Adaptive receive side equalization for improved intelligibility in automotive hands-free systems

Tobias Wolff and Markus Buck

Acoustic Speech Enhancement Research, Nuance Communications Aachen GmbH, D-89077 Ulm, Germany,
Email: {tobias.wolff, markus.buck}@nuance.com

Introduction

The intelligibility of the received audio signal in automotive hands-free systems typically suffers from limited bandwidth as well as from being rendered in noisy environment. Additionally, a hands-free system is often operated in different vehicles, and therefore also with different amplifiers and loudspeakers. A car-specific tuning including the final sound system is, however, expensive and not always possible. Therefore, the amplifier-loudspeaker-enclosure (ALE) system changes in different applications. As a consequence, the perceived quality of the presented speech signal may vary depending on the utilized audio system and the acoustics of the car cabin. In order to increase the intelligibility, artificial bandwidth extension (BWE) may be employed [1]. Automatic volume control and adaptive equalization are further means to enhance the SNR for the listener, whereas the equalizer can also be used to equalize the ALE system.

Fig. 1 shows an overview of the proposed system. Firstly, the receive path contains an automatic gain control (AGC) to establish a reasonable signal level. Secondly, BWE is applied which, ideally, goes along with an increased intelligibility. On the other hand, BWE may introduce artifacts so the quality can actually be degraded. The third stage in the receive processing chain is an automatic volume control to make up for changing noise levels. As described in [2], the volume control is driven by the echo-to-noise power ratio (ENR). The adaptive equalizer (AEQ), which is applied next, is also based on power spectral density (PSD) estimates of the noise and the echo signal, which are available from the send path as originally proposed in [2]. In addition, the AEQ explicitly exploits the acoustic echo cancellation (AEC) filter. Finally, soft-limiting is applied to avoid clipping.

Adaptive Equalizer

The frequency response of the AEQ has three parts:

$$H_{\mathsf{AEQ}}(\Omega) = H_{\mathsf{ENR}}(\Omega) \cdot H_{\mathsf{D}}(\Omega) \cdot H^{-1}_{\mathsf{ALE}}(\Omega) \quad \in \mathbb{R}. \tag{1}$$

Here, $H_{\mathsf{ENR}}(\Omega)$ depends on the spectral ENR and is used to improve the intelligibility while keeping BWE artifacts masked by local noise. $H_{\mathsf{D}}(\Omega)$ represents the time-invariant desired response of the system which shall be effective for $H_{\mathsf{ENR}}(\Omega) = 1$. To this end the equalizer inverts the characteristics of the ALE system represented by $H_{\mathsf{ALE}}(\Omega)$. An estimate for the latter can be obtained using the transfer function of the AEC filter:

$$\hat{H}_{\mathsf{ALE}}(\Omega) = |H_{\mathsf{AEC}}(\Omega)| \cdot |H_{\mathsf{MIC}}(\Omega)|^{-1}. \tag{2}$$

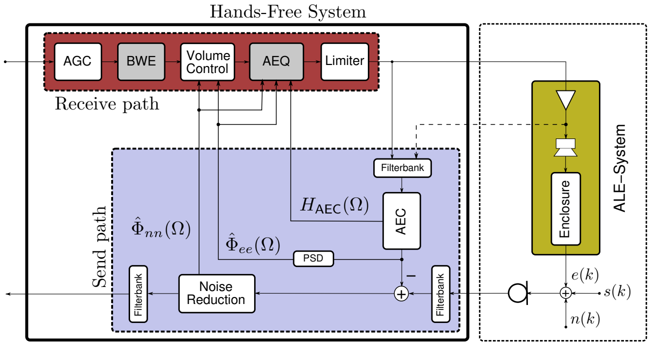

Figure 1: Block diagram of the proposed system. The receive path processing exploits PSD estimates available from the send path as well as the transfer function of AEC filter.

The idea is thus to measure the frequency response of the microphone $|H_{\mathsf{MIC}}(\Omega)|$ during the tuning process and to use the AEC transfer function to achieve a consistent sound impression when used with different ALE systems. This is particularly important in combination with BWE. In order to introduce only little processing delay, the equalization filter is implemented in time domain using a low-order FIR approximation of $H_{\mathsf{AEQ}}(\Omega)$. In principle, the FIR coefficients can be obtained by applying the Levinson-Durbin recursion to the IFFT of $H^{-2}_{\mathsf{AEQ}}(\Omega)$ [2].

As a matter of fact, $\hat{H}_{\mathsf{ALE}}(\Omega)$ refers to the position of the hands-free microphone. For the equalization according to Eq. 1 to be successful, it is required that $\hat{H}_{\mathsf{ALE}}(\Omega)$ is at least representative for the driver's position. To evaluate this experimentally, ALE transfer functions $H_i(\Omega)$ referring to 20 randomly distributed positions i in the vicinity of the driver's and the co-driver's position have been measured. For each position, $|H_i(\Omega)|^2$ and its equalized version is looked at. The result of the Levinson-Durbin algorithm to the IFFT of $\hat{H}^2_{\mathsf{ALE}}(\Omega)$ (Eq. 2), is used as FIR equalization with frequency response $Q_{\mathsf{FIR}}(\Omega)$. Note that this inverts $\hat{H}_{\mathsf{ALE}}(\Omega)$ and refers to the hands-free microphone. Fig. 2 (top) shows the spectral flatness [3]

$$SF_i = \frac{\exp\left(\frac{1}{2\pi}\int_{-\pi}^{\pi} \ln\left(\,|Q_{\mathsf{FIR}}(\Omega)|^2 \cdot |H_i(\Omega)|^2\right)\, d\Omega\right)}{\frac{1}{2\pi}\int_{-\pi}^{\pi} |Q_{\mathsf{FIR}}(\Omega)|^2 \cdot |H_i(\Omega)|^2 \, d\Omega} \tag{3}$$

in dB averaged over the driver and co-driver positions respectively for different FIR orders N_{FIR}. It can be seen that for both, driver and co-driver position, the spectral flatness can be increased if the FIR order is chosen between $N_{\mathsf{FIR}} = 3$ and $N_{\mathsf{FIR}} = 7$. An example of the frequency responses and the corresponding equalization filter $Q_{\mathsf{FIR}}(\Omega)$ for $N_{\mathsf{FIR}} = 5$ can be seen in Fig. 2 (bottom).

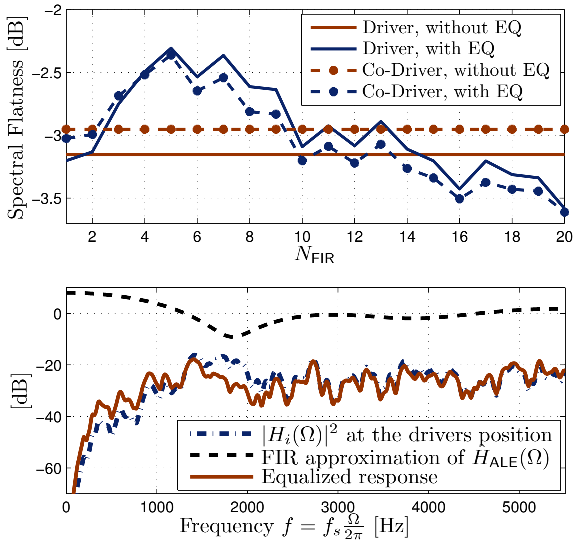

Figure 2: Top: Average spectral flatness as a function of the equalizer filter-order N_{FIR} for the driver and co-driver position. The spectral flatness may be increased with a low-order FIR approximation of Eq. 2. **Bottom**: Example frequency responses for $i = 2$ and $N_{\mathsf{FIR}} = 5$.

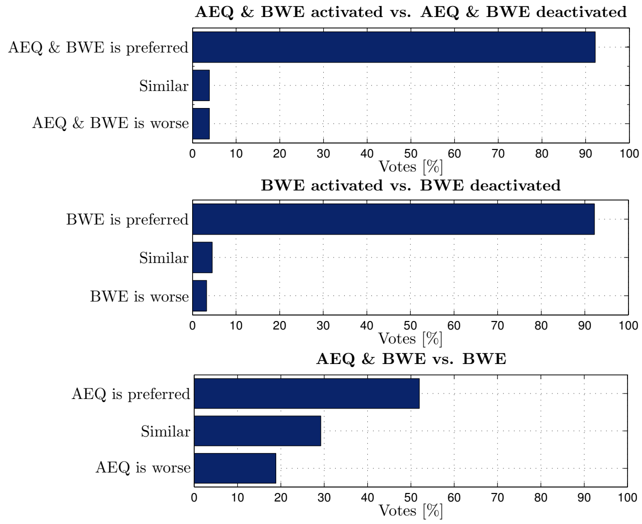

Figure 3: Results of the subjective evaluation. **Top**: Proposed system versus the reference (AGC, volume control and Limiter). **Middle**: AEQ deactivated versus reference. **Bottom**: Direct comarison AEQ & BWE versus BWE.

As for the ENR-dependent part $H_{\mathsf{ENR}}(\Omega)$ in Eq. 1, it has been proposed in [2] to either define some desired ENR, or to use equal-loudness contours. In order to introduce as little distortions as possible, we propose to control $H_{\mathsf{ENR}}(\Omega)$ such that the spectral ENR does not drop below a certain minimum level $ENR_{\mathsf{MIN}}(\Omega)$. Hence, the volume control may achieve this on its own, and the AEQ only pronounces certain frequencies if necessary. Additionally, $H_{\mathsf{ENR}}(\Omega)$ is not allowed to exceed a predefined maximum value. $ENR_{\mathsf{MIN}}(\Omega)$ allows to control the AEQ-characteristic with respect to BWE during high-noise situations. Pronouncing the lower frequencies should generally be handled with care, as this may lead to humming basses when the low frequency noise is rather loud already. A larger minimum ENR, on the contrary, can be chosen in the upper frequencies. The upper bound for choosing $ENR_{\mathsf{MIN}}(\Omega)$ in the higher frequencies is mostly determined by the BWE-artifacts. Ideally, those should be kept below their masking thresholds.

Subjective Evaluation

The system described above has been evaluated in a subjective test, wherein 10 speech signal processing experts have been asked to compare binaural recordings of differently processed receive signals. AGC, volume control and soft-limiting were always activated and this version served as a reference. The reference was to be compared to signals processed with activated BWE as well as both, BWE & AEQ, active. These two versions were also to be compared with each other. The subjects task was to pick the preferred version, whereas each subject was instructed to make his / her judgements by trading off intelligibility and speech quality. The test set was composed of 25 short samples from female and male speakers. A short trial stage was offered to accustom the subjects to the task and the presentation order was completely randomized. The results for a noise level corresponding to 160 km/h can be seen in Fig. 3. Both algorithms under test, BWE and BWE & AEQ are clearly preferred over the reference. The direct comparison indicates a tendency towards the version with activated AEQ. For 120 km/h our results show basically the same trend, whereas the direct comparison is less pronounced. This can be expected as $H_{\mathsf{ENR}}(\Omega)$ changes the spectrum only moderately in this case. Due to space limitations these results are omitted here.

Conclusions

An adaptive receive equalizer is proposed to automatically control the playback of an audio signal in noisy environment. Besides its ENR characteristic the equalizer is designed to produce a desired frequency response when used with different amplifier-loudspeaker-enclosure systems. This is achieved by using the transfer function of the acoustic echo canceler. Our analysis indicates that the system's frequency response can be modified towards the desired response because the ALE system can be flattened to some extend. The effect has been achieved for the driver and co-driver position using low-order FIR filtering. Furthermore, subjective tests indicate a preference for the proposed system.

References

[1] Larsen, E. and Aarts, R.: Audio Bandwidth Extension, John Wiley, 2004.

[2] Iser, B. and Schmidt, G.: Receive side processing for automotive hands-free systems. Proc. Joint Workshop on Hands-Free Speech Communication and Microphone Arrays (HSCMA 2008), Italy, 2008.

[3] Jayant, N.S. and Noll, P.: Digital Coding of Waveforms, Prentice Hall, 1984.