Audio Signal Dynamics Expansion

Martin Pospisilik, Milan Adamek

***Abstract*—**Currently, audio recordings are usually edited in order their dynamic range was decreased. Such signal processing brings several advantages. When applied, the recordings sound more loudly even on cheap and weak reproducing devices. The compressed recordings act compactly and can also be listened in noisy environment. On the other hand, too much compression makes the recording contrived. Therefore the authors of this paper decided to create a software-based audio signal dynamics expander and, after gaining the experience, to discuss further approach to the audio signal dynamics expansion. Both issues are described in this paper.

***Keywords*—**Audio dynamics, Expansion, Digital signal processing, Software, Expander.

I. Introduction

The audio signal compression processed in the recording studios is well justified. The human perception of an audio recording loudness is determined not by its maximum power but by its mean power. According to [6] the mean power of the signal can be defined according to its mean amplitude that is defined as follows:

$$\widetilde{s(t)} = \frac{1}{t_b - t_a}\int_{t_a}^{t_b} s(t)dt \tag{1}$$

The replacing the signal s(t) by its mean amplitude between the time intervals t_a and t_b is depicted in Fig. 1.

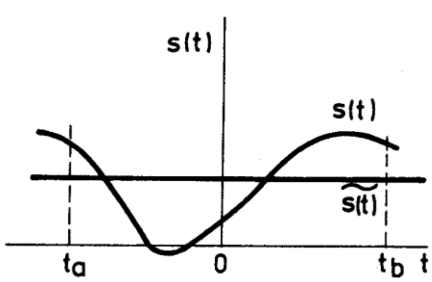


Fig. 1 – General signal s(t) and its replacing by its mean amplitude $\widetilde{s(t)}$.

Martin Pospisilik is with Faculty of Applied Informatics, Tomas Bata University in Zlin, Namesti Tomase Garrigua Masaryka, 76001 Zlin, Czech Republic. (phone: +420 57-603-5228; e-mail: pospisilik@fai.utb.cz).

Milan Adamek is with Faculty of Applied Informatics, Tomas Bata University in Zlin, Namesti Tomase Garrigua Masaryka, 76001 Zlin, Czech Republic. (e-mail: adamek@fai.utb.cz).

This paper is supported by the Internal Grant Agency at TBU in Zlin, project No. IGA/FAI/2012/056 and by the European Regional Development Fund under the project CEBIA-Tech No. CZ.1.05/2.1.00/03.0089.

The instantaneous power of the signal is defined as the squared function describing the signal:

$$p(t) = |s(t)|^2 \tag{2}$$

The mean power of the signal can then be expressed according to the following equation:

$$P = \widetilde{p(t)} = |\widetilde{s(t)}|^2 = \frac{1}{t_b - t_a}\int_{t_a}^{t_b} p(t)dt = \frac{1}{t_b - t_a}\int_{t_a}^{t_b} |s(t)|^2 dt \tag{3}$$

The problem is that in audio recordings the mean power of the signal is usually much lower than the power spikes p_{max}:

$$P \ll p_{max} \tag{4}$$

In practice there can only one or several high power spikes occur in the recording, due to which the total signal loudness must be decreased in order the peak-to-peak amplitude range of the processing devices and recording mediums was not exceeded.

In Fig. 2 this situation is well depicted. For rock music recordings this can be considered as a typical s(t). The x axis of the graph represents time while the y axis represents the amplitude of the signal. Isolated power spikes can be easily distinguished here.

Fig. 2 – Example of audio recording with a minor compression [7]

As obvious from Fig. 2, most of the amplitude reaches 30 % of maximum but several spikes reach up to 85 %. Therefore it can be said that the highest power spike p_{max} is 8 times higher than the mean power of the signal. The practical consequence is that even if the listener listens at the loudness level corresponding to the mean power of 10 W, the spikes can reach up to 80 W of power, requiring at least 100 W power amplifier.

It is obvious that in order the dynamics of the real audio signal could be obtained, the amplitude range of the audio signal processing devices as well as of the recording media would have to be great. Therefore before halve of the last century, methods of the dynamic compression became to be a subject of research. The recording media used in those times as well as the processing devices did not allow achieving dynamic range higher than approximately 40 dB.

Currently, the characteristics of the media and the processing devices are more promising, but still the smooth compression of the signal is justified, because the circumstances under which the recordings are listened to are usually different from the circumstances under which the music is really played. At least in block of flats the heavy power spikes usually cannot be reproduced so that the neighbors were not disturbed etc.

Unfortunately, current studio software equipment allows the sound engineers to process the compression too heavily and ultimately.

The effort to signal dynamics processing degenerated into the “loudness war” of the sound engineers. In Fig. 3 the same recording as depicted in Fig. 2 can be found, after its “advanced” dynamics range processing.

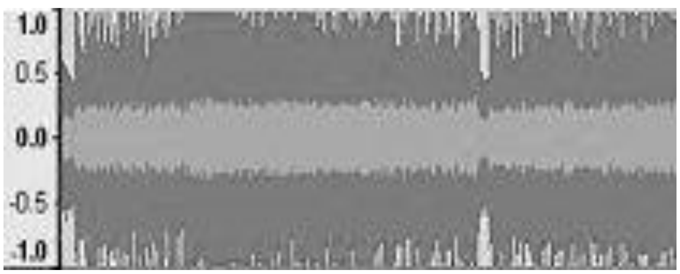


Fig. 3 – Example of audio recording with a heavy compression

As obvious from Fig. 3, after “remastering” the power spikes disappeared at all and the amplitude could have been increased up to the maximum. The result is, that with the 10 W amplifier this recording appears to be as loud as the recording depicted in Fig. 2 with the 80 W amplifier. Or, with the amplifier of same power it appears much louder. Anyway, for this illusion a considerable tax must have been paid in the form of losing the powers spikes and therefore the sound of percussions. The sound is heavily distorted at all. In order the mean power of the signal could be increased even more, clipping of the power spikes has been introduced. Briefly expressed, if the spike cannot be compressed enough, it is replaced by a more flat curve. The example of the clipping is depicted in Fig. 4. When clipped, the recording is irreversibly distorted and cannot be repaired by any dynamics expander. The information on dynamics disappeared forever. On the other hand, when the dynamics compression is not too heavy and a compressor with reasonable time constants was employed at the recording processing, the rest of the dynamics information can be identified and according to it the backward expansion of the recording’s dynamics can be processed.

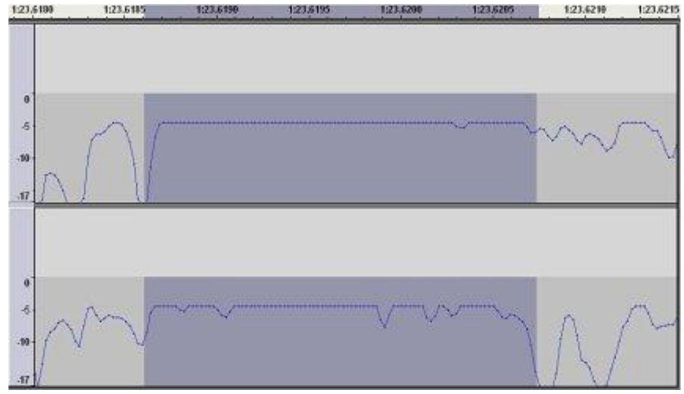


Fig. 4 – Example of the audio recording clipping

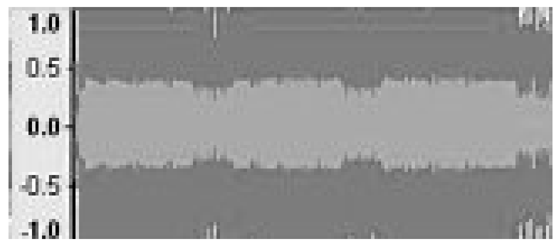


Fig. 5 – Example of the heavily clipped audio recording

II. Motivation

According to the facts mentioned above, it can be stated that the dynamics compression is useful when processed moderately. Unfortunately, the effort of the recording studios to produce as loud recording as possible undoubtedly damages the quality of the recordings. The dynamics range of the recording media is not utilized at all. While the ordinary CD can offer the dynamics range up to 96 dB, the recording studios produce audio recordings with the dynamics range of only 3 dB or less. Such recordings are well optimized for portable players as I-phones, mp3 players in cars etc., but when listened on a standard reproducing device they sound hollow. Therefore several attempts to increase the signal dynamics were made, but this increase was always partial and inaccurate because in heavily compressed recordings the information of dynamics is irreversibly lost.

In the framework of one of such attempts a software dynamics expander has been created and on the basis of the experience with its performance further discussion on hardware device that would be capable of the same operation was settled.

III. Software-based audio dynamics expander

The Dynamic Expander application serves for audio signal dynamics increasing. Under several circumstances the recordings created with too high compression ratio can be processed in order the better performance was gained. This issue is described in [3] together with the proposals on designing the voltage controlled amplifier to process the dynamics increasing. The software application described in this paper has been created in order to prove not only the influence of changing the amplification factor on the subjective perception of the dynamics by the listener, but also in order to debug and tune the algorithm that controls the dynamics processing.

The application can process 16 bit two channel PCM encoded wav files with the bitrate of 44.1 kHz. In order the intermodulation effect was supressed the processed frequency range is divided into 4 sub ranges with the boundaries predefined subjectively on the basis of perception of the basses, vocal band, upper midrange and high end. At the frequency range sub-sectioning, software IIR filters are implemented. Because the separated processing can result in changing of the sound colour, the bands are backward mixed with adjustable mixing ratio. This can be also done automatically by evaluating the average energy of the signal in each bands and comparing it to the energy of the processed signal separately for each of the bands.

A. Application description

The main window screenshot is depicted in Fig. 1. The frequency sub ranges are denoted as Band 1, Band 2, Band 3 and Band 4. For each of the bands the attack and decay time in milliseconds can be adjusted separately as well as the mixing ratio. The boundary frequencies between the bands are 150 Hz, 1 kHz and 8.5 kHz. Parameters Slope, Threshold and Level are adjusted for all of the bands at all. The description of the controls is as follows:

Attack times can be adjusted within the range from 1 to 100 ms. The attack time is a time period between the signal peak occurrence and the point at which the amplification factor has been increased to the appropriate level. High attack times result in short peak damping. Short attack times may result in signal distortion.

Decay times can be adjusted within the range from 5 to 500 ms. The decay time is a period between the signal peak disappearance and the point at which the amplification factor has been decreased to the initial level. High decay times result in short peak damping and may result in perceptible volume intensity wobbling. Short decay times may result in signal distortion.

The Final Band Mixing sliders are used to adjust the levels of the signal in each of the bands. The adjustment range is ± 10 dB. Usually there is a need to increase the level of that

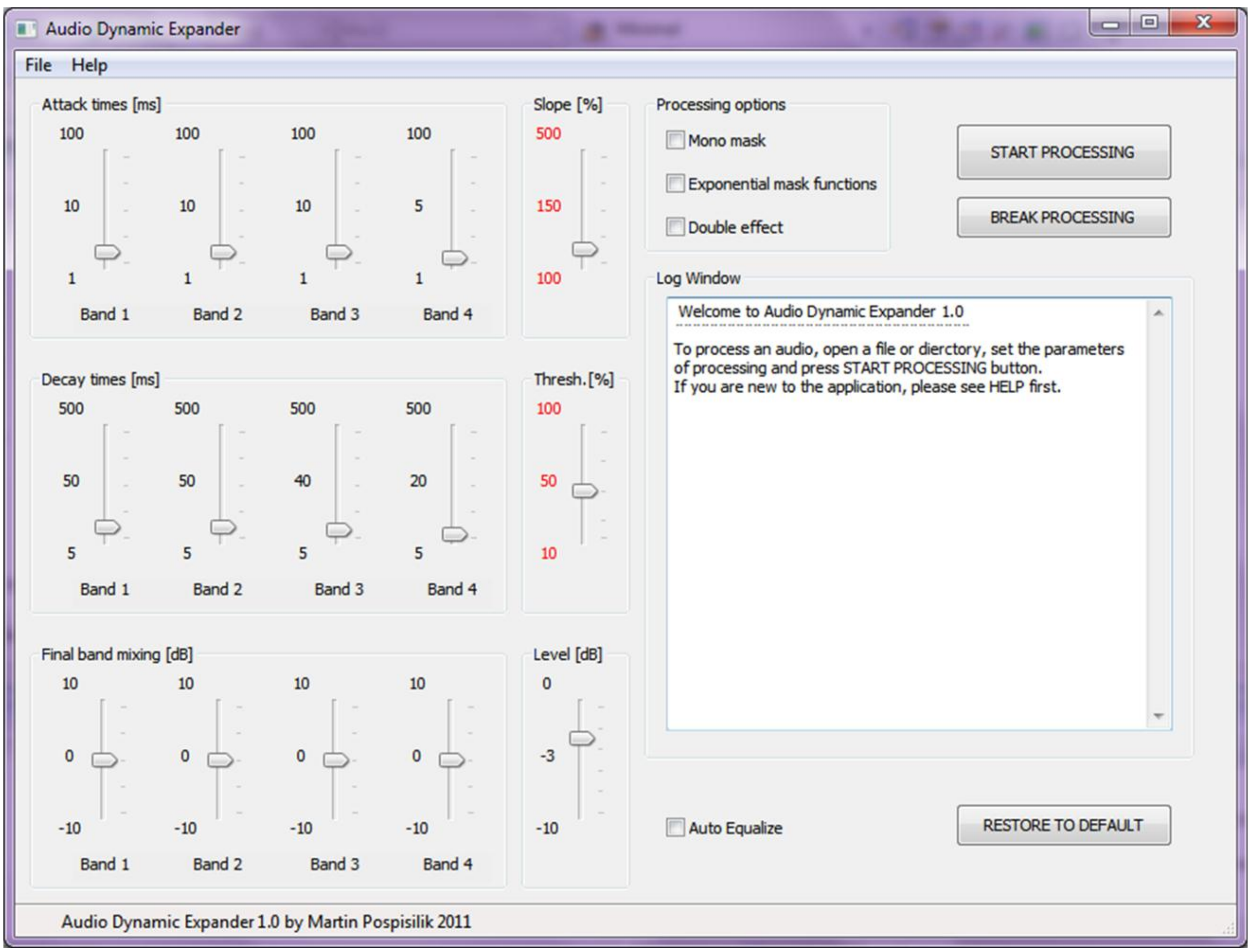


Fig. 6 – Main window of the application

band that includes short and steep peaks because these are emphasized by decreasing the level of the low-level signal parts which subjectively results in the volume decrease in the appropriate band.

The Slope slider serves for defining the dynamics expansion slope. The adjustment range is from 100 % (1:1) to 500 % (5:1). This means that if the signal level has exceeded the threshold, the amplification factor is adjusted prospectively. For example, when 250 % slope is set and the signal level is 30 % above the threshold level, by the time set by the Attack Time slider it will have been amplified in order it would reach the level 95 % above the threshold.

The Threshold slider serves for defining the threshold above which the dynamics expansion is processed. The adjustment range is from 10 % (- 20 dB) to 100 % (0 dB). When set to 10%, all the signals with the level above -20 dB are processed. On the other hand when set to 100% no expansion is performed as none of the peaks can exceed 100 % level.

The Level slider adjusts the level of the highest peak in the processed file. There are several players at which problems occur when the level is 0 dB. For this reason the user can decrease the level to leave sufficient headroom.

The Mono Mask function creates an average from the left and right expansion mask and applies the resultant mask on both channels simultaneously. This option can be used if there are significant differences between the stereo signal channels resulting in subjective stereo balance wobbling.

If the Exponential Mask Functions option is active, the expansion masks are processed as exponential functions. This should better fit the human perception and emulate the charging and discharging of the RC integrator that is usually used to delay the hardware amplifier response.

If the Double Effect function is active, the expansion masks are powered by two in order stronger performance was achieved.

The Auto Equalization function works with the average signal levels in each band. When turned on, the average signal level in each band is processed for the signal before and after the expansion and then the mixing levels of the bands are adjusted in order their average levels before and after the processing was achieved. Furthermore, the user can still make corrections by the Final Band Mixing sliders to achieve the best performance.

In the File menu, the user can choose a single file to be processed or more files in order to process batch processing. The processing is then started by clicking on the Start Processing button and can be any time suspended by clicking on the Break Processing button. By clicking on the button Restore to Default all sliders are set to the default value (predefined by the programmer) and all options are unchecked.

During the processing, in the Log Window a log information is displayed. The log example is depicted in Fig. 7.

```
** Multiple files opened [number of files]: 1
C:\Users\Uzivatel\Documents\Any Audio Converter\01
audio_file.wav


  *** Processing started!! ***

 > File length [kB]: 42336

 > Band splitting [%]:  10  20  30  40  50  60  70
80  90

 > Peak values (Band 1 Left, Band 2 Left, ... , Band
4 Right:
0.642517, 0.931869, 1.148988, 1.067740, 1.227088,
1.177637, 0.665140, 0.621575,

 > Average band levels:
 Total [dB]:  -17.843761
 Band 1 [dB]:  -22.632248
 Band 2 [dB]:  -28.120147
 Band 3 [dB]:  -27.258794
 Band 4 [dB]:  -25.785936

 > Determining adequate threshold values:
 Band 1 threshold:  0.393596
 Band 2 threshold:  0.554182
 Band 3 threshold:  0.601181
 Band 4 threshold:  0.321679

 > Generating mask file [%]:  10  20  30  40  50  60
70  80  90

 > Getting equalisation values [%]:  10  20  30  40
50  60  70  80  90

 > Average band levels before equalisation:
 Total [dB]:  -23.484207
 Band 1 [dB]:  -32.886144
 Band 2 [dB]:  -28.869457
 Band 3 [dB]:  -28.551068
 Band 4 [dB]:  -38.437798

 > Applying mask [%]:  10  20  30  40  50  60  70
80  90

 > Peak values before normalization:
 Left channel:  0.953737
 Right channel:  1.196172

 > Average band levels:
 Total [dB]:  -20.955638
 Band 1 [dB]:  -35.450243
 Band 2 [dB]:  -31.433555
 Band 3 [dB]:  -31.115167
 Band 4 [dB]:  -41.001897

 > Equalisation & Saving [%]:  10  20  30  40  50
60  70  80  90
****** PROCESSING FINISHED ******
```

Fig. 7 – Log example

B. Code description

The core of the application has been debugged in Maple mathematics software and converted into C language consequently. The basic idea of dynamics expansion is based on the algorithm that is reciprocal to the compression used at the final records mixing. How the algorithm works in general is obvious from Fig. 8.

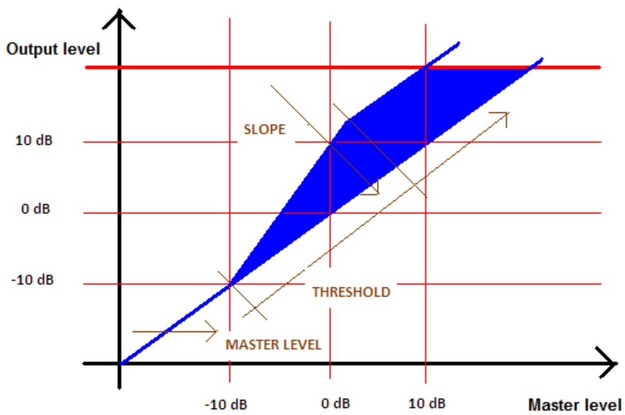


Fig. 8 – Dynamics expansion principle

The colored area in Fig. 8 symbolizes the difference between the master level and the processed level. As obvious, above the threshold level the signal is amplified according to the setting of the Slope parameter. Because in digital audio files the highest level cannot exceed 0 dB, the levels must be normalized with respect to the highest peak prospectively. Moreover, if the output level followed the master level immediately, significant distortion would occur. Therefore, delays generally known as Attack time and Decay time must be employed. To ensure the length of the processed file was unlimited (of course, with respect to the wav files specifications), several temporary files are created on the hard drive and all the processes are splatted into blocks by 80 kB with storing the intermediate results on the hard drive. The processes applied to the processed files are described below.

C. Algorithms

When the processing starts, a temporary file “bands.tmp” is created. The PCM data stream from the opened wav file is encoded into floating point numbers and 3rd order IIR filters are applied on it in order to split the processed signal into four frequency bands. The discrete IIR bandpass filters are designed according to [4] and [5]. In the temporary file for each time sample 8 values are stored (4 bands, 2 channels).

Based on the data in the temporary file, peak values are found for each of the 8 streams. As can be seen in Fig. 2,

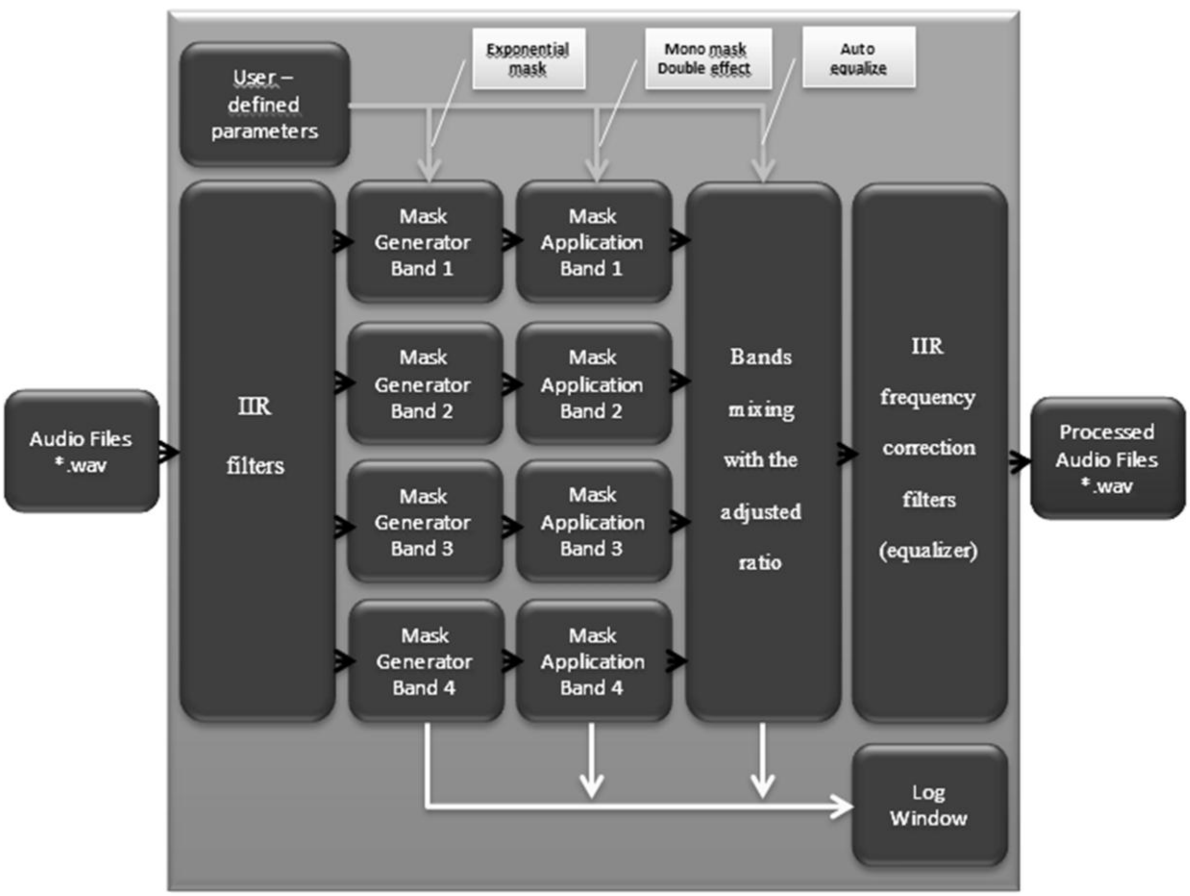
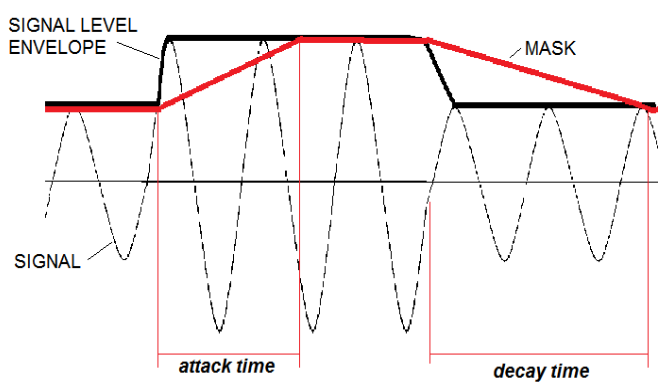


Fig. 9 – General block diagram of the dynamics expansion processing

despite the fact that the binary coded PCM values cannot exceed the ± 1 value, some of the peak values are greater. This is caused by the non-flat response of the IIR filters applied to the data stream. According to the data stored in the temporary file, average levels for each band are calculated as well as the threshold levels above which the expansion will be processed. Consequently, another temporary file called “masks.tmp” is created in order to store the masks which are then applied to the appropriate bands. These masks are generated according to the parameters defined by the user. As default, the masks are generated for each channel separately, but when Mono Mask option is checked, these masks are averaged and the resulting average is applied on both channels simultaneously. A visualization of how the masks are created can be found in Fig. 10.



.

Fig. 10 – Generating a (linear) mask

The mask copies the signal level envelope, that can be in the easiest approximation expressed as abs(signal) with a predefined time delay. In Fig. 4 the mask transitions are linear but as state above, they can be exponential as well. In “mask.tmp” another signal is stored, describing the time sequence of the envelope. This signal consists solely of positive values. The upper limit is 1, the lower refers to the threshold. If the level drops below the threshold, the mask levels are equal to the threshold.

In the following steps the samples stored in “bands.tmp” are multiplied by the appropriate samples stored in “masks.tmp”. Because the mask values are from the range <threshold, 1>, the highest peaks remain unchanged while the parts of the signal that are of lower levels are appropriately attenuated. The results of the multiplication are stored in the third temporary file called “merged.tmp”. At this point the levels of each band are adjusted by the appropriate sliders. Moreover if the Auto Equalization is turned on, additional steps are processed. The statistics is processed and the equalization coefficients are calculated. Then the final merging is processed again with employing of the equalization coefficients.

Finally second equalization must be applied. Because of the IIR filters phase response, the splatted bands cannot be merged together without the negative effect on the total frequency response. For this reason a 2^{nd} order equalizer with the fixed coefficients was implemented into the software in order to compensate the frequency response deviations from the ideal condition caused by adding signals with different phase responses. The coefficients of this equalizer have been determined according to the total frequency response measurements processed with the aid of Maple mathematical software that has not been published yet. Simultaneously with the second equalization the floating point values are converted into the fixed point PCM data stream and the resulting file is stored on the hard drive with the name of the master file appended with “_processed”. When finished, the temporary files are deleted.

The block diagram of the processing algorithm in general can be found in Fig. 9.

D. Results

The application has been proven by subjective listening tests. The recordings of several music genres were processed at different adjustment settings and the time consummation was checked as well. Only 2 channel and 16 bit recordings with the sample rate of 44.1 kHz can be processed. Generally, the older recordings that tend to sound flatly were improved when Double Effect option was checked together with low adjustment of the slope and Auto Equalization. However, in some recordings which were not mixed perfectly there occurred volume wobbling of some of the instruments’ volume as a result of bad mixing of the master recording. The improvement of contemporary recordings has been quite negligible, probably because the compression of these recordings is usually too heavy and the compressors are too efficient leaving no information about the previous recording dynamics. However, even the contemporary recordings can benefit from increased voice intelligibility when the expansion parameters are adjusted properly.

1) Expansion of moderately compressed recording

Firstly, a moderately compressed recording was processed. A short time period of the result in comparison with the source signal is displayed in Fig. 11. The x-axis symbolizes time and the y-axis amplitude of the signal in percents of the maximum. Moreover, indicative thin lines are also drawn in the figure in order to set an approximate border between the mean power value and the power spikes. As can be seen, the approximate mean power level of the source signal is 58 % and the spikes reach up to 100 %. The difference between the highest spike power and the mean power is approximately 2.4 dB. After processing, the mean level of the signal was decreased to the level of only 20 % while the spikes reach up to 62 %. This refers to the indicative difference between the highest spike power and the mean power to be approximately 4.9 dB. Moreover, sufficient headroom up to 7 dB is obtained for more powerful spikes. However, when explored carefully, it can also be seen in Fig. 11 that not all of the spikes are emphasized properly. This is probably caused by the fact that it is impossible to set the attack and decay times accurately because there is no information on how the compressor was set when processing the recording.

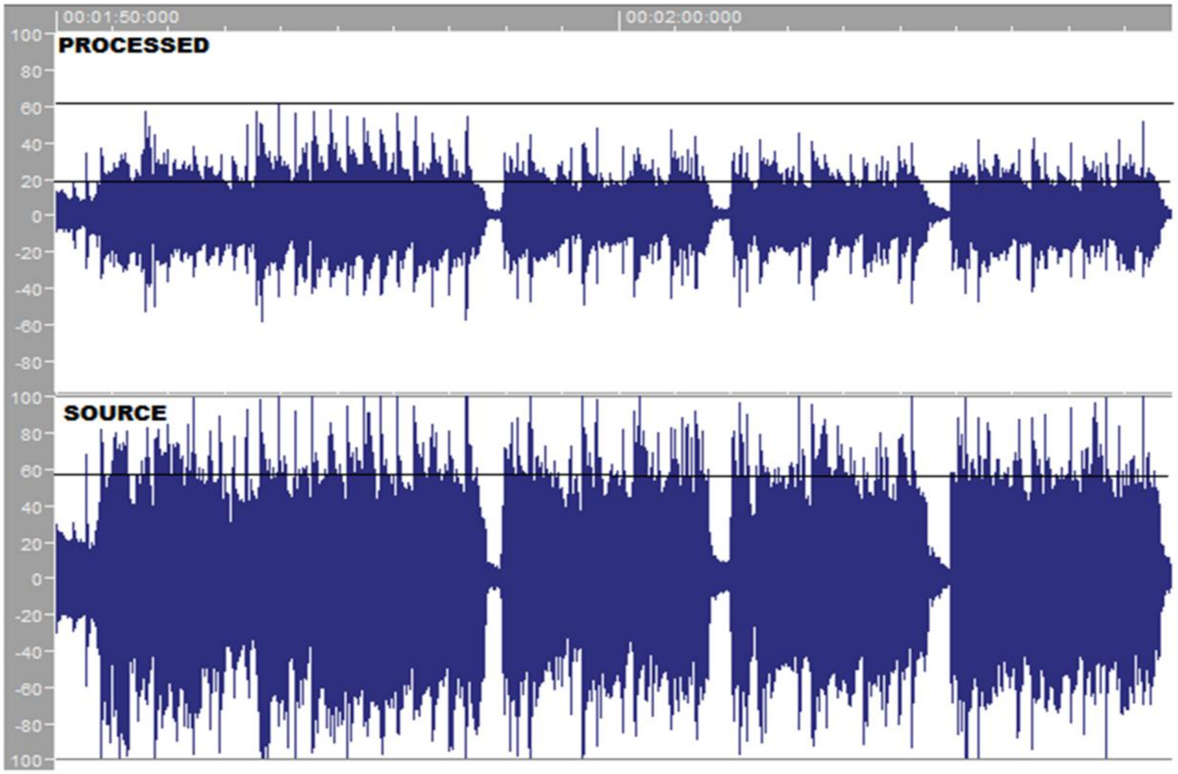


Fig. 11 – Comparison of the source signal and the output of the dynamics expander (linear scale of the y-axis), when the moderately compressed recording was processed

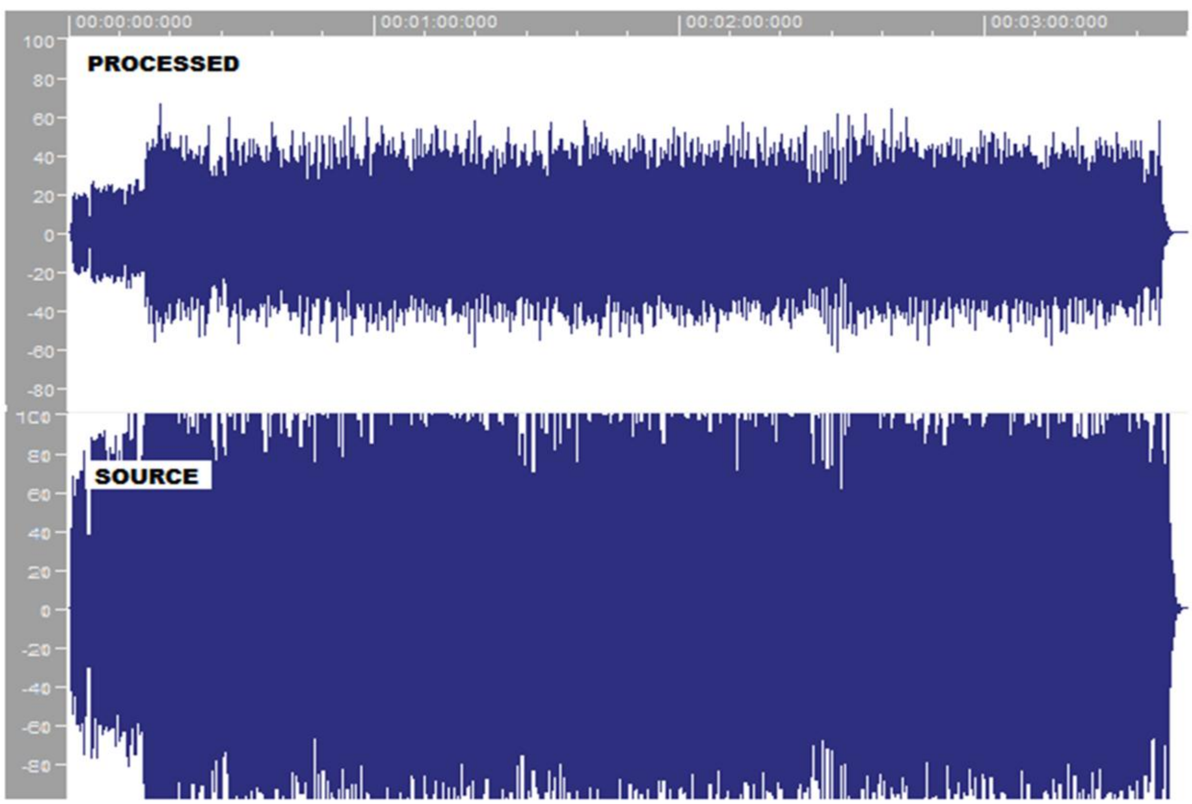


Fig. 12 - Comparison of the source signal and the output of the dynamics expander (linear scale of the y-axis), when the heavily compressed recording was processed

2) Expansion of heavily compressed and clipped recording

Secondly, the function of the dynamics expander was proven on the heavily compressed and clipped recording. The result is depicted in Fig. 12. At first glance it may seem that a miracle occurred and the clipped signal was restored into the spikes that are present in the processed signal. Unfortunately the result is not so optimistic because in the original signal there is only small information on the dynamics that gives only a slight chance to compute the signal envelope correctly. Statistically, the dynamics of the recording was increased, but subjective listening tests did not show a reasonable improvement. Instead, the character and color of the sound was changed.

IV. Experience Discussion and Conclusion

The software-based dynamics expander has been created in order it could be decided whether it is convenient to start development of the hardware expansion unit. However, the results are partly ambiguous. The sensible recording improvement was indicated at those recordings that were compressed only in a moderate way. Mostly these recordings were older, processed on the old analogous devices that did not give the sound engineers so many possibilities to compress and clip the signal heavily. When processing such recordings, the expander gives satisfactory results. Unfortunately, if the audio recording is compressed too heavily, the information on the dynamics of the signal is lost and the methods of obtaining the signal level envelope described in this paper and employed in the presented algorithm are inoperative. The further research of better methods of the signal level envelope obtaining is needed. However, it was proven that it is possible to partially increase the dynamics of the signal when not compressed too heavily. From the experience with the software presented in this paper the following recommendations arise:

- The dynamics expansion processor should provide real time preview of the effects on the expansion in order the setting of the expansion parameters could be done more accurately.
- More than 4 frequency bands should be employed.
- A possibility to select the cross-over frequencies among the frequency bands should be provided.
- The expander should provide the possibility of different slope and threshold settings for different frequency bands because especially at high frequencies the expansion is usually too heavy compared to the other frequency bands.
- In case of hardware construction, voltage-driven amplifiers as for example described in [3] can be used being driven by the output of the signal processor that analyses the characteristics of the processed signal.
- The signal level envelope should be processed more sophistically, distinguishing different characteristics of different sounds (percussion, voice, etc.).
- For older noisy recordings the expander should be also provided with the possibility of attenuation of those frequency bands that are not utilized at the moment.

Practical subjective tests shown that even the dynamics was increased heavily the listeners were not sensitive directly to this increase but to the sound color and characteristics that were changed by the expansion. The subjective perception of the dynamics improvement was different for different recordings. Therefore the expander should provide a real-time preview of how the result is with the parameters set by the user.

References

[1] Mastorakis N, New General Transformations for 2-D FIR and IIR Filter's Design, WSEAS Transactions on Circuits and Systems, Vol.10, No. 1, 2011, pp. 1-9, ISSN: 1109-2734.

[2] Smart, J., Hock, K., Csomor, S., Cross-platform GUI programming with wxWidgets, Pearson Education, USA, 2006, ISBN 0-13-147381-6.

[3] Pospisilik, M., Adamek, M., Simulation and Design of Voltage Controlled Amplifier for Dynamic Expanders, International Journal of Mathematics and Computers in Simulation [online], 2011, Vol. 5, Issue 5 [cit. 2011-29-11], s. 446-453, accessible from WWW < http://www.naun.org/journals/mcs/17-129.pdf>, ISSSN 19980159

[4] Uhlíř, J., Sovka, P., Čmejla, R.: Introduction to digital signal processing [Úvod do číslicového zpracování signálů]. Czech Technical University in Prague, Praha, 2003.

[5] Interactive filter design package, Online: http://www-users.cs.york.ac.uk/~fisher/mkfilter/trad.html [cited 03-01-2012]

[6] Vejražka, F., Signals and systems [Signály a soustavy], Czech Technical University in Prague, 1992

[7] Ježek, D. Road to Hell - How the publishers affects the technical quality of the music. *Deep in it* [online]. 2010 [cit. 2012-07-26]. Accessible from: http://diit.cz/clanek/road-to-hell-aneb-jak-vydavatele-poskozuji-technickou-kvalitu-hudby

[8] Biolek, D., Olsak, M., Biolkova, V. Optimization of elliptic leap-frog CDBA-based filters. Contribution to the book "Computational Methods in Circuits and Systems Applications", WSEAS Press, Electrical and Computer Engineering Series, 2003, p. 221-225.

[9] Mahesh S. Chavan, R.A. Agarwala, M.D. Uplane, "Application of Chebyshev II digital filter for noise reduction in ECG Signal", WSEAS transactions on Circuits and Systems, October 2005, Volume 4, Issue 10, pp. 1260-1267

[10] Mahesh S. Chavan, R.A. Agarwala, M.D. Uplane, "Real Time application of Digital Chebyshev I Filter for removal of interference in the Electrocardiogram Signal", WSEAS transactions on Biology and Biomedicine, July 2005, Volume 2, Issue 3, pp. 280-289.