Proceedings of 20th International Congress on Acoustics, ICA 2010

23-27 August 2010, Sydney, Australia

Automatic Loudness Control

Gottfried Behler, Martin Guski and Michael Vorländer

Institute of Technical Acoustics, RWTH Aachen University, Aachen, Germany

PACS: 43.60.Dh

ABSTRACT

In vehicles music is often disturbed by background noise which sometimes even drowns out quiet passages completely. An adjustment by the driver is not a good solution, because it occurs always too late and the driver might cause an accident. The aim of this work was to develop an Automatic Loudness Control (ALC), which takes music and background noise into account. An algorithm compares the specific loudness of the music with the background noise and approximates masking effects by using a developed model. To implement the ALC an estimation of the background noise is required. Echo compensation was developed, which separates the background noise from the music. The operating data of the vehicle is a second source for this estimation. For practical use the complete system has to be able to operate in real time. To achieve this multiple measures for optimization were implemented. At the end an objective and subjective analysis of the developed system was performed which proved the correct functionality of this system.

INTRODUCTION

In modern cars the transfer of information and the rendition of entertainment are united by the infotainment system. The audio signal in the vehicle is disturbed by several sources of background noise like tires, road, wind and the engine of the vehicle itself or other vehicles. A change of the noise level, for example caused by a change of the driving condition, can provoke the driver to adjust the level of the infotainment system to reobtain the former subjective relation between the level of audio signal and noise level.

Audio content with a high dynamic range, like classical music or jazz, can also cause an intervention of the driver even if the noise level is constant. In a transition from a pianissimo to a fortissimo passage the music can turn unpleasant loud while during a transition into the other direction the music can be disturbed by the background noise so that it is not completely audible.

These two scenarios - change of music and noise level - necessitate the analysis of the noise as well as the audio.

AUTOMATIC LOUDNESS CONTROL

There are several automatic volume control systems available for modern cars. The widely used GALA system for example is a speed-dependent volume control that turns up the volume with increasing speed. However, the system is based only on a rough approximation of the background noise. The actual noise level in the vehicle has other influencing variables besides speed, like load and revolutions of the engine, street conditions or external noise sources. Furthermore the loudness of the audio signal (usually music) is not considered.

There are further approaches to adaptive volume control like from the company Volkswagen AG® [2] or the system AutoPilot® [3] form the company Bose®. However, the internal operation principles are unknown to the authors.

Nicolas Heyde has shown that a level controlling system which based on a psychoacoustic model provides better results than energy based models [1]. In this work the loudness model by Zwicker (defined in ISO 532 [4]) is used, according to Heyde.

Ratio of Total Loudness

The algorithm presented in this paper is based on the assumption that the driver wants to have a constant ratio between the subjective perceived volume of the audio signal and the background noise. In the total loudness (TL) algorithm the controlled process variable is the ratio V between the total loudness of the audio signal N_x and the total loudness of the background noise N_b:

$$V = \frac{N_x}{N_b}$$

The ratio V is an appropriate user interface, because a statement like '*The music should be two times as loud as the background noise.*' is intuitively understandable.

For a given target value V and a calculated actual ratio the assumption can be made weather to increase, decrease or hold the volume. Though it is not possible to compute the physical factor by which the audio signal has to be modified because of the nonlinearity of the loudness calculation. To avoid an iterative calculation an empirical estimator was determined. The limited accuracy is small compared to the deviation between target and actual value caused by the post-processing (see Subsection Post-Processing).

Ratio of weighted Specific Loudness

The ratio of the total loudness does not take the spectral composition of the two signals into account. That is why occurring masking effects cannot be included. This effect has a huge impact because of the typical spectral distribution of the background noise (see Figure 1). In this example a large part of the background noise (approximately above 5 Bark) is masked by the audio signal and cannot be perceived. This part is responsible for the deviation of the calculated ratios of signals from the subjective impression. Accordingly the presumption follows that the TL algorithm calculates often too high amplification.

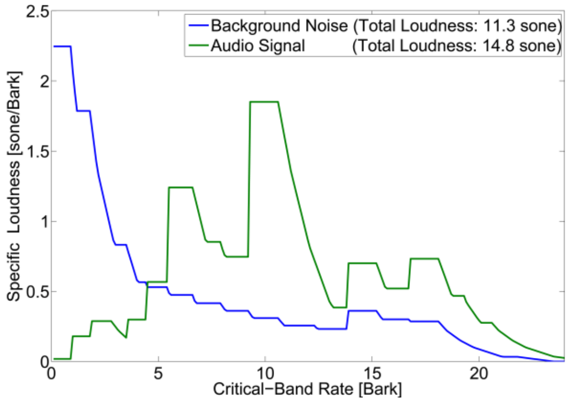

Figure 1. Example of the specific loudness of music signal N'_x (green) and background noise N'_b (blue).

The TL algorithm is extended by using the specific loudness $\boldsymbol{N'}$of the signals (SL algorithm). A simple weighting function was created to approximate the mutual masking effects: Two sounds with the same specific loudness do not mask each other, while one sound is masked completely if the specific loudness of the second sound is twice as high. A linear interpolation between these two cases leads to the weighting function shown in Figure 2.

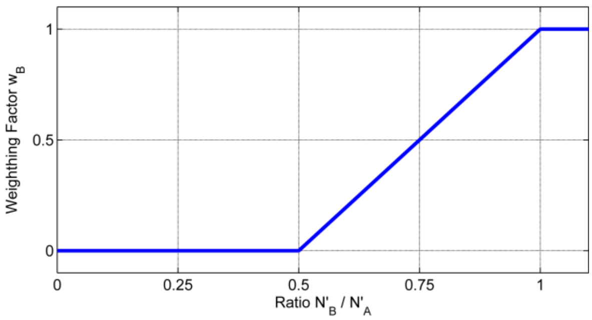

Figure 2. Weighting factor w_B as function of the specific loudness ratio $\frac{N'_B}{N'_A}$.

By applying the function to both signals the specific loudness is reduced in the frequency parts that cannot (completely) be heard by the driver. The spanned area of the weighted specific loudness $\boldsymbol{\widetilde{N}'_x}$ is equal to the weighted total loudness $\boldsymbol{\widetilde{N}_x}$

The ratio

$$\widetilde{\mathrm{V}} = \frac{\widetilde{\mathrm{N}}_\mathrm{x}}{\widetilde{\mathrm{N}}_\mathrm{b}}$$

is the new controlled process variable. An iterative search makes it possible to determine the correct amplification. The search crawls are limited to guarantee the real-time capability.

After applying the weighting function to the examples from Figure 1 the calculated ratio $\boldsymbol{\widetilde{V}}$ matches better with the subjective impression. The parts of the specific loudness of the background above 5 sone have nearly no influence on the weighted total loudness $\boldsymbol{\widetilde{N}_b}$ (see Figure 3).

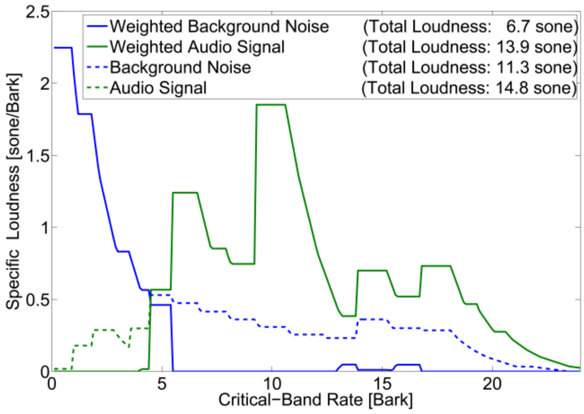

Figure 3. Specific Loudness after application of the weighting function.

A comparison of the calculated gain of the two concepts shows that the TL algorithm applies a higher amplification. This confirms the assumption that the gain of the TL algorithm tends to be too high.

Another possibility for the target value is to define a percent value for the total loudness of the original signal that should remain unmasked by the background noise.

Post-Processing

Although a definite target value is given, the aim is not to achieve this value exactly. The extreme fast adaption of the amplification factor would distort the signal so that the listener would not accept it. Therefore the target value is interpreted as a guide value and the calculated amplification is temporally smoothed before appliance. The smoothing is carried out in two steps. First, the calculated gain is exponentially smoothed and then the slew rate is limited.

At the same time the calculated amplification is analyzed and enduring changes in the loudness of the signals are detected. This strategy should avoid unpleasant situations produced by the ALC algorithm: Granted that there is a sudden change form quiet to loud music and in the quiet part the ALC applied amplification. The temporal smoothing would cause an amplification of the first part of the loud music and thus increase the difference of the change.

Additionally to the presented system, that performs the task of a dynamic compression, there is a second control loop: the basic control loop (BCL). The BCL operates slower than the dynamic compression and its task is to compensate different modulation amplitudes of different media so that the operating point of the dynamic compression is controlled.

BACKGROUND NOISE ESTIMATION

The background noise estimation is carried out by four parallel working modules (see Figure 4): Acoustic echo cancellation (AEC), subtraction of power spectra, acoustic model of the vehicle, and direct interpretation of the microphone signal. The four modules will be presented in detail in the following section.

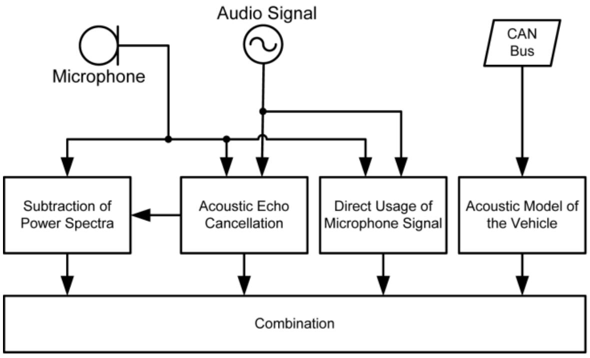

Figure 4. Structure of the background noise estimation system

Acoustic Echo Cancellation

The background noise in the vehicle is not that easy accessible. A microphone that is attached to the vehicle interior records a signal $\boldsymbol{r(t)}$ that consists of a combination of the signal played back by the infotainment system $\boldsymbol{d(t)}$ and the background noise $\boldsymbol{b(t)}$

$$\mathbf{r(t) = d(t) + b(t)}$$

Here $\boldsymbol{d(t)}$ is the audio signal $\boldsymbol{x(t)}$ that was influcenced by the tranferfunctions of loudspeaker, room and microphone. For an accurate estimation of the background based on the microphone signal the audio signal $\boldsymbol{d(t)}$ has to be eliminated from the recording. An incorrect estimation of the background noise that includes parts of the audio signal could lead to instability of the system:

In case of loud background noise the algorithm will increase the level of the audio signal. Due to the also increased remaining part of the signal $\boldsymbol{d(t)}$ in the noise estimation the algorithm will again increase the level. This feedback continues until the maximum level of the audio system is reached.

The echo cancellation is able to solve this problem. Therefore the loudspeaker-room-microphone (LRM) system $\boldsymbol{h(t)}$ is estimated by the evaluation of the correlation of the loudspeaker and the microphone signal. The estimated audio signal $\boldsymbol{\hat{d}(t)}$ can be calculated and subtracted from the recording to get the estimated background noise (see Figure 5):

$$\boldsymbol{\hat{b}(t) = r(t) - \hat{d}(t) = b(t) + d(t) - \hat{d}(t)}$$

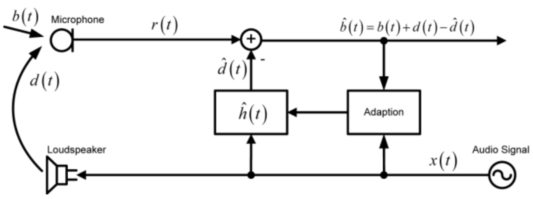

Figure 5. Schematic structure of the acoustic echo cancellation system.

In this work the Fast Least-Mean-Square (FLMS) algorithm is used to estimate and continuously adapt the LRM system [5]. The aim of the algorithm is to minimize the energy of the remaining-echo-signal. The implementation of the algorithm in the frequency domain provides a significant performance improvement due to the use of the Fast Fourier Transformation.

The choice of the speed of the adaptation is of great importance and depends on the noise level in the vehicle. This unknown noise level is estimated by observing the coherence between microphone signal $\boldsymbol{r(t)}$ and source signal $\boldsymbol{x(t)}$ and the coherence between microphone signal $\boldsymbol{r(t)}$ and estimated audio signal $\boldsymbol{d(t)}$ [6].

Subtraction of Power Density Spectra

The echo cancellation is very sensitive to phase errors. A phase error of 1° for example limits the maximum attenuation to 35 dB. Because of this sensitivity and the fact that the phase is irrelevant for the loudness calculation a different approach is taken to estimate the background noise. The correct phase subtraction is replaced by a subtraction of the power density spectra:

$$\begin{aligned}\left|\widetilde{\mathbf{B}}\right|^2 &= |\mathbf{R}|^2 - \left|\widehat{\mathbf{D}}\right|^2 \\ &= |\mathbf{D}|^2 + \mathbf{2} \cdot \mathbf{Re}\{\mathbf{DB}^*\} + |\mathbf{B}|^2 - \left|\widehat{\mathbf{D}}\right|^2\end{aligned}$$

The error term $\boldsymbol{Re\{DB^*\}}$ is the real part of the cross-correlation between audio signal at the microphone $\boldsymbol{D}$ and background noise $\boldsymbol{B}$. Although it can be assumed that $\boldsymbol{D}$ and $\boldsymbol{B}$ are uncorrelated, the error term is not negligible because of the limited observation period. The error term can be reduced by using a buffer structure to expand the observation period. The drawback of this technique is the increased latency. The buffer size has to be a trade-off to minimize the impact of the disadvantages.

First tests showed that the error term from the subtraction of power spectra has less negative influence on the background noise estimation than the occurring phase errors.

Acoustic Model of the Vehicle

Mostly the background noise is caused by the vehicle itself. It stands to reason to use an acoustic model of the vehicle to estimate this background noise. The CAN-bus can provide the input data like speed, number of revolutions and engine load. This technique is used in automobile industry for authentic and realistic audio simulations. In this work only the third octave band levels are needed, which simplifies the situation. Additional data from shock absorbers or rain sensors can help to obtain more information about the street conditions and thus improve the background noise estimation.

Direct Usage of the Microphone Signal

In the case that no audio signal is played back the recording of the microphone can be directly interpreted as noise. This can for example be useful if there is no estimation of the impulse response of the echo cancellation yet. This situation occurs when estimating the transfer function was launched or relaunched. In this situation the reliability of the acoustic echo cancellation is small.

EVALUATION AND OUTLOOK

The evaluation of the system serves to verify the intended functionality of the ALC algorithm and to consider the acceptability by the user.

To analyse these findings, a computer simulation was carried out. The advantages of this approach are the comparability of different versions (without ALC, TL- and SL-algorithm) and the repeatability.

Objective Evaluation

The simulation system was as well adapted to the reality as possible to create realistic and authentic conditions. A test drive was made to record the background noise in different situations. The transfer functions from infotainment system to a dummy head h_{DH} and to the ALC microphone h_{mic} were determined in order to use these in the simulation, so that the mixture of the audio signal and the background noise could be created for the computer simulation. There were no variations of the transfer functions in this setup so that the task was simplified for the AEC.

In the objective evaluation different combinations of audio and background signals were tested to verify the correct functional basis. Figure 6 shows the situation at an accelerated ride on a cobblestone street while playing back Pop-Music. The loudness of the background noise increases over the time while the loudness of the original music signal remains constant up to about 42 seconds. The ALC-algorithm adjusts the level of the music so the ratio of music and background remains constant.

The music becomes suddenly louder at about 42 seconds. The post-processing detects this increase in loudness and deactivates the temporal smoothing of the calculated amplification. The sudden volume change is prevented, which would become more extreme due to the amplification.

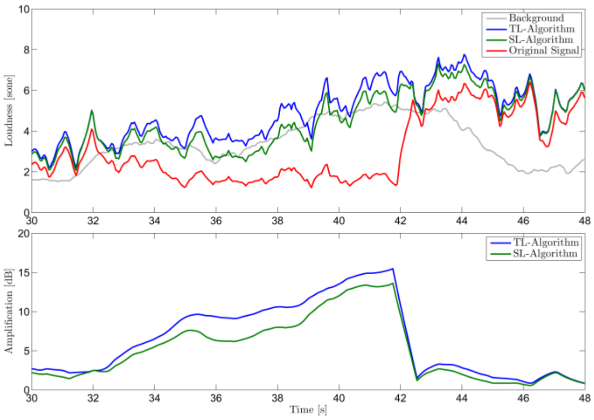

Figure 6. Loudness and amplification factor of the ALC-System. Audio Signal: Pop Music, Background Noise: Ride on a cobblestone street.

Subjective Evaluation

As a subjective evaluation hearing tests were carried out. The aim was to verify that the modification of the signal has no negative effects on the listener.

The structure of the simulation system is as far as possible adapted to the real scenario (see Figure 7). Due to the recorded background noise it is possible to compare different versions (with and without ALC) of exactly the same situation.

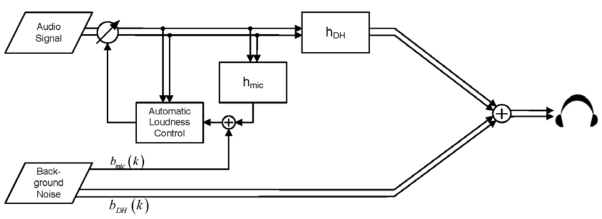

Figure 7. Schematic structure of the simulation system. The background noise was recorded at a possible position for a ALC microphone b_{mic} and with a dummy head on the front passenger seat b_{DH}.

There were two conflicting requirements when choosing the length of the test sounds. On the one hand the sounds should be short, so that the test person could still remember the first sound while hearing the second. On the other hand the effects of the ALC algorithm can only be properly recognized if the test sounds are long enough.

To carry out a proper hearing test a graphical user interface was developed that provides the possibility to switch between the versions time-synchronous. With this technique it is possible compare even long test sounds without any problem.

The hearing test showed that users listening to the music preferred the ALC versions to the original versions. The evaluation of the average rank correlation of the subjects showed that the preferences of the ALC versions are statistically significant. Exceptions were the speech signals that had a very low rank correlation. The Evaluation of the comments of the test persons indicated, that most people could not notice a difference between the three Versions.

In one audio example some subjects noticed the intervention of the ALC system and evaluated the ALC versions as worse than the original. This suggests that the various parameters of post-processing not yet been set optimally. With further hearing test is necessary to check whether better adjustment can be made. Otherwise there is the possibility of making some parameters accessible to the user.

Outlook

The echo cancellation will be subject of further research including on the one hand the test of the system under realistic conditions and on the other hand the expansion of the system to stereo respectively 5.1 signals.

A problem which may occur only under real conditions might be speaking noise of the passengers in the car. Instead of decreasing the audio level to enable a conversation, the speech is interpreted as noise and the audio signal is increased. A possible solution might be to integrate speech activity detection.

REFERENCES

1 N. Heyde, “Loudness Control of Level in Automotive Infotainment Systems” *Diploma Thesis Institute of Technical Acoustics, RWTH Aachen University* (2008)

2 Volkswagen AG, Patent “DE 100 52 104 A 1” (2000)

3 Bose® AudioPilot® Technology “http://www.bose.com/controller?url=/automotive/innovations/audiopilot.jsp“

4 International Organization for Standardization, *ISO532: Acoustics – Method for Calculating Loudness Level* (1975)

5 P. Vary and R. Martin, *Digital Speech Transmission: Enhancement, Coding and Error Concealment* (John Wiley & Sons, 2005)

6 J. Vogel, M. Heckman and K. Kroschel, “Frequency Domain Step-Size Control in Non-Stationary Environments” *Asilomar Conference 2000. Vol. 1, p. 212-216* (2000)