A Combined Implementation of Echo Suppression, Noise Reduction and Comfort Noise in a Speaker Phone Application

Christian Schüldt[a], *Fredric Lindstrom*[a], *Ingvar Claesson*[b]

[a]Konftel AB, Research and Development, Box 268, SE-90106, Umeå, Sweden, {csc,fli}@konftel.com
[b]Blekinge Institute of Technology, Department of Signal Processing, SE-37225, Ronneby, Sweden

Abstract- **Echo suppression, noise reduction and comfort noise are desirable features in loudspeaker phone products. This paper proposes a set of algorithms for a combined, subband based, implementation of these three processing blocks. The proposed algorithms are verified by evaluation of a fix-point real-time implementation.**

I. INTRODUCTION

In a speaker phone, an echo is generated as the loudspeaker signal passes the loudspeaker-enclosure-microphone (LEM) system. This echo can be removed using echo suppression or e.g. a combination of echo suppression and beamforming/acoustic echo cancellation (AEC) methods [1]. Additionally, noise reduction can be used to reduce the disturbance from background noise. Several echo suppression and noise reduction algorithms, as well as joint processing approaches, have been proposed [2]. Since the echo suppression will modulate the background noise, comfort noise injection is also desirable.

The contributions of this paper include concurrent use of an estimated noise floor parameter in the three signal processing blocks, a version of the minimum statistics method for noise floor estimation, a partion of the echo suppression gain in order to reduce complexity and signal delay, and an estimation method for residual echo based on acoustic coupling.

II. PROPOSED ALGORITHM

A. Filterbank

In this paper, I finite impulse response (FIR) filters, $\mathbf{h}_i = [h_{0,i}, \cdots, h_{N-1,i}]^T$, all of length N are used to construct a uniform analysis filter bank, where i denotes subband index, see figure 1. The filterbank is used to divide the microphone signal $m(k)$ into i suband signals $m_i(k) = e_i(k) + n_i(k) + s_i(k)$, where $e_i(k)$, $n_i(k)$, and $s_i(k)$ are subband signals of echo, background noise and near-end speech, respectively. No downsampling is used, eliminating the need of synthesis filters. This maintains the low signal delay.

Thanks to the Swedish Knowledge Foundation (KKS) for funding.

B. Echo Suppression

The total echo suppression is composed of a full-band gain $g_{\mathrm{es},f}(k)$ and gains $g_{\mathrm{es},i}(k)$, operating the in individual frequency bands, see figure 1. It is assumed that the echo is not disturbing if the power of the echo signal is lower or at the same level as the power of the noise floor in the corresponding frequency band, i.e. if

$$g_{\mathrm{es},f}(k) g_{\mathrm{es},i}(k) \leq C_{\mathrm{es}} \frac{\hat{P}_{n_i}(k)}{\hat{P}_{e_i}(k)}, \tag{1}$$

where C_{es} is a constant and $\hat{P}_{n_i}(k)$ and $\hat{P}_{e_i}(k)$ are estimations of the noise floor and the squared echo powers, respectively, see section III. The gains are distributed so that the gain in every band is limited by a lowest level G_{es}, where $g_{\mathrm{es},i}(k) > G_{\mathrm{es}}$. This limit implies that the stopband requirements in the filterbank filters are relaxed, and the filter order can be reduced, maintaining low complexity and low signal delay.

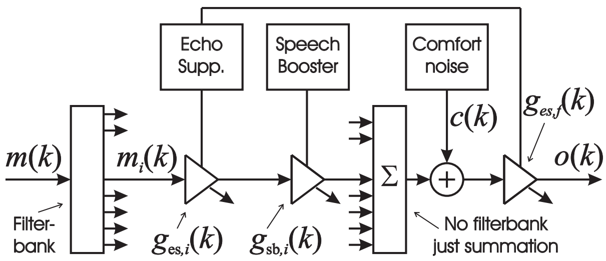

Fig. 1. A simplified schematic of the arrangement.

C. Noise Reduction

In this paper a speech booster approach for noise reduction is used [3]. The speech boosting gain $g_{\mathrm{sb},i}(k)$ is given by

$$g_{\mathrm{sb},i}(k) = \begin{cases} G_{\mathrm{sb}} & \text{if } \frac{\overline{m}_i^2(k)}{\hat{P}_{n_i}(k)} < C_{\mathrm{sb}} \\ 1 & \text{if } \frac{\overline{m}_i^2(k)}{\hat{P}_{n_i}(k)} > \frac{C_{\mathrm{sb}}}{G_{\mathrm{sb}}} \\ \frac{G_{\mathrm{sb}}}{C_{\mathrm{sb}}} \frac{\overline{m}_i^2(k)}{\hat{P}_{n_i}(k)} & \text{otherwise} \end{cases} \tag{2}$$

where $\overline{m}_i^2(k) = (1-\gamma_m)\overline{m}_i^2(k-1) + m_i^2(k)$ and $G_{sb} < 1$ and C_{sb} are constants. This assures that the gain is kept low

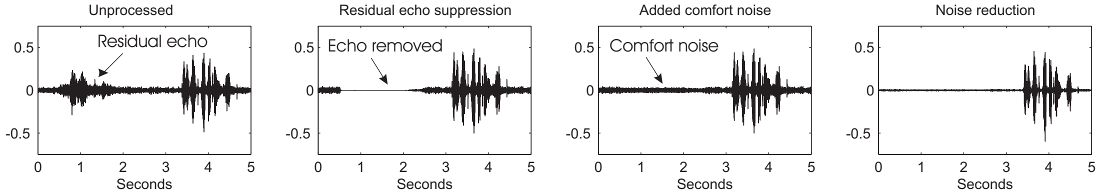

Fig. 2. The output signal $o(k)$ for successive onset of processing blocks.

when the short-time signal power $\overline{m}_i^2(k)$ is close to the noise floor power and increased up to a limit of 0dB for higher values of $\overline{m}_i^2(k)$. If echo suppression is active the signal should not be boosted. Thus, in that case $g_{\mathrm{sb},i}(k) = G_{sb}$.

D. Comfort Noise

A "white" pseudo noise signal $w(k)$ is first generated using linear recursive sequences. The comfort noise $c(k)$ is then obtained through

$$c(k) = \sum_{i=0}^{I-1}(1 - g_{\mathrm{es},f}(k)g_{\mathrm{es},i}(k))G_{\mathrm{sb}}\hat{P}_{n_i}^{\frac{1}{2}}(k)\mathbf{w}(k)^T\mathbf{h}_i, \quad (3)$$

where $\mathbf{w}(k) = [w(k), \cdots, w(k-N+1)]^T$, i.e. $c(k)$ is similar to the noise that was removed by the echo suppression.

III. PARAMETER ESTIMATION METHOD

A. Estimating the Echo

The echo is estimated from the loudspeaker signal $l(k)$, according to

$$\hat{P}_{e_i}(k) = \begin{cases} \beta(k)l_i^2(k) & \text{if } l_i^2(k) \geq \hat{P}_{e_i}(k) \\ (1-\gamma_s)\hat{P}_{e_i}(k-1) + \gamma_s\beta(k)l_i^2(k) & \text{otherwise,} \end{cases} \quad (4)$$

where $l_i(k)$ is the i:th subband signal of the loudspeaker signal $l(k)$, $\beta(k)$ is the estimated coupling factor [2], and γ_s is a "slow decrease" averaging constant. The signals $l_i(k)$ can be obtained by e.g. using the filters $\mathbf{h}_i$ or directly from a subband AEC. The averaging factor γ_s is used to model the character of a typical LEM, i.e. the remaining echo after the offset of the loudspeaker signal.

B. Estimating the Noise Floor

A block-processing minimum statistics method is used to estimate the noise floor power, in order to reduce computational complexity. For every T sample (when $k = jT$ for $j = 1, 2, \cdots$), the block power

$$P_{m_i}(j) = \frac{1}{T}\sum_{t=0}^{T-1} m_i^2(jT - t), \quad (5)$$

is calculated. The difference of the parameters $P_{\max}(j)$ and $P_{\min}(j)$, given by

$$P_{\max}(j) = \max\{P_{m_i}(j), \cdots, P_{m_i}(j - L + 1)\} \quad (6)$$

$$P_{\min}(j) = \min\{P_{m_i}(j), \cdots, P_{m_i}(j - L + 1)\}, \quad (7)$$

are then compared with a constant C_n

$$P_{\max}(j) - P_{\min}(j) \leq C_n, \quad (8)$$

and if the condition in equation (8) is satisfied, the background noise floor power estimate is updated according to

$$\hat{P}_{n_i}(k) = \begin{cases} (1-\gamma_n)\hat{P}_{n_i}(k-1) + \gamma_n P_{\min}(j) & \text{if (8) is true} \\ \hat{P}_{n_i}(k-1) & \text{otherwise,} \end{cases} \quad (9)$$

where γ_n is an averaging constant.

IV. REAL-TIME IMPLEMENTATION

The proposed algorithms were implemented in real time on a fix-point digital signal processor [4]. The sampling rate was 8khz and the parameter settings were $I = 32$, $N = 48$, $C_{es} = 2$, $G_{es} = -18$dB, $C_{sb} = 2$, $G_{sb} = -10$dB, $\gamma_s = 0.08$, $T = 256$, $L = 8$, $C_n = 0.08$, and $\gamma_n = 0.016$. The setup was done in a small office with a distance between the microphone and loudspeaker of approximately 40cm. Shown in the first plot from the right in figure 2 is the unprocessed output signal $o(k)$ consisting of residual echo followed by a burst of near end speech. The second plot shows $o(k)$ after echo suppression is turned on, the third when comfort noise is added, and the forth when all three processing are active. The proposed algorithms were also implemented as post processing unit together with an AEC in a commercial conference phone product. Subjective tests showed that comfortable, noise reduced, full duplex operation could be achieved without introducing audible artifacts.

V. CONCLUSION

This paper presented a combined implementation of echo suppression, noise reduction and comfort noise for a speaker phone implementation. The main contributions of the paper are the combined use of the estimated noise floor in all three processing blocks and the partition of the echo suppression gain in order to reduce filter bank complexity. Finally, the proposed algorithms were verified in a real-time fix-point implementation.

VI. REFERENCES

[1] S. Haykin, *Adaptive Filter Theory*, 4th ed., Prentice, 2002.

[2] E. Hänsler and G. Schmidt, *Acoustic Echo and Noise Control: A Practical Approach*, Wiley, 2004.

[3] N. Westerlund, M. Dahl, I. Claesson, "Speech Enhancement for personal communication using an adaptive gain equalizer", *Signal Processing*, vol. 85, pp. 1089-1101, 2005.

[4] *ADSP-BF533 hardware reference*, Analog Devices, 2005.