OPEN ACCESS Freely available online
Journal of the Audio Engineering Society
Vol. 64, No. 6, June 2016 (© 2016)
DOI: http://dx.doi.org/10.17743/jaes.2016.0014

Adaptive Personal Tuning of Sound in Mobile Computers

ANDRZEJ CZYZEWSKI,[1] *AES Fellow*, **ANDRZEJ CIARKOWSKI**[1], **BOZENA KOSTEK,**[2] *AES Fellow*,
(ac@pg.gda.pl)

JOZEF KOTUS,[1] *AES Member*, **KUBA LOPATKA**[1], **AND PIOTR SUCHOMSKI**[1]

[1]*Gdansk University of Technology, Faculty of Electronics, Telecommunications and Informatics, Multimedia Systems Department, Narutowicza 11/12 80-233 Gdansk, Poland*
[2]*Gdansk University of Technology, Faculty of Electronics, Telecommunications and Informatics, Audio Acoustics Laboratory, Narutowicza 11/12 80-233 Gdansk, Poland*

An integrated methodology for enhancing audio quality in mobile computers is presented. The key features are adaptation of the characteristics of their acoustic track to changing acoustic conditions of the environment and to users' individual preferences. Signal processing algorithms are introduced that concern: linearization of frequency response, dialogue intelligibility enhancement, and dynamics processing tuned up to the users' hearing characteristics. The description of the algorithms implemented in the C++ programming language is provided. The processing is performed utilizing custom Audio Processing Objects (APO) installed in the Windows sound system. The sound enhancement package is managed with the user interface enabling a control over the sound system. The results of subjective evaluation of the sound processing methods and algorithms introduced to mobile computer devices are discussed.

0 INTRODUCTION

Mobile computer devices—tablets, ultrabooks, netbooks—are now frequently used both for work and for entertainment purposes. Sound quality is an important factor influencing the user's experience in such use cases as: VoIP conversations, listening to music, or film playback. Meanwhile, the small size and low production cost of the above said devices often lead to deterioration in the quality of sound. The factors causing the degraded sound quality are related mostly to their mechanical form, i.e., the limited size of loudspeakers, the presence of resonances of the casing, etc.

The manufacturers of mobile devices aim to improve the sound quality by two types of means. The first approach is installing better quality transducers. There are devices available on the market that contain loudspeakers produced by manufacturers of hi-end audio devices. The second approach is employing a software process whose purpose is to improve the perceived quality of sound. Turnbull et al. noticed this problem and had employed signal processing algorithms to enhance the sound quality in mobile phones and game consoles [1]. Several models of laptops or ultrabooks are equipped with sound enhancement bundles by such companies as Dolby [2] or Waves [3].

In general, the solutions described in the literature incorporate the following features:

- Low frequency enhancement [1][2][3][4],
- Spatialization [2][3][4][5][6],
- Improvement of dialogue clarity [2][3],
- Dynamics processing (most frequently in the form of loudness maximization) [1][2][3],
- Compensation of irregular frequency response characteristics [1][2][3][5].

In our research we developed our own approaches to some of the above mentioned problem solving. In this work we introduce our methods for: linearization of frequency response, improvement of dialogue intelligibility (also referred to as Smart Dialogue), and adjusting the dynamics of sound (also referred to as Ear Tune Up). In our related work the topic of low frequency enhancement is also addressed [7][8]. The novelty of our approach lies in the adaptive and in the same time personalized approach. We aim to adjust the sound to the changing conditions and to individual preferences of the listener. We also propose a novel approach to dynamic range adjustment, in which we do not maximize the loudness of sound, but we rely on investigated user's preferences concerning loudness. The algorithms are

implemented as Audio Processing Objects (APOs) in the Windows audio subsystem. This approach allows for real-time operation mode, independently of the source of audio data and it remains transparent to the user. This article is partially based on some previous convention papers [9][10]. Compared to the convention papers' content, the description of the methods were extended, the adaptive linearization was added, and new evaluation results were introduced.

All experiments and practical objective and subjective tests were performed not only on a laptop platform (HP Pavilion G6) but also on a portable device (Dell All-In-One). Algorithms applied during performed tests and obtained results are discussed in this paper.

The remainder of the paper is organized as follows. In Sec. 1 we introduce the method for linearization of frequency response. In Sec. 2 we describe the algorithm for enhancement of dialogue intelligibility. The methodology for personalized adjustment of the dynamics features is discussed in Sec. 3. In Sec. 4 the results of evaluation of the proposed methods are discussed. Sec. 5, concluding the article, contains some general comments pertaining the achieved results.

1 LINEARIZATION OF FREQUENCY RESPONSE

The purpose of the introduced linearization algorithm is to compensate for the irregular frequency response of the device. Let us assume that the digital signal $x(n)$ is played back by the device. After the playback a distorted signal $x^*(n)$ is obtained:

$$x^*(n) = x(n) * h(n) \tag{1}$$

where: $*$ denotes convolution and $h(n)$ is the impulse response of the audio playback system combined with the impulse response of the listening room. The aim of the linearization algorithm is to calculate the estimate of the impulse response $\hat{h}(n)$. Hence, the inverse form of this characteristic $\tilde{h}(n)$can be computed, such that:

$$\hat{h}(n) * \tilde{h}(n) = \delta(n) \tag{2}$$

or in the frequency domain:

$$\hat{H}(f) \cdot \tilde{H}(f) = 1 \tag{3}$$

Two methods are employed in order to estimate the characteristics of the playback system. In the static approach a calibration application is used to measure H(f). In the adaptive approach an adaptive filtration algorithm is employed to estimate the characteristics of the device on a frame-by-frame basis.

1.1 Static Approach

In this approach the auto-calibration method is utilized to adjust the linearization algorithm for current acoustic conditions. The details of the method have been presented already in related papers [10][11].

The linearization algorithm is the method for correction of amplitude response of the built-in speakers. In this approach we assumed that for given acoustic conditions the frequency response $H(f)$ of the speakers is constant. The auto calibration algorithm is utilized once to adjust the linearization algorithm for current acoustic conditions [11]. It is important to emphasize that the user should run the auto calibration process in order to take into account any changes in the acoustic conditions that may occur during the listening to music. Consequently, the calculated linearization filter characteristic will reflect not only the influence of ambient noise properties but also the current playback level. This type of linearization algorithm is realized according to the diagram in Fig. 1. The original signal $x(n)$ is modified using $\tilde{H}(f)$ function. After performing this operation we obtained the signal $\tilde{x}(n)$. This modified signal is played back through loudspeakers.

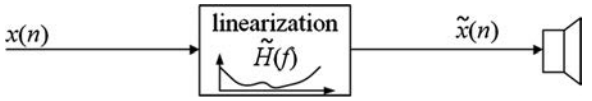


Fig. 1 Diagram of the static linearization algorithm.

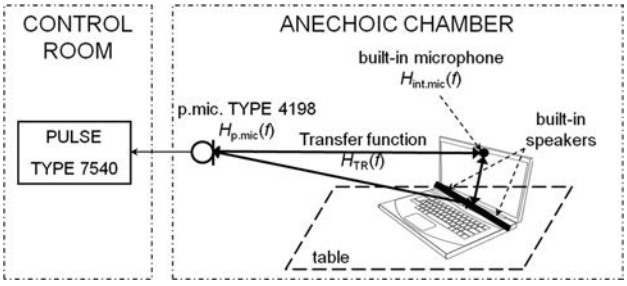


Fig. 2 Setup utilized for measurement of mobile device characteristics.

For the practical realization of the algorithms the acoustical features of the built-in loudspeakers and microphone are needed to be known. For considered computer devices these properties were determined during frequency response measurements. The details of this process are given in the next subsection.

1.1.1 Frequency Response Measurement

The proposed setup (presented in Fig. 2) enables the measurements of the frequency response of the device in the user's head position. $H_{p.mic}(f)$ denotes the response registered with the PULSE measurement system and $H_{\text{int.mic}}(f)$ denotes the frequency response of the internal microphone. It comprises the computer device (mobile or portable), measurement microphone connected to the measurement system PULSE, and the internal sound system of the mobile device (built-in microphone and speakers).

Based on the obtained responses: $H_{p.mic}(f)$ and $H_{int.mic}(f)$ it is possible to calculate the differential characteristic between them. This kind of a characteristic is called transfer function $H_{\text{TR}}(\text{f})$ and is given by Eq. (4):

$$H_{TR}(f) = H_{p.mic}(f) - H_{int.mic}(f) \tag{4}$$

Using the $H_{TR}(f)$ characteristic and the characteristic obtained by means of the built-in microphone it is possible to predict the amplitude characteristics of the computer

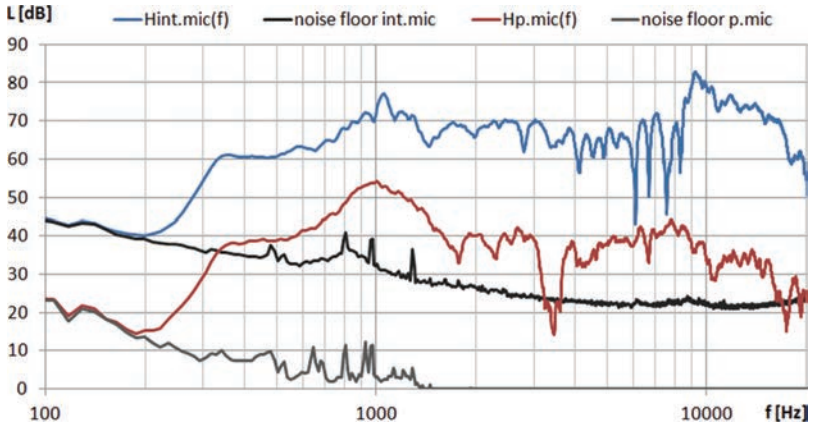


Fig. 3 Characteristics measured using the PULSE system ($H_{p.mic}(f)$ - red line) employing the built-in microphone ($H_{int.mic}(f)$ - blue line) are plotted. The noise floor for both microphone positions are shown as well.

speakers in the user's head position. The Transfer Function $H_{TR}(f)$ can be applied for self-calibration of the mobile device and for updating the linearization filter.

The measurement procedure was as follows. The examined device was placed on a table. The measurement microphone was placed approx. 0.5 m in front of the device and 0.5 m above the table. Its placement reflects the typical position of the user's head. The authors considered an application of a head and torso simulator to the calibration process, but it is a rather complex device with embedded, thus remaining beyond a control, characteristics of artificial ears that may differ from characteristics of the users' hearing. Moreover, the application of a single microphone for the calibration purpose is also the simplest and the best choice for practical reasons, therefore it remains quite common in measurements of loudspeaker characteristics. All measurements were conducted in an anechoic chamber (free field). The main purpose of measurement of the transfer function in the anechoic room was to avoid unwanted reflections and additional noise, which could occur in other conditions. Another reason is that the transfer function should be recorded as precisely as possible. It will be applied many times during the auto calibration process that can be performed by the user in different acoustic conditions.

The devices and software used during measurements included: Bruel & Kjær PULSE recorder with dedicated software (type 7540), Bruel & Kjær measurement microphone (type 4189-A-021), Bruel & Kjær acoustic calibrator (type 4231). Examined computer devices were: HP Pavilion G6, Dell All-In-One. The sampling rate of the recorded signals was equal to 48 kSa/s. The FFT analysis was performed using 4096-sample window length.

The measuring signal was played back through the computer sound system. The measuring signals radiated through the computer speakers are recorded by the measuring microphone and by the built-in microphone of the computer device. The results obtained for the laptop (HP Pavilion G6) are presented in Fig. 3.

Based on those signals the characteristics: $H_{int.mic}(f)$ and *noise floor int.mic* were calculated for the internal microphone and characteristics: $H_{p.mic}(f)$ and *noise floor p.mic* for the external microphone, respectively (see Fig. 3).

The external microphone was calibrated by means of the acoustic calibrator type 4231. Next, the sound levels for both signals were calculated. The correction factor was computed as a difference between the obtained sound levels. The reference value was the sound level measured with a measurement microphone. White noise was used to obtain the amplitude characteristics of the speakers and the transfer function.

The obtained characteristics show not only the frequency responses of the built-in computer device speakers, but also the influence of its casing and sound propagating conditions (i.e., comb filtration occurring in the acoustical environment).

The characteristic $H_{int.mic}(f)$ is rescaled in reference to the characteristics $H_{p.mic}(f)$ in that way that levels of both characteristics are equal for 1 kHz, thus we obtained the characteristic *int.mic.cal*. in this way (see Fig. 4). This action allows for a direct comparison of both characteristics ($H_{int.mic.cal.}(f)$ and $H_{p.mic}(f)$). In the next step, the characteristics are smoothed out. The moving average is used for this purpose. The window length of the moving average is dependent on the frequency value. The calculation of window length is based on the modified equivalent rectangular bandwidth (*ERB*) scale [12]

The original *ERB* value can be calculated according to Eq. (5) as follows:

$$ERB = 24.7 \cdot (4.37 \cdot F + 1) \tag{5}$$

where F is the frequency expressed in kHz.

In the practical realization of the moving average information about the beginning and end of the window length expressed by indexes of FFT components is needed. For that reason the length of the moving average (marked as Δf) is calculated according to Eq. (6) (the application of square root was verified in practice, since it brought proper results in the averaging process):

$$\Delta f = \sqrt{ERB \cdot \frac{f}{1000}} \tag{6}$$

where f is the frequency expressed in Hz.

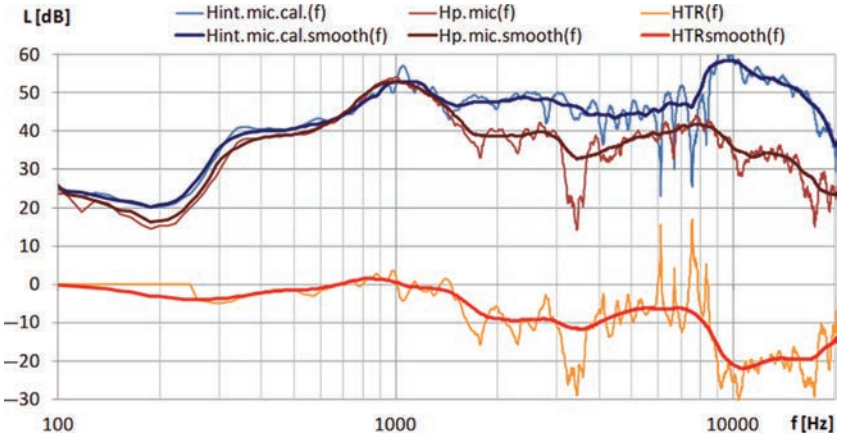


Fig. 4 Rescaled and smoothed characteristics measured using PULSE system and the built-in microphone. The calculated transfer function (red curve) was also shown.

After that, based on Δf value the indexes of begin (i_{begin}) and end (i_{end}) are calculated according to Eqs. (7) and (8):

$$i_{begin} = \left(i_f - \Delta f/2\right) \tag{7}$$

$$i_{end} = \left(i_f + \Delta f/2\right) \tag{8}$$

where: i_f is the number of the frequency component for the FFT frame (values from 0 up to 2047).

The characteristics: $H_{int.cal.smooth}(f)$ and $H_{p.mic.smooth}(f)$ were calculated in this way. Moreover, based on the smoothed characteristics $H_{p.mic}(f)$ and $H_{int.mic}(f)$, the transfer function $H_{TR}(f)$ is calculated and smoothed ($H_{TRsmooth}(f)$). The obtained results are shown in Fig. 4. After these steps the calculation of the linearization filter and the autocalibration of the computer device became possible.

The procedure of the automatic determination of the linearization filter response using the internal microphone built-in computer device is shown in Fig. 5. The whole process is based on the $H_{int.mic}(f)$ and $H_{TR}(f)$ characteristics. If $H_{int.mic}(f)$ and $H_{TR}(f)$ are known, the H(f) can be calculated using the formula in Eq. (9):

$$H\ (f) = H_{\text{int.mic}}\ (f) + H_{\text{TR}}\,(f) \tag{9}$$

In the block 3 determination of effective frequency range of the built-in computer device speakers, employing ambient noise is performed. As was shown before, the amplitude characteristic of the computer speakers measured in the users' head position is strongly irregular. For this purpose the frequency range is calculated on the basis of characteristic $H(f)$ (calculated in the block 2) and the ambient noise distribution (while the difference between $H(f)$ and ambient noise for a given frequency is greater than 10 dB). In our experiments we observed that the linearization range for the device HPG6 was: 300 Hz–15 kHz and 180 Hz–15 kHz for the Dell All-In-One computer.

In Fig. 6 the determination of the L_{MAX} and L_{ref} indicators were shown (block 4). L_{ref} indicator is determined according to Eq. (10):

$$L_{ref} = L_{\text{MAX}} - N \tag{10}$$

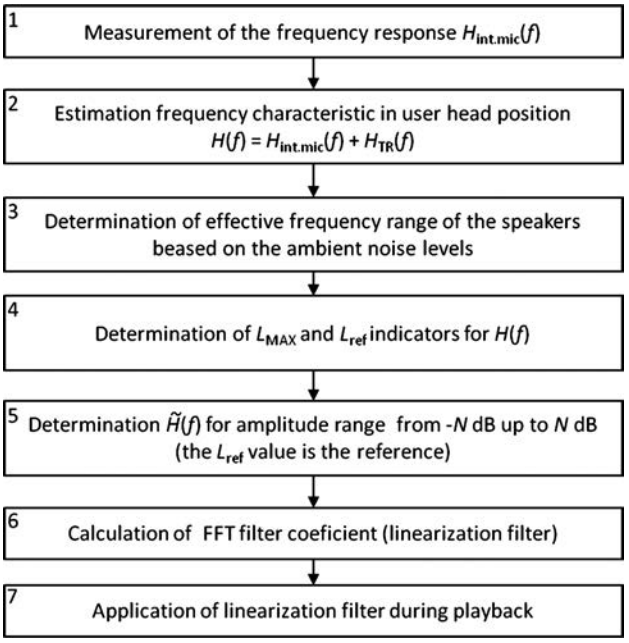


Fig. 5 The algorithm of the linearization filter characteristic determination.

where: L_{MAX} is the maximum value of the estimated frequency characteristic in the user's head position $H(f)$.

Linearization depth can be determined using N factor (block 5). For practical reasons N factor was selected as equal to 12 dB in our implementation. Next, the reverse H(f) characteristic is calculated (blue line in Fig. 6), according to Eq. (11):

$$\tilde{H}\ (f) = L_{\text{ref}}\ - H\,(f) \tag{11}$$

Finally, the magnitude response of the linearization filter $\tilde{H}(f)$is calculated and smoothed out using moving average described above. In Fig. 7 the theoretical average characteristic after an application of the linearization filter is shown (black line). The original frequency response of the built-in

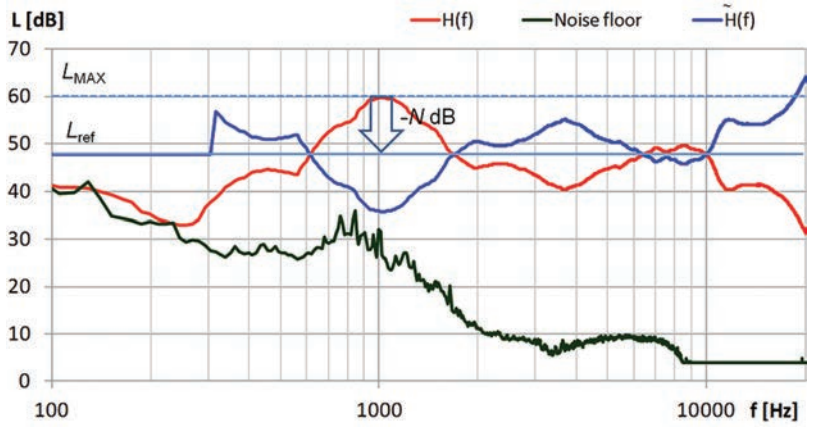


Fig. 6 Determination of the L_{MAX} and L_{ref} indicators.

speakers is also presented (violet line). The linearization filter characteristics (red line) was obtained for characteristics measured by the external microphone. Such a characteristic is used for a reference during objective evaluation of the calculation of $\tilde{H}(f)$result on the basis of $H_{int.mic}(f)$ and $H_{TR}(f)$ (Sec. 1.1.2).

The same measurement procedure was applied for the second considered computer device: Dell All-In-One. The final results including amplitude characteristics, linearization filter, and expected frequency response after an application of the designed linearization filter are presented in Fig. 8.

1.1.2 Objective Evaluation of Static Linearization

A methodology applied for the linearization of the frequency response of the built-in mobile device speakers described in Sec. 1.1.1 has been practically implemented and objectively evaluated on the basis of the measurement

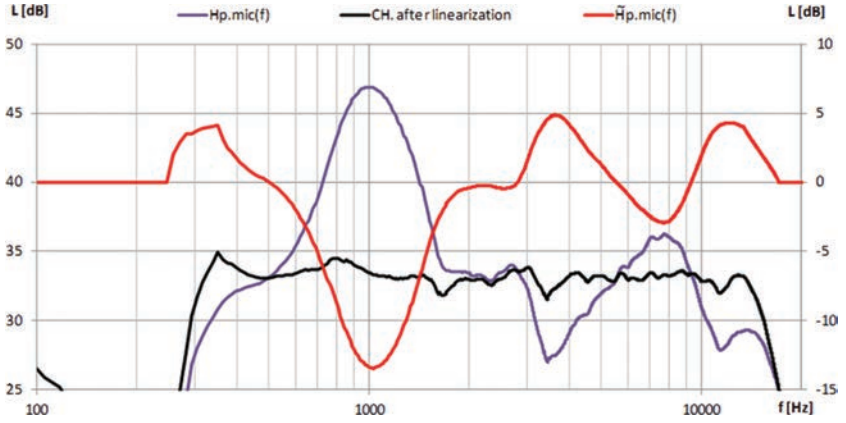


Fig. 7 Amplitude characteristic of the linearization filter for the HP Pavilion G6 device.

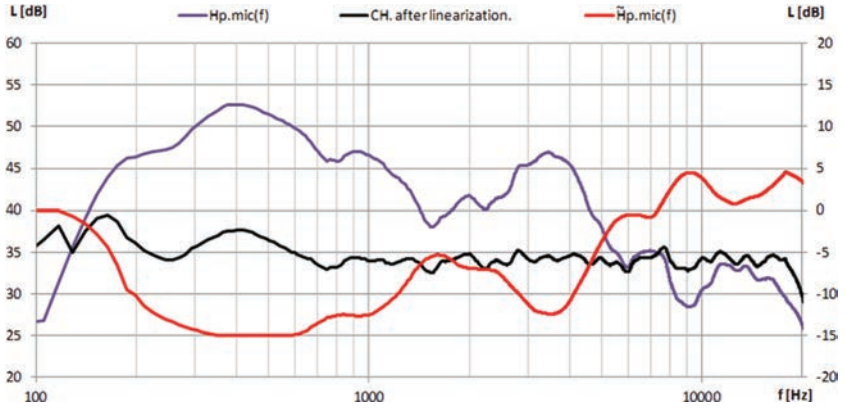


Fig. 8 Amplitude characteristic of the linearization filter for the Dell All-In-One device.

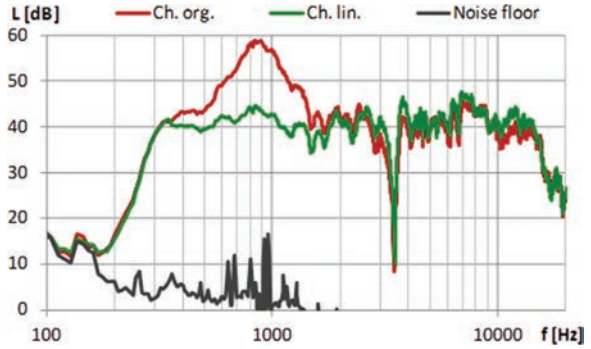


Fig. 9 Amplitude characteristic for HP Pavilion G6 computer device after linearization process (green line). The original amplitude characteristic is also presented (red line).

results. The mobile device was placed on the desk. Measurements were performed in a typical living room. Frequency response was measured in the users' head position by means of the PULSE measuring system.

In Fig. 9 the obtained characteristics of the mobile device audio system are shown. The measurement microphone was placed in a typical users' head position. First, the original characteristic of mobile device speakers was measured (*Ch.org.*). The second characteristic was obtained after an application of the linearization process (*Ch.lin.*).

In Fig. 10 the original frequency response obtained using the measuring microphone placed at the user's head position ($H_{p.mic}(f)$ - red line) and frequency response obtained using internal microphone and transfer function $H(f)$ are shown. Both curves are very common; it means that the internal (built-in) microphone of the mobile device can be used during the auto calibration process. In Fig. 11 the linearization filter responses calculated for both kind of measuring signals are shown.

It is clearly noticeable that after the application of the linearization filter the frequency response becomes more uniform. For a better illustration of the impact of the linearization process on the final amplitude characteristic of the speakers some additional calculations were done. The level distributions for both characteristics were calculated. The standard deviations of signal levels were also determined. The obtained results are shown in Fig. 12 and Fig. 13. For all kind of analyses the linearization process provides better results: a more narrow level distribution, a more rapid slope

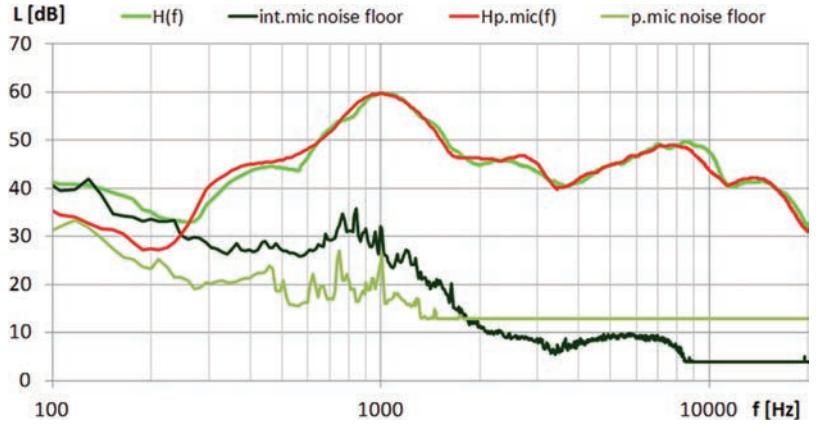


Fig. 10 The original frequency response obtained using the measuring microphone placed at the user's head position (red line) and frequency response calculated according to Eq. (9) using an internal microphone and the transfer function (green line).

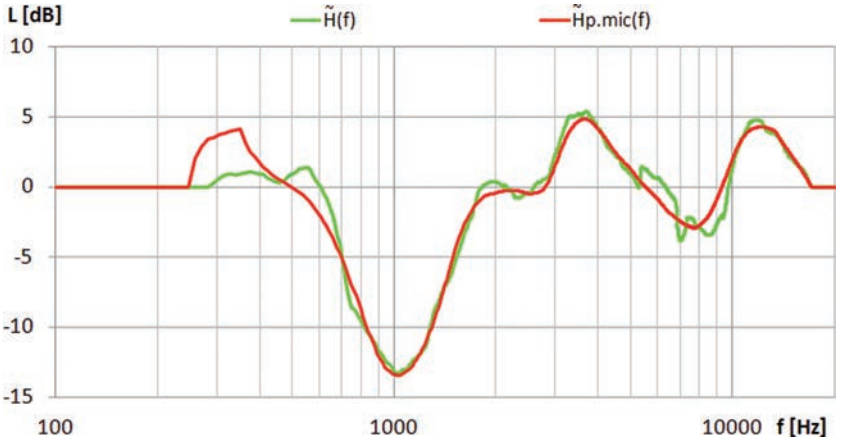


Fig. 11 Linearization filter responses: original (red line) and obtained by means of the transfer function and the internal microphone signals (green).

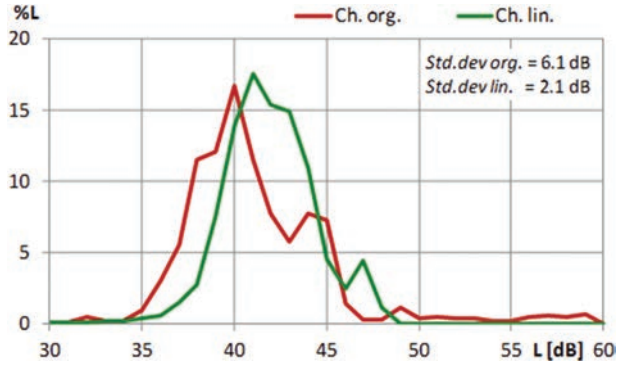


Fig. 12 Level distribution for both amplitude characteristics (original: red line and after linearization: green line).

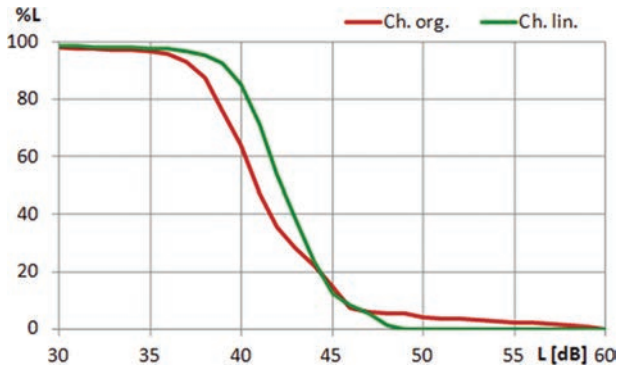


Fig. 13 Cumulative distribution for both amplitude characteristics (original: plotted with red line and obtained after linearization: plotted with green line).

for cumulative distribution, and a lower value of standard deviation.

1.2 Adaptive Approach

The above-described static linearization approach is capable of addressing the transfer function perturbations resulting from the manufacturer's choice of enclosure shape and materials, transducers, and other factors directly related to the design of the device. It cannot however compensate for the external factors, e.g., the room/environmental acoustics, which change over time. The adaptive linearization algorithm presented in this section goes a step forward, by a continuous monitoring of the instantaneous frequency characteristics of the sound emitted by the device under given conditions. Since the subject is quite broad and the adaptive filtration algorithms adopt some rather complex principles, the proposed solution will be outlined here only, with included references to details discussed in other papers.

1.2.1 Problem Statement

The use of adaptive filtering for linearization of audio systems is known from the literature [13], however, the proposed algorithm introduces a novel approach, which utilizes digital audio watermarking techniques for the realization of a module, which detects whether there are additive disturbances present in the recorded signal, which could negatively impact on the adaptation process. This idea closely follows previous research of the authors' in the field of Acoustic Echo Cancellation, wherein a similar method was used for the purpose of double-talk detection (DTD) [14].

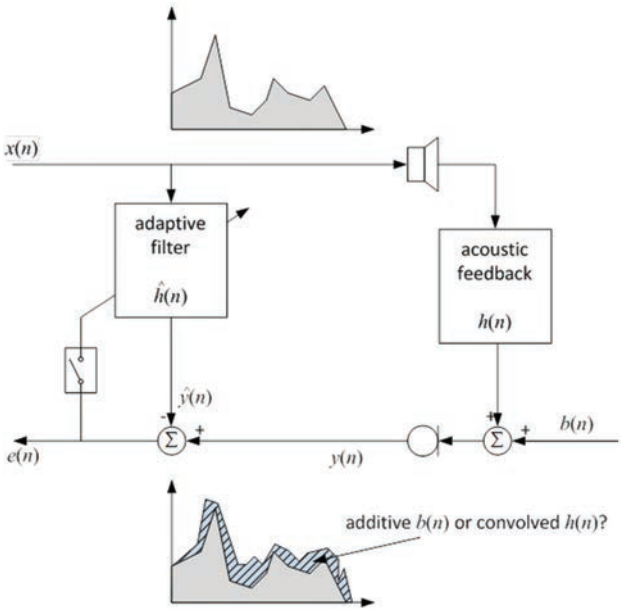


Fig. 14 Adaptive self-linearization problem illustration.

In the proposed application, the semi-fragile audio watermarking is used, enabling the monitoring of the local speaker-microphone loop. This allows for an accurate detection of the situation where additive distortions are present in the signal recorded by the microphone, rendering the watermarking signature undetectable. Such a distortion could disturb the adaptive algorithm used for estimating the characteristics of the loop. A reader interested in the discussion regarding the choice of the watermarking method and the detailed signature embedding/detection algorithms can find them in the aforementioned paper [14].

The purpose of the introduced adaptive filtering is to achieve the most faithful sound reproduction regardless of changing acoustic conditions. The means to achieve that is the self-linearization algorithm based on the use of adaptive filter, which continuously estimates the changes in the characteristics of a distortion and controls the playback system to compensate for them. In the case of the echo cancellation algorithms, the adaptation process is sensitive to the presence of additive distortions originating externally to the transducer-housing-environment system. Such a distortion from the viewpoint of the adaptive algorithm introduces some specific changes in the spectrum of the recorded signal, for which the algorithm will also attempt to compensate. This will lead to a divergence of the adaptive algorithm from the actual characteristics and also will produce a significant estimation error, resulting in the ineffective correction. This problem is illustrated in Fig. 14.

The application of the equalization applied to the linearization of the characteristics of the audio path is

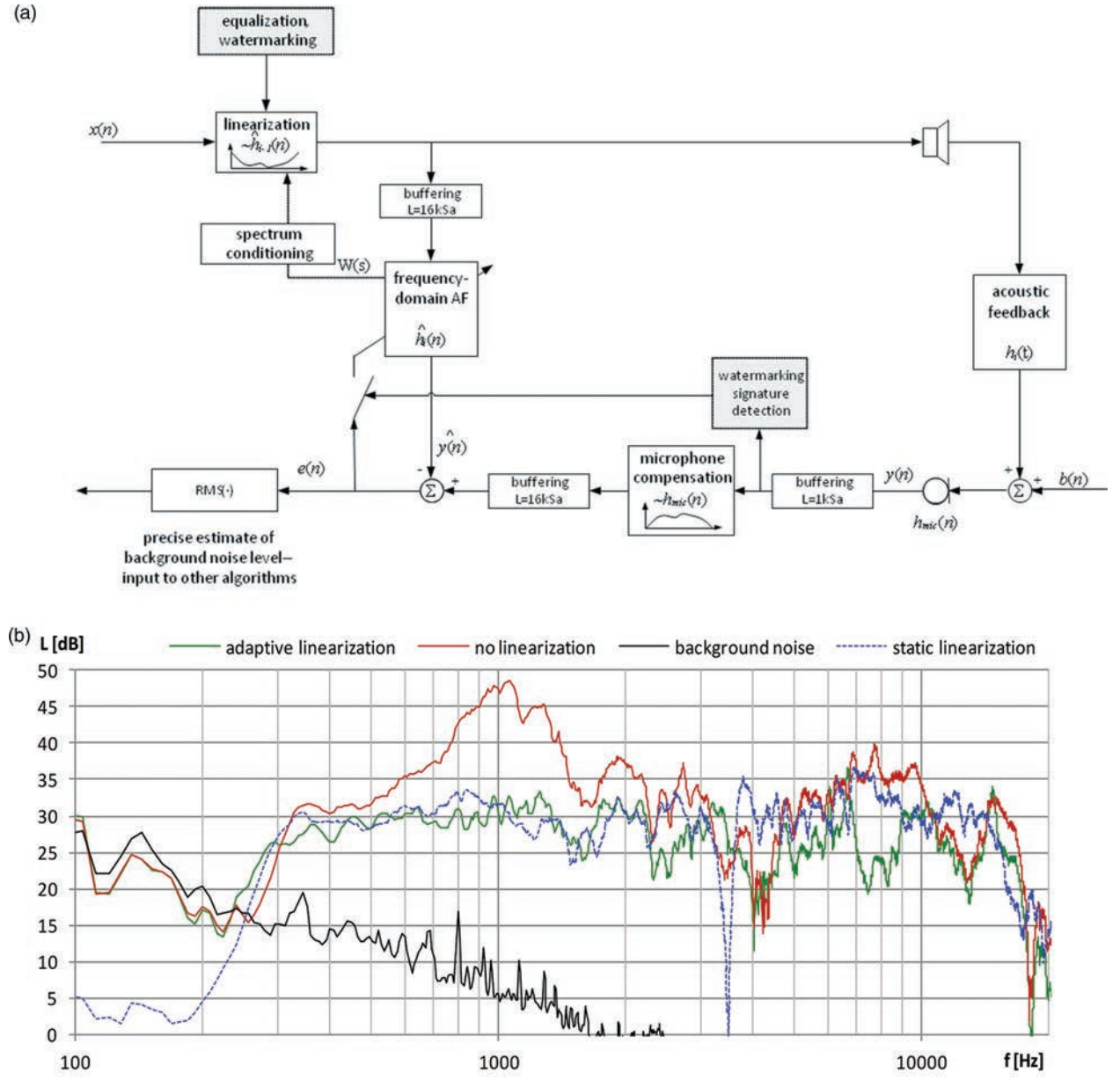


Fig. 15 Block diagram of the adaptive linearization algorithm (a) and its objective performance measurement result (b), static linearization result was also presented.

often used, among others, in the concert hall sound systems, where the graphical equalizer is utilized to offset the frequency response in a number of sub-bands, based on indications of a spectrum analyzer. This kind of a correction, however, is done manually and incidentally, and its precision is limited by the number of equalizer bands, which is fixed. There are known implementations of such devices in which the process of tuning up the equalizer has been automated by combining the equalizer with the system generating a signal of known characteristics (e.g., the "sweep" signal or pink noise). Recording the signal using a dedicated microphone and spectral analysis allows for adjusting the characteristics of the equalizer to the measured signal properties. This approach is not suitable, however, for the correction in the continuous manner, because in this case the measuring signal interferes with the normal sound reproduction.

1.2.2 Algorithm Design

The operation of the adaptive self-linearization algorithm is illustrated by the diagram presented in Fig. 15(a). The digital audio signal x(n) intended for reproduction is supplied to the adaptive filter and then used as a reference signal providing a basis for the estimation of the system's transfer function. Prior to this, the signal is processed in the linearization block, which applies a digital filtering employing an inverse characteristic to the one estimated by the adaptive filter in the previous calculation step (Eq. (12)). For simplicity, the equations describing these operations are given using the time-domain notation, however the transition is straightforward and it was already introduced in Sec. 1. The linearization block performs also the embedding of the watermarking signature. The conditioned signal $x_c(n)$

is then subjected to digital-analog conversion and played back by the loudspeaker, as in Eq. (12):

$$x_c(n) = x(n) * \sim \hat{h}_{i-1}(n) \tag{12}$$

In the acoustic field the signal is degraded by time-varying distortions of linear nature, which make the result of bandwidth-limiting loudspeakers and the influence of housing and environment (reverberation), which all together constitute the audio path transfer characteristics h_i(t). The signal is also distorted with various additive perturbations stemming from the noise and additional sounds (e.g., conversation) represented by $b(t)$ in Eq. (13):

$$y\ (t) = x_c\ (t) * h_i\ (t) + b\ (t) \tag{13}$$

The distorted signal is recorded by a microphone with the known, measured ahead and being immutable over the time digitized characteristics h_{mic}(n). The way of measuring of this characteristics and its influence on the recorded signal were discussed in Sec. 1.1.

The recorded and digitized microphone signal y(n) is the input to the watermark signature detection block, exploiting the properties of the semi-fragile watermarking method. When a significant additive distortion appears in the input, the signature is not possible to be recovered. Then, the adaptation process is stopped.

The microphone signal is filtered through a microphone compensation characteristic $\sim h_{mic}$(n), providing an inverse of the measured microphone characteristics. The filtered signal is fed into the adaptive algorithm and used there for obtaining the estimate $\hat{h}_i(n)$ of the speaker-housing-environment system characteristics. This response is subject to conditioning and transformation procedure leading to the update of current compensation characteristics applied in the linearization block.

The spectrum conditioning is a multi-step process, whose purpose is to transform the spectrum of the audio path response estimate calculated by the adaptive filter into its stable and causal inverse form. Hence, it includes the following steps:

- Discarding the phase response;
- Smoothing of the magnitude response. The smoothing is performed with a parameterized factor, expressed as 1/N-octave (1/N-octave smoothing; typically 1/3rd - 1/48th of octave) as in Eq. (14):

$$W_s\ (k) = \sqrt{\frac{1}{b-a+1} \cdot \sum_{i=a}^{b} W\ (i) \cdot W^*\ (i)} \tag{14}$$

where: $a = round(k \cdot 2^{-\frac{1}{2n}})$, $b = round(k \cdot 2^{\frac{1}{2n}})$

- Discarding (zeroing) magnitude response factors outside the specified frequency range bins;
- Inverting the magnitude response;
- Limiting the magnitude to the parametrized range (typically +/–15dB) in order to safeguard the inverse response from an uncontrolled gain stemming from possible $\hat{h}_i(n)$ zeros presence. The limiting is performed using tanh() characteristics;
- Synthesis of the causal/linear-phase phase response.

1.2.3 Implementation of the Adaptive Filter and Results

The most important factor affecting the design and the implementation of the above-described algorithm is the numerical complexity of the performed operations. In order to be able to reliably estimate the acoustic path transfer response, the adaptive filter must have a length of few hundred milliseconds (typically >200 ms). In turn, in order to be able to perform calculations in the real-time mode on a typical mobile hardware, and to keep the CPU load low enough to avoid an increased power consumption and not to disturb foreground tasks, it was decided that all the filtering operations should be performed in the frequency domain, making use of the FFT algorithm. This led to the choice of the OLA (Overlap-and-Add) structure for the linearization and for the microphone compensation filter as well. Similarly the adaptive filter was implemented in the frequency-domain—NLMS FDAF using Overlap-Save sectioning. In the typical operating conditions (Fs = 44.1kSa/s, FDAF length 16384Sa, that corresponds to 372 ms) this allows for a reduction of the numeric cost in the order of 400 times while compared to time-domain implementation. The above choice has also another advantage, namely all the spectrum-conditioning operations don't require any additional transformation from the time domain and back in this case. Thus, the compensation characteristic is directly synthesized in the frequency domain and implemented into the OLA filter.

In order to maintain a low latency of the algorithm both compensation filters operate on blocks of the length of 1024Sa; therefore to account for the difference in the block size, the FDAF was implemented as to operate on the blocks of equal size. The actual design of the FDAF was based on the methods known from the literature [15], discussed in Haykin's book [16] on pp. 360–368, for the case when the filter length differs from the operating block size. The implementation of the employed DSP algorithms was made available by one of this paper's authors (A. Ciarkowski) in the Internet repository [17]. In order to keep the paper concise, the authors, instead of reminding the complex matrix equations here, which explain the operation of the aforementioned Overlap-Save FDAF, refer the interested reader to the cited bibliography positions.

The effectiveness of the implemented algorithm was assessed positively on the basis of initial subjective tests, however in our hitherto performed experiments we concentrated first of all on the objective validation of the algorithm performance (as illustrated in Fig. 15(b)). The experience gained during the experiments with the algorithm led us to the following observations:

- The algorithm performs best with magnitude limiter set to +/– 10–15 dB range and frequency smoothing factor ≥1/24 octave,

- Setting the proper bandwidth of the linearization is crucial, too broad bandwidth leads to increased gain in subbands that are not transmitted by a speaker-microphone system, so in consequence, to overdriving the signal,
- Increasing linearization depth beyond +/–15 dB brings no improvement in performance, but increases the risk of overdriving and it requires a higher headroom,
- Performance heavily depends on the accuracy of the captured microphone transfer characteristics,
- If the microphone is unable to record the signal with a sufficient level, the adaptation is slow or it is unable to converge—fallback to non-adaptive method should be used in this case (the same holds true for very noisy environments),
- Both linearization methods (static and dynamic) may be combined together with weights depending on the acoustic conditions in the listening room,
- Adaptation is slower in the frequency bands that are missing in the signal.

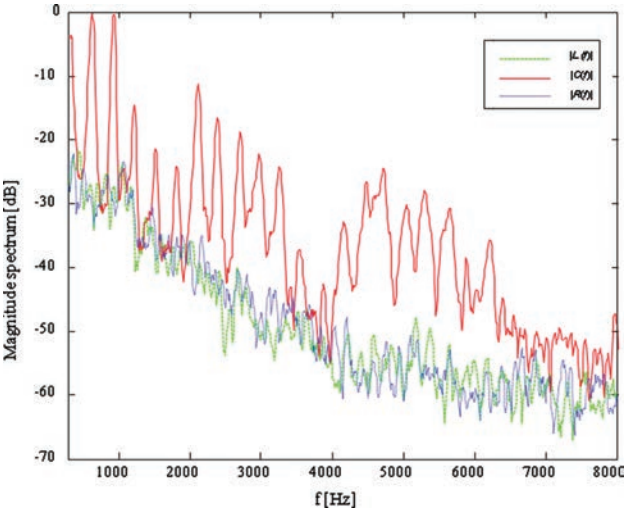


Fig. 16 Example magnitude spectra of left, right, and center channels.

2 DIALOGUE INTELLIGIBILITY ENHANCEMENT

Low clarity of dialogue during movie playback is a common problem experienced by many users. In case of playback on portable devices the low quality of installed loudspeakers and the presence of external noise are factors that make the dialogue in movies difficult to understand. Hence, it motivates for developing methods that aim to improve dialogue clarity. Most of the methods rely on dialogue detection to identify the parts of the soundtrack in which dialogue is present. Subsequently, the dialogue can be boosted or filtered to ensure increased clarity. Our approach is based on the frequency-domain disparity analysis of the signals in front channels of the 5.1 mix (left, right, and center) and on selective boosting of the frequency components that are identified as dialogue. The details of the algorithm were presented in previous publications on the subject [18][19][20].

Solutions for the problem of low dialog intelligibility are widely discussed in the literature. Fuchs divides the approaches to the enhancement of dialogue clarity into three groups [21]. The first one is to provide different versions of the mix with different dialogue gain. The second one is to make the source sounds available at the user's side. The new standard for object audio coding introduced by MPEG facilitates such an approach [22]. The third group comprises the algorithms that operate on the original mix and use signal processing to extract the dialogue channel. Our approach falls into this last category. The advantage of such methods is that they are compatible with all existing soundtracks. Several techniques for dialogue extraction were proposed in the literature. Kotti et al. utilized a neural network classifier [23]. Lee et al. utilized Independent Component Analysis for speech extraction [24]. Our algorithm is much less complicated. It does not require a signal model or complex processing, which favors our method to be used as an APO in real-time operation in the audio engine. For maintaining simplicity and low CPU load we choose to implement a technique relying on the dialogue being panned in the center. It is a known approach to analyze the channels in the frequency domain and to identify the frequency components, which are panned at a certain angle. Barry et al. proposed a similar algorithm for the separation of musical instruments [25]. ITU-R provided a technical standard concerning downmixing of multichannel sound [26]. Avendano and Jot aimed at separating the dialogue from other sounds in the 2.0 stereo mix using a coherence measure [27]. The difference of our approach is that we employ a measure based on the magnitude difference and we consider primarily the 5.1 mix.

2.1 Interchannel Disparity Analysis

Let us consider a 5.1-channel soundtrack. The channel layout is as follows: left *l,* right *r,* center *c,* low frequency *lfe,* left surround l_s, right surround r_s. The low frequency channel is not considered, since it cannot be properly projected in portable devices. The proposed method relies on the assumption that the signal components that are present in the center channel and are absent in the side channels (left, right) relate to dialogues. The goal is to extract these components and to boost them in order to achieve increased dialogue clarity. To identify the dialogue components the frequency analysis of the left, right, and center channels is performed. The signals are analyzed in blocks of 2048 samples (42.6 ms at 48000 samples per second) with 50% overlap. For each block the Fast Fourier Transform (FFT) is applied and the spectra $L[k]$, $R[k]$, and $C[k]$ are obtained. From this point the magnitudes of the complex spectrum are considered. The spectral disparity function is defined as follows (Eq. (15)):

$$V[k] = \frac{C[k] - L[k]}{C[k] + L[k]} \cdot \frac{C[k] - R[k]}{C[k] + R[k]} \tag{15}$$

The example magnitude spectra and the resulting disparity function are plotted on Fig. 16 and Fig. 17 respectively. High values of V indicate that the frequency component is

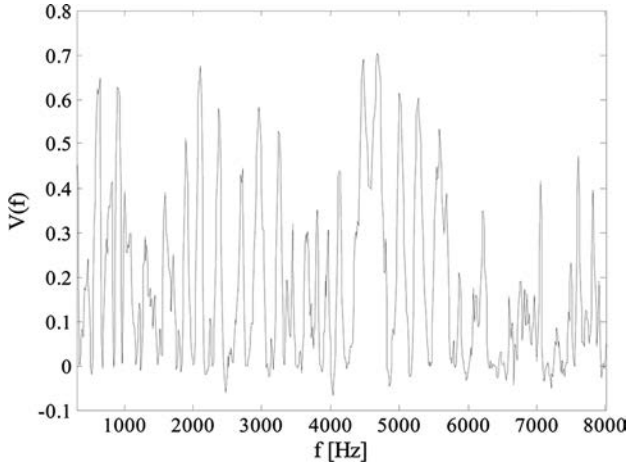


Fig. 17 Example of the disparity function obtained from the spectra shown in Fig. 16.

present in the center channel only and it is considered as being related to the dialogue.

2.2 Dialogue Extraction

The interchannel disparity function calculated according to Eq. (15) is used for calculating the dialogue extraction mask. The mask $m[k]$ is defined as follows (Eq. (16)):

$$m[k] = \begin{cases} 1 & if\ V[k] \geq t_d\ \wedge\ k_1 <; k <; k_2 \\ 0 & if\ V[k] <; t_d \end{cases} \tag{16}$$

where: t_d is the threshold for dialogue extraction and k_1, k_2 are the frequency bin limits. We assume that the extracted dialogue components are located between 300 Hz and 16000 Hz. The threshold is constrained to the interval (0;1) and it can be adjusted by the end user of the algorithm.

We found as a result of research experiments that the dialogue extraction mask obtained in such a way may have some discontinuities because the values change from 0 to 1. This can result in the presence of audible discontinuities in the audio signal. Therefore, the mask is smoothed out in order to reduce above undesirable effect. First, the frequency domain smoothing is applied with the moving average filtration as in Eq. (17):

$$m_s[k] = \frac{1}{L}\sum_{i=1}^{L_f} m[k-i] \tag{17}$$

where: L_f is the length of the moving average filter employed, which in this case contains five spectral bins. Subsequently, time averaging of the dialogue mask is performed, as expressed in Eq. (18):

$$\mathbf{m} = \mathbf{m_{new}} \cdot \alpha + \mathbf{m_{old}} \cdot (1-\alpha) \tag{18}$$

where: $\mathbf{m_{new}}$ is the current mask value, $\mathbf{m_{old}}$ represents $\mathbf{m}$ from the previous frame and α is the averaging constant.

The constant α translates to the time constant as in Eq. (19):

$$T_c = \frac{N}{SR \cdot \alpha} \tag{19}$$

where: N is the block size and SR is the sampling rate (in our experiments we used 48 kSa/s). The time averaging constant equals 200 ms, which was determined to be an optimum value in pilot experiments. Having obtained the smoothed dialogue mask the spectrum of the dialogue $D[k]$ is obtained as in Eq. (20):

$$D[k] = m_s[k] \cdot C[k] \tag{20}$$

2.3 Dialogue Boosting

To achieve increased dialogue clarity the extracted dialogue components are boosted according to the formula (Eq. (21)):

$$C_m[k] = C[k] \cdot (1 + g \cdot m_s[k]) \tag{21}$$

where: $C_m[k]$ is the spectrum of the modified center channel signal and g is the gain applied to dialogue components.

The gain parameter should be set in line with the listening comfort of the user. In our experiments 10 dB gain was applied. Different values of gain were evaluated in the previous work by means of PESQ calculation [28]. It was shown that for 6 dB and 10 dB boost an increase in the objective speech intelligibility measure is perceived. The finally assumed value was determined during a subjective evaluation that preceded the listening tests reported in this work. The listeners indicated that the 10 dB gain leads to an optimum aural experience, since the increase in dialogue volume becomes clearly audible, whereas no annoying distortions are perceived.

The modified center channel signal is transformed back after boosting into the time domain using inverse FFT, thus obtaining the signal $c_m[n]$. In order to play the 5.1 channel soundtrack on internal speakers of a portable device the downmix operation has to be performed. The standard ITU downmix equation is employed known from the literature [26]:

$$\begin{aligned} l_t[n] &= l[n] + 0.707 \cdot c_m[n] + 0.5 \cdot l_s[n] \\ r_t[n] &= r[n] + 0.707 \cdot c_m[n] + 0.5 \cdot r_s[n] \end{aligned} \tag{22}$$

As discussed in Sec. 4, in employing the above non-complex algorithm we obtained quite satisfying results.

3 PERSONALIZED DYNAMICS PROCESSING

The ability to personalize settings of audio processing in consumer digital devices is becoming now more common. Most solutions focused primarily on the smart correcting of the frequency characteristics of the audio devices (EQ) fitted to the user's hearing preferences [29].

In addition to spectral characteristics, the dynamic characteristics of sound is one of the most important sound parameters, which significantly affects the subjective perception of audio quality. Audio dynamics processing has been the subject of many studies [30]. Unfortunately, today a trend prevails on increasing the loudness of the sound at the expense of narrowing its dynamic range. There are publications whose authors criticize such trends and they also propose some methods for reconstructing the natural dynamics of sound [31].

The perception of sound dynamics is linked closely with the hearing dynamics. Generally, the sound is perceived as comfortable when it matches the dynamic characteristics of the listener's hearing system. There are also well known methods allowing for fitting the dynamics of the signal to the dynamic range of the transmission path. While the estimation of the signal dynamics is a relatively simple task, the estimation of the hearing dynamics appears to be a much more complex problem. Meanwhile, a correct estimation of the dynamic properties of the listener's hearing is essential for personalizing the dynamic characteristics of the audio signal.

A method for fitting of the sound dynamics to the hearing dynamics of the users of mobile devices is presented in this section.

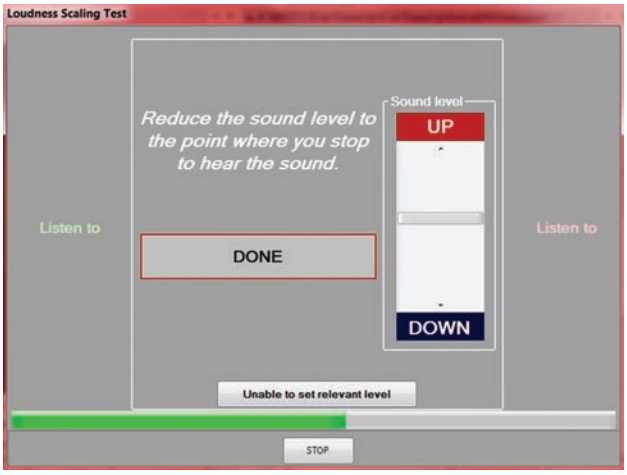


Fig. 18 Graphical user interface for loudness levels setting.

3.1 Loudness Scaling Test

To determine the hearing dynamics it has to be evaluated how the user perceives loudness, i.e., which sounds are perceived as soft, comfortable or loud. To assess the impression of loudness in audiology the loudness scaling tests are employed. The results of these tests show how the impression of loudness depends on both sound level and frequency [32][33].

An example of such a test is the LGOB test (Loudness Growth in ½ Octave Bands) [34]. The test signals in the test are in the form of narrow-band noise (half octave width) with center frequencies: 500 Hz, 1000 Hz, 2000 Hz, 4000 Hz. The test is carried out using the calibrated headphones. The level of the test signal varies in the range from 20 dB SPL to 120 dB SPL in the steps of 5 dB. Test signals are played back at a random order, while the task of the examined person is to assess the loudness impression related to the test signals. For loudness assessment seven loudness categories are used: I CAN'T HEAR; VERY SOFT; SOFT; COMFORTABLE; LOUD; VERY LOUD and TOO LOUD. In the basic version of the LGOB method the results can be utilized to determine both the hearing dynamics characteristics and dynamics processing characteristics, which can compensate for the hearing impairment [32].

A loudness scaling test is rather difficult to use in the context of fitting the audio dynamics range to hearing preferences of users of mobile devices. It is because the LGOB test for one ear takes approximately 10 minutes, the test signals are artificial, and they sound unpleasant to listen to. Moreover, the whole procedure requires a lot of attention from the examined person. Therefore, the direct application of the test known from the domain of audiology to this practical case is hardly feasible.

Nevertheless, it was decided to use the knowledge and the experience related to the audiology domain in order to propose a new test procedure. The first element that was modified was the test signals selection. Consequently, it was decided to replace the narrow-band noise by signals that are more pleasant to the listener, some sounds of musical instruments were chosen. These sounds are not only more pleasant to listen to but also their level and frequency characteristics are easy to control. In order to keep the natural sound it was decided to limit the bandwidth of test signals to one octave only while the filter slope was not greater than 6 dB/oct. The following sounds were selected:

- Drums—width of one octave bandwidth the center frequency of 500 Hz,
- Piano—width of one octave band with the center frequency of 1000 Hz,
- Electric guitar—width of one octave band with the center frequency of 2000 Hz,
- Violin I—width of one octave band with the center frequency of 4000 Hz,
- Violin II—width of one octave band with the center frequency of 8000 Hz. This higher frequency band has been added due to the applications of the developed method to speech and music, while the original LGOB method was developed for speech-related frequency band only.

The second element, which required a modification, was the scale of the loudness categories. Consequently, in the first step, it was decided to simplify the scale used in the LGOB test. The discrimination between loudness categories VERY SOFT and SOFT, as well as LOUD and VERY LOUD is difficult to make, therefore it was decided to reduce the scale to 5 loudness categories: I CAN'T HEAR; SOFT; COMFORTABLE; LOUD; TOO LOUD.

Initially, the modified loudness scaling method was tested with the participation of 70 students at our University. The obtained results were correct from the audiological point of view [25]. However, in the application context of the developed method above modifications proved to be insufficient. The test duration was still too long and the proposed loudness scale was still difficult to be interpreted by a typical user. Therefore, the procedure was simplified even further. The scale of loudness categories was replaced with a slider for the sound level setting by the user (see Fig. 18). The user's task was to set for each test signal three sound levels that corresponded to the following loudness impression categories: COMFORTABLE; I CAN'T HEAR, and TOO LOUD.

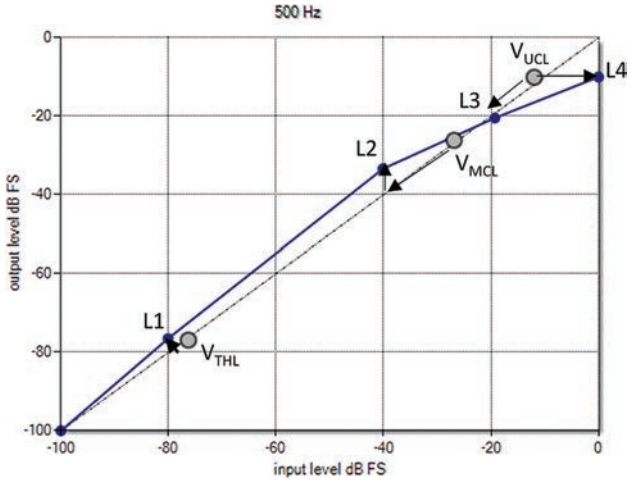


Fig. 19 Determining of characteristic for dynamics processor.

The developed method allows for obtaining information on the perception of sound in the analyzed frequency bands in a relatively short period of time (on average approximately 2 minutes). Information about levels of sound "too soft," "comfortable," and "too loud" allows for building dynamic characteristics in order to fit the dynamics of sound to the user's hearing preferences.

However, it was found during the pilot studies that after performing dynamics processing based on the results of the analysis of dynamics in particular frequency bands, that there is a risk that the processed sound can be too loud (especially in the case of processing loud sounds). Therefore, as the last element of the audio dynamics processing chain a wideband dynamics compressor was added, with settings calculated on the basis of the sound level selected for the wideband test signal.

3.2 Calculation of Dynamics Curves

The shape of the characteristics of the audio dynamics compressors was developed on the basis of both the experience gained during the implementation and as a result of carrying out numerous audiological tests. As is seen in Fig. 19, four points are considered: L1, L2, L3, and L4. Each point is related to the input level (e.g., $L1_{in}$) and the output level (e.g., $L1_{out}$). During the determination of the dynamics characteristics information about sound level specified by the user as too soft is interpreted as the hearing threshold (V_{HTL}). The value of this threshold is then related to the reference data (obtained in other collected test results). If the obtained value is higher than the adequate reference value (R_{HTL}), a reference signal level is amplified by a value that represents the difference between the level set by the user and the reference level. Considering above denotations, the first point (L1) of the static dynamics characteristic in a given frequency band is calculated according to Eqs. (23) and (24):

$$L1_{in} = R_{THL} \tag{23}$$

$$L1_{out} = V_{HTL} \tag{24}$$

Other information obtained using the developed procedure is the sound level value (V_{MCL}), which is associated with the comfortable loudness. The next point (L2) of characteristic is calculated as follows (Eqs. (25) and (26)):

$$L2_{in} = V_{MCL} - 15\,dB \tag{25}$$

$$L2_{out} = L2_{in} + 5\,dB \tag{26}$$

The last point of the characteristic is obtained based on the value of sound level that the user has set as too loud (V_{UCL}). To ensure that the processed audio levels will not exceed the designated level V_{UCL} two points L3 and L4 are created in the static characteristic, whereas the L3 point is a typical audio dynamics compression threshold. The value of this threshold is equal to V_{UCL}, which value is in turn reduced of 15 dB (as in Eq. (27)). The L4 point limits the characteristic of the audio dynamics compressor. Its input level is 0 dB FS (full scale), while the output level is equal to V_{UCL} (as in Eqs. (28), and (29)):

$$L3_{in} = L3_{out} = V_{UCL} - 15\,dB \tag{27}$$

$$L4_{in} = 0\,dB \tag{28}$$

$$L4_{out} = V_{UCL} \tag{29}$$

An example dynamics characteristic for a selected frequency band is shown in Fig. 19.

During the development of the method few preliminary tests were performed in which a group of 10 audio experts (researchers from the Multimedia Systems Department of Gdansk University of Technology) took part. The first results of those tests showed that the developed method tends to produce too loud sounds. As a result, the obtained dynamics characteristic yielded worse effects for louder sounds than for quieter sounds. It turned out, that after mixing the narrowband audio with the dynamically processed signals with the wideband signal, the resultant audio signal sounds louder than it was expected. To solve this problem a wideband audio dynamics compressor was added to perform the multiband audio dynamics processing. The results of some next tests showed that adding the wideband compressor has helped to solve this problem [35].

The characteristics of the wideband dynamics processor are obtained in an analogous manner to the narrow-band characteristics, yet in the case of the characteristics of wideband audio dynamics compression only the points L3 and L4 are calculated.

During experiments with processing of multiband dynamics it was observed that the point L3 presence was often source of distortions. Therefore the point L3 was omitted during further calculation of multiband dynamics characteristics.

3.3 Multiband Dynamics Processor

This section describes the design of the multiband dynamics processor algorithm, which is a main executive block of the developed method.

The general overview of the algorithm is depicted in Fig. 20. The processing consists of several stages, which are performed in a sequence. The detailed description of

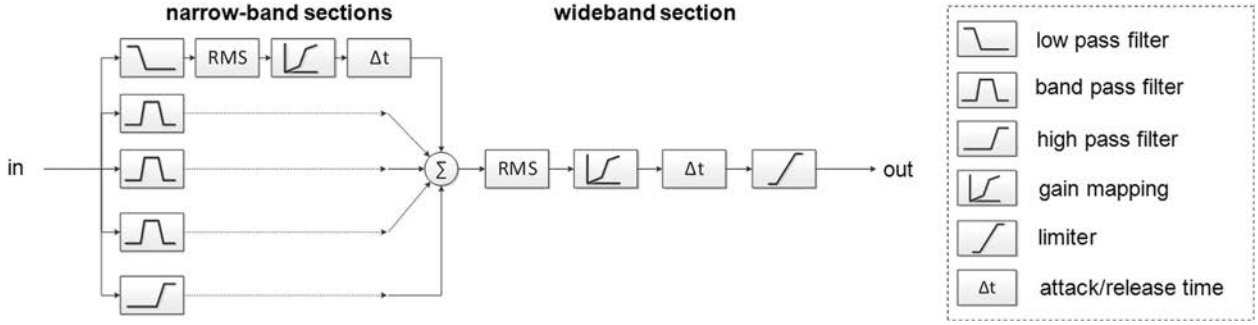


Fig. 20 Block diagram of the multiband dynamics processor algorithm evaluation.

the successive stages will be the subject of the following paragraphs.

Since the compressor operates in several independent bands, the first stage of the processing is splitting the input signal into the respective sub-band signals. This task is realized through a bank of band-splitting filters. The number of bands is configured to 5, however the method is flexible as far as the number of frequency bands is concerned. The filter bank consists of low-pass, high-pass (outer), and band-pass (inner) filters. The filters employed are of IIR type. They were implemented as a cascade of second-order-sections. In case of low-pass and high-pass filters, the cascade consists of 3 SOS-es (6^{th} order filter), and 4 for band-pass implementations (8^{th} order). The following cross-over frequencies are preconfigured: 707, 1414, 2828, 5656 Hz, which correspond central frequencies of the following frequency bands: 500, 1000, 2000, 4000, 8000 Hz, complying with the frequencies of the signals utilized in the developed loudness scaling test. The calculated filter coefficients are normalized in order to get a unity gain at the central band frequency. The SOS cascade filters are realized as Direct-Form II, time-domain, floating-point implementation.

The next step after the band splitting is the calculation of RMS power in each of the frequency bands. The temporary RMS value is used as an input to gain-mapping procedure, which calculates the resulting gain through the application of dynamics processing curve that resulted from the loudness scaling procedure. The final step is summing up the sub-band signals and applying soft-knee limiter to prevent signal clipping. The limiter equation is given as follows (Eq. (30)):

$$y = \begin{cases} x, \ if \ |x| <; t \\ x \cdot \frac{t+\tanh\left(\frac{|x|-t}{1-t}\right)\cdot(1-t)}{|x|}, \ otherwise \end{cases} \tag{30}$$

where: t equals the limiting threshold and x represents the sample value.

4 EVALUATION

The listening tests were conducted according to the ITU-T recommendation [36]. The Comparison Category Rating test (CCR) was employed. The sound samples processed with the designed algorithms were compared with non-processed samples. The order of samples in pairs was not known to the subjects. According to the standard, a 7-grade scale was used from –3 through 0 to 3, where –3 means that sample A is much better than B, and 3 means that B is much better than A. Four tests were performed that evaluate static linearization, dialogue intelligibility enhancement, and dynamics processing. In fact, dynamics processing algorithm was applied in two modes: in the first test as a part of the developed Ear tune up method (dynamics characteristics was calculated adequately to the loudness sensation assessment) and in the second case as loudness maximizer (dynamics processors were set according to the fixed characteristics). This approach was justified by the fact that the Ear tune up method may cause a significant increase of the time duration of the other tests, as well as it could influence their results in an uncontrolled manner.

Table 1 Number of listeners participating and selected as experts in respective tests

device	algorithms	n° of all listeners	n° of selected listeners
All-in-one	Linearization	34	12
All-in-one	Ear tune up.	30	10
Laptop	Lin. + dyn.proc.	38	17
Laptop	Dialogue + dyn. proc.	30	25

Two devices were considered: a laptop and an all-in-one computer. Types of the performed tests were presented in Table 1. The test evaluating dialogue intelligibility enhancement was performed in the presence of ambient noise. The samples employed in the listening tests are listed in Table 2. A selection of musical tracks was used, as well as soundtracks and speech samples. The musical samples are typically compressed (in mp3 or in AAC format). Some excerpts were downloaded from a music streaming service. For each test the most suitable samples were selected.

The methodology of conducting the listening tests is as follows:

1) Listeners are recruited from subjects both having background in acoustics and not experienced in this domain, however all of them were not familiar with the details of the processing;
2) Dedicated computer software created in the research project was used to present the samples. The software tools serve the purpose of playing back the test files mentioned in Table 2 and the purpose of

Table 2 Samples used in the subjective listening tests

Sample	Artist - title	Content	duration
rock	Alice Cooper - Poison	heavy rock band with male vocals	00:15
pop	Adele - Set Fire to the Rain	pop music with female vocals	00:14
pop2	Nelly Furtado - Say it Right	R&B music with female vocals	00:15
polrock	Czesław Niemen - Dziwny jest ten świat	rock band with male vocals	00:15
elektro	Tiesto - Sweet Things	Electro pop with female vocals	00:10
electric	Hybrid - Formula of Fear	Electronic music, instrumental	00:15
Classic	P. Tschaikovsky - Capriccio Italien	Classical music, symphonic, instrumental	00:13
Jazz	Till Bronner - Have You Met Chet	Smooth jazz, male vocals	00:15
Rocknroll	Jerry Lee Lewis - Great Balls of Fire	rock'n'roll, male vocals, band-limited	00:12
GDT	Girl with the Dragon Tattoo	music, male and female voices	00:15
S_Ryan	Saving Private Ryan	war noises, shots, explosions, male voices	00:17
BHD	Black Hawk Down	helicopter noise, music, male voices	00:17
2012	2012	music, male voices	00:18
Skype_male		male voice via Skype	00:09
Skype_female		female voice via Skype	00:09
male_voice		male actor's voice	00:10

online processing the audio in the audio driver of the operating system;

3) Samples were arranged in two sessions, 10 pairs per each session. Each pair contained processed vs. unprocessed sample, apart from 1–2 null pairs, which consisted of two unprocessed samples. The second sessions contained the same samples arranged in a different order to evaluate the listeners' consistency;
4) Listeners performed the test and gave their responses on the paper sheet provided;
5) Results were then manually input to a computer;
6) Selection of reliable listeners (experts) was performed according to the following criteria:

- Each listener who gave a score greater than 1 or less than –1 in a null pair was rejected;
- Each listener who gave opposite sign scores of magnitude 2 or 3 to the same sample in more than 2 pairs was rejected.

The details of the number of listeners participating in each test and the number of listeners kept after the selection is provided in Table 1;

7) The results were analyzed with the ANOVA test. The type of processing was considered an independent variable, whereas the score was considered a dependent variable;
8) Tests of linearization and dynamic processing for both computer devices were conducted in the office environment (see Fig. 21). Tests conditions of dialogue enhancement and dynamics processing algorithms were presented in detail in Sec. 4.4

4.1 Linearization: All-in-One Computer Case

A total number of 34 listeners took part in the linearization test employing the All-in-One computer. Twelve listeners were selected as reliable experts. In Table 3 the mean scores, standard deviations, and results of ANOVA analysis are presented for each sample. All CMOS scores have positive values. It means that the listeners in general preferred the linearized sample. The standard deviations are relatively high. It can be observed that the samples that obtained a CMOS indicator greater than 1 yield lower values of p, what can be interpreted that the difference between the processed and unprocessed sample is statistically significant. For samples whose CMOS is closer to 0, the statistical significance is lower (e.g., for samples "*elektro_LIN*" and "*rock_LIN*"). The boxplots that visualize the distribution of the scores are presented in Fig. 22.

Table 3 Results of subjective test evaluating linearization on All-in-One computer

sample	CMOS	σ	p
classic_LIN	1.333	1.761	3.84E-06
electric_LIN	1.500	1.414	2.74E-09
elektro_LIN	0.875	1.849	1.99E-03
jazz_LIN	1.375	1.408	2.06E-08
polrock_LIN	1.583	1.283	4.59E-11
pop2_LIN	1.500	1.560	2.95E-08
pop3_LIN	1.875	1.191	2.39E-14
rock_LIN	0.542	1.668	2.92E-02
rocknroll_LIN	1.625	1.345	8.51E-11

4.2 Ear Tune-Up—All-in-One Computer Case

The final method was evaluated during tests employing 30 subjects. Most of them were students aged 21–23 years. At the beginning each of them performed the developed loudness scaling test, then the characteristics of the audio dynamics compression was calculated automatically. The validation process ended with the pairwise comparison test in which each expert choose the signal that was closer to their hearing preferences. The test signals represented fragments of musical tracks (rock, jazz, pop) and also one sample of speech (female voice in the Skype quality). The sound level of the signals was normalized in the test to one of three levels: –11 LUFS (high level), –40 LUFS (medium level), and –58 LUFS (low level). Thus, 12 different test signals were obtained. The processed test signal and the original test signal were presented as a pair in random order.

Fig. 21 Subjective test setup: a) for all-in-one computer, b) for laptop.

In the developed method a model was assumed to represent the determined dynamic characteristics. In the model four points (L1, L2, L3, L4) were defined. It turned out during experiments that the point L3 is unnecessary because it makes often bending of the characteristics, which is a source of distortions. Therefore, in order to avoid the risk of distortions, in the calculation of dynamics characteristics this point is omitted. Despite such a simplified model of compression characteristics application, the developed method still allows for obtaining varied dynamics characteristics fitted to the hearing preferences. The diversity of the achieved characteristics was illustrated on the basis of values of both variance and standard deviation (Table 4). The values are much higher than zero, which means that the obtained characteristics are diversified.

The scores obtained by individual samples are shown in Table 5. The boxplot in Fig. 23 shows the distribution of the scores. From the results it can be concluded that the sounds at the middle level (MID) and the low level (LOW) have higher scores than those at the high level. For these cases, the dispersion of scores is also much smaller. In case of loud sounds the results were slightly worse. In general, the average score of all LOW samples was 1.216, MID samples yielded 1.557, and the HIGH samples brought the score of 0.864. There were no apparent differences in the results for the different types of test signals. It is also important that

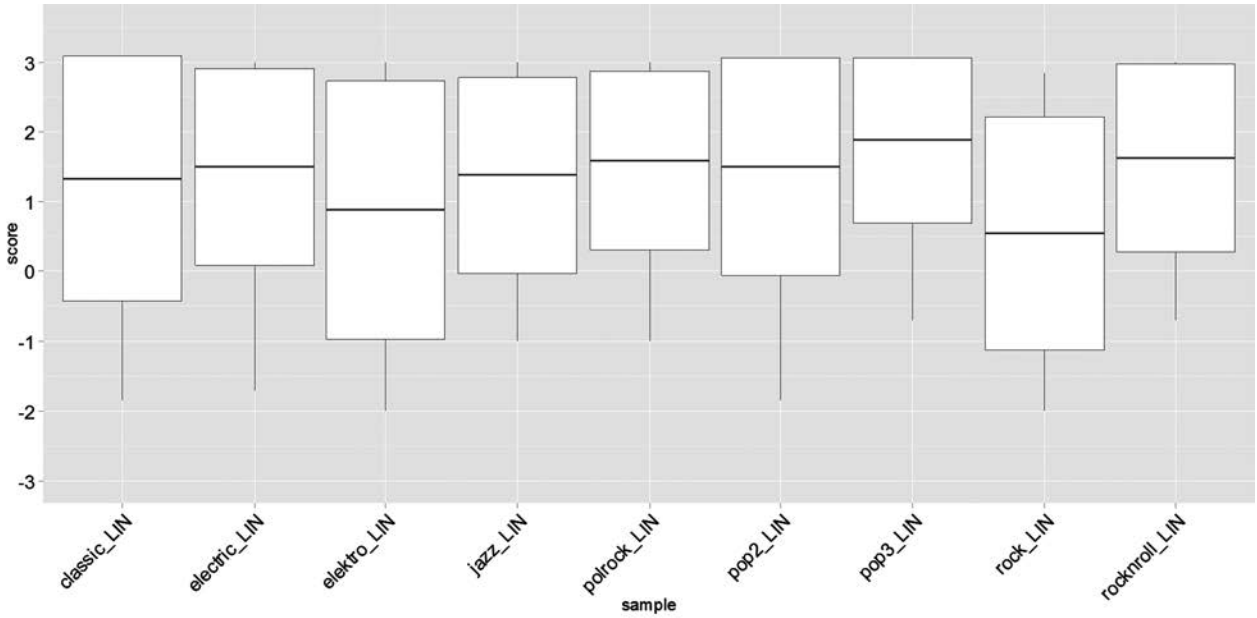


Fig. 22 Scores in linearization test on All-in-One computer. Boxes indicate mean and standard deviation, whiskers indicate 5 and 95 percentiles.

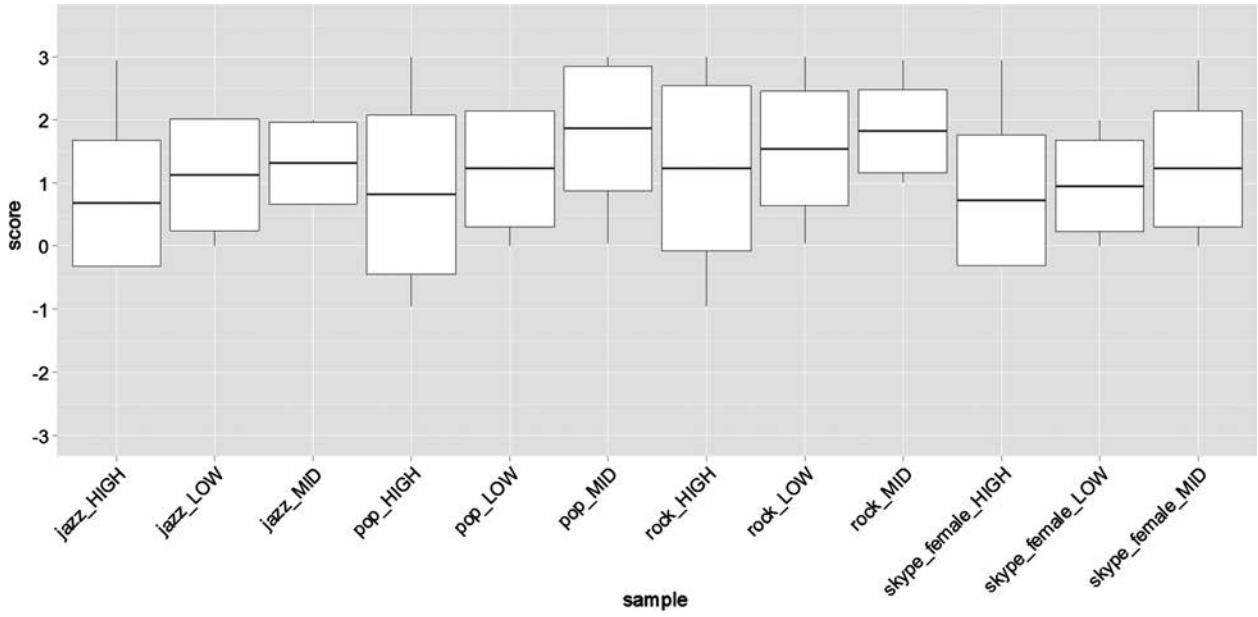


Fig. 23 Scores in Ear tune-up test on all-in-one computer. Boxes indicate mean and standard deviation, whiskers indicate 5 and 95 percentiles.

the proposed method does not degrade speech intelligibility, which can be concluded from the results obtained for the speech sample.

The tests show that the proposed dynamics processing algorithm can improve the subjective quality of sound and is able to adapt it to the user's listening preferences, but this effect is not common for all users. Since the processing result depends on how the listener responded during the loudness scaling test, the method will be more useful to the users who perform the test with understanding of the underlying principles, or to those who spend some time experimenting with the method. This possibility, however, could not be achieved in such a brief form of testing that underlies our method.

4.3 Linearization and Dynamics Processing—Laptop Case

The evaluation was performed with the participation of 54 listeners, 25 of which were selected as experts. The dynamics processing algorithm was used with fixed dynamics curves designed with a view to maximize the loudness. The statistics of the scores are shown in Table 6. The distributions are visualized on a boxplot in Fig. 24. The dispersion of results is larger than in case of All-in-One device. A positive CMOS was achieved for 11 of 16 samples. In case of 5 samples the listeners preferred the unprocessed signal. The linearized samples (LIN) was evaluated positively in 5 out of 6 samples. The samples processed with the dynamics processing (ETU) brought slightly lower scores. The combination of both algorithms (LIN_ETU) yielded positive scores for 5 out of 6 samples.

It can be concluded that the linearization of the frequency response of the device yields a good acoustic effect. However, the dynamics processing was not always positively assessed. One of the possible reasons is that increasing the loudness of the signal does not necessarily improve the subjective quality. Another explanation is that when the dynamics processing algorithm was used alone, it revealed the imperfections of the frequency response of the device. Hence, the combination of dynamics processing and

Table 4 Diversity of calculated dynamics characteristics

		L1		L2		L4	
		in	out	in	out	in	out
500 Hz	variance	0.00	27.59	44.80	44.80	0.00	38.04
	σ	0.00	5.25	6.69	6.69	0.00	6.17
1000 Hz	variance	0.00	72.39	24.79	24.79	0.00	48.81
	σ	0.00	8.51	4.98	4.98	0.00	6.99
2000 Hz	variance	0.00	27.92	18.79	18.79	0.00	45.36
	σ	0.00	5.28	4.33	4.33	0.00	6.74
4000 Hz	variance	0.00	30.47	23.95	23.95	0.00	35.44
	σ	0.00	5.52	4.89	4.89	0.00	5.95
8000 Hz	variance	0.00	28.79	34.70	34.70	0.00	59.33
	σ	0.00	5.37	5.89	5.89	0.00	7.70

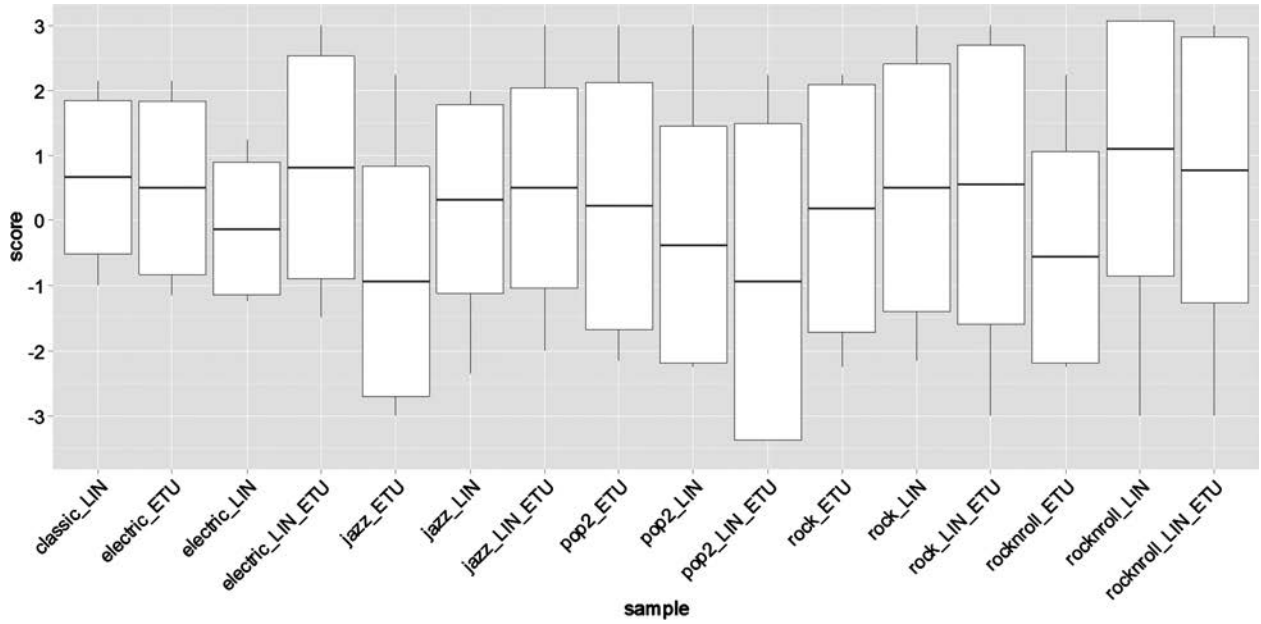


Fig. 24 Scores in linearization and dynamics processing test on laptop. Boxes indicate mean and standard deviation, whiskers indicate 5 and 95 percentiles.

Table 5 Results of subjective test evaluating Ear tune-up method on All-in-one computer

sample	CMOS	σ	p
jazz_LOW	1.136	0.889	1.21E-10
jazz_MID	1.318	0.646	6.64E-17
jazz_HIGH	0.682	0.995	4.56E-05
pop_LOW	1.227	0.922	4.06E-11
pop_MID	1.864	0.990	5.47E-13
pop_HIGH	0.818	1.259	9.59E-05
rock_LOW	1.545	0.912	3.02E-14
rock_MID	1.818	0.664	1.71E-21
rock_HIGH	1.227	1.307	1.86E-07
skype_female_LOW	0.955	0.722	4.90E-11
skype_female_MID	1.227	0.922	4.06E-11
skype_female_HIGH	0.727	1.032	3.04E-05

Table 6 Results of subjective test evaluating linearization and dynamics processing on a laptop

sample	CMOS	σ	p
classic_LIN	0.667	1.188	1.90E-03
electric_ETU	0.500	1.339	3.18E-02
electric_LIN	−0.125	1.025	4.95E-01
electric_LIN_ETU	0.813	1.721	1.21E-02
jazz_ETU	−0.938	1.769	5.42E-03
jazz_LIN	0.324	1.451	7.04E-02
jazz_LIN_ETU	0.500	1.543	6.03E-02
pop2_ETU	0.222	1.896	4.87E-01
pop2_LIN	−0.375	1.821	2.53E-01
pop2_LIN_ETU	−0.938	2.435	3.74E-02
rock_ETU	0.188	1.905	5.82E-01
rock_LIN	0.500	1.917	1.27E-01
rock_LIN_ETU	0.559	2.149	3.57E-02
rocknroll_ETU	−0.563	1.632	6.06E-02
rocknroll_LIN	1.111	1.967	1.79E-03
rocknroll_LIN_ETU	0.778	2.045	2.89E-02

linearization yielded better scores. It was also noticed that the combination of these two methods will perform well in the presence of external noise.

4.4 Dialogue Enhancement and Dynamics Processing in Noise

In the previous published work the dialogue intelligibility enhancement algorithm was evaluated in *clean* listening conditions [20]. It was reported that the employed processing contributes to a significant increase of perceived dialogue clarity.

In this test a typical use case is considered in which the listener is located in a noisy space (in this case inside an airplane cabin) and he or she watches a movie on a portable computer using headphones. The setup of this test is presented in Fig. 25. Four loudspeakers are employed to emit noise. The power spectral density function of the added noise signal is presented in Fig. 26. The signals emitted by the four loudspeakers were shifted in time (decorrelated) in order to avoid interferences. The level of the noise in the listening room was equal to 77 dBA, which is comparable with the noise levels measured in aircraft cabins [37]. The playback level of the computer was set to maximum and the sound was presented over headphones. The listeners were asked to evaluate the quality of the samples with regard to speech clarity. The goal of the test was to show that the engineered signal processing algorithms described in Sec. 2 enable an increase in speech clarity in difficult listening conditions.

The results of the test are presented in Table 7. The algorithms evaluated in this test were the introduced dialogue intelligibility enhancements (denoted SD) and dynamics processing (ETU). The dynamics processing curves were prepared by the experimenter and fixed during the

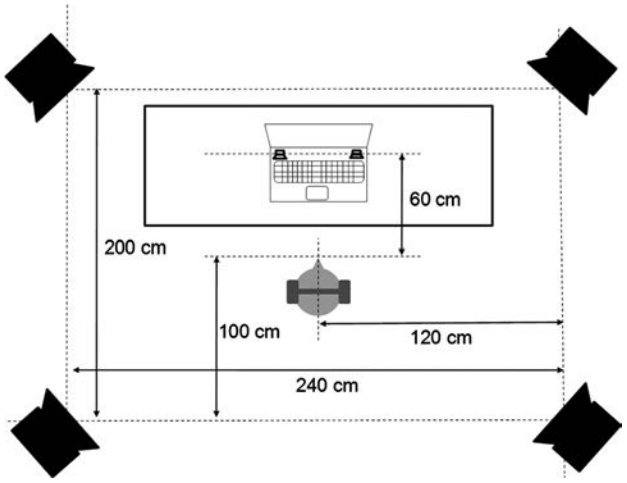


Fig. 25 Setup of the listening test for evaluation of dialogue enhancement and dynamics processing.

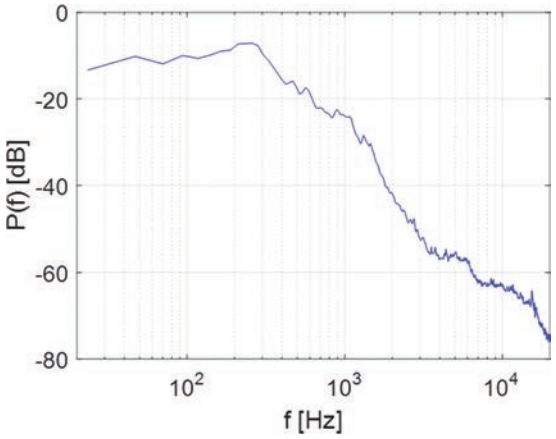


Fig. 26 Averaged power spectral density of noise employed in the experiment.

Table 7 Results of subjective test evaluating dialogue intelligibility enhancement and dynamics processing

sample	CMOS	σ	p
2012_ETU	2.393	0.685	2.54E-32
2012_ETU_SD	2.682	0.477	7.80E-34
2012_SD	1.840	0.955	4.36E-35
BHD_ETU	2.727	0.456	6.29E-35
BHD_ETU_SD	1.536	1.071	5.30E-15
BHD_SD	2.100	0.678	1.94E-52
GDT_ETU	2.818	0.501	7.87E-34
GDT_ETU_SD	2.143	0.803	1.37E-26
GDT_SD	2.020	0.915	7.78E-40
male_voice_ETU	1.591	0.908	1.05E-14
S_Ryan_ETU	2.393	0.786	2.28E-29
S_Ryan_ETU_SD	1.393	1.449	1.99E-09
S_Ryan_SD	1.227	1.343	3.23E-07
Skype_female_ETU	1.364	0.790	1.66E-14
Skype_male_ETU	−0.071	1.631	7.44E-01

experiment in order to maximize the loudness in frequency bands related to speech. The boxplots of the scores assigned to each sample are shown in Fig. 27. It is visible that the mean scores (CMOS) are in all cases but one larger than 1. It can be understood that the listeners perceived a strong increase in signal quality. The test for statistical significance with the ANOVA method leads to rejection of the null hypothesis (on the condition $p < 0.05$) for all samples except the sample "Skype_male_ETU." As far as the soundtracks are concerned, the samples processed with the *Ear Tune Up* algorithm tend to bring higher scores than those processed with *Smart Dialogue* or a combination of *Ear Tune Up* and *Smart Dialogue*. The exception is the sample "2012_ETU_SD" that was rated higher than "2012_ETU." This finding can be explained by the fact that the dynamics processing algorithm maximizes the loudness of the sample, thus also contributing to an increase of speech clarity, while maintaining the original balance of the sounds in the soundtrack. The *Smart Dialogue* method, on the other hand, modifies the proportions of sounds in the mix. The results of the test show that the listeners generally prefer to listen to the original mix, with the maximized loudness. Nevertheless, processing with the *Smart Dialogue* algorithm also leads to a significant increase in the perceived signal quality. Moreover, the advantage of the *Smart Dialogue* method is that it does not require performing a loudness scaling test prior to listening. It should be also noted that for speech samples processed with *Ear Tune Up* ("male_voice_ETU," "Skype_female_ETU") an increase in signal quality was observed. For the sample "Skype_male_ETU" the test result was inconclusive.

The standard deviations of the results are small, compared to the previous three test results. The number of selected listeners is relatively high (25 out of 30). This means that the listeners had a clear impression that the processed samples are easier to perceive and the dialogue is easier to understand. This result proves the usefulness of the developed algorithms. Therefore, it was decided that the engineered algorithms will be compiled as an integrated software provided with an elaborated GUI (Fig. 28).

The prepared experimental software bundle integrates all algorithms described in the paper plus the bass enhancer [7][8] and the loudness maximizer (multi-band compressor) omitted in order to limit the manuscript length.

5 CONCLUSIONS

Methods for improving the sound quality in computer devices by personalized and adaptive processing were introduced in this work. The algorithms for linearization of frequency response, dialogue intelligibility enhancement, and personalized dynamics processing were presented. The introduced methods and algorithms were tested on two different computers (All-in-One and laptop), both located in a quiet office-like condition and in the presence of strong noise. The methods were evaluated by means of subjective listening tests. The analysis of the results leads to the following conclusions:

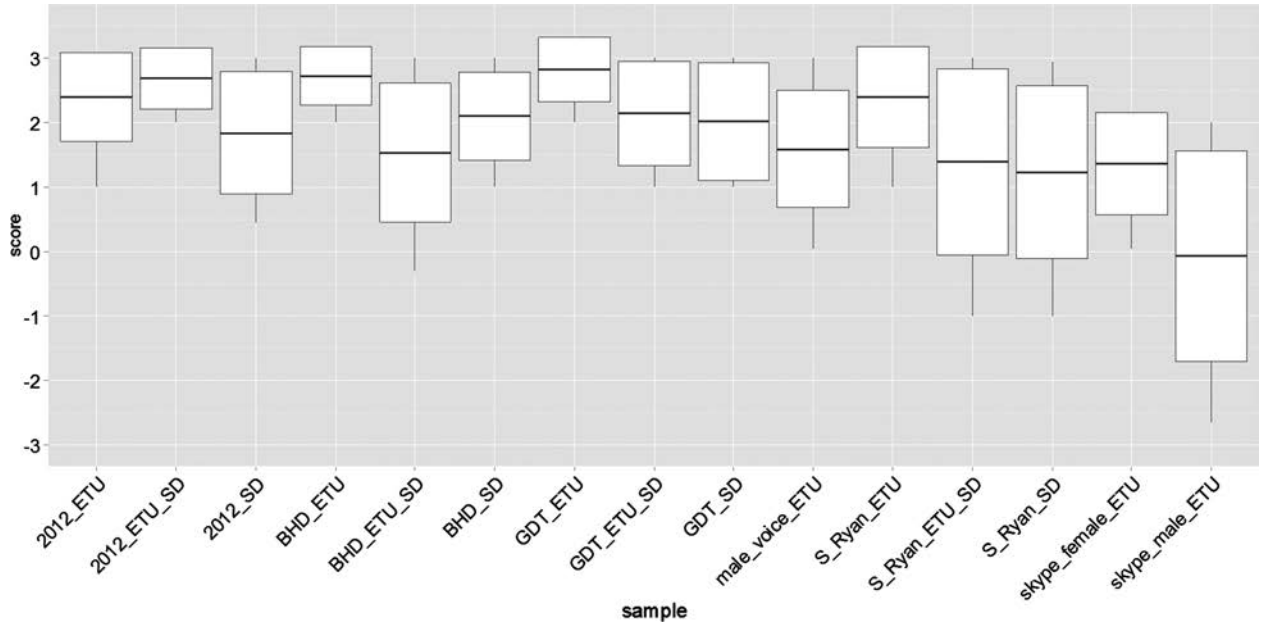


Fig. 27 Scores in dialogue intelligibility enhancement and dynamics processing test. Boxes indicate mean and standard deviation, whiskers indicate 5 and 95 percentiles.

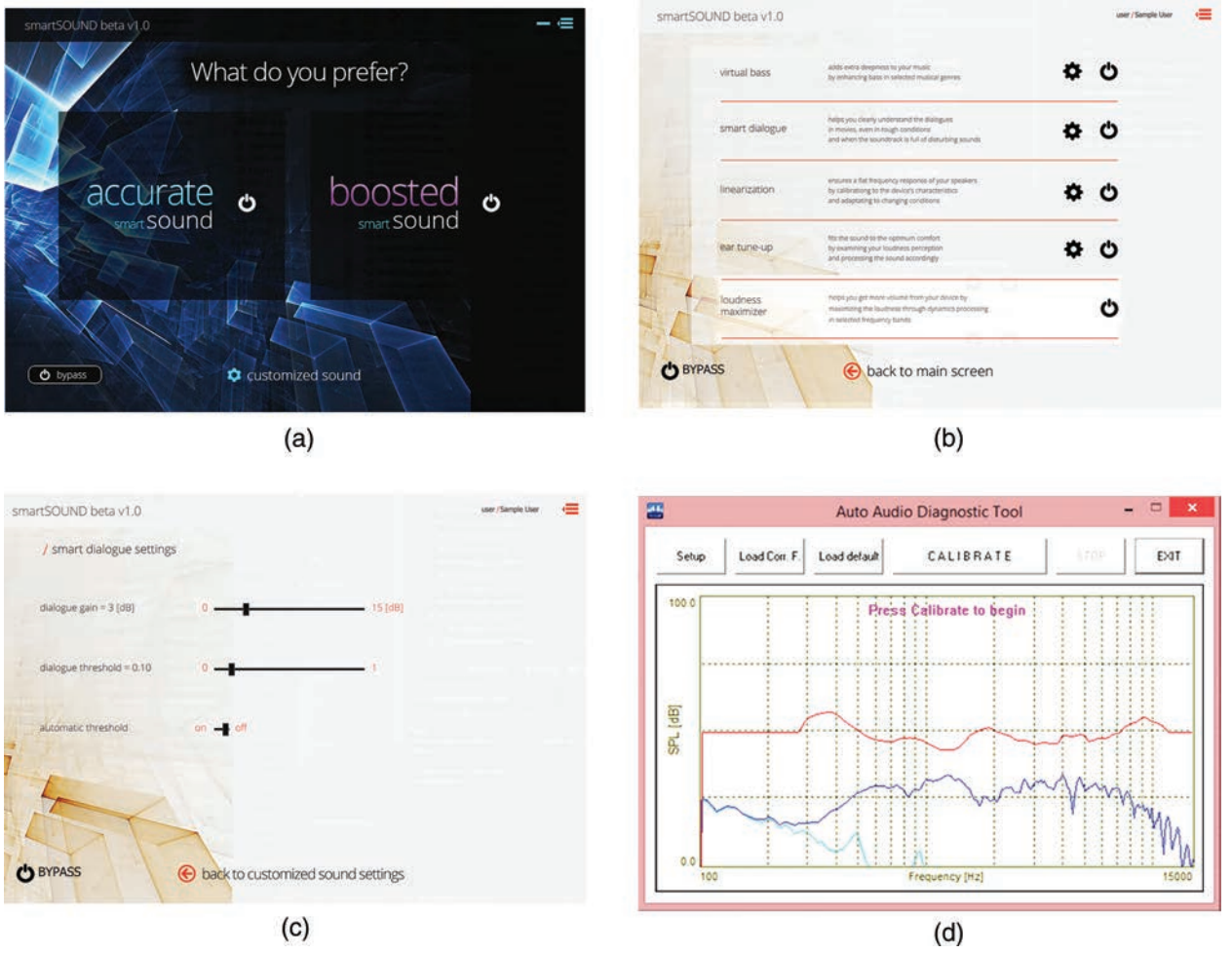


Fig. 28 Selected screens of the multi-layer GUI of experimental "SmartSound" software developed for Windows® OS: start screen (a), algorithms selection (b), settings of "Smart Dialogue" (c), autodiagnostic tool of linearization algorithm (d).

1) The majority of mean comparative scores (CMOS) brought positive values, which means that the listeners perceive the effect of processing as desirable and could be interested in using the proposed methods;
2) The CMOS scores greater than 1 can be considered a significant improvement. Such scores were assigned to numerous samples;
3) The linearization algorithm enables an improvement in the perceived spectral quality of sound. In the test of the linearization algorithm on the All-in-One computer all samples obtained positive CMOS scores and 7 out of 9 samples were given a score greater than 1;
4) The effect of the developed algorithm for the personalized dynamics processing depends on the level of the input signals. The performance is best for the samples with low or medium level. Samples with high level were also evaluated positively, however they obtained slightly lower scores than low- and medium-level samples. It proves that the developed method serves the purpose of adjusting the dynamics of sound to the user's hearing preferences by raising the level of the fragments of the signal that are too soft;
5) The combination of linearization and dynamics processing performed better than dynamics processing alone while testing on the laptop. The explanation is that the introduced loudness maximization underlines the irregularities in the frequency response of the device. Thus, the samples with loudness maximization obtained negative CMOS scores. However, if the linearization was added, the frequency response was improved and the acoustic effect was assessed as a positive one;
6) The test for speech clarity in noise (dialogue enhancement and dynamics processing) yielded the highest scores for the processed samples. Both the dialogue enhancement method and the loudness maximization led to a significant improvement of the signal quality, however the loudness maximization performed slightly better in this case;
7) Most of the listeners were untrained, i.e., not skilled in audio engineering and not knowing the principles of the algorithms. It is known and expected phenomenon that the deviation of scores in such a case can be large. This fact has been confirmed in numerous studies [38][39][40][41], including the presented one;
8) The dispersion of scores is smaller in the case of the the All-in-One computer, because the overall sound quality of this device is better than in the case of the laptop. Some listeners commented that they found it difficult to decide which sample is better, even though they heard a difference quite clearly. It shows that for untrained listeners it is a difficult task to align the subjective impression with the objective sound quality. It was also observed that some listeners prefer the sound without any linearization, probably due to their listening habits, i.e., being used to the colorized sound typical for mobile computers. A significant number of results were discarded with regard to the above-mentioned reasons, because the responses given by the listeners were inconsistent;
9) The evaluation was performed in the form of the so-called "blind test," i.e., the listeners did not know which sample is processed currently. It has to be noted that the perceived sound quality can be influenced by many other factors beyond the characteristics of the signal. It was shown in the literature that the appearance of the audio processing software GIU may have an influence on the subjective impression of the listeners [42]. The manufacturers of audio post-processing software often utilize catchy brand names to bias the listener towards liking the introduced algorithm. Our evaluation was free of such suggestions, therefore it can be considered an unbiased evaluation of the subjective audio quality.

In general, the tests results show that the introduced audio processing methods make useful tools for the improvement of the sound quality in compact computers. It was presented how the engineered algorithms successfully raise the perceived quality of sound in such aspects as frequency characteristics, dynamics, and speech clarity.

6 ACKNOWLEDGMENTS

This work was supported by the grant No. PBS1/B3/16/2012 entitled "Multimodal system supporting acoustic communication with computers" financed by the Polish National Centre for Research and Development.

7 REFERENCES

[1] R. Turnbull, P. Hughes and S. Hoare, "Audio Enhancement for Portable Device Based Speech Applications," presented at the *124th Convention of the Audio Engineering Society* (2008 May), convention paper 7350.

[2] "Dolby PC Entertainment Experience v4," White Paper, Dolby Laboratories, San Francisco (2010).

[3] Maxx by Waves Maxx Audio Technology, http://www.maxx.com/technologies/maxxaudio/

[4] M. Arora, H. G. Moon, and S. Jang, "Low Complexity Virtual Bass Enhancement Algorithm for Portable Multimedia Device," presented at the *AES 29th International Conference: Audio for Mobile and Handheld Devices* (2006 Sep.), conference paper 4-3.

[5] S. Cecchi, M. Virgulti, A. Primavera, F. Piazza, F. Bettarelli, and J. Li, "Investigation on Audio Algorithms Architecture for Stereo Portable Devices," *J. Audio Eng. Soc.,* vol. 64, pp. 75–88 (2016 Jan./Feb.). http://dx.doi.org/10.17743/jaes.2015.0084

[6] K. Drossos, S. I. Mimilakis, A. Floros, and N. G. Kanellopoulos, "Stereo Goes Mobile: Spatial Enhancement for Short-Distance Loudspeaker Setups," *(IIH-MSP) 2012, Eighth International Conference on Intelligent Information Hiding and Multimedia Signal Processing*, pp. 432–435 (2012 July). http://dx.doi.org/10.1109/iih-msp.2012.111

[7] P. Hoffmann, T. Sanner and B. Kostek, "Smart Virtual Bass Synthesis Algorithm Based on Music Genre Classification," *The 18th IEEE SPA Conference*, Poznan (2014 Sep.).

[8] P. Hoffmann and B. Kostek, "Evaluation of a Novel Approach to Virtual Bass Synthesis Strategy," presented at the *138th Convention of the Audio Engineering Society* (2015 May), eBrief 184.

[9] K. Lopatka, J. Kotus, P. Suchomski, A. Czyżewski, and B. Kostek, "Personal Adaptive Tuning of Mobile Computer Audio," presented at the *139th Convention of the Audio Engineering Society* (2015 Oct.), convention paper 9455.

[10] J. Kotus, A. Ciarkowski, and A. Czyżewski, "Auto Adaptation of Mobile Device Characteristics to Various Acoustic Conditions," presented at the *136th Convention of the Audio Engineering Society* (2014 Apr.), eBrief 136.

[11] J. Kotus, "Application of Auto Calibration and Linearization Algorithms to Improve Sound Quality of Computer Devices," *IEEE SPA Conference*, Poznan (2015). http://dx.doi.org/10.1109/spa.2015.7365151

[12] B. R. Glasberg and B. C. J. Moore, "Derivation of Auditory Filter Shapes from Notched-Noise Data," *Hearing Research*, vol. 47, no. 1-2, pp. 103–138 (1990). http://dx.doi.org/10.1016/0378-5955(90)90170-t

[13] F. Gao and W. Snelgrove, "Adaptive Linearization of a Loudspeaker," *1991 International Conference on Acoustics, Speech, and Signal Processing*, vol. 5, pp. 3589–3592 (1991 Apr.). http://dx.doi.org/10.1109/icassp.1991.150251

[14] G. Szwoch, A. Czyzewski, and A. Ciarkowski, "A Double-Talk Detector Using Audio Watermarking," *J. Audio Eng. Soc.*, vol. 57, pp. 919–926 (2009 Nov.). http://dx.doi.org/10.1121/1.3508209

[15] J. J. Shynk, "Frequency-Domain and Multirate Adaptive Filtering," *Signal Processing Magazine, IEEE*, vol. 9, no. 1, pp. 14–37 (1992 Jan.), http://dx.doi.org/10.1109/79.109205

[16] S. Haykin, *Adaptive Filter Theory*, 5th ed. (Harlow, England, Pearson, 2014).

[17] A. Ciarkowski, dsp++: library of portable DSP algorithms using modern C++ language, https://bitbucket.org/andrzejc/dsp (visited: 2015 Dec.).

[18] K. Lopatka, B. Kunka, and A. Czyzewski, "Novel 5.1 Downmix Algorithm with Improved Dialogue Intelligibility," presented at the*134th Convention of the Audio Engineering Society* (2013 May), convention paper 8931.

[19] K. Lopatka, "Detection of Dialogue in Movie Soundtrack for Speech Intelligibility Enhancement," *7th International Conference on Multimedia Communications Services and Security*, vol. 429, pp. 149–158, Kraków (2014 June). http://dx.doi.org/10.1007/978-3-319-07569-3_12

[20] K. Lopatka, A. Czyzewski, and B. Kostek, "Improving Listeners' Experience for Movie Playback through Enhancing Dialogue Clarity in Soundtracks," *Digital Signal Processing*, vol. 48, issue C, pp. 40–49, (2016 Jan.), http://dx.doi.org/10.1016/j.dsp.2015.08.015.

[21] H. Fuchs, "Clean Audio," ITU Workshop on "Making Media Accessible to all: the Options and the Economics" (2013 Oct.).

[22] J. Herre et al., "MPEG Spatial Audio Object Coding - The ISO/MPEG Standard for Efficient Coding of Interactive Audio Scenes," *J. Audio Eng. Soc.*, vol. 60, pp. 655–673 (2012 Sep.).

[23] M. Kotti, E. Benetos, C. Kotropoulos, and I. Pitas, "A Neural Network Approach to Audio-Assisted Movie Dialogue Detection," *Neurocomputing*, Elsevier, vol. 71 (1-3), pp. 157–166 (2007). http://dx.doi.org/10.1016/j.neucom.2007.08.006

[24] T.-W. Lee, M. S. Lewicki, M. Girolamo, and T. J. Sejnowski, "Blind Source Separation of More Sources than Mixtures Using Overcomplete Representations," *Signal Processing Letters, IEEE*, vol. 6, no. 4, pp. 87, 90 (1999 April). http://dx.doi.org/10.1109/97.752062

[25] D. Barry, R. Lawlor, and E. Coyle, "Real-Time Sound Source Separation: Azimuth Discrimination and Resynthesis," presented at the *117th Convention of the Audio Engineering Society* (2004 Oct.), convention paper 6258.

[26] standard Technical: "ITU-R B S.775-3 - Multichannel stereophonic sound system with and without accompanying picture," *International Telecommunication Union* (2006).

[27] C. Avendano and J-M. Jot, "A Frequency-Domain Approach to Multichannel Upmix," *J. Audio Eng. Soc.*, vol. 52, pp. 740–749 (2004 Jul./Aug.).

[28] K. Lopatka, "Detection of Dialogue in Movie Soundtrack for Speech Intelligibility Enhancement," *7th International Conference on Multimedia Communications Services and Security*, vol. 429, pp. 149–158, Kraków (2014 June). http://dx.doi.org/10.1007/978-3-319-07569-3_12

[29] B. Prando, D. Little, and D. Gergle, "Building a Personalized Audio Equalizer Interface with Transfer Learning and Active Learning," *Proc. of 2nd International ACM Workshop on Music Information Retrieval with User-centered and Multimodal Strategies*, pages 13–18, New York (2012). http://dx.doi.org/10.1145/2390848.2390852

[30] G. W. McNally, "Dynamic Range Control of Digital Audio Signal," *J. Audio Eng. Soc.*, vol. 32, pp. 316–327 (1984 May).

[31] M. Walsh, E. Stein, and J. M Jot, "Adaptive Dynamics Enhancement," presented at the *130th Convention of the Audio Engineering Society* (2011 May), convention paper 8343.

[32] B. Kostek, P. Suchomski, and P. Odya, "Loudness Scaling Tests in Hearing Problems Detection," presented at the *AES 58th International Conference: Music Induced Hearing Disorders* (2015 June), conference paper 1-3.

[33] C. Elberling, "Loudness Scaling Revisited," *J. Am. Acad. Audiol.*, no. 10, pp. 248–260 (1999).

[34] J. B. Allen, J. L. Hall and P. S. Jeng, "Loudness Growth in ½ Octave Bands (LGOB) – A Procedure for the Assessment of Loudness," *J. Acoust. Soc. Am.*, vol. 88, no. 2, pp. 745–753 (1990). http://dx.doi.org/10.1121/1.399778

[35] P. Suchomski and B. Kostek, "Fitting of the Sound Dynamics Characteristics to the Hearing Preferences of the User of Mobile Devices," *Przegląd Telekomunikacyjny + Wiadomości Telekomunikacyjne*, no. 8-9, pp. 1360–1364, Kraków (2015).

[36] "ITU-T Recommendation P.800, Methods for Subjective Determination of Transmission Quality," ITU (1996).

[37] H. Kurtulus Ozcan and S. Nemlioglu, "In-Cabin Noise Levels during Commercial Aircraft Flights," *J. Canadian Acous. Assn.*, vol. 34, no. 4, pp. 31–35 (2006).

[38] F. Rumsey, "Subjective Assessment of the Spatial Attributes of Reproduced Sound," *AES 15th International Conference: Audio, Acoustics & Small Spaces* (1998 October), conference paper 15-012.

[39] F. Rumsey, "Spatial Quality Evaluation for Reproduced Sound: Terminology, Meaning and a Scene-Based Paradigm," *J. Audio Eng. Soc.*, vol. 50, pp. 651–666 (2002 Sep.).

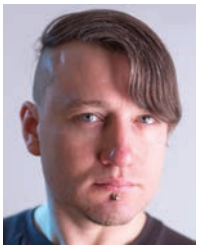
[40] S. Zielinski, "On Some Biases Encountered in Modern Listening Tests," *Spatial Audio & Sensory Evaluation Techniques*, Guildford, UK (2006 April). http://dx.doi.org/10.17743/jaes.2015.0094

[41] I. McGregor, P. Turner, and D. Benyon, "Using Participatory Visualization of Soundscapes to Compare Designers' and Listeners' Experiences of Sound Designs," *J. Sonic Studies*, vol. 6, no. 1 (2014 Jan.).

[42] M. Lech and B. Kostek, "Testing a Novel Gesture-Based Mixing Interface," *J. Audio Eng. Soc.*, vol. 61, pp. 301–313 (2013 May).

THE AUTHORS


A. Czyżewski


A. Ciarkowski


B. Kostek


J. Kotus


K. Łopatka


P. Suchomski

Andrzej Czyzewski is a full professor and Head of the Multimedia Systems Department and is author of more than 500 scientific papers in international journals and conference proceedings. He has led more than 30 R&D projects funded by the Polish Government and participated in 7 European projects. He is also author of 12 Polish patents and 7 international patents. He has extensive experience in soft computing algorithms and their applications to sound and image processing. He is a recipient of many prestigious awards, including two First Prizes of the Prime Minister of Poland for research achievements (in 2000 and in 2015) and the Audio Engineering Society Fellowship in 2000.

•

Andrzej Ciarkowski was born in Gdansk in 1979. He received M.Sc. degree in telecommunications from the Technical University of Gdansk in 2003, majoring in sound and vision engineering. Soon after he joined the Multimedia Systems Department team and was working on research projects in the field of intelligent surveillance, archival record reconstruction, and quality improvement of mobile devices audio. His main scientific interest is audio signal processing. He is an author and co-author of over 30 published papers and the developer behind Open Source dsp++ library project.

•

Bozena Kostek holds professorship at the Faculty of Electronics, Telecommunications and Informatics, Gdansk University of Technology (GUT), Poland. She is Head of the Audio Acoustics Laboratory. She received her M.Sc. degrees in sound engineering (1983) and organization and management (1986) from GUT. She also received postgraduate DEA degree (1988) from Toulouse University, France. In 1992 she supported her Ph.D. thesis with honors at GUT, and in 2000 her D.Sc. degree at the Research Systems Institute, Polish Academy of Sciences. In 2005 the President of Poland granted her the title of Professor. She published over 500 scientific papers in journals and at international conferences. She serves as the Editor-in-Chief of the *Journal of the Audio Eng. Soc.* since 2011. She was the recipient of many prestigious awards for research, including those of the Prime Minister of Poland for outstanding research achievements. She received the Audio Eng. Soc. Fellowship Award in 2010 and the AES Citations in 2013 and 2015. In 2013 Prof. Kostek was been elected as a member of the Polish Academy of Sciences. In 2003–2007 and 2009–2011, she was elected Vice-President of the Audio Engineering Society for Central Europe. In 2007–2009 and in 2011–2013 she served as the AES Governor. She holds the title Fellow of the Audio Engineering Society.

•

Jozef Kotus graduated from the Faculty of Electronics Telecommunications and Informatics, Gdansk University of Technology in 2001. In 2008 he completed his Ph.D. under the supervision of prof. Bozena Kostek. His Ph.D. work concerned issues connected with application of information technology to noise monitoring and prevention of the noise-induced hearing loss. He is a member of the Audio Engineering Society (AES) and European Acoustics Association (EAA). Until now he is an author and co-author more than 50 scientific publications, including 14 articles from the ISI Master Journal List and 32 articles in reviewed papers. He has extensive experience in sound and image processing algorithms.

•

Kuba Łopatka graduated from Gdansk University of Technology in 2009, majoring in sound and vision engineering. He completed his doctoral studies in 2013 at the Multimedia Systems Department and in 2015 he supported his Ph.D. dissertation on detection and classification of hazardous acoustic events. His scientific interest lies in audio, signal processing, speech acoustics, and pattern recognition. He is an author or co-author of over 30 published papers, including 5 articles in journals from the ISI master journal list. He has taken part in various research projects concerning intelligent surveillance, multimodal interfaces, and sound processing.

•

Piotr Suchomski was born in Gdansk in 1973. He received his M.Sc. degree in informatics from the Technical University of Gdansk in 1997. His specialty was multimedia techniques. In 2005 he received his Ph.D degree. His Ph.D thesis was devoted to developing of the new hearing aid fitting methodology and system on the basis of examination of speech in noise signal perception. His research interest is in multimedia systems (applications) developing and programming, audio-video signal processing, and modern audio-video postproduction techniques.