Signal Processing for In-Car Communication Systems

Christian Lüke, Halil Özer, Gerhard Schmidt, Anne Theiß, Jochen Withopf

Christian-Albrechts-Universität zu Kiel, Germany
E-mail: cl/hao/gus/ath/jow@tf.uni-kiel.de

Abstract

Communicating inside a car can be difficult because there is usually a high level of background noise and also the talking passengers do not face each other as they would do in a natural conversation. In-car communication (ICC) systems are a solution to this problem. They record the talkers’ speech signal by means of microphones and reproduce it over loudspeakers that are located close to the listening passengers. However, such systems operate in a closed electro-acoustic loop which significantly limits the gain that can be introduced by the system. In order to improve this gain margin and to achieve additional signal enhancement, several signal processing techniques are applied in ICC systems. Special care has to be taken about the signal delay: If it is too large, the reverberation inside the car is increased considerably and the speech over the loudspeakers might be perceived as an echo.

In this paper, an overview of the signal processing components of an ICC system is given. The necessary signal processing steps are explained and approaches to implement these are shown, especially with a focus on low processing delays.

Keywords In-car communication, low-delay filter banks, feedback, echo cancellation, noise reduction.

1. INTRODUCTION

The communication in cars often lacks of quality in the sense of intelligibility. Especially at higher speed, the conversation comfort is reduced due to the high background noise (engine, wind, tire noise, etc.). Also the sound absorbing materials in the car which should reduce the noise inside the passenger compartment degrade the speech intelligibility. In large vehicles, for instance minivans and buses, there is also a considerable attenuation of the acoustic signals due to the distance between the talking and the listening passenger. The usual reaction is that the rear passenger speaks louder and leans forward to the front passenger. This problem increases in the communication between the front and rear passengers as the front passenger talks towards the windshield. The front passenger usually turns over which is uncomfortable for a longer time and, in addition, a security risk if the driver does so.

To overcome these problems, in-car communication systems record the talking passengers and distribute the seat-dedicated microphone signals to the loudspeakers [1, 7]. However, this technical support of the conversation contains some challenges due to the interfering signals (noise, music, etc.) and the closed-loop operation. Various signal processing techniques are required to reduce feedback, echo, and noise as well as to prevent system instability.

If the system delay exceeds 10 to 15 ms, passengers start to perceive the additional playback as a separate source [1]. The system delay consists of the delay caused by the analog-digital and digital-analog converters, amplifiers, block based signal transport on the car’s signal processing hardware, acoustical paths and also the signal processing. Subsequently, all algorithms should be designed to cause as little delay as possible. However, selected loudspeaker signals might also be delayed on purpose in order to overcome a localization mismatch between the acoustically perceived talker location and the actual one.

The ICC-system that is presented in this paper has been implemented in the Kiel Real-Time Audio Toolkit (KiRAT) in the programming language C. For testing the algorithms in a car, the software runs on a PC-platform with audio connections over low-delay ASIO soundcards. The delay (without that originating from signal processing) of this configuration is approximately 5.7 ms.

Sec. 2 gives an overview of the ICC-System and briefly explains the contained components and how they interact. More details about the algorithms employed in certain modules are given in Sections 3 to 8. Examples are shown after the algorithms are introduced in order to demonstrate the performance. Finally, conclusions are drawn in Sec. 9.

2. OVERVIEW

Figure 1 shows an overview of the signal processing in an ICC system containing the essential components. First, pre-processing is applied to each microphone signal. This contains a signal analysis, where, e.g., clipping or complete blackout of a microphone is detected. In automotive environments, usually the background noise is dominating the speech components at low frequencies. For this reason, there is a highpass filter to remove these frequencies of bad signal quality. The highpass filter that is used in the presented system has Butterworth characteristic and is of second order (two poles and two zeros). The 3 dB cutoff-frequency is set to 200 Hz, but this value depends on the user preferences as well as on the properties of the vehicle. Most of the remaining signal processing takes place in the frequency domain which
allows for reduced computational complexity. The next block is therefore an analysis filter bank (see Sec. 3 for details) that computes a subband-signal representation.

All signal spectra of microphones that are assigned to talking passengers are then enhanced in terms of their signal quality. This mainly consists of noise- and feedback reduction by a Wiener type filter as explained in Sec. 6. For this filter, noise and feedback estimates have to be computed as presented in Sec. 4 and 5, respectively.

The remaining part of the signal processing is concerned with the distribution of signals from the input microphones to the output loudspeakers and adjusting the signals for good playback quality. If multiple microphones are available for one talking passenger, first one signal per talker has to be extracted. This can be done by combining the signals, e.g., by beamforming where knowledge about the position of the talkers can be exploited. Another method is to detect which microphone offers the best signal quality in terms of SNR. Any method used here should work adaptively because the noise level might change, e.g., when a window is opened or the ventilation is turned on. Based on the output signal of this signal-combination module, a voice activity detection (VAD) as shown in Sec. 7 is necessary to determine the active talking passenger for correctly managing the subsequent steps of the signal distribution.

With the information of the VAD, the signals of non-active talkers are attenuated by a unit called loss control. Then, the talker signals are mapped to the listening passengers and further mapped to the loudspeakers that are available for a specific listener. In this last mapping, the gain of each signal is adjusted according to the background noise level. While no support of the system is usually needed during standstill, more gain is required with increasing speed. Because the noise might vary considerably between the seats of a car, each listener can be assigned one or more microphones that are used to estimate the noise level at his position. This noise estimate is then converted into a gain factor by the noise dependent gain control (NDGC, see Sec. 8). This gain factor is computed individually for each loudspeaker of a listener because, due to their position, some loudspeakers are more critical in terms of feedback.

Finally, the block Loudspeaker-specific processing shown in Figure 1 contains some processing units which enhance the loudspeaker-dedicated signals for playback. Two different equalizers are implemented to improve the sound impression, but also to optimize the feedback properties of the system by attenuating those frequencies that exhibit the largest coupling to the microphones. The first one works in the frequency domain and provides zero-phase equalization with low computational complexity. After this frequency-domain equalizer, the signals are transformed back to the time-domain by a synthesis filter bank. A so-called peak-filter equalizer can be used to realize narrow band corrections of the frequency response. Setting such narrow notches or peaks would not be possible with the frequency-domain equalizer. Other components contained in the postprocessing are a gain/delay element that can be used to adjust the spatial hearing impression and a limiter to prevent clipping of the digital/analog converters. Because the estimation of the feedback component needs information about the loudspeaker signals and operates in subband-domain, another analysis filter bank that also contains the preprocessing which is applied to the input microphones is needed.

3. ANALYSIS AND SYNTHESIS FILTER BANK

Filter banks provide a conversion between time and frequency domain. Both parts, the analysis and the synthesis filter bank need to be matched for proper operation. Their performance can be improved by applying pre- and de-emphasis filters before the analysis and after the synthesis stage.

\[1\] If the delay and the computational load of the analysis and synthesis filter banks are neglected.
3.1. Modified Overlap-Add Structure

The analysis filter bank computes the DFT \( X(\mu, k) \) of a segment of the signal \( x(n) \) which is windowed by the analysis window \( h_{\text{ana}}(n) \):

\[
X(\mu, k) = \sum_{n=0}^{N_{\text{DFT}}-1} h_{\text{ana}}(n) \, x(n + kR) \, e^{j \frac{2\pi}{N_{\text{DFT}}} n}. \tag{1}
\]

This short-time Fourier-transform is evaluated every \( R \) samples. Thus, \( R \) is often referred to as the frameshift or the subsampling rate. The variable \( k \) is the frame index and \( \mu \) the index of the subband. After arbitrary manipulations of the spectrum, the frequency domain signal \( Y(\mu, k) \) is obtained which is to be transformed back into the time domain by the synthesis filter bank. A common way for doing so is the overlap-add (OLA) method [3] where first the inverse Fourier-transform

\[
y_k(n) = \begin{cases} 
\text{IDFT}\{Y(\mu, k)\}, & \text{if } n = 0, 1, \ldots, N_{\text{DFT}} - 1 \\
0, & \text{else}
\end{cases}
\tag{2}
\]

of frame \( k \) is computed. All overlapping time domain signal snippets are then weighted by the synthesis window \( h_{\text{syn}}(n) \) and added to form the filter bank output

\[
y(n) = \sum_{k=-\infty}^{\infty} h_{\text{syn}}(n - kR) \, y_k(n - kR), \tag{3}
\]

where the synthesis window \( h_{\text{syn}}(n) \) is padded with zeros for \( n < 0 \) and \( n \geq N_{\text{syn}} \). The longer the synthesis window is, the more frames are overlapping and thus, the more delay is introduced by the synthesis filter bank. Hence, the delay can be reduced by shortening the synthesis window. The most extreme case, where this length is \( N_{\text{syn}} = R \), is known as the overlap-save (or overlap-scrap, OLS) filter bank [3]. However, this approach has some drawbacks [5]:

- Projection filters must be used to suppress artifacts of cyclic convolution.
- When adaptive algorithms (e.g., noise reduction) are applied, their parameters must be smoothed in order to avoid echoes.

Since these problems do not occur with OLA filter banks, we propose to trade off the OLS drawbacks against delay by reducing the length of the synthesis window to, e.g., \( N_{\text{syn}} = N_{\text{DFT}}/2 \).

In our implementation, we have chosen the parameters

- Frameshift \( R = 32 \) samples
- Length analysis window \( N_{\text{ana}} = 256 \) samples
- Length synthesis window \( N_{\text{syn}} = 128 \) samples
- Sampling rate \( f_s = 22050 \) Hz.

3.2. Window Design

When designing proper pairs of analysis an synthesis windows, several aspects have to be taken into account:

- The windows have to ensure perfect reconstruction [3]. Perfect reconstruction is achieved when the analysis and synthesis windows fulfill the condition

\[
\sum_{k=0}^{\left\lfloor N_{\text{DFT}}/2 \right\rfloor - 1} h_{\text{ana}}(n - kR) \cdot h_{\text{syn}}(n - kR) = 1, \tag{4}
\]

where \( \lceil \cdot \rceil \) denotes rounding to the next greater integer.
- Aliasing distortions should be kept as low as possible. During the design, this can be verified, e.g., by running an echo canceler based on the normalized least mean square (NLMS) algorithm [8] and evaluating its performance.
- Cyclic artifacts can be minimized by tapering the analysis window stronger towards one end.

With these criteria in mind, the windows shown in Fig. 2 have been developed.

3.3. Pre- and De-Emphasis Filters

Due to the limited amount of subbands, the resolution of a filter bank is limited. Even with a proper design of the analysis window, aliasing in the frequency-domain cannot be avoided totally. Therefore, a pre-emphasis filter is used to whiten the signal and thus achieve an approximately constant power of the aliasing distortion over the subbands.
Because the input signals are instationary, the desired de-
correlation of the time-domain signals cannot be achieved ex-
actly with a fixed pre-emphasis filter. However, it can be used to remove the high-frequency roll-off that is common to all speech signals. This means that low filter orders are sufficient. After the synthesis filter bank, a de-emphasis filter has to be applied in order to undo the filtering introduced in the pre-
emphasis stage. One method is to design a prediction error filter for the pre-emphasis, as these filters are always minimum phase and thus straightforward to invert [9].

4. FEEDBACK ESTIMATION

In order to obtain sufficient system gain, it is necessary to investigate the electro-acoustic feedback loop. One possibility to attack the feedback problem is to estimate the feedback component for every microphone and suppress it with a frequency-dependent attenuation factor as described in Secti-
on 6.

The model for estimating the power spectral density (PSD) of the feedback from microphone $m$ to loudspeaker $l$ in frame $k$ and subband $\mu$ is

$$
\hat{S}_{ff}^{(lm)}(\mu, k) = \alpha_{lm}(\mu) \cdot \hat{S}_{ff}^{(lm)}(\mu, k-1) + \beta_{lm}(\mu) \cdot \hat{S}_{yy}(\mu, k-d_{lm}),
$$

where the quantities are

- $\hat{S}_{ff}^{(lm)}(\mu, k)$: estimated feedback PSD,
- $S_{yy}(\mu, k)$: PSD of loudspeaker $l$,
- $\beta_{lm}(\mu)$: room coupling factor,
- $\alpha_{lm}(\mu)$: attenuation factor,
- $d_{lm}$: signal delay in frames.

The loudspeaker PSD $S_{yy}(\mu, k)$ can be estimated from the loudspeaker signal by computing the squared magnitude $|Y^{(l)}(\mu, k)|^2$. According to this first order infinite impulse response (IIR) model, the feedback component is comprised of the previous estimate, weighted by the attenuation constant $\alpha_{lm}(\mu)$ which describes how fast the feedback decays in subband $\mu$. This system is driven by the loudspeaker output signal, delayed by the length of the acoustic path $d_{lm}$ between loudspeaker $l$ and microphone $m$ and weighted by the coupling $\beta_{lm}(\mu)$.

The complete feedback PSD $\hat{S}_{ff}^{(m)}(\mu, k)$ at microphone $m$ can be estimated by summing over all contributions of the $N_{lp}$ loudspeakers:

$$
\hat{S}_{ff}^{(m)}(\mu, k) = \sum_{l=0}^{N_{lp}-1} \hat{S}_{ff}^{(lm)}(\mu, k).
$$

All model parameters of Equation (5) can be estimated from the impulse responses which describe the feedback paths. The attenuation factor $\alpha_{lm}(\mu)$ can also be converted to the more familiar reverberation time $T_{60}$ in seconds by

$$
T_{60} = \frac{60}{20 \log_{10} (\alpha_{lm}(\mu)) \frac{f}{R}}.
$$

The reverberation time is the time it takes an impulse response to decay by 60 dB. For cars, the $T_{60}$ is usually around 50 ms and the coupling $\beta_{lm}(\mu)$ for typical loudspeaker and microphone positions between 0 and $-60$ dB. Especially the coupling depends heavily on the frequency and is usually larger for low frequencies. Figure 3 shows values for the reverberation time $T_{60}$ and the coupling $\beta_{lm}(\mu)$ that have been measured inside a car for one feedback path. These parameters could also be updated and adapted to changing environments during operation by estimating the impulse responses online. This is of particular interest if the ICC-system is also equipped with echo-compensation algorithms where the needed measurements are already available.

5. NOISE ESTIMATION

It cannot be avoided that, besides the desired speech signal, the microphones also pick up background noise. If this background noise would be played back over the loudspeakers, the overall noise level in the car would be increased which is of course undesirable. The noise reduction algorithm described in Sec. 6 needs an estimate of the background noise PSD $S_{bb}(\mu, k)$ which can be obtained for (nearly) stationary noise processes in a rather simple way.

First, the magnitude of the input spectrum $X(\mu, k)$ is smoothed over time with a first order IIR filter:

$$
|X(\mu, k)|^2 = \beta_{sm} |X(\mu, k)|^2 + (1 - \beta_{sm}) |\bar{X}(\mu, k-1)|^2.
$$

The smoothing time constant $\beta_{sm}$ describes how fast the smoothed magnitude $|\bar{X}(\mu, k)|^2$ may vary over time. Since its value depends on the sampling rate $f_s$ and the frame shift $R$, it is convenient to define it in the physical unit of dB/s by the conversion

$$
\beta_{sm} = 20 \log_{10} (1 - \beta_{sm}) \frac{f_s}{R}.
$$

From now on, the tilde is used to annotate these “user friendly” variables.

Fig. 3. Reverberation time $T_{60}$ and the coupling $\beta_{lm}(\mu)$. 

Lüke, Özer, Schmidt, Theiß, Withopf – 4
A time constant of, e.g., $\beta_{\text{sm}} = 300 \text{ dB/s}$ helps to remove outliers efficiently.

The smoothed short-term power estimate $[X(\mu, k)]^2$ is then compared to the previous estimate of the noise PSD $\hat{S}_{bb}(\mu, k - 1)$ to update the estimated value:

$$\hat{S}_{bb}(\mu, k) = \left\{ \begin{array}{ll}
\gamma_{\text{inc}} \hat{S}_{bb}(\mu, k - 1), & \text{if } [X(\mu, k)]^2 > \hat{S}_{bb}(\mu, k - 1) \\
\gamma_{\text{dec}} \hat{S}_{bb}(\mu, k - 1), & \text{else.}
\end{array} \right. \quad (10)$$

The increment and decrement time constants could be chosen, e.g., like $\gamma_{\text{inc}} = 3 \text{ dB/s}$ and $\gamma_{\text{dec}} = -10 \text{ dB/s}$. If $\gamma_{\text{inc}}$ is chosen much higher, the noise estimate will increase too fast during speech periods, if it is set too small, the noise estimator cannot follow changes in the noise power fast enough. Usually, the decrement is set to a “faster” value than the increment. The noise estimator is initialized to a rather high value because the estimate drops faster and thus reaches the correct value earlier after the estimation procedure is started. Of course, more sophisticated noise estimation schemes like, e.g., minimum statistics [6] could be used.

6. NOISE AND FEEDBACK REDUCTION

For suppression of the undesired background noise and feedback components, the microphone signal $X(\mu, k)$ is multiplied with a frequency-dependent attenuation factor $G(\mu, k)$ to form the enhanced spectrum

$$X_{\text{enh}}(\mu, k) = X(\mu, k) \cdot G(\mu, k). \quad (11)$$

The attenuation coefficients are found by a modified Wiener characteristic

$$G(\mu, k) = \max \left\{ G_{\text{min}}, 1 - \frac{\beta_b \hat{S}_{bb}(\mu, k) + \beta_f \hat{S}_{ff}(\mu, k)}{\hat{S}_{xx}(\mu, k)} \right\}, \quad (12)$$

where $\hat{S}_{bb}(\mu, k)$ and $\hat{S}_{ff}(\mu, k)$ are estimates for the background noise and feedback PSDs, respectively. $\hat{S}_{xx}(\mu, k)$ is the microphone signal PSD of the current frame $k$ and can be estimated as squared magnitude of the microphone $|X(\mu, k)|^2$ spectrum. The overestimation factors $\beta_b$ and $\beta_f$ are used to correct or to intentionally introduce a bias in the estimates. Values greater than one make the filter more “aggressive”, i.e., the filter attenuates more often. Subsequently, a compromise between suppression of unwanted signal components and speech distortion introduced by extensive filtering has to be found. An overview over the noise and feedback reduction for a single microphone channel including the estimators is shown in Fig. 6.

When the filter attenuates randomly for a short time and only at some subbands, this can be heard as so-called musical tones. They can be avoided (or masked) if some residual noise is allowed by introducing the maximum attenuation $G_{\text{min}}$ which is typically set to values $-15 \text{ dB} < G_{\text{min}} < -9 \text{ dB}$.

Fig. 4 shows an example for noise reduction only (i.e., $\beta_f = 0$): The upper plot shows the spectrogram of a signal recorded in a car moving at a speed of 100 km/h after the noise reduction coefficients, shown in the plot below, have been applied according to Eq. (11). Blue color indicates the maximum attenuation of $G(\mu, k) = G_{\text{min}} = -9 \text{ dB}$, red color no attenuation $G(\mu, k) = 0 \text{ dB}$. The plot of the attenuation coefficients clearly show, where the speech components are.

An example for feedback reduction only ($\beta_b = 0$) is shown in Fig. 5. The ICC-system was operating at a maximum gain and the feedback reduction is turned off around 3 and 6.5 seconds. The upper plot shows the output signal of a loudspeaker in the time domain and it can clearly be seen that the signal energy increases considerably in these time intervals. The spectrogram below reveals that the system starts oscillating at a frequency of approximately 500 Hz. In the lower plot, the attenuation coefficients are depicted. Again, red color indicates no attenuation and the two periods when the feedback attenuation is switched off can be readily identified. The howling stops almost immediately after the
Fig. 5. Example for the feedback reduction: Loudspeaker output signal (upper and middle plot) and feedback reduction coefficients (lower plot). The feedback reduction is switched off between 2.5 and 3.5 and between 6 and 7 seconds.

Fig. 6. Structure of noise and feedback reduction with the necessary estimation schemes for a single microphone channel. Feedback reduction is switched on again.

7. VOICE ACTIVITY DETECTION

For the voice activity detection (VAD), a noise estimation has to be computed for the talker signals. This is done in $N_{\text{VAD}}$ frequency bands whose lower and upper cut-off frequencies can be set arbitrarily. It is also possible to exclude certain frequency ranges, e.g., if they are known to be heavily corrupted by noise.

For the decision of voice activity, two conditions are tested for each noise estimation band:

1. Does a talker achieve a minimum SNR?
2. Does the large SNR originate from a neighboring talker?

If a condition is met for talker $p$, this is rewarded by the increase of a counter by

$$c^{(p)}(k) = \min \left\{ 1, c^{(p)}(k - 1) + \Delta_{\text{inc}} \right\}. \quad (13)$$

If a condition is missed, it is penalized in a similar manner:

$$c^{(p)}(k) = \max \left\{ 0, c^{(p)}(k - 1) - \Delta_{\text{dec}} \right\}. \quad (14)$$

Additionally, the counter is limited to the interval $c^{(p)}(k) \in [0, 1]$. The counter changes should be normalized to the number of noise estimation bands, e.g., $\Delta_{\text{inc}} = 1/N_{\text{VAD}}$.

The first condition is, whether a minimum SNR is achieved, i.e., if

$$\hat{S}_{xx}^{(p)}(i, k) > \hat{S}_{bb}^{(p)}(i, k) \cdot \text{SNR}_{\text{min}}. \quad (15)$$

If this is true for the noise estimation band $i$, the counter $c^{(p)}$ is increased according to Eq. (13) and the second condition – if the hight SNR for talker $p$ actually originates from talker $q$’s speech – is tested. A good estimator for the signal PSD needed in Eq. (15) is the short-term power $|\bar{X}^{(p)}(\mu, k)|$ which is available as a byproduct of the noise estimation procedure of Sec. 5.

Before comparing the signal PSDs $\hat{S}_{xx}^{(p)}(i, k)$ of all talking passengers, they are normalized to the background noise level in order to remove differences in the signal power that stem from inaccuracies in the hardware, e.g., different gain settings in the microphone pre-amplifiers. Therefore, first the mean noise level over all $N_{\text{talk}}$ talking passengers

$$\bar{S}_{bb}(i, k) = \frac{1}{N_{\text{talk}}} \sum_{p=0}^{N_{\text{talk}}-1} \hat{S}_{bb}^{(p)}(i, k) \quad (16)$$

is calculated in all noise estimation bands $i$. This mean noise level is then used to find the normalization factor

$$\alpha^{(p)}_{\text{norm}} = \max \left\{ N_{\text{min}}, \min \left\{ N_{\text{max}}, \frac{\bar{S}_{bb}(i, k)}{\hat{S}_{bb}^{(p)}(i, k)} \right\} \right\}, \quad (17)$$

Lüke, Özer, Schmidt, Theiß, Withopf – 6
8. NOISE DEPENDENT GAIN CONTROL

The noise dependent gain control (NDGC) adjusts the playback volume to the noise level inside the vehicle. This is done for each listener and loudspeaker individually in order to exploit the gain-before-feedback margin as much as possible.

8.1. Basic Principle

The basic principle of the NDGC is depicted in Fig. 7: the noise estimate \( \hat{N}(k) \) is mapped onto an instantaneous gain factor \( \hat{g}(k) \) using a piecewise linear characteristic made up of \( N_{\text{map}} \) pieces. In order to avoid abrupt changes in the gain factor, the actual gain

\[
g(k) = \begin{cases} 
\eta_{\text{inc}} g(k-1), & \text{if } \hat{g}(k) > g(k-1) \\
\eta_{\text{dec}} g(k-1), & \text{else},
\end{cases}
\]

is computed by incrementing or decrementing the previous value. The corresponding time constants \( \eta_{\text{inc}} \) and \( \eta_{\text{dec}} \) can be defined in dependence of the current gain value \( g(k) \). This is useful, e.g., when the microphones should be muted during standstill. A faster increase for the low-gain case would then allow to reach an appropriate system gain within a reasonable time when the noise level increases.

8.2. Loudspeaker and Frequency Dependent NDGC

The NDGC concept explained so far can be extended to a loudspeaker and frequency dependent design which allows better adaption to the conditions of a given vehicle. Fig. 8 shows how the gain vector \( g(\mu, k) \) for a certain loudspeaker of a listener is computed. Several noise estimates \( \hat{N}_i(k) \) can be processed simultaneously by applying Eq. (14). The second condition tests if the signal to interference-ratio (SIR) between talker \( p \) and talker \( q \) (considered to be an interferer) is greater than a threshold:

\[
\alpha_{\text{norm}} \cdot \hat{S}_{xx}(i, k) > \alpha_{\text{norm}} \cdot \hat{S}_{xx}(q)(i, k) \cdot \text{SIR}_{\text{min}}.
\]

If the inequality (18) does not hold, this is penalized by decreasing the counter of talker \( p \) by applying Eq. (14).

After all noise estimation bands have been evaluated for updating the counters of all talkers, the score is compared to a threshold \( VAD_{\text{min}} \) to decide whether talker \( p \) is active or not

\[
VAD^{(p)}(k) = \begin{cases} 
1, & \text{if } \varepsilon^{(p)}(k) > VAD_{\text{min}} \\
0, & \text{else},
\end{cases}
\]

where \( VAD^{(p)}(k) = 1 \) denotes speech activity. By deciding in this fashion, it is possible to classify multiple talkers as active.
be obtained in $N_{\text{acc}}$ noise estimation bands. In the example of Fig. 8, $N_{\text{acc}} = 3$ noise estimation bands are used. These can be specified by their lower and upper cut-off frequencies $f_{i1}$ and $f_{iu}$ and might be overlapping or with gaps in between to exclude certain frequency bands totally. Each of the noise estimates is input to a set of $N_{\text{mel}}$ mapping characteristics of the type of Fig. 7 to obtain preliminary gain values $\hat{g}_i(m, k)$, where $i \in [0, N_{\text{acc}} - 1]$ and $m \in [0, N_{\text{mel}} - 1]$. In Fig. 8, $N_{\text{mel}} = 5$ melbands have been chosen. To obtain one gain factor for each melband, the preliminary gains of the same melbands are added:

$$\hat{g}(m, k) = \sum_{i=0}^{N_{\text{acc}}-1} \hat{g}_i(m, k).$$  \hspace{1cm} (21)

These factors $\hat{g}(m, k)$ are assigned to the subbands by

$$g(\mu, k) = \sum_{m=0}^{N_{\text{mel}}-1} a_{\mu, m} \cdot \hat{g}(m, k),$$  \hspace{1cm} (22)

where $a_{\mu, m}$ are overlapping triangular weighting functions for the extrapolation from melbands to subbands as schematically sketched in Fig. 8. The widths of the triangles are chosen according to the mel-scale, i.e., they are increasing towards higher frequencies.

This scheme has been successfully used in practice with $N_{\text{acc}} = 1$ and $N_{\text{mel}} = 2$. Since the maximum possible gain in the test car was about 4 dB higher at low frequencies, some extra boosting could have been applied there when very high system gain was required. Further degrees of freedom could be added to fine-tune the system.

9. CONCLUSIONS

In this contribution we presented an ICC-system for increasing the quality of a conversation inside a car. The individual algorithmic components have been presented in an overview followed by a more detailed description of most of the signal processing modules. Examples for suitable parameterizations of these algorithms have been given and also some processed data has been presented to demonstrate the functioning of the algorithms. All results have been obtained from an implementation of the ICC-system within the KiRAT framework.

Informal tests were made in a car equipped with our ICC-system consisting of low-latency sound cards, a PC for signal processing based on the presented algorithms and amplifiers for driving loudspeakers. These tests showed the ICC-system increases speech intelligibility and communication comfort at medium and high driving speed. The feedback reduction helps to improve the gain-before-feedback margin significantly. When the system operates at maximum gain and the feedback reduction is switched off, howling occurs almost instantly.

But even before the system starts oscillating, the signal quality is degraded due to an increase of reverberation caused by the feedback. The concept of the frequency and loudspeaker dependent NDGC helps to adapt the system to a given vehicle and to exploit the gain resources as good as possible.

At very high noise levels, even more gain than the system can provide currently might be desired. One way to improve the gain margin is to apply a feedback cancellation which works similar to the echo cancellation algorithms known from hands-free telephony. However, in the ICC scenario difficulties arise in continuously estimating the required impulse responses. Another issue of the presented system is that the noise estimation cannot handle highly instationary noise that occurs, e.g., when a window is opened. As a consequence, many subsequent components like the noise reduction, microphone selection, VAD or the attenuation control cannot work properly. Therefore, a detection of instationary background noise is desirable to increase the overall system performance.

10. REFERENCES