Audio Engineering Society

Convention Paper 6960

Presented at the 121st Convention
2006 October 5–8 San Francisco, CA, USA

Automatic Volume and Equalization control in mobile devices

Alexander A Goldin [1], Alexey Budkin [2], Sergey Kib [3]

[1] Alango Ltd, Haifa, Israel
alexander.goldin@alango.com

[2] Alango Ltd., St.Petersburg, Russia,
alexey.budkin@alango.com

[3] Alango Ltd., St.Petersburg, Russia,
sergey.kib@alango.com

ABSTRACT

Noise spectrum and level are changing dynamically in mobile environments. Speaker volume comfortable in quite conditions becomes too low when the ambient noise level increases significantly. Speaker volume adjusted for good intelligibility in high ambient noise becomes annoyingly loud in quite. Automatic Volume Control may compensate for different levels of ambient noise by increasing or decreasing the speaker gain accordingly. However, if the noise and sound spectra are very different, such simple gain adjustment may not work well. More advanced technology will dynamically equalize reproduced sound so that it exceeds the noise level by a specified ratio all over the frequency range. This paper describes principals and practical aspects for Automatic Volume and Equalization Control in mobile audio and communication devices.

1. INTRODUCTION

Mobile audio and voice communication devices are often used in noisy environments where noise level and its spectrum are changing dynamically. Consequently, speaker volume comfortable in quite conditions becomes too low when the ambient noise level increases significantly. From the other hands, speaker volume adjusted for good intelligibility in high ambient noise becomes annoyingly loud in quite. Implemented correctly, Automatic Volume Control (AVC) may compensate for different levels of ambient noise by increasing or decreasing the speaker signal level accordingly. However, if the noise and sound spectra are very different, such simple gain adjustment may not provide satisfactory results. For example, car noise is mainly low frequency noise with most energy below 500Hz. If the whole signal is amplified to compensate for noise masking properties in the low frequency region, high frequency sounds may become too loud. To provide equal perceptual loudness and intelligibility, different frequency regions of the reproduced sound should be amplified with different gains depending on the spectrum of the ambient noise and the sound itself.

General principals of sound equalization in noisy environments were outlined in a paper presented at 110th AES convention [1]. Automatic Volume and Equalization control (AVQ) technology developed in Alango Ltd. provides an efficient implementation of ideas outlined in [1]. AVQ may give significant benefits in a variety of practical applications such as mobile phones, Bluetooth headsets, hands-free car kits, portable music players, GPS navigation devices (voice prompts), industrial intercoms, public announcement systems and many others. This paper analyses both theoretical and practical issues of implementing Automatic Volume and Equalization control technology in mobile audio and voice communication devices.

2. GENERAL SCHEME

Automatic Volume and Equalization control is based on continuous monitoring of ambient noise and modifying the loudspeaker signal accordingly. Microphone used for monitoring may be a special reference microphone or, in many cases, the microphone already present in a voice communication terminal. Figure 1 illustrates the concept.

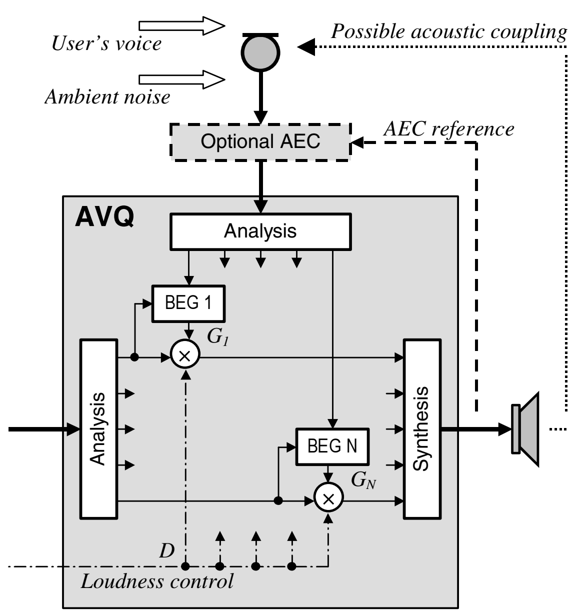

Figure 1 General scheme of Automatic Volume and Equalization control

Acoustic Echo Cancellation (AEC) is optionally performed on the reference microphone signal before AVQ processing. The purpose of echo cancellation is to clean the signal from the speaker signal possibly present in it due to acoustic coupling between the speaker and the microphone.

The speaker and the reference microphone signals are divided into equivalent frequency subbands. The corresponding subband signals are fed into Band Equalization Gain (BEG) blocks. BEG blocks compute gains $G_1 \ldots G_N$ by which each subband signal is multiplied to provide the same sound pressure level as the ambient noise. Each subband signal is further multiplied by the loudness control factor D that defines the required signal to noise ratio.

In each subband the processing is done independently as follows. The current subband speaker (LS_B) and noise (LN_B) signal levels are estimated. As it will be explained in the next section, algorithms for these estimations are principally different. The subband signal levels LS_B, LN_B measured by AVQ algorithm are signal levels inside a DSP implementing the technology. However, equalizing these levels by setting the subband gain as $G_B = LN_B / LS_B$ is not what we need. Our ultimate interest is to know how to equalize the speaker and noise Sound Pressure Levels (SPLs) at a certain point of interest. This requires multiplying the above ratio by an equalization factor

$$G_B = \frac{LN_B}{LS_B} \cdot Q_B \qquad (1)$$

Equalization factors Q_B depend on the analog gains inside the system, the speaker and microphone frequency responses as well as the distances between the point of interest, reference microphone, speaker and the noise source. As it will be shown below, in many practical cases factors Q_B may be computed in a simple calibration procedure performed once in the system lifetime.

3. PRACTICAL CONSIDERATIONS

3.1. Estimation of equalization factors Q_B

The concept of estimating the equalization factors Q_B is explained bellow and illustrated by Figure 2. The point where the speaker and noise SPLs are of interest is, naturally, user's ear. For simplicity, the subband indices "B" are not shown on Figure 2 but used in the following equations.

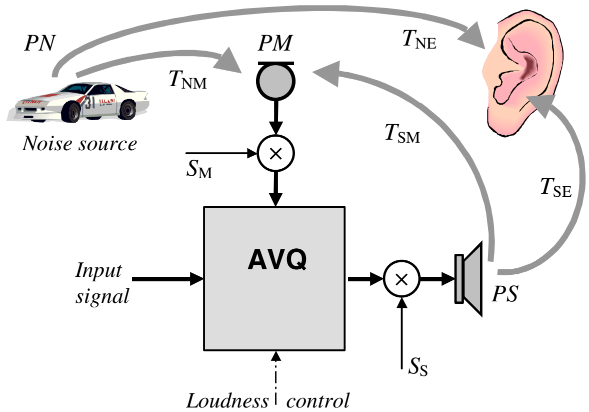

Figure 2 AVQ at user's ear

The subband SPL level at the speaker output PS_B may be represented as the speaker subband signal level LS_B multiplied by the speaker channel sensitivity $S_{S,B}$ in the corresponding frequency subband. In its turn the subband noise SPL at the microphone PM_B may be computed as the noise signal level LN_B divided by the microphone sensitivity $S_{M,B}$ in the corresponding frequency subband.

$$PS_B = LS_B \cdot S_S \qquad PM_B = LN_B / S_M \tag{2}$$

The subband speaker SPL at user's ear is computed as SPL at the speaker multiplied by the subband transfer coefficient between the speaker and the ear $T_{SE,B}$. Similar, the subband noise SPL at user's ear is computed as the noise SPL at the noise source multiplied by the subband transfer coefficient $T_{NE,B}$. In its turn, the subband noise SPL at the noise source may be estimated as the noise SPL at the microphone divided by the corresponding transfer coefficient between the noise source and the microphone $T_{NM,B}$. Since the subband gain shall equalize the noise and speaker sound pressure levels, the corresponding equation is

$$\frac{PM_B \cdot T_{NE,B}}{T_{NM,B}} = PS_B \cdot T_{SE,B}$$

Combining with Eq.2 we get

$$\frac{LN_B \cdot T_{NE,B}}{S_{M,B} \cdot T_{NM,B}} = G_B \cdot LS_B \cdot S_{S,B} \cdot T_{SE,B}$$

so that the required subband gain may be computed as

$$G_B = \frac{LN_B}{LS_B} \cdot \frac{T_{NE,B}}{T_{NM,B}} \cdot \frac{1}{S_{M,B} \cdot S_{S,B} \cdot T_{SE,B}} = \frac{LN_B}{LS_B} \cdot Q_B$$

where factor Q_B (see also Eq.1) is given as

$$Q_B = \frac{T_{NE,B}}{T_{NM,B}} \cdot \frac{1}{S_{M,B} \cdot S_{S,B} \cdot T_{SE,B}}$$

The above expression contains many unknown coefficients that can only be measured experimentally and may change dynamically. Fortunately, the equation above may be significantly simplified with the assumption that the transfer coefficients between the noise source and the microphone $T_{NM,B}$ and the noise source and user's ear $T_{NE,B}$ are equal. This assumption is valid for distant or diffused noise sources and omnidirectional reference microphones. It is also valid in most practical cases (e.g. car hands-free or office speakerphones) where directional microphones may be used

$$Q_B = \frac{1}{S_{M,B} \cdot S_{S,B} \cdot T_{SE,B}}$$

If the analog channel gains do not change, the factors Q_B may be estimated during a simple calibration procedure. For this procedure to be valid, it is assumed that the transfer coefficients between the speaker and user's ear $T_{SE,B}$ and the speaker and the microphone $T_{SM,B}$ are similar. If a calibration signal with the subband signal level LCS_B produces the subband microphone signal level LCM_B then the subband equalization factor Q_B is computed as

$$Q_B = \frac{LCS_B}{LCM_B} \cdot \tag{3}$$

The simplest calibration signal that works well is a white noise. The assumption $T_{SE,B} = T_{SM,B}$ is valid in many practical situations. For example, for a hands-free system installed in a car the microphone is generally placed as close as possible to user's head while the speaker(s) are relatively far from user's head. If this is not the case, the reference (or equivalent) microphone may be temporarily placed close to user's ear or in another place where the transfer coefficients may be considered equal.

The calibration procedure is needed only when the system analog gains or transfer coefficients are unknown or change. After the calibration procedure is performed, the equalization coefficients Q_B may be stored in a non-volatile memory and loaded on the system power up. For example, in case of a hands-free car kit the calibration procedure is performed after the installation at a service center. This is necessary since the cars may differ in acoustics. In case of a mobile phone, the calibration is performed once for each phone model and the same equalization coefficients are hard-coded or stored during the manufacturing process.

3.2. Estimations of signal and noise levels LS_B, LN_B

The subband speaker signal level LS_B is an estimation of the subband speaker signal amplitude when the signal is active. For example, for a voice communication terminal "active" means when the far participant is talking. In a music player or car stereo it may be the subband short time average signal power or envelop.

The subband noise signal level LN_B is an estimation of the subband microphone signal amplitude when the signal contains noise only. Since the microphone signal may also contain voices of people using the device or being close to it as well as other non-stationary sounds, the noise level estimator must know how to differentiate between "noise only" and other time intervals. This may be achieved by methods used in stationary noise reduction algorithms for voice signals. These algorithms are capable of differentiating between stationary (noise) and non-stationary (speech) intervals in each frequency subband by exploiting specific properties of human speech. Commonly, subband noise levels are estimated in flat, stationary valleys between speech amplitude or energy peaks on the time axis. As such, for pure voice cases the system will not be confused neither by user's speech nor by speech in the speaker signal. However, music possesses a potential threat, as its spectral properties may be relatively stationary and thus similar to noise. If the reference microphone signal contains music that does not present in the speaker signal, the speaker signal will be eventually amplified to be above the music level. This is good as, for example, the user may be able to understand his phone conversation partner even there is some background music. However, if music is present in the speaker signal itself, part of it may be fed back to the microphone signal due to acoustic coupling between the speaker and the microphone. If this feedback signal is confused with ambient noise, the speaker signal will be amplified thus creating a positive feedback with the speaker signal amplifying itself. To cancel presence of speaker signal in the microphone signal and prevent the positive feedback, an acoustic echo canceller (AEC) block is used on te microphone signal before it is divided on the frequency subbands (see Figure 1).

4. AVQ TECHNOLOGY AS A PART OF ALANGO VOICE COMMUNICATION PACKAGE

Automatic Volume and Equalization control technology is a standalone technology that can be used in a variety of applications. There may be variations of the technology each tailored for a specific application. For example, music applications are more demanding compared to voice for such parameters as tolerable amount of non-linear distortions allowed as well as self-noise. MIPS and memory requirements are other constraints especially in battery-powered applications such as mobile phones, Bluetooth headsets and others.

A specific version of AVQ technology is integrated into Voice Communication Package (VCP) developed by Alango Ltd. VCP can be used in a variety of voice communication devices such as mobile phones, Bluetooth headsets, hands-free car kits, industrial intercom and others where the background noise level and spectrum may change dynamically. Besides AVQ the package integrates other technologies:

- Acoustic Echo Cancellation with Residual Echo Suppression (microphone channel);
- Automatic Gain Control and Speech Enhancement (microphone channel)
- Noise Reduction (both microphone and speaker channels);

VCP structure and signal flow is shown on Figure 3. The microphone signal and the processed speaker signals are divided on complex frequency subbands with downsampling by Analysis-M blocks. A modification of the method described in [2] is used. The number of subbands is configurable and can be 8, 16 or 32. The subband signals are then processed by Subband Adaptive Filtering (SAF) block that clear out the linear part of the acoustic feedback due to acoustic coupling between the speaker and the microphone. The non-linear part is blocked by Subband Residual Echo Suppressor (SRES) block. Microphone Subband Noise Reduction (SNR-M) block suppresses stationary noises in the microphone signal by dynamically reducing subband gains for non-active subbands. The echo and noise cleared microphone subband signals are combined into the full band by Synthesis-M block. The full band microphone signal is finally processed by Automatic Gain Control (AGC) and Speech Enhancement (SE) block before it enters the voice communication network processing depending on the application.

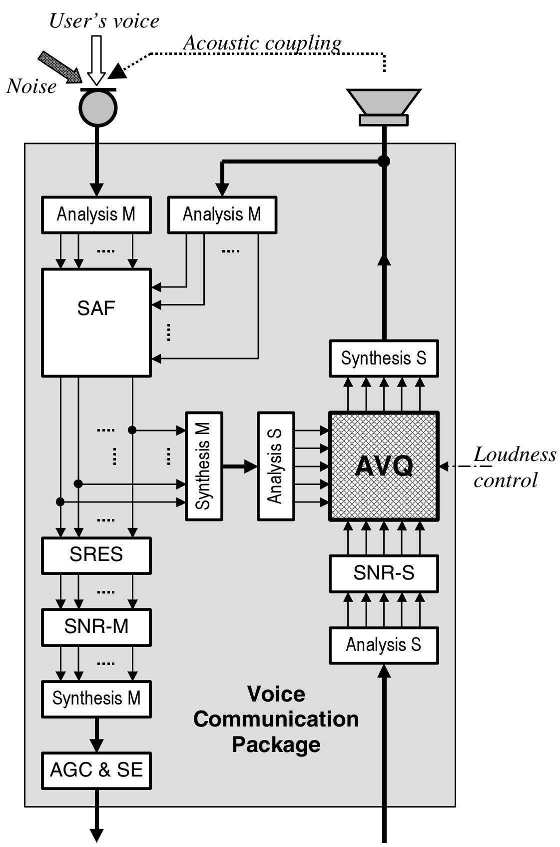

Figure 3 VCP structure and data flow (MIPS and data memory sensitive application)

The output subband signals from SAF block are also immediately recombined into the full band signal by another Synthesis-M block. This full band signal is again divided on frequency subband by Analysis-S block where the subbands constitute the reference input for Automatic Volume and Equalization control (AVQ) block. This block also takes subband signals from speaker channel Noise Reduction block (SNR-S) operating on the speaker channel subbands. In the current VCP version Analysis-S blocks are composed of five, simple band-pass IIR, phase-matching filters. No downsampling is made on the subband signals in the speaker channel. This reduces Synthesis-S blocks to simple summation.

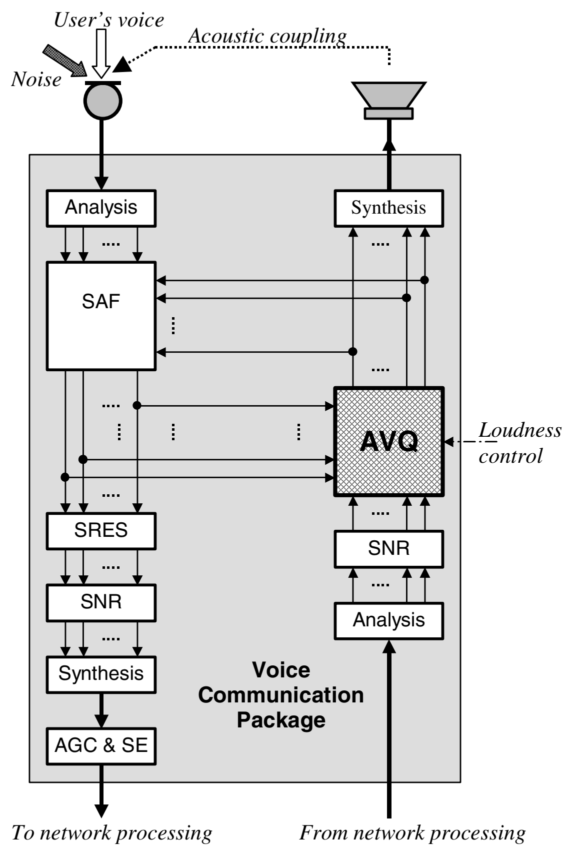

Figure 4 VCP structure and data flow (program memory sensitive application)

The structure on Figure 3 shows that the output of SAF block is combined into the full band signal and then immediately divided on the subbands again. The same is true for the SAF block reference input which is the output of AVQ block. Such synthesis/analysis scheme allows using different subband schemes in the microphone and speaker channels. The structure is chosen for optimization purposes in very MIPS/memory sensitive applications.

Adaptive filtering performed on the microphone channel is a “heavy” computational task that is greatly simplified by using subband scheme [3]. The larger number of bands is used, the more computational saving is. As such, options for 8, 16 or 32 bands are implemented. From the other hand, processing on the speaker channel is often considered as optional and must be simplified as much as possible. Five-band scheme without downsampling provides the necessary tradeoff between the quality and computational resources required.

If program memory is a concern, than the structure shown on Figure 4 has an advantage. In this structure the same frequency decomposition scheme is used on both microphone and speaker channels. This allows “reuse” of Analysis, Synthesis and SNR blocks on the speaker channel thus saving a significant part of the program memory. Beside, this scheme has an advantage of better frequency resolution for the speaker SNR block providing better noise suppression quality.

5. REFERENCES

[1] Alexander Goldin, “Sound Equalization in Noisy Environments”, presented at the 110th Convention of Audio Engineering Society (Amsterdam, 2001).

[2] P. P. Vaidyanathan, *Multirate Systems and Filter Banks*. Englewood Cliffs, NJ: Prentice-Hall, 1993

[3] Alexander Goldin, Alexey Budkin “Challenges of Acoustic Echo Cancellation in low cost applications”, presented at the 118th Convention of Audio Engineering Society (Barcelona, 2005).