Hands-Free Audio and its Application to Telecommunication Terminals

Martin Schönle[1], Christophe Beaugeant[1], Kai Steinert[1], Heinrich W. Löllmann[2], Bastian Sauert[2], and Peter Vary[2]

[1]*Siemens AG, Corporate Technology, Munich, Germany*

[2]*Institute of Communication Systems and Data Processing, RWTH Aachen University, Aachen, Germany*

Correspondence should be addressed to Martin Schönle (`martin.schoenle@siemens.com`)

ABSTRACT

In this paper we focus on the enhancement of speech quality by hands-free audio systems in modern, multimedia enabled telecommunication terminals. Requirements like the compatibility with wideband audio and full-duplex hands-free operation demand for sophisticated system designs. As a starting point we introduce a state-of-the-art system developed for mobile phones, considering practical aspects of algorithm design. Algorithmic extensions to this solution are presented, supporting speech enhancement for the user of the terminal or reducing the round-trip delay. In addition, more advanced concepts are discussed. They offer the capability of joint design for the most important algorithms and provide the possibility to exploit psychoacoustic properties of the human ear.

1. INTRODUCTION

Along with the tremendous success of mobile telephony during the last decade the request for convenient, easy-to-use, and multimedia enabled terminals has gained high importance. The original application of pure circuit-switched speech telephony is extended by multimedia services. For many of these applications, like video telephony or teleconferencing, hands-free operation of the terminals is essential. As a result sophisticated audio and acoustic systems are needed not only to enable an almost natural conversation between the user of the terminal and the communication partner at the far-end, but also to support features like speech recognition or man-to-machine dialog systems.

Within the scope of this paper are the algorithms required for the enhancement of speech quality. Our system, designed for telecommunication terminals, is based on a state-of-the-art solution, dealing with transducer equalization, acoustic echo cancellation, and noise reduction (Section 2). Complementary, more innovative algorithms like downlink speech intelligibility enhancement or low-delay synthesis are added to the core functions (Section 3). Furthermore system concepts beyond state-of-the art are discussed (Section 4) in order to point out the way to future hands-free solutions which can cope with upcoming requirements like compatibility for wideband audio or full-duplex capability. It is shown how the trade-off between sufficient echo suppression and good double-talk properties can be resolved by system designs in the frequency domain. The basic algorithmic relationships can be transferred to the time-frequency domain, approximating the psychoacoustic properties of the human ear by non-uniform low-delay filter bank systems.

2. STATE-OF-THE-ART HANDS-FREE SYSTEM

Figure 1 shows a state-of-the-art hands-free system as it is used in many mobile, cordless, and fixed-network terminals. The representation of the algorithms is only schematic, more details can be found in the literature. Via the signals $s_{UL}(n)$ and $s_{DL}(n)$ the hands-free system interfaces to the speech codecs. The signal direction from the microphone to the speech encoder and via some network to the far-end user is called uplink (UL) throughout the paper, the opposite direction from the far-end user via network and speech decoder to the near-end loudspeaker is called downlink (DL).

2.1. Interface to Audio Front-End

Due to miniaturization and mass-production the acoustic front-end of modern telecommunication terminals is often sub-optimal and does not conform to the requirements of telecommunication standards or to the user's demand regarding audio quality. As a result, a panel of basic algorithms is required to optimize the properties of

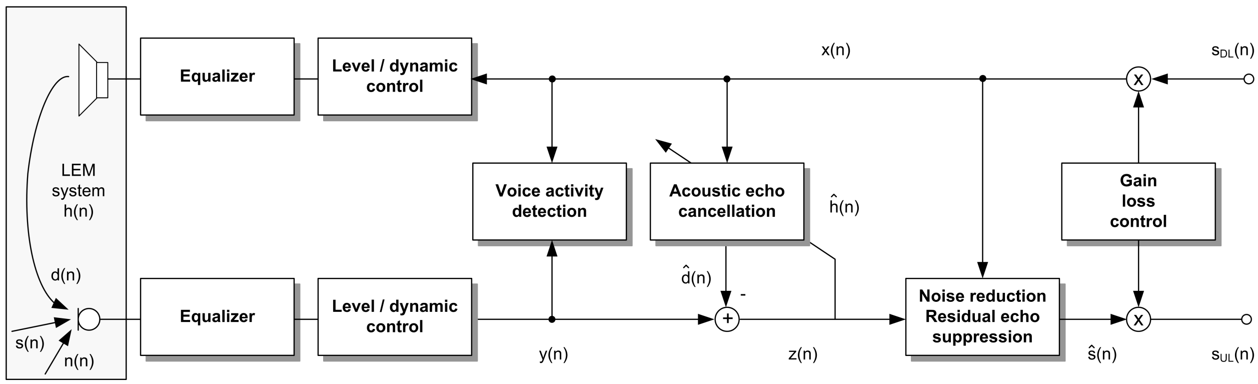

Fig. 1: State-of-the-Art Hands-Free System

the acoustic front-end. Equalizer filters in UL direction are typically used to compensate non-ideal microphone frequency responses and to meet the requirements defined by the telecommunication standards. In DL direction the frequency range of the loudspeaker is compensated, either by attenuating frequencies that could damage the device or by intentionally amplifying resonance frequencies to gain more sound pressure level.

Another disturbing effect for terminals used in outdoor environment is the pick-up of wind noise by the microphone. As this kind of noise is typically a low frequency signal, it can be detected by observing the energy below 400 Hz. The amount of energy is then used to control the edge frequency of a high-pass filter which can adaptively be inserted into the signal path.

Sound quality is also influenced by the distance between transducers and users. Especially in hands-free operation the distance between microphone and near-end user cannot be kept small and constant. Non-linear algorithms are used to provide constant mean signal levels to the hands-free algorithms and to limit the dynamic range. A simple example is

$$\begin{aligned} s_{out}(n) &= s_{in}(n)\,\frac{\alpha}{\alpha+\bar{s}(n)}, \qquad (1)\\ \bar{s}(n) &= \bar{s}(n-1)(1-\beta)+|s_{in}(n)|\,\beta, \end{aligned}$$

which provides a normalization to the smoothed magnitude of the input signal $s_{in}(n)$, with $0<\alpha\leq 1$ being an attenuation constant and $0<\beta\leq 1$ a smoothing constant.

In DL direction, enhancement of the audio quality is performed by a more sophisticated dynamic compression and limitation (DCL) algorithm. By amplification of medium signal levels speech intelligibility is improved. Music signals sound more ‘direct’ and ‘present’ by the reduction of the dynamic range and limitation of the signal to a maximum level protects the loudspeaker from mechanical clipping.

When such non-linear algorithms are used within a linear signal processing system, the correct place for their integration is not trivial. Integration before or after the acoustic echo cancellation (AEC) in UL or DL will influence the convergence and compensation properties of this algorithm. A trade-off has to be found between the performance of the echo canceller and the maximum achievable signal level in both directions. In Figure 1 the non-linearity in the UL influences the echo canceller directly while the dynamic control in the DL is placed after the branch for the echo canceller. The non-linearly modified DL signals are not available to the algorithm. In this case the non-linearity must be considered as part of the subsequent acoustic system (see next section). However, this configuration seems to be a very common one as a clear separation between algorithms dedicated to the optimization of the acoustic front-end and block-based algorithms for speech enhancement is possible.

While in the previous example non-linearities are inserted intentionally, sometimes non-linearities have to be compensated by dedicated algorithms. Often these non-linear effects are caused by mechanical and acoustical coupling due to small terminal housings. They can also arise from loudspeaker volumes being too low to linearly reproduce high sound pressure levels. An example for the cancellation of whistling caused by the combination of digital signal feedback (side tone) and mechanical coupling can be found in [1].

It can be seen in this section that many intrinsic properties of modern terminals and additive disturbances require dedicated algorithmic solutions for their compensation. However, these algorithms are not independent of each other, but are highly interactive and sometimes even contradictive. The tuning of such algorithms towards an optimum audio quality leads typically to many trade-offs.

2.2. Echo Reduction

Hands-free operation of terminals introduces several challenges to reduce echo and - at the same time - maintain a good double-talk quality [2]. Increasing the overall signal level compared to the standard handset operation causes stronger acoustical and mechanical coupling, very often being of non-linear nature. As a result the far-end user hears not only the voice of his near-end partner but also his own delayed voice. In theory, for perfect double-talk only acoustic echo cancellation should be applied. As it is depicted in Figure 1, an adaptive filter, represented by the vector of filter coefficients $\hat{\mathbf{h}}(n)$, is used to provide an estimate of the loudspeaker-enclosure-microphone (LEM) system $\mathbf{h}(n)$. As already mentioned in the previous section, equalizer and level control functions have to be considered as part of the LEM system, if they are placed after the branch for the AEC in the DL path. After the filter has converged, ideally the filter output signal $\hat{d}(n)$ is equal to the echo signal $d(n)$ picked up by the microphone. The most common method to compute updates for the adaptive filter $\hat{\mathbf{h}}(n)$ is the normalized least mean square algorithm (NLMS) [3], which is obtained from minimizing the expectation of the squared error signal $e(n) = d(n) - \hat{d}(n)$:

$$\hat{\mathbf{h}}(n+1) = \hat{\mathbf{h}}(n) + \mu \frac{\mathbf{x}(n)e^*(n)}{\|\mathbf{x}(n)\|^2}, \tag{2}$$

$\mathbf{x}(n)$ representing the vector of the most actual input samples in DL direction and μ being the adaptation step-size.

Several ideal assumptions have to be made during the derivation of this formula, which cannot be fulfilled under practical conditions. First of all the signal vector $\mathbf{x}(n)$ is assumed to be white, which is not true for speech signals. Secondly (2) only works fine, if $y(n) = d(n)$, i.e., the input signal to the echo canceller at the UL side only contains the acoustic echo signal. The useful near-end speech signal $s(n)$ and the environment noise $n(n)$ are interfering inputs for the echo canceller, which slow down the convergence and affect stability. Therefore a voice activity detection (VAD) unit is usually required to control the filter update and to differentiate near-end, far-end, and double-talk speech periods. In addition, the correct adjustment of the step size parameter μ is difficult. In practice a trade-off between convergence speed and stability has always to be found. During double-talk periods μ has to be decreased or the adaptation has to be stopped at all. There is a huge amount of literature dealing with solutions for the mentioned problems. Not all algorithms are that simple as the 'whitening' of the input signals by pre-emphasis filters and often overextend the computational capabilities of mobile terminals.

An aspect which cannot be neglected for the usual fixed-point implementations on DSPs is the problem of number representation. This applies especially for the computation of the update equation (2) and for the representation of the filter coefficients which model the tail of the LEM impulse response $\mathbf{h}(n)$. One has to take care that the quantization of these coefficients is fine enough, so that they represent an accurate estimation of $\mathbf{h}(n)$ and no algorithmic noise is added in this stage. Finally, $\mathbf{h}(n)$ and the transfer functions of dynamic control algorithms are not linear in practice. The resulting signals $d(n)$ and $y(n)$ may thus contain non-linear echo portions which can only partially be cancelled by a linear echo canceller. These non-linear fractions either have to be removed by dedicated algorithms, e.g., [4, 5], or by subsequent processing stages.

For these reasons the performance of linear time-domain based echo cancellation algorithms is not sufficient under real conditions and they have to be complemented by additional echo suppression methods. A solution working in the frequency domain is described in the next section. A very common method is the usage of gain loss controls in time-domain which - often controlled by the VAD - reduce the signal level in the less active signal path, thereby severely influencing double-talk quality. A simple example for the implementation of a gain loss control in UL direction is obtained if the signal $s_{in}(n)$ in the second line of (1) is replaced by the DL signal $s_{DL}(n)$. Double-talk quality can be controlled by applying an adjustable maximum threshold to the denominator term.

2.3. Noise Reduction / Echo Suppression

The algorithms for noise reduction and residual echo suppression are computed in the frequency domain. The commonly used processing scheme employs spectral analysis/synthesis and a spectral gain computation, introduced for noise reduction in [6] and used for echo suppression in [7–9]. In this scheme, the time signal is divided into frames which are transformed to the frequency domain through Fast Fourier Transform (FFT). For each frame l and each frequency bin k, the FFT of $z(n)$, denoted by $Z(l,k)$, is assumed to be the sum of the useful speech $S(l,k)$, of the residual echo $E(l,k)$, and of the noise $N(l,k)$. The core function of the spectral gain rule is common to noise reduction and echo reduction. It is based on a biased Wiener filter:

$$G_P(l,k) = \frac{\sum_{i=l_0}^{l} w_z(i) \cdot |Z(i,k)|^2}{\sum_{i=l_0}^{l} \left(w_z(i) \cdot |Z(i,k)|^2 + w_p(i) \cdot |P(i,k)|^2\right)} \tag{3}$$

where the index P stands for the type of perturbation ($P = N$ for noise reduction, $P = E$ for echo reduction), $w_z(l)$ and $w_p(l)$ are temporal smoothing factors. In our practical implementation, the estimation of $|N(i,k)|^2$ is obtained through smoothing of $|Z(i,k)|^2$, with a smoothing factor depending on the time variation of the noisy signal between two consecutive frames [10]. The estimation of $|E(i,k)|^2$ is obtained via intercorrelation between the loudspeaker signal and the microphone signal in a similar way as in [8]. If the echo reduction is computed by direct filtering through the weighting factor $G_E(l,k)$ of (3), the noise reduction can use the output of the Wiener filter to compute a masking threshold $\gamma_T(l,k)$ with the output of the Wiener filter being considered as the masker signal. This threshold is used for the computation of a psycho-acoustically motivated spectral gain [9], where the desired amount of noise reduction in the psycho-acoustical sense is defined by a scalar noise attenuation factor $\varsigma(l,k)$. According to this rule, the weighting factor $G_{PSY}(l,k)$ is chosen in such a way that all components of the residual noise, which exceed the desired amount, are just 'hidden' below the estimated threshold. This leads to the following weighting rule:

$$G_{PSY}(l,k) = \min\left\{1, \sqrt{\frac{\gamma_T(l,k)}{|N(l,k)|^2}} + \varsigma(l,k)\right\} \tag{4}$$

As shown in [9], this method results in the smallest possible speech distortion using spectral weighting for the desired amount of noise reduction $\varsigma(l,k)$.

In a last step G_E and G_{PSY} are limited by a post filter in order to reduce the disturbance on the useful signal in case of estimation errors. The limitation is based on the following philosophy: If the output power of the filter operation is smaller than a certain amount of the noise power, then the filter gain is limited in a way that the processed signal power is equal to this amount. Accordingly, the weighting rules are constraint by a noisy spectral floor, so that artifacts are perceptually hidden by a 'natural' residual noise.

3. ALGORITHMIC EXTENSIONS

3.1. DL Speech Intelligibility Enhancement

Most of the algorithms described in the last sections are dedicated to improve speech quality in the UL direction, that is, for the *far-end* communication partner. But not only speech intelligibility at the far-end is affected if the terminal is used in a very noisy environment. In fact it may become even more difficult for the *near-end* user to understand his far-end partner. Therefore algorithms have been designed to improve speech intelligibility in the DL path and offer an added value for the user of the terminal in such situations.

However, for this problem, as opposed to the task of UL noise reduction, the noise signal itself can not be influenced because the person is located in the noisy environment and, thus, the noise reaches the ears without any possibility to intercept. The fundamental idea of our approach is a time-adaptive and frequency-dependent amplification of the received far-end speech signal to reestablish a minimum distance between the average measured speech spectrum and the average measured noise spectrum. The frequency domain gain $G_{SE}(l,k)$ is chosen to recover a certain signal-to-noise ratio ξ in terms of the short-term power spectral densities $\Phi_{XX}(l,k)$ and $\Phi_{NN}(l,k)$

$$G_{SE}(l,k) = \max\left\{\sqrt{\xi \cdot \frac{\Phi_{NN}(l,k)}{\Phi_{XX}(l,k)}}, 1\right\}. \tag{5}$$

In order to prevent hearing damage and overload of sound equipment, the dynamic power range is compressed afterwards, e.g., using the DCL described in Section 2.1. For further details and an example algorithm refer to [11].

3.2. Low-Delay Synthesis

In telecommunication systems round-trip delay is always a critical design parameter. While in current circuit-switched systems (GSM, UMTS, satellite systems) round-trip delays of more than 200 ms are usual, this value may even be increased by packet-switched systems using VoIP transmission. Therefore the additional round-trip delay introduced by hands-free speech enhancement systems should be kept to a minimum.

A commonly used algorithmic procedure is spectral weighting by means of the overlap-add method, as it was mentioned in section 2.3. One approach to reduce the signal delay is to use a lower frequency resolution or to shorten the overlap between consecutive signal frames. A different approach to achieve a low signal delay is to perform time-domain filtering with coefficients adapted in the frequency domain. One realization of this concept used here is the recently proposed filter bank equalizer [12, 13], which allows filtering with uniform and non-uniform frequency resolution (see also Section 4.3). Its algorithmic signal delay is determined by the prototype filter for the analysis filter bank. Therefore, a lower delay than for a corresponding analysis-synthesis filter bank (including the overlap-add method as a special case) can be achieved.

A technique to further reduce the signal delay of the filter bank equalizer in a simple and flexible manner has been proposed in [14]. The original time-domain filter of the filter bank equalizer is approximated by a filter of lower degree. For this, an all pole filter can be used whose coefficients can be efficiently calculated by the Levinson-Durbin recursion. The obtained filter is of minimum phase and achieves a signal delay of only a few samples. By employing this filtering concept, a very low signal delay is achieved while obtaining a similar (subjective) quality for the enhanced speech than with the common spectral weighting method.

4. NEW SYSTEM CONCEPTS

4.1. Wideband Audio

The emergence of wideband speech coding with the adoption of the AMR-WB speech codec by 3GPP [15] has raised a fundamental question: Should the general principles of algorithmic audio systems be modified when they are migrated from 8 kHz to 16 kHz sampling frequency - or are some basic adaptations sufficient? Concrete solutions have already been proposed, like noise reduction for wideband signals in [16], although they are far away from providing a complete answer to all the questions.

At a first glance, the basic operating mode of most algorithms can be seen as independent of the sampling frequency f_s. Indeed, nowhere in the previous sections, the term sampling frequency has been mentioned in the equations describing the proposed solution. The influence of f_s is in fact hidden as only the intrinsic variables are affected. Only a few aspects are mentioned here, a more detailed discussion can be found in [17]: The design of any first-order low-pass filter is highly depending on the sampling frequency; the length of the echo cancellation filter has to be doubled in wideband, resulting in a fourfold increased computation load; the resolution of frequency analysis is linked to the number of frequency bins and the length of the analysis window, both bound to f_s.

Accordingly, if the same mathematical description of an algorithm can be implemented whatever the sampling frequency is, an optimal solution would involve many parameters to be taken as variables depending on f_s. This would lead to a flexible but complex design of the algorithms, where any influenced parameter would be modifiable. In order to avoid such difficult system architectures, a simpler solution is based on the usage of filter banks splitting the signal in two bands (0-4 kHz, 4-8 kHz). The proposed approach provides the advantage of compatibility with both sampling frequencies, as the wideband part can just be plugged to a narrowband system, leading to a simple scalable processing scheme. In addition, the algorithms in the sub-bands can be designed and tuned independently of each other, e.g., different lengths of the adaptive filters can be used for echo cancellation or different weighting rules can be applied for noise reduction. In the 8 kHz case only the narrowband branch is computed without the filter bank processing blocks. For 16 kHz sampling frequency the wideband branch and the filter bank processing blocks are plugged in. No redesign of the existing narrowband solution is required.

4.2. Frequency Domain Based Systems

The LMS adaptive filter, a variant of the NLMS algorithm described in Section 2.2, can be realized efficiently as block adaptation algorithm in the frequency domain [18]. The resulting algorithm is commonly referred to as frequency-domain adaptive filter (FDAF) and offers the possibility of a frequency-selective time-varying step size control [18]. A major advantage is the easy integration of spectral weighting rules for residual echo suppression and/or noise reduction [19] since the filter output error signal is available in the frequency domain. For example the following simple relation applies between an optimum frequency-selective step size $\mu(l,k)$ (at frame l and in frequency bin k) and a Wiener post filter $G(l,k)$

for the suppression of residual echo [19–21]:

$$\mu(l,k)+G(l,k)=1 \tag{6}$$

This interdependency of step size control and spectral weights results in a complementary behavior of echo cancellation and post filter. If $\mu(l,k)$ is small, e.g. after the system has approximately converged, $G(l,k)\approx 1$, and the influence of the post filter is small. Reversely, for $\mu(l,k)\approx 1$ the system may not have converged yet and the residual echo is nearly completely suppressed by the post filter. A double-talk detector, which is normally needed for a step size control to mitigate divergence, can even be avoided if the squared frequency-domain convergence state $|H(l,k)-\hat{H}(l,k)|^2$ is calculated in a recursive manner as shown in [21] and an optimum step size [21–23] is used. Thus the drawbacks of time-domain gain loss controls working independently of the adaptive filter are avoided, resulting in improved duplex capability.

4.3. Filter Bank Based Systems

The frequency-based algorithms of the last chapter provide uniform frequency resolution. However, looking at the distribution of echo energy over time and frequency in real rooms or at psychoacoustic findings, a sub-band structure with a non-uniform frequency resolution provides several advantages. First of all, the frequency resolution (and vice versa the corresponding time resolution) can be matched to the Bark scale that approximates the logarithmic properties of the human ear. Secondly, a sub-band structure allows for a different treatment of each band, e.g., different lengths of the adaptive filters used for echo cancellation to simulate the physical behavior of real rooms. In addition, by omitting the redundancy inherent to uniform frequency-based systems, the computation effort can be reduced by appropriate design of the filter bank systems. Nevertheless the advantage of interdependent control of step size and post filter *within* the sub-bands as given in (6) is maintained [20].

Usually sub-band structures are realized by critically sub-sampled analysis-synthesis filter banks. The drawbacks of critical decimation are aliasing effects and processing delay, the former affecting the performance of algorithms, the latter being a critical parameter in the design of hands-free systems. We propose to replace the uniform FFT analysis by a non-uniform one, e.g., the allpass transformed FFT analysis stage of the filter bank equalizer [12] introduced in Section 3.2. No sub-sampling is applied to the sub-band signals and, if the prototype filter of the analysis stage is designed as M^{th}-band filter, perfect amplitude reconstruction is achieved by addition of the sub-band signals. The filter bank delay is determined by the analysis stage delay and by the delay of a subsequent phase equalizer, by which near perfect signal reconstruction can be achieved [13]. With appropriate design, the delay can be reduced in comparison to a critically sub-sampled filter bank where constraints apply due to interdependent design of the analysis/synthesis filters.

For echo cancellation with adaptive filters in the sub-bands the UL signal has to be transformed to the sub-band domain. Accordingly, the computational efficient synthesis concept of the non-uniform filter bank equalizer [13] using time-domain filtering with coefficients adapted in the frequency domain cannot be applied directly. Two solutions are possible: First of all the wanted signal can be resynthesized after echo cancellation, the post filter coefficients are computed in the sub-band domain and applied in a subsequent time-domain post filter [20] using the filter bank equalizer concept. A second approach is to apply the post filter coefficients directly in the sub-band domain and then synthesize the fullband signal. In both approaches the adjustment of delay elements balancing the different transient behavior of echo cancellation and post filter algorithms is critical for the performance of the system.

5. CONCLUSION

We have considered several aspects of algorithm design for hands-free audio systems in telecommunication terminals. Starting from a state-of-the-art solution new algorithmic concepts have been discussed which tackle challenges like dual-mode solutions for narrowband/wideband audio, low round-trip delay, and full-duplex hands-free telephony. The advantages of joint design for echo cancellation and post filter algorithms in the frequency domain can be transferred to filter bank based systems. The involvement of psychoacoustic aspects enables a new dimension in the design of speech enhancement systems for telecommunication terminals, particularly with regard to the ambition of natural, hands-free communication.

6. REFERENCES

[1] M. Schönle, V. Gilg, "An Algorithm for Cancellation of Sidetone Oscillations," International Workshop for Acoustic Echo and Noise Control, Kyoto, September 2003.

[2] ITU-T, P.340, "Transmission characteristics and speech quality parameters of hands-free terminals," International Telecommunication Union, 2000.

[3] S. Haykin, "Adaptive Filter Theory ", Prentice Hall, Englewood Cliffs, New Jersey, 2002.

[4] W. Frank, "An Efficient Approximation to the Quadratic Volterra Filter and Its Application in Realtime Loudspeaker Linearization," Signal Processing, vol. 45/1, 1995, pp. 97-113.

[5] A. Stenger, W. Kellermann, "Adaptation of a memoryless preprocessor for nonlinear acoustic echo cancellation," Signal Processing, vol. 80, 2000, pp. 1747-1760.

[6] S. F. Boll, "Suppression of Acoustic Noise in Speech Using Spectral Subtraction," IEEE Trans. Acoust., Speech, Signal Processing, vol. 27, April 1979, pp. 113-120.

[7] S. Gustafsson, R. Martin, and P. Vary, "Combined Acoustic Echo Control and Noise Reduction for Hands-free Telephony," Signal Processing, vol. 64, no. 1, January 1998, pp. 21-32.

[8] C. Beaugeant, P. Scalart, "Combined systems for noise reduction and echo cancellation," European Signal Processing Conference, 1998, pp. 957-960.

[9] S. Gustafsson, R. Martin, P. Jax, and P. Vary, "A Psychoacoustic Approach to Combined Acoustic Echo Cancellation and Noise Reduction," IEEE Trans. on Speech and Audio Processing, vol. 10, no. 5, July 2002, pp. 245-256.

[10] C. Beaugeant, V. Gilg, M. Schönle, P. Jax, and R. Martin, "Computationally Efficient Speech Enhancement Using RLS and Psycho-Acoustic Motivated Algorithm," World Multi-Conference on Systemics, Cybernetics, and Informatics, July 2002.

[11] Bastian Sauert and Peter Vary, "Near End Listening Enhancement: Speech Intelligibility Improvement in Noisy Environments," Intl. Conf. on Acoustics, Speech, and Signal Processing, May 2006.

[12] P. Vary, "An adaptive filter bank equalizer for speech enhancement", Signal Processing, Elsevier, Special Issue on Applied Signal and Audio Processing, 2006.

[13] H. W. Löllmann and P. Vary, "Efficient Non-Uniform Filter-Bank Equalizer," European Signal Processing Conference, Antalya, Sept. 2005.

[14] H. W. Löllmann and P. Vary, "Low Delay Filter for Adaptive Noise Reduction", International Workshop on Acoustic Echo and Noise Control, Eindhoven, Sept. 2005.

[15] I. Varga, R. Drogo De Iacovo, P. Usai, "Standardization of the AMR Wideband Speech Codec in 3GPP and ITU-T," IEEE Comm. Mag., May 2006.

[16] C. Beaugeant, I. Varga, T. Lotter, P. Jax, P. Vary, "Noise Reduction for AMR-WB," ETSI STQ Workshop on Wideband Speech Signal Processing, June 2005.

[17] C. Beaugeant, M. Schönle, I. Varga, "Challenges of 16 kHz in Acoustic Pre- and Post-Processing for Terminals," IEEE Comm. Mag., May 2006.

[18] J. J. Shynk, "Frequency-Domain and Multirate Adaptive Filtering," IEEE Signal Processing Magazine, January 1992, pp. 14-37.

[19] G. Enzner, R. Martin, and P. Vary, "Partitioned Residual Echo Power Estimation for Frequency-Domain Acoustic Echo Cancellation and Postfiltering," European Transactions on Telecommunications, vol. 13, no. 2, March/April 2002.

[20] E. Hänsler and G. Schmidt, "Hands-free telephones - joint control of echo cancellation and postfiltering," Signal Processing, vol. 80, 2000, pp. 2295-2305.

[21] G. Enzner and P. Vary, "Robust and Elegant, Purely Statistical Adaptation of Acoustic Echo Canceller and Postfilter," International Workshop on Acoustic Echo and Noise Control, Kyoto, Japan, Sept. 2003.

[22] B. H. Nitsch, "Adaptive Filter im Frequenzbereich für Freisprecheinrichtungen," PhD thesis, in German, Fortschritt-Berichte VDI, Reihe 10, Nr. 623, 2000.

[23] K. Steinert and T. Fingscheidt, "On a Robust FDAF Step Size Control," ITG Fachtagung Sprachkommunikation, Kiel, April 2006.