OPEN ACCESS Freely available online
Journal of the Audio Engineering Society
Vol. 63, No. 12, December 2015 (© 2015)
DOI: http://dx.doi.org/10.17743/jaes.2015.0087

Bass Enhancement Settings in Portable Devices Based on Music Genre Recognition

PIOTR HOFFMANN, *AES Member,* **AND BOŻENA KOSTEK,** *AES Fellow*
(phoff@sound.eti.pg.gda.pl) (bokostek@audioakustyka.org)

Audio Acoustics Laboratory, Faculty of Electronics, Telecommunications and Informatics, Gdansk University of Technology, Narutowicza 11/12, 80-233 Gdansk

The paper presents a novel approach to the Virtual Bass Synthesis (VBS) applied to mobile devices, called Smart VBS (SVBS). The proposed algorithm uses an intelligent, rule-based setting of bass synthesis parameters adjusted to the particular music genre. Harmonic generation is based on a nonlinear device (NLD) method with the intelligent controlling system adapting to the recognized music genre. To automatically classify music genres, the k-Nearest Neighbor classifier combined with the Principal Component Analysis (PCA) method is employed. To fine tune the SVBS algorithm, the MUSHRA test is performed. Subjects are presented with music excerpts belonging to various genres, unprocessed and also processed by SVBS and a conventional bass boost algorithm. Listening tests show that subjects in most cases prefer the SVBS strategy developed by the authors in favor of both the conventional bass boost algorithm and the unprocessed audio file. Furthermore, the listeners indicated that perception of the SVBS-processed music excerpts is similar for several types of portable devices.

Improving the low frequency sound of mobile devices is a problem that appears in many studies [1, 2]. Most of the algorithms increase the power of low frequencies to enhance low frequency sound. However, in mobile technology, increasing the signal power may easily lead to the signal overdriving. An additional problem of mobile devices is the limited frequency range of speakers in the lower frequency bands. The existing solutions often require the user to adjust the low frequency manually to adapt it to particular conditions, which may be a task of great difficulty for less-skilled users. Therefore, the users of electronic devices increasingly expect automatic, intelligent solutions that can aid them in configuring their software. Automation is designed to help the user in the process of calibration and adjusting equipment for their needs. This leads to better exploitation of the potential of the device by users without technical expertise required for manual configuration.

In the rapidly expanding mobile market, the users expect devices to be constantly on the edge of the technological development and rich in sophisticated functionalities. More and more versatile devices extend their range of applications in everyday life. In addition, users often employ the same device for both work and entertainment. Then again, the growing interest in mobile devices makes the manufacturers increase versatility of new models and adapt them to the needs expressed by the market.

One of the key aspects in multimedia applications is the user listening experience. However, in most mobile devices, the quality of the built-in speakers and the sound they produce are very low. This results from the fact that small speakers of poor quality do not transfer low frequencies and significantly distort the reproduced content due to their physical limitations. This problem is particularly significant in the case of listening to music, when a listener expects true reproduction of sound. Physical limitations of mobile devices can even make certain parts of the musical instrumentation completely inaudible. An additional difficulty in listening in such conditions is the unfavorable structure of the device, which is not sufficiently conductive to transmit low frequencies. This is due to very small dimensions of the speakers which limit the formation of reinforcement of the low-frequency resonances.

The proposed solution, called Smart Virtual Bass Synthesis (SVBS), uses an intelligent adaptation of parameters of the algorithm to improve the low frequency sound quality depending on the music genre being reproduced [3]. Content of audio files processing using the Smart VBS algorithm is transparent to the user. The user's participation in the process is not necessary. This means that no special technical skills are required to benefit from the algorithm performance.

The paper first provides a short theoretical and algorithmic background of Virtual Bass Synthesis (VBS) methods, then presents the Smart VBS solution. Next, the conducted subjective tests along with the analysis of the results are described. To obtain listeners' rates and opinions the

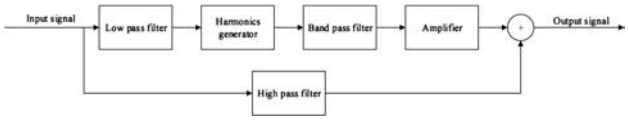

Fig. 1. General block diagram for enhancement the low-frequency by nonlinear element method.

MUSHRA test methodology was utilized. A group of 31 persons participated in the evaluation of the Smart VBS strategy. Subjects in most cases prefer the Smart VBS (SVBS) strategy developed by the authors in favor of both the conventional bass boost algorithm and the unprocessed audio file.

1 TECHNICAL BACKGROUND

1.1 Missing Fundamental Frequency Phenomenon

The missing fundamental frequency phenomenon was first described by Seebeck in 1841. He discovered that pitch is perceived more strongly in the case of complex sound than in the case of simple tones [4]. First descriptions of the phenomenon of the missing fundamental frequency appeared in the 1930s. Schouten et al. published a paper on this subject [5]. The most well-known work related to this phenomenon was published in 1962 [6]. The authors made an observation that the complex sound pitch is perceived resulting from the distribution of harmonics in the spectrum band rather than from the lowest frequency present in the spectrum.

Due to the above observation, a new interest in this phenomenon of pitch was formulated. Terhardt distinguishes real pitch—spectral and virtual—the residual [7]. Residual pitch refers to the human ability to hear the first partial of a harmonic sound even if it is not present within the spectrum of the sound. Real pitch is perceived when the sound spectrum is in a frequency band corresponding to the perceived pitch. In case where the frequency spectrum does not occur the perceived amount is an amount of the resulted virtual one.

On the basis of this phenomenon described in the literature, a family of algorithms designed to enhance the low frequency sound by adding harmonics to the sound of the bass was proposed. In case where a low fundamental frequency sound cannot be reproduced because of physical imperfection of the speaker, this is compensated by the frequency of occurrence of this impression, formed as virtual pitch.

Virtual bass algorithm implementations present in the literature use the nonlinear element method, the method of phase vocoder, or the hybrid method. The NLD (Non-Linear Devices) method operates in the time domain and introduces harmonic distortion in accordance with the nonlinear function used. Thorough research studies were carried out to describe mathematically the structure of the harmonics generated by various functions [8–10]. Resulting from these studies a variety of mathematical functions were introduced and used. The NLD method faces a major problem with generated harmonic structure that changes depending on the amplitude of the input signal. This is a very undesirable feature. The NLD method achieves the best results for the transient state, since it has a high temporal resolution. The phase vocoder (PV) method is based on the signal processed in the frequency domain and allows for precise control of various harmonics in both amplitude and frequency, and even phases. What is important, this method is insensitive to changes of the amplitude of the input signal. The disadvantage of PV is that to obtain suitable frequency resolution, it is required to provide a relatively extensive analysis window in time and a relatively high bass sound length (about 100 ms). Hence, the method is unsuitable for fast processing of transient changes, while giving good results in the case of the sound with a long steady state.

In order to exploit the advantages of both methods hybrid systems are proposed in the literature [1, 2]. In the method proposed by Hill and Hawksford [2] the main point is the detection of transient states. Then, the signal is subjected to processing by the PV or NLD method. If the input signal is transient in nature, it is preferred to process the signal using the NLD method, and when the signal is in a steady state, the PV method is preferred. The method uses a transient detector, which operates in CQT transform. On the basis of the analysis, appropriate weights to the signals processed by the NLD and PV methods are selected [2]. A new approach that first separates musical signals into transient and steady-state components, and then applies NLD and PV on the separated signals was proposed by Mu et al. [1]. The authors of this solution tested the proposed hybrid approach against the NLD and PV methods using MUSHRA environment. The results of this approach are very promising.

In the literature, the following steps have been described that should be performed to obtain an increase in the perception of a low-frequency signal [11]. The following description is consistent with the blocks in Fig. 1.

The first step concerns low pass filtering (1). The task is to select a filter bandwidth that is not efficiently processed by the speaker. Frequencies of the filter should be matched to the type of speaker [11]. In addition, bandwidth should not be too wide because of aliasing that will be produced by additional harmonics [11]. In addition, an important element is to remove frequencies below 20 Hz [12]. Commonly, IIR filters are used in the third or fourth order of [12]. Generation of additional harmonics (2) is provided by the method of nonlinear elements (NLD) or a phase vocoder (PV). When processing a low frequency bandwidth, it is recommended to use several sub-bands, which should be handled by separate sections generating harmonics. There exist many non-linear functions. Desirable features include low computational complexity and independence of the structure resulting from the amplitude of the audio input signal [12]. Band-pass filtration (3) operation is to choose the right amount of harmonics that are added to the input signal. Amplification (4) of the generated harmonic amplitude for the right output level is then made. High-pass filtering (5) of the input signal is processed in

such a way as to reduce the power needed for processing by the low-frequency loudspeaker. This block does not always have to be used [12]. Diagram of the signal processing algorithm using the virtual bass synthesis is shown in Fig. 1.

1.2 Music Information Retrieval System

Music information retrieval systems (MIR) require interdisciplinary knowledge for the preparation of well-functioning practical solutions [13, 14], especially in the case when a delay caused by music excerpt processing and classifying may be an issue. Music signal processing methodology exploits such issues as databases storing music tracks, parametrization, decision algorithms, presentation of the results, copyright management system, and some other important topics. For the purpose of this article a description of the system for recognizing musical genres prepared by the authors will shortly be presented with regard to the above mentioned notions [15–18].

The main component of musical genre recognition systems is the optimized parametrization block. The prepared feature vector (FV) in this block should have a very good separability between parameters. Taking into account these assumptions, the feature vector containing 173 elements was conceived in earlier research studies carried out by the authors [13, 17, 19]. A collection of 52532 music excerpts described with a set of descriptors obtained through the analysis of mp3 recordings was gathered in a database called SYNAT. The SYNAT database was realized by the Gdansk University of Technology (GUT) [13, 15]. For the recordings included in the database, the analysis band is limited to 8 kHz due to the music excerpts format, this means that the frequency band used for the parametrization is in the range from 63 to 8000 Hz. The prepared feature vector is used to describe parametrically each signal frame. The database stores 173-feature vectors, which in majority are the MPEG-7 standard parameters [20]. The vector has additionally been supplemented with 20 Mel-Frequency Cepstral Coefficients (MFCC), 20 MFCC variances, and 24 time-related "dedicated" parameters. The vector includes parameters associated with the MPEG-7 standard, melcepstral (MFCC) parameters and is enlarged by the so-called dedicated parameters that refer to the temporal characteristic of the analyzed music excerpt; their names are included in Table 1. The list of parameters and their definitions were shown in the earlier study [15], however, it is worth noting that the proposed FV was used in the ISMIS 2011 contest in which there were over 120 participants [17]. The best contest result returned almost 88% of accuracy [17], and later in the authors' own study gained even better effectiveness [21].

The 173-element vector generates a very large amount of information describing a given track. As a consequence, this leads to an extensive amount of data undergoing classification, which in the context of using the *k*-Nearest Neighbor classifier is important. Therefore, Principal Component Analysis (PCA) was applied to reduce the data redundancy as it transforms a number of possibly correlated variables into a smaller number of variables called principal components [22]. PCA is defined as a non-parametric method of extracting relevant information from complex data sets. This is to identify patterns in the data and present them in such a way as to indicate their similarities and differences. To quantify the redundancy between data is to use the variance of the data to prepare a new set of parameters. The new components are a linear combination of parameters that carry most information about the test set, thus they no longer refer to descriptors contained in the original feature vector. The PCA method can shorten the feature vector of 173 elements to 19 components, which significantly reduces the computation time. Furthermore, the use of the described analysis can increase classification efficiency, as shown in the earlier paper of the authors [21].

Recognition of music genres started with the *k*-Nearest Neighbors algorithm (*k*-NN), as it is one of the most commonly used classification algorithms [23]. An object is classified by a majority vote of its neighbors, with the object being assigned to the class most common among its *k* nearest neighbors measured by a distance function. To calculate the distance function, the Euclidean metric is often used:

$$d(A, B) = \sqrt{(x_a - x_b)^2 + (y_a - y_b)^2} \tag{1}$$

where:

x, y – parameters value for A and B objects,
d – Euclidean distance metric.

The *k*-NN algorithm effectiveness was compared with the Bayesian networks algorithm and Support Vector Machines using Sequential Minimal Optimization (SMO) [16] to obtain the highest efficiency of classification. In Fig. 2, the results of the classification of genres obtained on 32110 tracks, selected from the SYNAT database, are shown. The database containing feature vectors is available at http://www.audioakustyka.org/modality-mir/. Classification involved 11 musical genres. The effectiveness of classification for the prepared feature vector with the PCA algorithm utilized is at a very good level of over 85% accuracy. The experiments were conducted using 10-fold cross validation. In 10-fold cross validation the entire collection of tracks is divided into 10 disjoint subsets that are used for learning and testing. For each data split the classifier is retrained with the training examples from individual subsets and estimated with the test examples. The training set is changed 10 times along with the test set [24].

The detailed analysis of the experiment results shown in Fig. 2 is presented in Fig. 3. Apart from the previously outlined accuracy results, Precision, Recall, and F-Measure metrics were analyzed. The obtained values of additional statistical metrics depend on the number of parameters used in the classification process. The analyzed algorithms return similar metrics while using the same number of parameters. Similarly to the overall effectiveness of the presented recognition system the highest values of the recall and precision were obtained from the algorithm *k*-Nearest Neighbors. Values of statistical metrics approach 90% starting

Table 1. The list of parameters within the SYNAT music database [15].

No.	Parameter
1	Temporal Centroid
2	Spectral Centroid
3	Spectral Centroid variance
4–32	Audio Spectrum Envelope for particular bands
33	ASE average for all bands
34–62	ASE variance values for particular bands
63	averaged ASE variance
64	average Audio Spectrum Centroid
65	variance of Audio Spectrum Centroid
66	average Audio Spectrum Spread
67	variance Audio Spectrum Spread
68–87	Spectral Flatness Measure for particular bands
88	SFM average value
89–108	Spectral Flatness Measure variance for particular bands
109	averaged SFM variance
110–129	Mel-Frequency Cepstral Coefficients for particular bands
130–149	MFCC variance for particular bands
150	number of samples exceeding RMS
151	number of samples exceeding 2×RMS
152	number of samples exceeding 3×RMS
153	mean value of samples exceeding RMS, averaged for 10 frames
154	variance value of samples exceeding RMS, averaged for 10 frames
155	mean value of samples exceeding 2×RMS, averaged for 10 frames
156	variance value of samples exceeding 2×RMS, averaged for 10 frames
157	mean value of samples exceeding 3×RMS, averaged for 10 frames
158	variance value of samples exceeding 3×RMS, averaged for 10 frames
159	peak to RMS ratio
160	mean value of the peak to RMS ratio calculated in 10 subframes
161	variance of the peak to RMS ratio calculated in 10 subframes
162	Zero Crossing Rate
163	RMS Threshold Crossing Rate
164	2×RMS Threshold Crossing Rate
165	3×RMS Threshold Crossing Rate
166	Zero Crossing Rate averaged for 10 frames
167	Zero Crossing Rate variance for 10 frames
168	RMS Threshold Crossing Rate averaged for 10 frames
169	RMS Threshold Crossing Rate variance for 10 frames
170	2×RMS Threshold Crossing Rate averaged for 10 frames
171	2×RMS Threshold Crossing Rate variance for 10 frames
172	3×RMS Threshold Crossing Rate averaged for 10 frames
173	3×RMS Threshold Crossing Rate variance for 10 frames

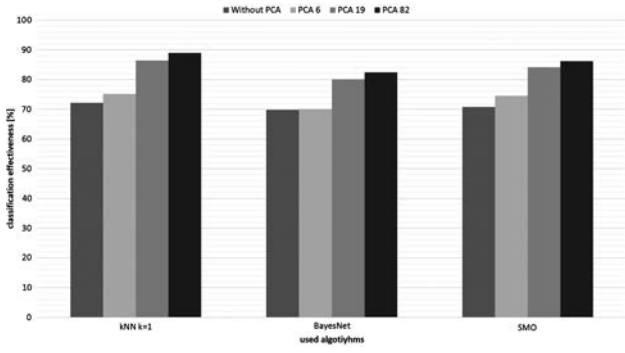

Fig. 2. The effectiveness of *k*-NN, Bayesian networks and SMO classification algorithms using the PCA method on 32110 tracks database.

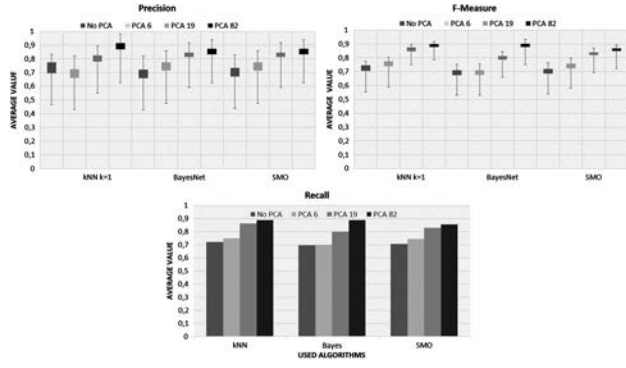

Fig. 3. Precision, Recall, and F-Measure values for genre recognition.

from 19 PCA components. The classification precision of the algorithms used in the process of recognizing musical genres is moderately accurate, a clear dispersion of the values within +/– 0.4 can easily be observed. Fluctuations of the obtained statistical values result from different levels of recognition acquired for different music genres. The values

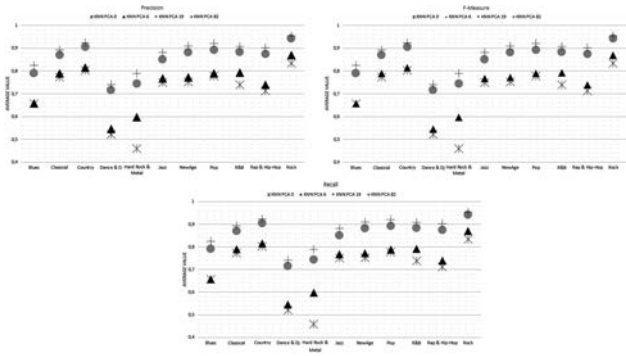

Fig. 4. Precision, Recall, and F-Measure metrics values for the k-NN algorithm.

of recall and precision are in most cases similar, and their values correspond to the total efficiency of the system that is confirmed by F-Measure.

For the *k*-NN algorithm, which is the most effective algorithm from the tested ones, detailed values of statistical measures with distinction drawn between different musical genres are shown in Fig. 4. In the developed genre recognition system the recall of the algorithm is closely associated with its precision. The obtained measures are closely correlated, which is confirmed by the F-Measure metric. Music genres recognized with the lowest recall and precision are the Hard Rock and Dance, and DJ. These genres are very similar to Rock and Pop and this negatively affects the classification process. It is also important to note that the number of samples in the learning process, in the case of genres Rock and Hard Rock, is almost four times higher than for the Rock genre. The best results were obtained for Country style. It can be seen that for most of the tests carried out retaining additional parameters in the PCA analysis improves the overall effectiveness of the music genre recognition process.

2 SMART VBS

The proposed solution boosts low frequencies in mobile devices, taking into account the content of the audio file. The proposed algorithm uses the NLD low frequency synthesis method for generating harmonics and adding them to the signal. Accurate synthesis parameters are determined based on automatically recognized genre. By exact matching the content of the synthesized signal and the device on which the file is reproduced, it is possible to introduce a minor distortion to the signal. The algorithm, in contrast to standard solutions, does not enhance the signal below the cut-off frequency of the speakers. Signal modification takes place in the band above the frequency limit, thus using the phenomenon of missing fundamental frequency. In Fig. 5, the chart showing the low frequency gain by traditional bass boost algorithm and the proposed Smart VBS is presented. The black line shows the measured cut-off frequency of the tested laptop speakers.

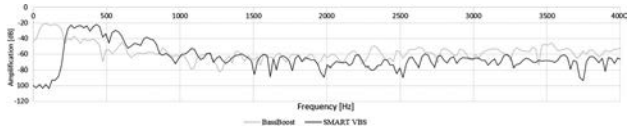

Fig. 5. Spectral characteristics of signal processed by bass boost algorithm and Smart Virtual Bass Synthesis.

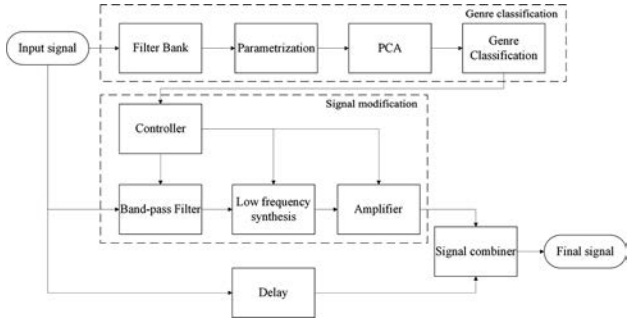

Fig. 6. Block diagram of proposed low frequency enhancement algorithm in a mobile device (Smart Virtual Bass Algorithm).

The signal processed by the VBS has additional harmonics in the band of 300–1000 Hz. In a typical low frequency gain algorithm, it can be seen that the modification of the signal was below the cut-off frequency of the speaker, indicating minor impact of the modification on the actual sound signal. In the next paragraph, an accurate description of the proposed solution together with the block diagram will be presented.

Fig. 6 is a block diagram of the proposed system that can automatically improve the quality of low-frequency sound. The algorithm is composed of three interrelated parts: classification of genres, modification of the signal, and the signal delay. At the input, the transmitted signal is divided into time frames of a predefined length in the range of 512–4096 samples.

The part associated with the classification of the signals consists of four blocks: a bank of filters, parametrization module, data reducer, and classifier. The main task of the block is to determine the type of reproduced track and send that information to the algorithm controller.

A filter bank processes the loaded signal according to parametrization module needs. On the output, the signal is split into bands corresponding to the specific parameters. Analysis frequency range is variable and can range from 31.5 Hz to 16 kHz. At this stage, the frequency range is limited by the parameters used for 63– 8000 Hz. The signal is transmitted after filtering the input signal to a parametrization module that calculates the parameters provided in the classification of musical genres.

At the entrance of the parametrization module, the input signal is analyzed at a predetermined sub-bands frequency and a set of timing, frequency, and time-frequency parameters are calculated that quantitatively describe the input signal. As mentioned earlier, the standard set of parameters used was from the MPEG 7 group with additional melcepstral parameters.

Table 2. NLD functions used in the Smart VBS.

Function name	Function equation
Exp1	$y = sign(x) \cdot \frac{(1-e^{-\lvert x\rvert})}{1-e^{-1}}$
Exp2	$y = sign(-x) \cdot \frac{(1-e^{\lvert x\rvert})}{e-1}$
Arctg	$y = 2,5 \cdot arctg(0,9x) + 2,5\sqrt{(1-0,9x)^2 - 2,5}$

Reduction of data redundancy is used to decrease the amount of information obtained during the parametrization operation. As mentioned before, for this purpose the Principal Component Analysis (PCA) method was used. The advantage of this method is a significant reduction in the number of parameters by limiting the transmitted information to parameters specific to particular music genre. This approach allowed for a significant reduction of the classification process [21], thus allowing the classification of music genres in real time based on the cached fragments of signal.

To perform musical genre classification, it is required to buffer at least 25 seconds of the song. This is needed to compute a feature vector to get reliable parameter values [25]. The predefined range of recognized musical genres is limited to six, but it can be expanded. Currently, the list of recognizable genres includes: classical, electronic, jazz, pop, R & B, and rock. To determine the type of music the nearest neighbor algorithm is used. It is determined by the set of parameters that best define a particular music genre, and it refers to the collection stored in the database. Information on the output is used in the modification process of sound and conditional on the further modification of the signal. The received information-based control system manages the process of sound enhancement. To determine the exact parameters of improving the sound quality, the mechanism uses fuzzy logic. The system incorporates the expert knowledge on the preferred system settings to improve sound quality for specific genres. The knowledge is retrieved from subjective listening tests carried out on a representative test group. The goal of the test was two-fold, i.e., to determine the level of gain and the NLD harmonic function. The test was conducted on a group of 31 people with no hearing problems reported. Respondents assigned the optimum synthesis parameters for the chosen song that represents the particular musical genre using a slider. Twenty-one audio excerpts from six genres were selected for the test. The obtained results confirmed that adding low frequency harmonics should depend on the music genre. Experts' answers have shown that the maximum magnitude of the virtual bass effect could be applied for rock genre and in the case of classical music should be kept to minimum. Other tested genres require intermediate values. The NLD functions chosen by the experts in subjective tests are presented in Table 2 [25].

The second part of the signal modification algorithm consists of three blocks: a band-pass filter, a low frequency synthesizer, and an amplifier. At the amplifier output, the signal is modified and adapted to sum with the delayed original signal derived from the third part of the algorithm.

The band-pass filter prepares the band of the input signal that will be subject to processing. Adding harmonics exclusively to the filtered band minimizes interference in the signal, making it possible to introduce only minor distortion. Signal filtration is performed in the frequency domain taking into account the Overlap and Add Method [3], thereby increasing the processing speed of the system. Further modification of the sound quality is limited to a filtered-out signal.

A low frequency synthesis block is the main element leading to enhancing low frequencies in the algorithm. It uses the previously described methods for the synthesis of additional harmonics in the signal. Low frequency synthesis method causes them to boost by adding additional harmonics to the signal. It is a well-known method widely described in the literature [2, 8–10, 12, 21]. The distinctive feature of the proposed solution is the intelligent control of existing synthesis parameters. The authors propose a modification of the parameters associated with the enhancement of the components and the non-linear function of the NLD. The optimal values of the parameters in the algorithm are selected on the basis of a specified genre. The various parameters on the basis of subjective tests are assigned to music genres and by using the controller transmitted to the synthesizer. At the output of the block signal, the band-pass is modified.

After passing through the synthesis block, the signal has the wrong level due to the added harmonics. It is therefore necessary to add a block responsible for adjusting the signal to the right, pre-modified level. For this purpose, a signal amplifier block is implemented. The block is designed to fit the signal level to the original. Enhancing takes into account the level of the reference signal. The level of the reference signal is sent from the controller. At the output, the signal is ready to be summed up with the delayed input signal.

The third part of the algorithm consists of two units: a delay circuit and summation. In the delay circuit, the raw signal is delayed for a defined number of samples. The delay is to synchronize the processed signal to the remaining part. The delayed signal is sent to the adder where the summation operation with the modified delayed signal is performed. At the output, the signal frame with enhanced lower frequency band is sent.

3 EXPERIMENTS

3.1 Test Preparation

The main aim of the study was to test the listeners' experience of the Smart VBS algorithm in relation to a traditional

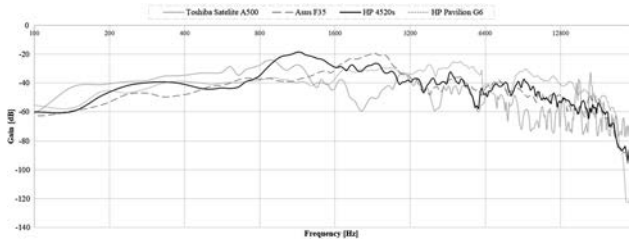

Fig. 7. Spectral characteristics of speakers of the laptops used.

low frequency enhancement method. To evaluate the quality of the proposed Smart VBS algorithm, listening tests were conducted following the principles of the MUSHRA methodology. Tests were conducted on a group of 31 people. Based on the listeners' answers, the statistical analysis was performed evaluating the usefulness of the algorithm as proposed for specific genres. The MUSHRA test utilized in this study works as JavaScript application with HTML5 elements [26]. The application runs in any browser and allows remote testing. Test application operates in accordance with the ITU-R BS.1534-2 recommendation [27].

Four laptops were utilized for the experiment: Toshiba Satellite A500 (device 1), Asus F35 (device 2), HP 4520s (device 3), HP Pavilion G6 (device 4). These are standard portable laptops with stereo speakers (the speakers' producers are unknown). The laptop drivers were set in the neutral mode and all sound effects were disabled in the tests. In Fig. 7 frequency characteristics of the equipment used, measured in the laboratory conditions, are presented. Large amplitude variations can be observed over frequency range and the low frequency response below 150 Hz is not adequate. The lowest cut-off frequency of speakers has the Toshiba Satellite A500 down to 150 Hz. The other speakers have the cut-off frequency above 250 Hz.

Twelve 10-second excerpts divided into six music genres—classical, jazz, pop, rap, rock, and electronic—were prepared. Excerpts were chosen to be a representative sample for a given music genre. According to the MUSHRA test procedure, a total of six 10-second excerpts were prepared, i.e., signal processed by the Bass Boost algorithm or SVBS-signal hidden references (original signal), and two anchor signals filtered with the low-pass filter of the frequency of 3.5 kHz and 7 kHz [27]. Listeners rated each music samples in the 0–100 scale. Tests were carried out in good listening conditions.

3.2 Results Analysis

A preliminary analysis of the results was performed to check the reliability of the listeners. The MUSHRA test provides the opportunity to check the reliability of listeners by using anchors placed in the test and hidden reference signals. Therefore, listeners' selection was based on their rates of hidden references and anchors. From a group of 31 listeners, 24 people were selected to ensure maximum reliability. The conducted analysis can be divided into three parts. The first stage was performed to analyze how the

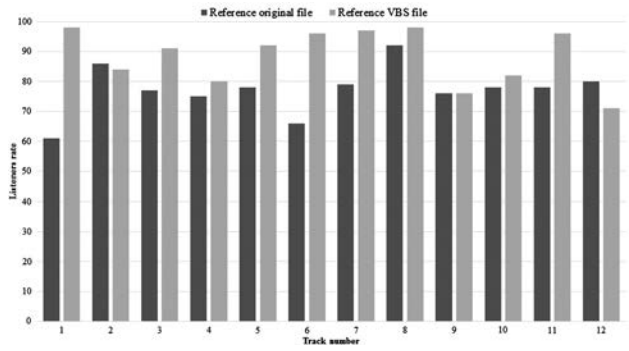

Fig. 8. The average listeners' ratings of the hidden reference signal.

listeners rated the reference signal. The reference signal was either the original signal or the signal with bass boost (SVBS) applied. In the second part, the listeners investigated quality of the Smart Virtual Bass Synthesis algorithm applied to the audio signal with regard to music genre. The last part of the analysis was related to the applicability of the proposed algorithm regardless of the device used. In the end, the results obtained in the subjective test was statistically checked to confirm their significance.

In the first stage, eight people participated in the MUSHRA test that was performed twice with two different reference signals. With regard to this stage, the average listener rates indicate an increase in grades for the SVBS enhanced signal served as a reference in comparison to unprocessed sample. The detailed results of this stage are shown in Fig. 8. This may be explained in terms of listening preference of the experts to choose the SVBS signal as a reference rather than the original signal. This indicated also that it might be advantageous to use the SVBS processed signal sample as a reference in a test in which the listeners determined quality of the SVBS processed signal in relation to the original signal. It occurred that an experienced listener rated the SVBS processed music excerpts in plus.

Further analysis of the results was carried out using grades assigned by a reliable 24-person group of subjects. The listeners rated quality of the original sample— unprocessed, processed using the SVBS algorithm, and the signal enhanced by the bass boost. The aim of the experiment was to test the SVBS algorithm on different devices in comparison to the traditional bass boost algorithm. SVBS algorithm was rated by listeners in relation to other signals.

On the basis of the results, the effectiveness measure was proposed to evaluate the tested algorithm in relation to other signals. Table 3 shows the effectiveness of the Smart VBS algorithm that was defined on the basis of subjective listeners' rates. Higher numerical value refers to higher efficiency of the algorithm performance with relation to the original signal or the one processed by the bass boost algorithm. The range of scale of the SVBS algorithm effectiveness is between –100 to 0 and 0 to 100. Getting –100 score means that SVBS is one time worse than the bass boost or the original signal. 0 means that listeners didn't hear any difference between the samples tested, and 100 score says that

Table 3. Effectiveness of using the SVBS algorithm comparing to the standard low-frequency enhancement and the original file.

Music genre	classical	jazz	pop	rap	rock	electronic
Smart VBS compared with the original signal	2.55	17.35	22.4	24.65	30.35	26.1
Smart VBS compared with the bass boosted signal	12.5	31.3	22.85	28.5	30.55	23.85

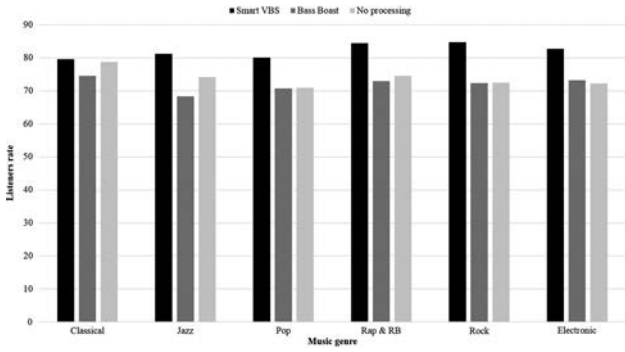

Fig. 9. The average scores with regard to the Smart VBS algorithm and music genres.

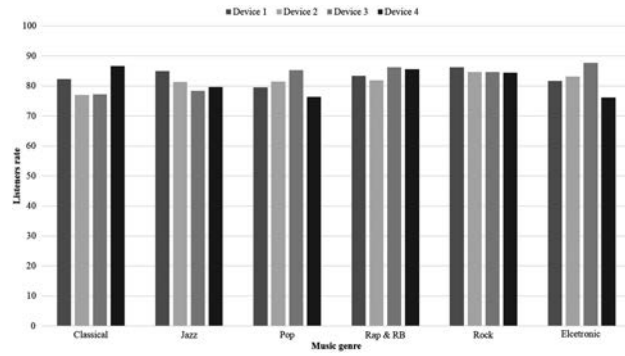

Fig. 10. The average rates obtained for the Smart VBS algorithm for four tested devices.

SVBS is one time better than the bass boost or the original file sample. The analysis of the results was carried out with regard to music genres. Listeners confirmed that the audio file processed using the SVBS algorithm returned better results than the bass boost algorithm. The smallest profit from using the SVBS algorithm was noted in classical music. In that case, listeners indicated a very small difference between the original signal and the file processed by Smart VBS. It should, however, be noted that the bass boost algorithm applied to classical music was less well-scored than SVBS.

For other genres, the SVBS algorithm processed samples were in most cases rated higher than the bass boosted or unprocessed samples. Evaluation results of excerpts divided into genres are presented in Fig. 9. An average score returned for the SVBS processed music genres, such as jazz, pop, rap, and electronic, assessed by the listeners was 82. Contrarily, there are only small differences in plus for the bass boost algorithm in comparison to the original sample sound.

The last stage of the analysis focused on the effectiveness of the algorithm SVBS tested using different portable devices. Four types of laptops were used to confirm the effectiveness of the SVBS algorithm independently of the device used. The results indicated that respondents noticed only a slight difference between the devices. The difference between devices is within the range of 10% in context of a given music genre. The exact users' rates for this case can be seen in Fig. 10.

To confirm the the statistical significance of the results, one-sample T-Student test was carried out. The analysis was based on an examination of certainty results with respect to the mean value. The basis for the test T-Student is the development of the null hypothesis to be accepted or rejected statistically. In the experiment, the tested hypothesis related to the lack of statistical significance obtained in tests of subjective results. The value of the T-student parameter above the 2.492 value will indicate that the null hypothesis can be rejected, thus the results are statistically significant. Statistical significance threshold was set at a typical value of 0.05.

In Table 4 the T-Student parameter values for all the experiments are presented. These are averaged values of the T-Student parameter for one-sample taken among all subjects.

All values obtained by the statistical significance test T-Student are above a particular threshold, moreover assumptions are met for even the strictest level of statistical significance of the parameter. Statistically, the most reliable results were obtained in experiments using the SVBS algorithm. This means that the spread between the individual subjective answers was statistically smallest.

4 SUMMARY

A Smart VBS algorithm with automatic adjustment of the main parameters of low frequency synthesis based on the music content of the reproduced audio material was

Table 4. Summary of the results of T-Student Test.

Music genre	1	2	3	4	5	6	7	8	9	10	11	12
Not processed file	3.89	3.49	3.57	3.86	7.70	3.90	3.98	3.91	3.82	4.36	3.80	3.78
SVBS	3.74	4.17	4.20	3.71	4.01	4.22	3.76	3.91	3.88	4.16	4.31	4.11
Bass Boost	4.13	3.98	3.87	4.10	3.81	3.83	4.06	3.69	3.73	3.93	3.75	3.71

presented. The effectiveness of the algorithm was proven in carried out subjective listening tests. The listeners confirmed the effectiveness of the algorithm compared to the conventional low frequency boost algorithm. By assigning similar rates regardless of the device used, listeners proved that the solution proposed may be adapted to different laptops.

5 ACKNOWLEDGMENTS

This research was partially founded by the grant no. PBS1/B3/16/2012 entitled "Multimodal System Supporting Acoustic Communication with Computers" financed by the Polish National Centre for R&D.

6 REFERENCES

[1] H. Mu and W.-S. Gan, "A Psychoacoustic Bass Enhancement System with Improved Transient and Steady-State Performance," *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing, ICASSP 2012*, pp. 141–144, Kyoto (2012 Mar.), http://dx.doi.org/10.1109/ICASSP.2012.6287837.

[2] A. J. Hill and M. O. J. Hawksford, "A Hybrid Virtual Bass System for Optimized Steady State and Transient Performance," *Computer Science and Electronic Engineering (CEEC) Conf.* (2010 Sept. 8–9). http://dx.doi.org/10.1109/CEEC.2010.5606489.

[3] P. Hoffmann and B. Kostek, "Smart Virtual Bass Synthesis Algorithm Based on Music Genre Classification," *18th IEEE SPA 2014, Signal Processing Conference*, Poznań, Poland (2014 Sept. 22–24).

[4] G. Bekesy, "The Missing Fundamental and Periodicity Detection in Hearing," *J. Acoust. Soc. Am.*, vol. 51, no. 2B, pp. 631–637 (1972).

[5] H. Mu, W.-S. Gan, and E.-L. Tan, "A Psychoacoustic Bass Enhancement System with Improved Transient and Steady-State Performance," *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing, ICASSP 2012* (2012 Mar.), http://dx.doi.org/10.1109/ICASSP.2012.6287837.

[6] J. F. Schouten, R.J. Ritsma, and B.L. Cardozo, "Pitch of the Residue," J. Acoust. Soc. Am., pp. 825–834 (1962).

[7] E. Terhardt, "Zur Tonhöhenwahrnehmung von Klängen. I. Psychoakustische Grundlagen," *Acustica*, vol. 26, pp. 173–186 (1972).

[8] N. Oo, W.-S. Gan, and W.-T. Lim, "Generalized Harmonic Analysis of Arc-Tangent Square Root (ATSR) Nonlinear Device for Virtual Bass System," *IEEE International Conference on Acoustics Speech and Signal Processing (ICASSP)* (2010 Mar.14–19), http://dx.doi.org/10.1109/ICASSP.2010.5495913.

[9] N. Oo and W.-S. Gan, "Harmonic Analysis of Nonlinear Devices for Virtual Bass System," *IEEE International Conference on Audio, Language and Image Processing (ICALIP)* (2008 July 7–9), http://dx.doi.org/10.1109/ICALIP.2008.4590264.

[10] W.-T. Lim, N. Oo, and W.-S. Gan, "Synthesis of Polynomial-Based Nonlinear Device and Harmonic Shifting Technique for Virtual Bass System," *ISCAS IEEE* (2009 May 24–27), http://dx.doi.org/10.1109/ISCAS.2009.5118144.

[11] E. Larsen and R. M. Aarts, "Perceiving Low Pitch Through Small Loudspeakers," presented at the *108th Convention of the Audio Engineering Society* (2000 Feb.), convention paper 5151.

[12] R. M. Aarts, E. Larsen, and D. Schobben, "Improving Perceived Bass and Reconstruction of High Frequencies for Band Limited Signals," *1st IEEE Beneloux Workshop on Model Based Processing and Coding of Audio (MPC 2002)*, Leuven, Belgium (2002 Nov. 15).

[13] B. Kostek, "Music Information Retrieval in Music Repositories," Ch. 17, A. Skowron, Z. Suraj (eds.), *Rough Sets and Intelligent Systems, ISRL 42*, pp. 463–489 (Springer Verlag, Berlin, Heidelberg, 2013).

[14] B. Kostek, "Music Information Retrieval—Soft Computing versus Statistics," *Computer Information Systems and Industrial Management, IFIP Internat'l Federation for Information Processing 2015, CISIM 2015*, K. Saeed, W. Homenda (eds.), LNCS No. 9339, pp. 36–47 (2015), http://dx.doi.org/10.1007/978-3-319-24369-6_3.

[15] B. Kostek, P. Hoffmann, A. Kaczmarek, and P. Spaleniak, "Creating a Reliable Music Discovery and Recommendation System," R. Bembenik et al. (eds), *Intelligent Tools for Building a Scientific Information Platform: From Research to Implementation, Studies in Computational Intelligence 541*, pp. 107–130 (Springer Intern. Publishing Switzerland, 2014), http://dx.doi.org/10.1007/978-3-319-04714-0_7.

[16] P. Hoffmann and B. Kostek, "Music Data Processing and Mining in Large Databases for Active Media," *The 2014 International Conference on Active Media Technology*, D. Slezak et al. (eds.), LNCS 8610, pp. 85–95 (Springer Intern. Publishing Switzerland, 2014).

[17] B. Kostek, A. Kupryjanow, P. Zwan, W. Jiang, Z. Ras, M. Wojnarski, and J. Swietlicka, "Report of the ISMIS 2011 Contest: Music Information Retrieval, Foundations of Intelligent Systems," *ISMIS 2011, LNAI 6804*, pp. 715–724, M. Kryszkiewicz et al. (eds.) (Springer Verlag, Berlin, Heidelberg, 2011).

[18] A. Rosner, B. Schuller, and B. Kostek, "Classification of Music Genres Based on Music Separation into Harmonic and Drum Components," *Archives of Acoustics*, vol. 39, no. 4, pp. 629–638 (2014), http://dx.doi.org/10.2478/aoa-2014-0068.

[19] B. Kostek and A. Kaczmarek, "Music Recommendation Based on Multidimensional Description and Similarity Measures," *Fundamenta Informaticae*, vol. 127, no. 1–4, pp. 325–340 (2013), http://dx.doi.org/10.3233/FI-2013-912.

[20] MPEG 7 standard, http://mpeg.chiariglione.org/standards/mpeg-7

[21] P. Hoffmann, B. Kostek, A. Kaczmarek, and P. Spaleniak, "Music Recommendation System," *J. Telecommunications and Information Tech.*, no. 2, pp. 59–69 (2014).

[22] L. J. Williams and H. Abdi, "Principal Component Analysis," Wiley Interdisciplinary Reviews: Computational Statistics, 2 (2010).

[23] P. Hall, B. U. Park, and R. J. Samworth, "Choice of Neighbor Order in Nearest-Neighbor Classification," *Annals of Statistics*, vol. 36, no. 5, pp. 2135–2152 (2008), http://dx.doi.org/10.1214/07-AOS537

[24] S. Arlot and A. Celisse, "A Survey of Cross-Validation Procedures for Model Selection," *Statistics Surveys*, vol. 4 pp. 40–79 (2010), http://dx.doi.org/10.1214/09-SS054

[25] P. Hoffmann, B. Kostek, and T. Sanner, "Intelligent Low Frequency Synthesis Applied to Mobile Devices," Telecommunication Rewiew + Telecommunication News (*in Polish*), 8–9, pp. 905–913 (2014).

[26] Mushra Test Software, https://code.soundsoftware.ac.uk/projects/sodamat/wiki/MUSHRA (March 2015).

[27] Recommendation ITU-R BS.1534-2, "Method for the Subjective Assessment of Intermediate Quality Level of Audio Systems," International Telecommunication Union (June 2014).

THE AUTHORS

Piotr Hoffmann

Bozena Kostek

Piotr Hoffmann was born in Gdansk in 1998. In 2012 he graduated from the Faculty of Electronics, Telecommunications and Informatics at Gdansk University of Technology. The subject of his B.Sc. engineer's thesis was a comparative analysis of decision algorithms applied in the context of the effectiveness of speaker recognition. The subject of his M.Sc. thesis was a sound reinforcement system designed for Gdansk Shakespeare Theater. Hoffmann received the award in the IEEE Gdansk Computer Science Chapter contest for the realization of his M.Sc. thesis. Since 2013 he is a Ph.D. student and a computer scientist in Multimedia System Department at Gdansk University of Technology. He is interested in intelligent signal processing including application to music information retrieval, sound perception, and sound reinforcements systems.

•

Bozena Kostek holds a 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 has published over 500 scientific papers in journals and at international conferences. Under her guidance, 13 Ph.D. students supported their doctoral theses and she supervised over 180 M.Sc. theses. She serves as the Editor-in-Chief of the *Journal of the Audio Eng. Soc.* since 2011. She was also the Editor-in-Chief of *Archives of Acoustics* (2007–2012). She was the recipient of many prestigious awards for research, including those of the Prime Minister of Poland (2000, 2014) for outstanding research achievements, prizes of the Polish Academy of Sciences and Ministry of Science, and the Bachelor Cross of the Polonia Restituta Order (2011). She also received the Audio Eng. Soc. Fellowship Award in 2010. In 2013 Prof. Kostek was elected as a member of the Polish Academy of Sciences. Her research activities are interdisciplinary, however the main research interests focus on musical informatics, audio signal processing, human-computer interaction, cognitive bases of sound and vision processing, as well as psychophysiology of hearing and vision.