REAL-TIME NEAR-END LISTENING ENHANCEMENT FOR MOBILE PHONES

Bastian Sauert, Florian Heese, and Peter Vary

Institute of Communication Systems and Data Processing (ind)
RWTH Aachen University, Germany
{sauert,heese,vary}@ind.rwth-aachen.de

ABSTRACT

Mobile telephony is often conducted in the presence of acoustical background noise such as traffic or babble noise. The near-end listener perceives a mixture of clean far-end (downlink) speech and background noise at the near-end side. Thus, the user experiences an increased listening effort and possibly a reduced speech intelligibility. This situation can be improved by adaptive signal processing of the far-end signal, which we call *near-end listening enhancement* (NELE).

In our real-time demonstration the visitor will experience the strong improvement in the acoustic conference/exhibition environment. The adaptive algorithm optimizes the intelligibility of the far-end speech even in instationary background noise with respect to the objective criterion *Speech Intelligibility Index* (SII). In contrast to state-of-the-art techniques, the presented method considers the requirements and restrictions of realistic scenarios such as in mobile phones.

1. INTRODUCTION AND MOTIVATIONS

With the advent of cellular phones, people often make phone calls in challenging acoustical environments where a conversation is eventually perceptually impossible.

In these situations, strong acoustical background noise, e. g., traffic or babble noise, is often present at the near-end side. This has three major implications:

- The near-end user modifies her/his speaking style as a consequence of the exposure to the near-end noise, an effect known as Lombard reflex.
- The near-end noise is captured by the microphone together with the near-end speech. Several noise reduction techniques have been proposed, to reduce this noise signal before speech coding and transmission.
- The near-end user perceives a mixture of the clean or noise reduced far-end speech and the local acoustical background noise at the near-end side. Thus, the user experiences an increased listening effort and possibly a reduced speech intelligibility, which is addressed in this contribution.

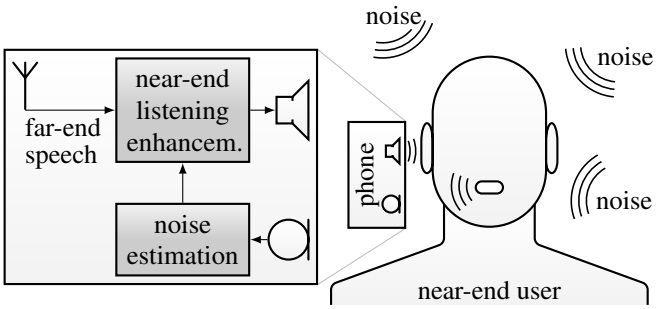

Fig. 1: Handset telephony with near-end listening enhancement (NELE).

The noisy environment can usually not be influenced, like car noise at a busy street or speech babble noise in a cafeteria. As there is no possibility to intercept the near-end noise, the manipulation of the far-end signal is the only way to effectively improve speech intelligibility for the near-end listener by digital signal processing. We call this approach *near-end listening enhancement* (NELE) and illustrate its application to handset telephony in the presence of acoustical background noise in Figure 1.

2. SCIENTIFIC AND TECHNICAL DESCRIPTION

In this demonstration, a sophisticated solution to the problem of near-end listening enhancement is shown. It optimizes the intelligibility of the far-end speech in local background noise with respect to the objective criterion *Speech Intelligibility Index* (SII). The presented method tackles the problem for the first time from the application point of view considering also the requirements and restrictions of realistic scenarios and devices such as mobile phones. It is of particular importance that the processing adapts dynamically to the time-varying characteristics of the ambient noise. Hence, an effective intelligibility enhancement is provided in the presence of background noise, while in silence *no* audible modification is applied. The utilized noise tracking algorithm estimates the noise spectrum blindly from the microphone signal captured by the mobile phone – the only access to the acoustical environment – and at the same time disregards the voice of the near-end user in

double-talk situations. Power limitation in critical bands ensures that the ear of the near-end listener is protected from damage and pain. Furthermore, the maximum thermal load of the micro-loudspeaker of the mobile phone is taken into account.

The basic concept of the demonstrated NELE algorithm consists of two steps: First, an optimum "speech spectrum level" in critical bands is determined which maximizes the SII under consideration of the current "disturbance spectrum level", i. e., the spectral characteristics of the ambient noise. Then, the subband weights are calculated to achieve this optimum speech spectrum level with the far-end speech at the ear of the listener.

It is a core part of the NELE algorithm to spectrally reallocate the audio power of the speech signal across critical frequency bands when necessary in order to improve intelligibility. This goes hand in hand with a moderate change of tone color of the speech through the influence of the noise spectrum. However, such a tone coloration is not perceived as distortion. If a (further) reallocation would *not* improve intelligibility, the tone color is preserved as much as possible.

2.1. Speech Intelligibility Index (SII)

The SII [1] is a standardized objective measure which correlates with the intelligibility of speech under a variety of adverse listening conditions. A detailed discussion of the calculation rules of the SII is given in [2, 4].

In short, the SII is based on the equivalent speech spectrum level[1] E_i as well as the equivalent disturbance spectrum level[1] D_i, which accounts for the noise as well as the masking of the speech. Given E_i and D_i, the Speech Intelligibility Index S is calculated as a weighted sum of the band audibility function $A_i(E_i, D_i)$ over all contributing subbands i:

$$S = \sum_i I_i \cdot A_i(E_i, D_i)\,, \tag{1}$$

where the band importance function I_i [1] characterizes the relative significance of each subband to speech intelligibility.

The band audibility function $A_i(E_i, D_i)$ specifies the effective proportion of the speech dynamic range within the subband that contributes to speech intelligibility. Its characteristics are sketched in Figure 2 for a low as well as a high disturbance scenario.

2.2. Near-End Listening Enhancement Concept

The basic idea of the following algorithm is to first determine an optimum speech spectrum level $E_i^{\text{opt}}(\kappa)$ for the half-overlapping frame of 20 ms length with index κ, which maximizes the SII S under consideration of the current disturbance

[1]The term "equivalent" is omitted in the following for the sake of clarity.

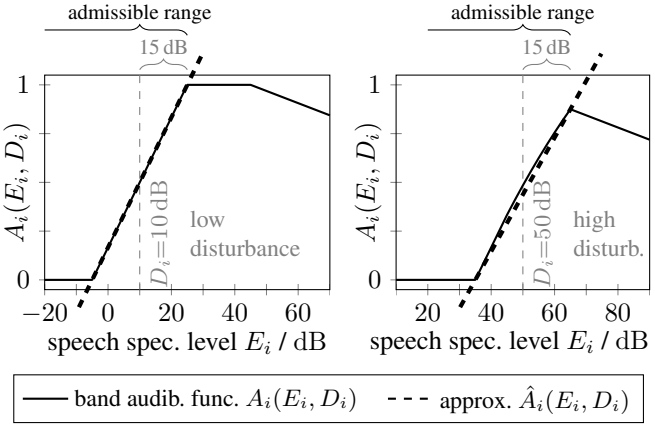

Fig. 2: Exemplary plot of band audibility function for low as well as higher disturbance case.

spectrum level $D_i(\kappa)$:

$$\underline{E}^{\text{opt}}(\kappa) = \arg\max_{\underline{E}} \sum_i I_i \cdot A_i\big(E_i, D_i(\kappa)\big) \tag{2}$$

subject to

$$\sum_i \hat{P}_{s,i} = \sum_i f_{\Delta,i} \cdot 10^{E_i/10} \overset{!}{\leq} \mathfrak{P}^{\max}(\kappa)\,, \tag{3}$$

where $\underline{E}$ denotes the vector of all contributing speech spectrum level $E_i = 10\log\{\hat{P}_{s,i}/f_{\Delta,i}\}$, $\hat{P}_{s,i}$ is the short-term subband power of the speech signal, and $f_{\Delta,i}$ is the frequency bandwidth of the i-th subband. The constraint limits the total audio power of the loudspeaker signal to a maximum allowed power $\mathfrak{P}^{\max}(\kappa)$. In this demonstrator, two variants are considered:

1. The loudspeaker signal power is constrained to the power of the original (input) signal, i. e., no additional audio power may be spent.

   This constraint is basically an extreme case for sound reproduction systems without head-room in terms of total output audio power. But it also proves useful for comparison with other NELE algorithms.

2. One major limitation of small loudspeakers used in mobile phones is the thermal load during continuous playback. In this realistic case, the maximum allowed total loudspeaker signal power $\mathfrak{P}^{\max}$ is constant and a parameter of the sound reproduction system, which could be derived during design of the device.

Next, the spectral weights $W_i(\kappa)$ which are necessary to achieve this optimum speech spectrum level at the ear of the listener are calculated. Assuming short-term stationary spectral weights, this finally leads to the spectral weights

$$W_i(\kappa) = 10^{\left[E_i^{\text{opt}}(\kappa) - E_i^{\text{in}}(\kappa)\right]/20}\,, \tag{4}$$

where $E_i^{\text{in}}(\kappa)$ denotes the current input speech spectrum level.

2.3. Recursive Closed-Form Optimization [3]

If $E_i = D_i + 15\,\mathrm{dB}$ in all subbands fulfills the constraint (3), the maximum SII can be reached. If not, all power must be used to maximize the SII. In this case, the solution lies within the admissible range $E_i^{\mathrm{opt}} \leq D_i + 15\,\mathrm{dB}$ and the inequality constraint (3) becomes an equality constraint

In order to facilitate the envisaged closed-form optimization, the band audibility function is approximated by a linear function $\hat{A}_i(E_i, D_i)$ as also depicted in Figure 2.

Then, the equality constrained nonlinear multivariate maximization problem (2) and (3) can be solved using the methods of Lagrange multipliers. As shown in [3] and [4], this finally leads to the closed-form solution

$$E_i^{(1)} = 10 \log \left\{ \frac{\Gamma_i}{\sum_{\zeta=1} \Gamma_\zeta} \cdot \frac{P_s^{\max}(\kappa)}{f_{\Delta,i}} \right\}, \qquad (5)$$

with the gradient Γ_i of the linear approximation $\hat{A}_i(E_i, D_i)$. This solution might, however, fall outside the admissible range. Therefore, further steps $\upsilon = 2, 3, \ldots$ are necessary, where the preceding solution $E_i^{(\upsilon-1)}$ is limited to $D_i + 15\,\mathrm{dB}$ and the closed-form solution (5) is repeated recursively until all subbands fulfill $E_i^{(\upsilon)} \leq D_i + 15\,\mathrm{dB}$.

Please refer to [3, 4] for further details of this algorithm.

3. IMPLEMENTATION AND USE

The described algorithm for near-end listening enhancement is implemented in C/C++ using the RTProc system for rapid development of real-time prototypes [5]. Figure 3 shows the graphical user interface of the demonstrator.

The demonstration is performed in real-time on a laptop using a real telephone handset. The environmental noise is captured with the microphone of the handset and its power spectral density is estimated using a noise tracking algorithm to enable double-talk interaction. In this implementation, both variants of maximum allowed total audio power are covered: the limitation to the power of the input signal as well as the restriction to a constant maximum allowed total audio power.

Due to its realistic nature, the demonstrator offers the interactive experience of the NELE problem and the proposed solution.

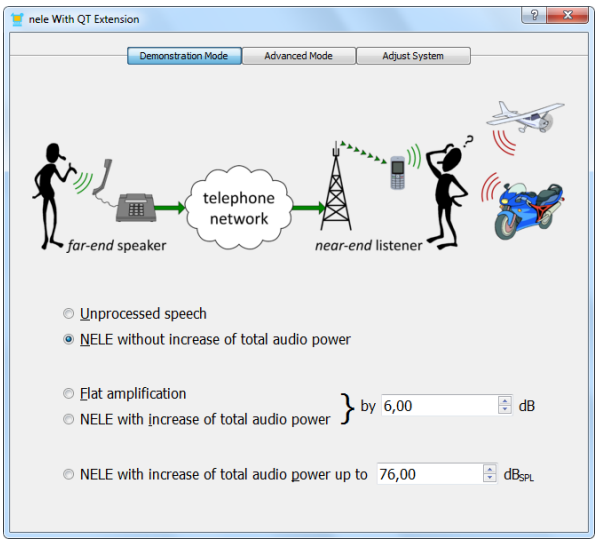

Fig. 3: Graphical user interface of the demonstrator.

4. CONCLUSIONS

This contribution demonstrates a sophisticated solution to the problem of near-end listening enhancement in mobile telephony. The intelligibility of the far-end speech signal which is perceived by the near-end user in local background noise environment is improved by adaptive processing.

The algorithm is presented in real-time under realistic circumstances.

References

[1] ANSI S3.5-1997. *Methods for the Calculation of the Speech Intelligibility Index*. American National Standards Institute, 1997.

[2] Bastian Sauert and Peter Vary. "Near End Listening Enhancement Optimized with Respect to Speech Intelligibility Index". In: *Proc. of European Signal Processing Conf. (EUSIPCO)*. (Glasgow, Scotland, Aug. 24–28, 2009). Vol. 17. European Association for Signal Processing (EURASIP). New York, NY: Hindawi Publ., Aug. 2009, pp. 1844–1848.

[3] Bastian Sauert and Peter Vary. "Recursive Closed-Form Optimization of Spectral Audio Power Allocation for Near End Listening Enhancement". In: *Proc. of ITG-Fachtagung Sprachkommunikation*. (Bochum, Germany, Oct. 6–8, 2010). Vol. 9. Berlin [u.a.]: VDE-Verlag, Oct. 2010. ISBN: 978-3-8007-3300-2.

[4] Bastian Sauert. "Near-End Listening Enhancement: Theory and Application". PhD thesis. RWTH Aachen University, 2014.

[5] Hauke Krüger et al. "RTPROC: Rapid Real-Time Prototyping for Audio Signal Processing". In: *Proc. of IEEE Intl. Conf. on Acoustics, Speech, and Signal Processing (ICASSP)*. (Taipei, Taiwan). Show and Tell Demonstration. IEEE. Apr. 2009. ISBN: 978-1-42442-354-5.