Automatic Loudness Control for Car Infotainment Systems

Martin Guski, Gottfried Behler

Institute of Technical Acoustics, RWTH Aachen University, 52066 Aachen, Germany, Email: mgu@akustik.rwth-aachen.de

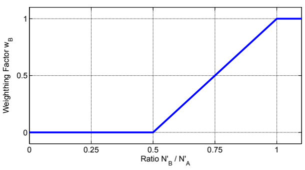
Introduction

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 subjective relation between the level of audio signal and noise level can provoke the driver to adjust the level of the infotainment system to reobtain the former ratio. Audio content with high dynamic range, like classical music or jazz, can cause such a change. The adjustment by the driver is an inconvenient solution, as it is always delayed and furthermore resulting in decreased attention to the traffic. A change of the driving conditions can also cause an intervention of the driver even if the music level is constant.

Automatic Loudness Control

Most automatic volume control systems are based only upon a rough approximation of the background noise by analyzing the velocity. However the actual noise level in the vehicle has other influencing variables like load and revolutions of the engine, street conditions or external noise sources. Furthermore in these systems the loudness of the audio signal (usually music) is not considered.
In a preliminary work on the subject Nicolas Heyde has shown that a level controlling system based upon a psychoacoustic model provides better results than energy based models [1]. In this work the loudness model by Zwicker is used, according to Heyde.
The algorithm presented in this paper is based on the assumption that the driver wants to have a constant relationship between the subjective perceived volume of the audio signal and the background noise. This ratio is a good user interface, because it is intuitively understandable. The ratio of the total loudness does not consider the spectral composition of the two signals. In this way occurring masking effects cannot be included. This effect has great impact because of the typical spectral distributions of the signals (see Fig. 1 dotted line). In this example a large part of the background noise (approximately above 5 Bark) is masked by the audio signal and because of that cannot be perceived. To improve the calculation the specific loudness N' of the signals is used.

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 yields the weighting function shown in the Fig. above. The integration of the weighted specific



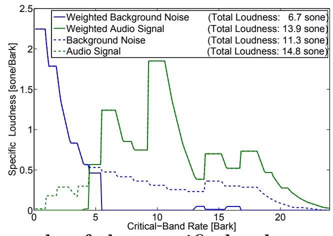


Figure 1: Example of the specific loudness of a music signal N'_x(green) and a background noise N'_b (blue).

loudness is the weighted total loudness $\widetilde{N}_x$ (or $\widetilde{N}_b$). The controlled process variable is the ratio

$$\widetilde{V} = \frac{\int\limits_{z\,=\,0\text{ Bark}}^{24\text{ Bark}} N'_x(z) \cdot w_x(z)\ \mathrm{d}z}{\int\limits_{z\,=\,0\text{ Bark}}^{24\text{ Bark}} N'_b(z) \cdot w_b(z)\ \mathrm{d}z} = \frac{\widetilde{N}_x}{\widetilde{N}_b}\ . \qquad (1)$$

The determination of the correct amplification successes using an iterative search. The search crawls are limited to guarantee the real-time capability. After applying the weighting function on the examples the parts of the specific loudness of the background noise above 5 sone have nearly no influence on the total loudness $\tilde{N}_b$ (see Fig. 1).

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 in such a way that the listeners would not accept it. Therefore the target value is interpreted as a guide value and the calculated amplification is temporal 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 an amplification. The temporal smoothing would cause an amplification of the first part of the loud music and in this way increase the difference of the change.

Background Noise Estimation

The background noise estimation is realized in four parallel working modules (see Fig. 2). The results of the modules are combined according to the accuracy of the estimation. One module directly evaluates the microphone signal if no audio signal is played back. Another module estimates the background by an acoustic model of the vehicle, which evaluates the data of the CAN-bus.

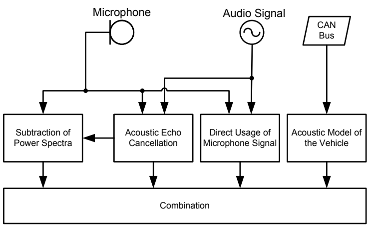


Figure 2: Structure of the background noise estimation system.

The main module of the background noise estimation is the Acoustic Echo Cancellation (AEC). A microphone in the interior of the vehicle records the signal $r(t)$ that consists of a combination of the signal played back by the infotainment system $d(t)$ and the background noise $b(t)$. The AEC is able to estimate the influence of the loudspeaker-room-microphone system on the audio signal and subtract it from the recording:

$$\hat{b}(t) = r(t) - \hat{d}(t) = b(t) + d(t) - \hat{d}(t) \tag{2}$$

In this work the Fast Least-Mean-Square (FLMS) algorithm is used to estimate and continuously adapt the LRM system. 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 microphon signal r(t) and source signal x(t) and the coherence between microphone signal r(t) and the estimated audio signal d(t) [2].

Because of the sensitivity of the echo cancellation to phase errors and the fact that the phase is irrelevant for the loudness calculation an extended approach is taken to estimate the background noise. The correct phase subtraction of (2) is replaced by a subtraction of the power density spectra:

$$|\widetilde{B}|^2 = |R|^2 - |\hat{D}|^2 \tag{3}$$
$$= |D|^2 + 2 \cdot \Re\{DB^*\} + |B|^2 - |\hat{D}|^2 \tag{4}$$

The error term $\Re\{DB^*\}$ is the real part of the cross-correlation between audio spectrum D at the microphone and background noise spectrum B. Although it can be assumed that D and 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.

Evaluation

The evaluation took place in a simulation on the PC and not in reality. The advantages of this approach are the comparability of different versions (with and without ALC) and the repeatability. 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 the infotainment system to a dummy head and to the ALC microphone were determined in order to use these in the simulation.

In the objective evaluation different combinations of audio and background signals were tested to verify the correct functional basis. Figure 3 shows the situation at an accelerated ride on cobblestones while playing back Pop-Music. The loudness of the background noise increases over time while the loudness of the original music signal remains approximately constant for the first 42 seconds. The ALC-algorithm adjusts the level of the music so that the ratio of music and background is approximately constant. The post-processing detects the music level shift at second 42 and deactivates the temporal smoothing of the calculated amplification. A sudden volume change is prevented, that would become more extreme due to the amplification. As a subjective evaluation hearing tests

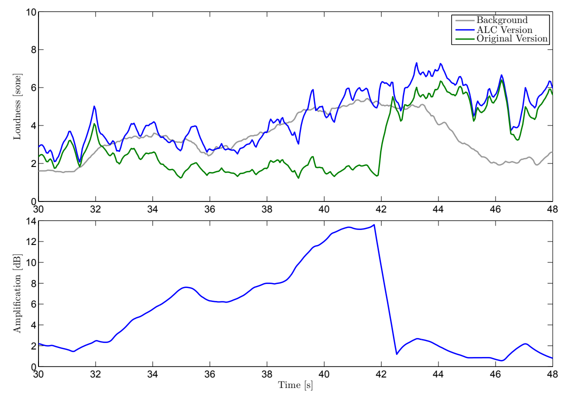


Figure 3: Loudness and amplification factor of the ALC-System. Audio Signal: Pop Music, Background Noise: Ride on a cubblestone street.

were carried out. The aim was to verify that the modification of the amplification has no negative effects on the listener. The hearing test showed that in case of music the ALC version was preferred. For speech signals the result was inconclusive. The evaluation of the average rank correlation showed the same result as the evaluation of the comments: Most of the test persons could not point out a difference for speech signals.

References

[1] Heyde, N. Loudness Control of Level in Automotive Infotainment Systems. *Diploma Thesis*, 2008, Institute of Technical Acoustics, RWTH Aachen

[2] Vogel, J., Heckmann, M., Kroschel K. Frequency Domain Step-Size Control in Non-Stationary Environments. In *Asilomar Conference 2000*. vol. 1