones along with a calibration unit, i.e. a small electronic device designed to control of sound level. However, in the case of a simplified form of loudness scaling test as shown above, there is no need to perform the calibration of sound level, because the obtained results do not require comparing them with the reference data. The results of loudness scaling test serve as input parameters for the algorithm processing the dynamics of sound according to the user's hearing preferences. This type of test can be implemented on any audio device, also mobile, e.g. smartphones or tablets. Moreover, as discussed by Udesen *et al.* (2015) sound-based psychoacoustic tests may be affected by the room environment in which the tests are conducted, thus this is another aspect that should be taken into account in such measurements.

Lastly, it seems that there is a possibility to use one category describing 'uncomfortable' sound level instead of LOUD and TOO LOUD categories. This was further investigated in the context of setting too soft, comfortable and too loud sound levels when listening to music using one's laptop. Resulting from subjective tests performed on a number of young people the user interface was designed (see Fig. 10), which in addition shortened the loudness test approx. to 2 minutes.

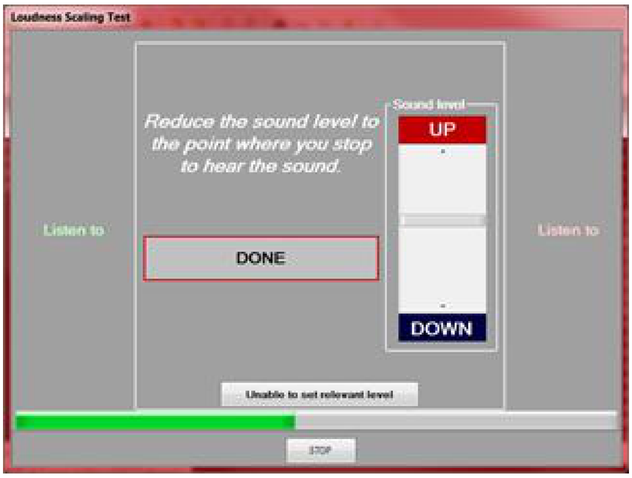


Fig. 10. Graphical user interface for loudness level setting.

Acknowledgments

This research was partially supported by the grant No. PBS1/B3/16/2012, entitled "Multimodal system supporting acoustic communication with computers", financed by the Polish National Centre for R&D.

References

1. Allen J.B., Hall J.L., Jeng P.S. (1990), *Loudness growth in 1/2 octave bands (LGOB) – A procedure for the assessment of loudness*, Journal Acoust. Soc. Am., **88**, 2, 745–753.
2. Al-Salim S.C., Kopun J.G., Neely S.T., Jesteadt W., Stiegemann B., Gorga M.P. (2010), *Reliability of categorical loudness scaling and its relation to threshold*, Ear Hear., **31**, 4, 567–578, doi: 10.1097/AUD.0b013e3181da4d15.
3. Brand T. (2000), *Analysis and optimization of psychophysical procedures in audiology*, Dissertation, Carl von Ossietzky Universität, Physics, Oldenburg, Bibliotheks- und Informations system der Universität Oldenburg.

4. Brand T. (2007), *Loudness Scaling*, 8th EFAS (European Federation of Audiological Societies) Congress/10th Congress of the German Society of Audiology, 1–7, Heidelberg, Germany, 6–9 June.

5. Brand T., Hohmann V. (2001), *Effect of hearing loss, centre frequency, and bandwidth on the shape of loudness functions in categorical loudness scaling*, Audiology, **40**, 2, 92–103.

6. Brand T., Hohmann V. (2002), *An adaptive procedure for categorical loudness scaling*, J. Acoust Soc. Am., **112**, 1597–1604.

7. BSA (2011), *Recommended Procedure Determination of Uncomfortable Loudness Levels*, British Society of Audiology, http://www.thebsa.org.uk/wp-content/uploads/2014/04/BSA_RP_ULL_FINAL_24Sept11.pdf (accessed Feb., 2016).

8. Bulla W.A., Hall J.W. III (1996), *Daily Noise-Level Exposures of Professional Music Recording Engineers*, 105th AES Convention, preprint no. 4792, San Francisco, USA.

9. Buus S., Florentine M. (2001), *Growth of Loudness in Listeners with Cochlear Hearing Losses: Recruitment Reconsidered*, JARO 03: 120–139, 2001, doi: 10.1007/s10162001008.

10. Czyżewski A., Kostek B., Suchomski P. (2000), *Expert System for Hearing Aids Fitting*, 108th Audio Eng. Soc. Convention, preprint 5094, 1–14, Paris, France, 19.2.2000–22.2.2000.

11. Dibble K. (1995), *Hearing Loss & Music*, J. of Audio Engineering Society, **42**, 251–266.

12. EBU – TECH 3341 (2010), Loudness Metering: ‘EBU Mode’ metering to supplement loudness normalization in accordance with EBU R 128, Geneva, Switzerland.

13. Elberling C. (1999), *Loudness scaling revisited*, J. Am. Acad. Audiol., **10**, 248–260.

14. Franks J.R. (2015), *Hearing Measurement*, National Institute for Occupational Safety and Health, http://www.who.int/occupational_health/publications/noise8.pdf (accessed Feb., 2016).

15. Hearing Loss Examples (2016), http://www.chime-health.co.uk/web/data/audiogram-hearing-loss-examples-2.pdf (accessed Feb., 2016).

16. Heller O. (1985), *Hörfeldaudiometrie mit dem Verfahren der Kategorienunterteilung (KU)*, Psycholog Beiträge, **27**, 478–493.

17. Hellbrück J., Moser L.M. (1985), *Hörgeräte Audiometrie: Ein computergestütztes psychologisches Verfahren zur Hörgeräteanpassung*, Psychologische Beiträge, **27**, 494–508.

18. Hood J.D. (1977), *Loudness Balance Procedures for the Measurement of Recruitment*, Audiology, **16**, 215–228.

19. HörTech Center of Competence, http://www.hoertech.de/web_en/produkte/messverfahren/skalierung.shtml (accessed Feb., 2016).

20. ISO 16832 (2006), Acoustics – Loudness scaling by means of categories. International Organization for Standardization, Geneva, Switzerland.

21. ISO 8253-1 (2010), Acoustics – Audiometric test methods. Part 1 – Basic pure tone air and bone conduction threshold audiometry, http://www.isa-audiology.org/standards.asp (accessed Feb., 2016).

22. ITU-R BS.1770 Rec. (2011), Algorithms to measure audio programme loudness and true-peak audio level International Telecommunications Union.

23. Jaroszewski A., Fidecki T., Rogowski P. (1998), *Hearing Damage from Exposure to Music*, Archives of Acoustics, **23**, 1, 3–31.

24. Keefer R. (2012), *Speed and Accuracy in Loudness Scaling*, The Hearing Review, http://www.hearing-review.com/2012/05/speed-and-accuracy-in-loudness-scaling/ (accessed Feb., 2016).

25. Klonari D., Pastiadis K., Papadelis G., Papanikolaou G. (2011), *Loudness Assessment of Musical Tones Equalized in A-weighted Level*, Archives of Acoustics, **36**, 2, 239–250, doi: 10.2478/v10168-011-0019-7.

26. Kostek B., Odya P., Suchomski P. (2015), *Loudness Scaling Tests in Hearing Problems Detection*, 58th Audio Eng. Soc. Conference, Aalborg, Denmark, June 28–30.

27. Kozlowski E., Mlynski R. (2014), *Effects of Acoustic Treatment on Music Teachers' Exposure to Sound*, Archives of Acoustics, **39**, 2, 159–163.

28. Krumhansl C.L., Iverson P. (1992), *Perceptual Interactions Between Musical Pitch and Timbre*, J. Experimental Psychology, Human Perception and Performance, **18**, 3, 739–751.

29. Lund T. (2007), Level and Distortion in Digital Broadcasting EBU Technical Review.

30. Marozeau J., Florentine M.J. (2007), *Loudness growth in individual listeners with hearing losses: a review*, J. Acoust. Soc. Am. Sep., **122**, 3, EL81.

31. Miskolczy-Fodor F. (1960), *Relation between loudness and duration of tonal pulses. III. Response in cases of abnormal loudness function*, J. Acoust. Soc. Am., **32**, 486–492.

32. Moore B.C.J. (2012), *An Introduction to the Psychology of Hearing*, 6th edition, Bingley: Emerald Group Publishing Ltd., Leiden Boston.

33. Moore B.C.J. (2007), *Cochlear Hearing Loss: Physiological, Psychological and Technical Issues*, J. Wiley & Sons, Chichester, England.

34. Moore B.C.J., Glasberg B.R., Stone M.A. (1999), *Use of a loudness model for hearing aid fitting. III. A general method for deriving initial fittings for hearing aids with multi-channel compression*, Br. J. Audiol., **33**, 241–258, doi: 10.3109/03005369909090105.

35. Oetting D., Brand T., Ewert S.D. (2014), *Optimized loudness-function estimation for categorical loudness scaling data*, Hear Res. 2014 Oct; **316**, 16–27. Epub, July 21.

36. Pawlaczyk-Luszczynska M., Zamojska M., Dudarewicz A., Zaborowski K. (2013), *Noise-Induced Hearing Loss in Professional Orchestral Musicians*, Archives of Acoustics, **38**, 2, 223–234.

37. Pettit C., Keefer R. (2011), *New computer-aided measurement system for hearing aid fitting*, Hearing Review., **18**, 5, 38–46.

38. Rasetshwane D.M., Trevino A.C., Gombert J.N., Liebig-Trehearn L., Kopun J.G., Jesteadt W., Neely S.T., Gorga M.P. (2015), *Categorical loudness scaling and equal-loudness contours in listeners with normal hearing and hearing loss*, J. Acoust. Soc. Am., **137**, 1899, doi: 10.1121/1.4916605.

39. Suchomski P., Kostek B., Czyżewski A. (2008), *Hearing Aid Fitting Method Based on Fuzzy Logic Processing*, Archives of Acoustics, **33**, 4, 153–158.

40. Udesen J., Piechowiak T., Gran F. (2015), *The Effect of Vision on Psychoacoustic Testing with Headphone-Based Virtual Sound*, J. Audio Eng. Soc., **63**, 7/8, 552–561, doi: 10.17743/jaes.2015.0061

41. Vogel I., Brug J., Hosli E.J., van der Ploeg C.P.B., Raat H. (2008), *MP3 Players and Hearing Loss: Adolescents' Perceptions of Loud Music and Hearing Conservation*, Journal of Pediatrics, **153**, 3, 400–404.