A Mixed Analog - Digital Hybrid for Speech Enhancement Purposes

Benny Sällberg, Henrik Åkesson, Mattias Dahl and Ingvar Claesson

Department of Signal Processing

Blekinge Institute of Technology

BOX 520, SE-372 25 Ronneby, Sweden

E-mail: bsa@bth.se

Abstract—This paper presents and evaluates a hybrid implementation of a low complexity algorithm for speech enhancement, the Adaptive Gain Equalizer (AGE). The AGE is a subband based time domain method for instantaneous boosting of speech. By combination of digital analysis and analog synthesis, main advantages of the digital domain and the analog domain are utilized. The hybrid solution is implemented on a printed circuit board and evaluated in real-time. The development time of the proposed solution was very short, the solution is flexible, robust, has high signal bandwidth in the signal chain and does not require a Voice Activity Detector (VAD). Furthermore, the solution is not restricted by quantization errors in the signal chain and does not require a high speed Digital Signal Processor (DSP) for analysis. Informal listening tests and Signal-to-Noise Ratio (SNR) measures verify excellent speech enhancement performance and quality.

I. INTRODUCTION

Many speech enhancement implementations of today are either digital or analog. Digital solutions are often superior in time-to-market, price per unit, structured and powerful development tools, flexibility, high degree of reconfiguration, robustness, the ability to use a Digital Signal Processor (DSP) for many tasks and the possibility to handle high complexity algorithms. Though many advantages, digital solutions might suffer from limitation in signal bandwidth, limited number of operations per second, and quantization errors. The drawbacks of digital solutions could be minimized by using high speed DSPs and longer word length. However, such preventive measures are likely to increase the total power consumption as well as the total price per unit. An algorithm intended for digital implementation is the famous spectral subtraction [1], [2]. Analog solutions often benefit from high signal bandwidth, continuous time signal processing, no quantization of data, and lower power consumption as opposed to corresponding DSP based solutions. On the contrary, analog solutions might require expensive simulation and design software and suffer from long time-to-market. Moreover, since analog solutions tend to be static, reconfiguration of analog solutions constitutes a troublesome task. An example of an analog microphone front-end for speech enhancement can be found in [3].

Many speech enhancement algorithms require so called Voice Activity Detectors (VAD) for identification of speech activity. The speech activity detection in turn controls the activity of the speech enhancement algorithm. The structure

of a VAD is often complex and solutions based on a VAD tend to be less robust compared to algorithms not requiring a VAD. Since speech enhancement algorithms are often applied in hand held battery powered applications, e.g. microphone front-ends, it is of highest importance to optimize the power consumption for battery life time. Speech enhancement algorithms should be flexible, versatile and adjustable to different scenarios. Furthermore, the algorithms should be adaptive, robust and of low complexity with a high level of speech enhancement quality and performance.

The proposed solution combines digital analysis and analog synthesis in a hybrid design. Thus, utilizing advantages of the two domains; the digital domain and the analog domain. The selected algorithm for implementation is the Adaptive Gain Equalizer (AGE) [4]. The AGE is a subband based time domain method for speech enhancement purposes. The algorithm is adaptive and does not require a VAD. Subband specific gain functions are employed for instantaneous *boosting of speech*. Each subband specific gain function constitutes a quotient of a short term average and a noise floor level estimate. The noise floor level estimate should be set to track slow changes in the background noise and the short term average should track the bursts of speech. A discussion on where to draw the analog and digital boundary in a hybrid design is given in [5]. However, the straightforward underlying idea of the AGE and its low complexity draws a natural analog/digital boundary by combination of digital analysis and analog synthesis. The proposed solution is implemented on a Printed Circuit Board (PCB) and evaluated in real time. The AGE has been successfully implemented in MATLAB¹ [6] and on analog circuitry [7].

The outline of this paper is as follows: In Section II the implemented AGE algorithm is defined. The implementation is presented in Section III. In Section IV the implementation is evaluated and results are presented in Section V. Finally, Section VI concludes this paper.

II. THE ADAPTIVE GAIN EQUALIZER

The AGE constitutes a filter bank for subband division of the input signal and an AGE kernel per subband, see Fig. 1. An AGE kernel is an operational block for calculating the

¹MATLAB is a trademark of the Mathworks, Inc.

Fig. 1. A K-band Adaptive Gain Equalizer (AGE) implementation. Each KERNEL_k block calculates the corresponding subband specific gain function, $G_k(t)$.

Fig. 2. Adaptive Gain Equalizer (AGE) KERNEL_k block for subband k .

subband specific gain function from the input subband signal, see Fig. 2. Assume having an input signal, $x(t)$, constituting of a speech signal, $s(t)$, with additive band limited noise, $v(t)$, according to $x(t) = s(t) + v(t)$. A division of the input signal gives a subband signal, $x_k(t)$, according to

$$x_k(t) = h_k(t) * x(t), \quad (1)$$

where $h_k(t)$ is the subband selecting filter, the convolution is denoted by $*$, and $k \in [0, K-1]$ is the subband index. The AGE relies upon the assumption that the time of stationarity of the noise is higher than the time of stationarity of the speech. Hence, two estimates are used in each subband: A Short Term Average (STA), $A_k(t)$, and a Noise Floor Level Estimate (NFLE), $\underline{A}_k(t)$. The estimates are applied to the full-wave rectified subband signal, i.e. $|x_k(t)|$, as integrators. The STA integration time, T_k , should be set close to the speech pseudo-stationarity time [2], i.e. 20–30 ms. The integration time of the NFLE, $\underline{T}_k$, should be set to track slow changes in the background noise, likely to be in the order of seconds instead of milliseconds. The STA and NFLE integrators could be expressed according to

$$\text{STA: } A_k(t) = (1 - \alpha_k)A_k(t - \Delta_t) + \alpha_k |x_k(t)| \quad (2)$$

$$\text{Prototype: } \underline{B}_k(t) = (1 - \beta_k)\underline{B}_k(t - \Delta_t) + \beta_k |x_k(t)| \quad (3)$$

$$\text{NFLE: } \underline{A}_k(t) = \begin{cases} \underline{B}_k(t), & \text{if } \underline{B}_k(t) < A_k(t) \\ A_k(t), & \text{if } \underline{B}_k(t) \geq A_k(t) \end{cases} \quad (4)$$

where α_k and β_k are parameters controlling the STA and NFLE integration times, the prototype variable $\underline{B}_k(t)$ is for temporary use, and Δ_t is a small time delay. Equation (4)

ensures that the NFLE never exceeds the STA. This limiting function is denoted the lower gain limit. The subband specific gain function, $G_k(t)$, is constructed by the quotient of the two estimates, STA and NFLE. To inhibit having too high gain function level, an upper limiter is imposed according to

$$G_k(t) = \begin{cases} A_k(t)/\underline{A}_k(t), & \text{if } A_k(t)/\underline{A}_k(t) < L_k \\ L_k, & \text{if } A_k(t)/\underline{A}_k(t) \geq L_k \end{cases} \quad (5)$$

where L_k denotes the upper gain limit for subband k . By imposing an upper and lower limit the gain function is limited to the range $1 \leq G_k(t) \leq L_k$. Having too high upper gain limit could result in distorted speech. Too low upper gain limit on the other hand, could result in weak speech enhancement performance. Empirical tests indicate a value of about 3–10 for the upper gain limit, L_k , to avoid unnaturally sounding enhanced speech. The output signal, $y(t)$, is constructed by summation of all subband input signals weighted by the corresponding gain functions, according to

$$y(t) = \sum_{k=0}^{K-1} G_k(t)x_k(t). \quad (6)$$

By sampling the above mentioned AGE structure, a corresponding digital implementation is gained, i.e. by letting $t = t_n = n/F_s$, where the sample index is denoted n and F_s is the sampling frequency (in [Hz]). The general formulation of the AGE has subband specific parameters. In [8] an analysis of subband dependent AGE-parameters is presented.

III. IMPLEMENTATION

The aim for the implementation is to utilize advantages from analog and digital solutions such as high signal bandwidth, no quantization of data, reconfigurability. Thus, the implementation is split into two parts; digital analysis and analog synthesis. A mapping function, $f_k\{\cdot\}$, maps the digital analysis gain function, $G_k(t_n)$, to a corresponding analog synthesis gain function, $r_k(t_n)$, according to $r_k(t_n) = f_k\{G_k(t_n)\}$. The structure of the mapping function depends on implementation specific parameters such as the maximal gain function value, L_k . The analog synthesis constitutes an analog signal chain of the AGE algorithm from input, $x(t)$, to output, $y(t)$. In Fig. 3 the proposed hybrid solution is illustrated.

A. Digital Analysis

The digital analysis is performed on a Texas Instrument Mixed Signal Processor (MSP) MSP430F149. A filter bank is implemented in the MSP for approximation of the analog filter bank in the synthesis (see Section III-B). Two-poles two-zeroes infinite impulse response filter sections form the digital filter bank. The digital subband selective filters, $\hat{h}_k(t_n)$, should be designed such that they match the corresponding analog subband selective filters, $h_k(t)$. All parts except for the actual summation of the output signal are implemented in the MSP, i.e. full wave rectifying, STA integrator, NFLE integrator, lower- and upper gain limiting, and gain function calculus. The integration time controlling constants are set to $\alpha_k = 1/(F_s T_k)$ and $\beta_k = 1/(F_s \underline{T}_k)$.

Fig. 3. Splitting of the Adaptive Gain Equalizer (AGE) into digital analysis and analog synthesis. An anti-alias low pass filter is situated prior to the Analog-to-Digital Converter (ADC).

Since the MSP is a low speed micro controller, a two-level priority scheme is required to ensure full functionality of the processor: A high priority stream and a low priority stream. The high priority stream is running at full sample rate, i.e. F_s . The operations subband filtering, full wave rectifying, STA estimation, and gain steering via a control interface are run with high priority. The low priority stream is using a Round-Robin time sharing algorithm where one subband is managed at a time. Calculation of the subband specific NFLE, upper- and lower gain limiting, and gain function calculus are time shared with low priority.

B. Analog Synthesis

The Mitsubishi Electric M5289P integrated circuit is used as an analog synthesis filter bank. The M5289P is an analog electronics Hi-Fi 7-element graphic equalizer and employs seven potentiometers for gain control in each subband. The gain of each subband of the M5289P can be controlled in a span of $10^{-13/20}$ to $10^{+13/20}$ by altering the value of the subband specific potentiometer. Additional capacitors and resistors are used in a subband filter selection network.

C. Control Interface

The digital potentiometer, X9C104P, from Xicor is used for individual control of the M5289P subband specific gains. The X9C104P has a resolution of one hundred steps spanning 100 k Ω and is controlled by the MSP via a three wire digital interface: $\overline{U/D}$ - Up/Down, $\overline{INC}$ - Increment, and $\overline{CS}$ - Chip Select. The analog synthesis gain function, $r_k(t_n)$, corresponds to a potentiometer value and is mapped from the digital analysis gain function, $G_k(t_n)$, by a mapping function, $f_k\{\cdot\}$.

D. Algorithm Parameter Selection

The implementation is optimized for band limited noise having 4 kHz bandwidth. Thus the sampling frequency of the digital analysis is set to $F_s = 8$ kHz. The implementation is using AGE parameter settings according to Table I. The parameters in Table I are determined by empirical research

TABLE I
DEFAULT ADAPTIVE GAIN EQUALIZER (AGE) PARAMETER SETTINGS.

Parameter	Default value
F_s	8000 Hz
T_k	30 ms, $\forall k$
$\underline{T}_k$	3 s, $\forall k$
L_k	13 dB = $10^{+13/20}$, $\forall k$
α_k	$1/(F_s T_k) = 1/240$, $\forall k$
β_k	$1/(F_s \underline{T}_k) = 1/24000$, $\forall k$

in combination with subjective tests, and have been shown suitable for many speech and noise scenarios.

IV. EVALUATION

The evaluation aims at comparing the proposed hybrid implementation to a corresponding digital MATLAB implementation. Transfer function measures show that the filter bank in the digital implementation matches the filter bank in the proposed hybrid implementation. Furthermore, the time constants of the STA and NFLE are in the same order of magnitude in both the hybrid implementation and the digital MATLAB implementation. A maximal gain of $10^{+13/20}$ is used in the digital MATLAB implementation, i.e. corresponding to 13 dB maximal gain in the proposed hybrid solution.

The performance and quality of a speech enhancement algorithm is not easily quantified. Several objective- and subjective tests exist today as presented in [9]. Examples of objective tests are: Itakura-Saito (IS) Distortion measure, Log-Likelihood Ratio (LLR) measure, Log-Area-Ratio (LAR) measure, and Segmental Signal-to-Noise Ratio (SNR) measure. Examples of subjective tests are: Modified Rhyme Test (MRT), Diagnostic Rhyme Test (DRT), and Mean Opinion Score (MOS). An objective test is selected for evaluation of the proposed hybrid speech enhancement implementation; an SNR measure. The SNR measure, $\Gamma_x(t_n)$, of a signal, $x(t_n)$, is herein defined by

$$\Gamma_x(t_n) = (1 - \gamma)\Gamma_x(t_n - 1/F_s) + \gamma |x(t_n)|^2, \quad (7)$$

where $\gamma = 1/(F_s T_\gamma)$ is a constant controlling the integration time, T_γ (in [s]), of the SNR measure. For comparing the signal before, $x(t_n)$, and after speech enhancement, $y(t_n)$, a SNR measure quotient is used, defined as $\Gamma_y(t_n)/\Gamma_x(t_n)$.

V. RESULTS

In Fig. 4 the speech enhancement performance of the proposed method is illustrated. The performance evaluation shows a speech enhancement of maximum 13 dB speech gain, confirmed by the SNR measure comparison in Fig. 5. A speech enhancement performance comparison of the proposed method and a corresponding digital implementation is illustrated in Fig. 6. The performance of the proposed method and the digital implementation are remarkably equal. Subjective listening

Fig. 4. Signal-to-Noise Ratio (SNR) measures. Unprocessed speech signal (solid), speech signal processed by the proposed hybrid implementation (dashed), where $T_\gamma = 25$ ms.

tests prove the performance and quality of the proposed speech enhancement implementation.

VI. CONCLUSIONS

Digital implementations for speech enhancement exhibit advantages in time-to-market, price per unit, flexibility, robustness, the ability to use the DSP for many tasks, and the possibility to handle high complexity algorithms. However, digital solutions tend to suffer from limited signal bandwidth, number of operations per second, and quantization errors. Analog solutions might benefit from high signal bandwidth, no quantization of data, and continuous time signal processing. However, analog solutions tend to be restricted in long time-to-market, static solutions, and requiring expensive simulation and design software. The proposed solution is a hybrid solution split into digital analysis and analog synthesis. The main advantages of the digital domain and the analog domain are utilized by the mixed domain solution. The development time of the proposed solution was very short, the solution is flexible, robust, has high signal bandwidth in the signal chain and does not require a VAD. Furthermore, the solution is not restricted by quantization errors in the signal chain and does not require a high speed DSP for analysis. Reproduction of the proposed method is probably a straightforward task due to the simplicity and robustness of the AGE algorithm. The mixed analog-digital solution is implemented on a printed circuit board and evaluated in real-time. The simplicity and robustness of the method allows for VLSI implementation. Informal listening tests in combination with SNR measures prove the proposed solution's excellent speech enhancement performance.

REFERENCES

- [1] S. F. Boll, "Suppression of acoustic noise in speech using spectral subtraction," *IEEE Trans. Acoust. Speech and Sig. Proc.*, vol. ASSP-27, pp. 113–120, April 1979.
- [2] J. R. D. Jr., J. G. Proakis, and J. H. L. Hansen, *Discrete time processing of speech signals*. Macmillan Publishing Company, 1993.
- [3] H. Yoo, R. Ellis, D. Anderson, P. Hasler, D. Graham, and M. Hans, "A continuous-time speech enhancement front-end for microphone inputs," Mobile and Media Systems Laboratory, HP Laboratories Palo Alto, HPL-2002-311, Technical Report, November 7th 2002.

Fig. 5. Quotient of speech signal SNR measures before- and after speech enhancement by the proposed method, where $T_\gamma = 25$ ms.

Fig. 6. Signal-to-Noise Ratio (SNR) measures. Speech signal processed by the proposed hybrid implementation (solid), Speech signal processed by a corresponding MATLAB implementation (dashed), where $T_\gamma = 25$ ms.

- [4] N. Westerlund, M. Dahl, and I. Claesson, "Speech enhancement using an adaptive gain equalizer," *DSPCS, Coolangatta, Australia*, September 2003.
- [5] P. Hasler and D. Andersson, "Cooperative analog-digital signal processing," *ISCAS*, vol. IV, pp. 3972–3975, 2002.
- [6] N. Westerlund, M. Dahl, and I. Claesson, "Real-time implementation of an adaptive gain equalizer for speech enhancement purposes," *WSEAS, Singapore*, September 2003.
- [7] B. Sällberg, H. Åkesson, N. Westerlund, M. Dahl, and I. Claesson, "Analog circuit implementation for speech enhancement purposes," *38th Asilomar Conference on Circuits, Systems and Computers*, Nov 2004.
- [8] N. Westerlund, M. Dahl, and I. Claesson, "Speech enhancement using an adaptive gain equalizer with frequency dependent parameter settings," *VTC04, Los Angeles, USA*, September 2004.
- [9] J. Hansen and B. Pellom, "An effective quality evaluation protocol for speech enhancement algorithms," *ICSLP, Sydney, Australia*, pp. 2819–2822, Dec 1998.