# Model-based adaptive pre-processing of speech for enhanced intelligibility in noise and reverberation

*Jan Rennies*[1], *Andreas Volgenandt*[1], *Henning Schepker*[2], *Simon Doclo*[1,2]

[1]Fraunhofer IDMT, Hearing, Speech and Audio Technology, and Cluster of Excellence Hearing4All, Oldenburg, Germany
[2]University of Oldenburg, Signal Processing Group and Cluster of Excellence Hearing4All, Oldenburg, Germany

{rns, vgt}@idmt.fraunhofer.de,{henning.schepker, simon.doclo}@uni-oldenburg.de

## Abstract

In this demonstrator we present the most recent advances in the development of the near-end listening enhancement algorithm *AdaptDRC*. The algorithm uses short-time estimates of the speech intelligibility index to control spectral shaping, dynamic range compression and/or level adjustment to achieve an adaptive enhancement of speech intelligibility in adverse listening conditions. Depending on the application scenario, the algorithm framework can take background noise and reverberation as well as different boundary conditions into account. The show and tell contribution comprises a real-time setup of the algorithm to demonstrate the sound modifications and the impact of the different parameters and boundary conditions. An accompanying poster shows results of formal listening tests evaluating the speech intelligibility improvement achieved by the algorithm for normal-hearing and hearing-impaired listeners.

**Index Terms**: speech intelligibility, noise, reverberation, near-end listening enhancement

## 1. Introduction

In many speech communication applications involving speech playback, such as public address systems or mobile telephony, speech perception can be substantially impaired by noise and reverberation in the listening environment, leading to reduced speech intelligibility and increased listening effort.

Near-end listening enhancement (NELE) algorithms aim at improving speech intelligibility in such adverse listening conditions by modifying the speech signal prior to playback. A straight forward solution to obtain a high speech intelligibility is to increase the speech level to achieve a good signal-to-noise ratio (SNR). However, this may not always be possible due to technical constraints and, consequently, other types of signal modification are required. Different types of pre-processing strategies have been proposed, involving time stretching, frequency shaping, or dynamic range compression (DRC). A systematic comparison of different NELE algorithms involving extensive formal listening tests was conducted in the Hurricane Challenge presented in [1]. It was generally observed that the largest improvement of speech intelligibility in fixed background noise was achieved by algorithms including DRC. One of these algorithms was the *AdaptDRC* algorithm proposed in [2]. Like all other algorithms compared in [1], AdaptDRC did not modify the speech level, i.e. the output SNR was the same as the input SNR. While this is a reasonable boundary condition for systematic comparisons of different algorithms, it is not the best option for practical solutions since, in many cases, an at least moderate increase in speech level is possible. Therefore, *AdaptDRC* has been extended to allow for an additional level increase. A further extension of the algorithm is proposed in [3]. In addition to background noise considered in previous studies, this version also takes detrimental effects of reverberation into account.

In this Show & Tell contribution, *AdaptDRC* as well as its extensions are presented as real-time audio demonstrator. Different algorithm versions as well as the perceptual impact of varying the algorithm parameters can be compared. Subjective data from different evaluation studies are shown on a supplementary poster to support the discussions during the session.

## 2. *AdaptDRC* algorithm

A schematic diagram of the *AdaptDRC* algorithm is shown in Figure 1. A detailed description of the processing steps is given in [2]. Generally speaking, the algorithm processes a target speech signal $\mathbf{s}[l]$ in short blocks (10 to 20 ms) indexed by $l$ to generate a modified speech signal $\tilde{\mathbf{s}}'[l]$. The processing is controlled adaptively by estimations of the speech intelligibility index (SII) [4]. For this purpose, the noise in the listening environment $\mathbf{n}[l]$ is estimated and both speech and noise are filtered in octave bands with center frequencies from 125 Hz to 16 kHz (if required by the application scenario, the highest center frequency can also be lower, e.g., at 4 or 8 kHz). The estimated SII (index between 0 and 1) controls the subsequent processing stages consisting of frequency shaping and DRC. If the estimated intelligibility is high (index approaching 1) no weighting or DRC is applied. Conversely, if the estimated intelligibility decreases, an increasing amount of frequency shaping (with increasing amplification for higher frequency components) and DRC is applied. The processed octave-band signals $\tilde{\mathbf{s}}_n[l]$ are summed, and the resulting reconstructed speech signal is normalized to restore the broadband input power $\Phi_s$. This equal-power constraint ensured that no level increase between input and output speech occurred [2].

Since practical applications often do not require this equal-power constraint, an extended version of *AdaptDRC* has been developed. This extension further processes the output of *AdaptDRC* and applies an adaptive broadband gain. The gain is controlled by the same SII estimation which also steers the processing stages of *AdaptDRC*, i.e., no gain is applied if the

This work was supported in part by the Research Unit FOR 1732 "Individualized Hearing Acoustics" and the Cluster of Excellence 1077 "Hearing4All", funded by the German Research Foundation (DFG).

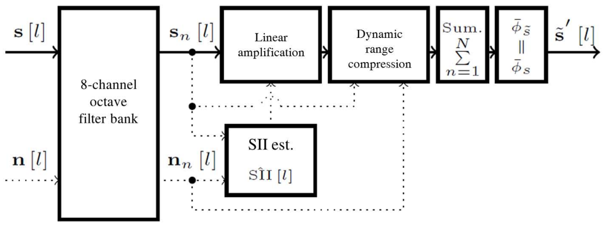


Figure 1: *Schematic diagram of the AdaptDRC algorithm as proposed in [2].*

estimated SII approaches 1, and the maximum allowed gain is applied if the index approaches 0. Instead of an equal-power constraint, the algorithm output is limited with respect to the peak level of the input signal.

These versions of *AdaptDRC* modify the speech depending on the amount of background noise. Possible detrimental effects of reverberation are not accounted for. To also include reverberation in the adaptive pre-processing, another algorithm extension has been developed and is described in [3]. Briefly, the reverberant part of the transmitted speech signal is transformed to an equivalent "effective" noise, and then added to the physical background noise for each octave band signal, following the concept of the speech transmission index [5]. The sum of effective noise and physical noise are then used for the SII estimation in the same way as described above to control frequency shaping and DRC.

## 3. Validation in formal listening tests

Several evaluation studies have been conducted to evaluate the gain in speech intelligibility achieved by the algorithm versions. In the listening tests conducted during the Hurricane Challenge [1], the original version *AdaptDRC* achieved up to 20.5% and 12.3% of speech intelligibility improvement for English everyday sentences in the presence of stationary speech-shaped noise and a competing talker, respectively. Since the algorithm did not increase the speech level, this improvement was not due to a better SNR, but due to the beneficial effects of frequency shaping and DRC. Even larger improvements were found for German speech material and different background noises (cafeteria, speech-shaped noise, driving noise, see Figure 2). Recent studies investigated the benefit of an adaptive level increase (Figure 2). It was found that both *AdaptDRC* (without level increase) and the extended version (with level increase) [6] achieved considerable speech intelligibility improvements. This observation was also made for hearing-impaired listeners which did not, however, benefit from the pre-processing to the same degree as listeners with normal hearing.

## 4. Show & tell at INTERSPEECH 2015

The demonstration at the Show & Tell session consists of a real-time setup of *AdaptDRC* and its different versions. A graphical user interface is used to select algorithm version and modify parameters to directly listen to the resulting perceptual changes. Environmental noise at the demonstration site is picked up by a microphone. If desired, additional noise (e.g., driving noise) can be added to the playback to explore the signal modifications in more listening scenarios. To illustrate the signal processing stages, internal algorithm parameters (such as estimated mo-

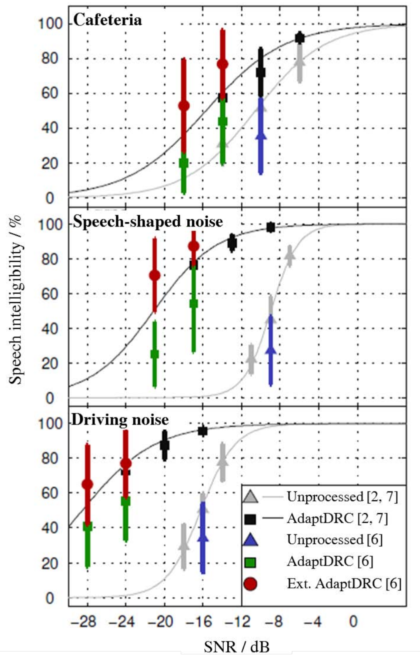


Figure 2: *Speech intelligibility as a function of SNR for unprocessed speech and speech processed by AdaptDRC [2, 7] and an extended version [6].*

mentary speech intelligibility, applied frequency shaping) and spectrograms of input and output signals are graphically displayed. The results of the subjective evaluation studies are summarized on an accompanying poster to encourage discussions about the potential and limitations of the different *AdaptDRC* versions.

## 5. References

[1] M. Cooke, C. Mayo, and C. Valentini-Botinhao, "Intelligibility enhancing speech modifications: the Hurricane challenge", *Proc. Interspeech, Lyon, France*, 2013, pp. 3552–3556.

[2] H. Schepker, J. Rennies, and S. Doclo, "Improving speech intelligibility in noise by SII-dependent preprocessing using frequency-dependent amplification and dynamic range compression", *Proc. Interspeech, Lyon, France*, 2013, pp. 3577–3581.

[3] H. Schepker, D. Hülsmeier, J. Rennies, and S. Doclo, "Model-based integration of reverberation for noise-adaptive near-end listening enhancement," *accepted for publication in Proc. Interspeech, Dresden, Germany*, 2015.

[4] ANSI S3.5-1997, "Methods for calculation of the speech intelligibility index," *American National Standard ANSI S3.5-1997, New York, USA*.

[5] IEC 60268-16, "Sound System Equipment - Part 16: Objective rating of speech intelligibility by speech transmission index," *International Electrotechnical Commission, 2003, Geneva, Switzerland.*

[6] J. Drefs, J. Rennies, H. Schepker, and S. Doclo, "Weiterentwicklung und Evaluation eines Algorithmus zur SII-basierten Sprachverständlichkeitsverbesserung in störgeräuschbehafteter Umgebung (Development and evaluation of an algorithm for SII-based intelligibility enhancement in noisy environments)," *Proc. of DAGA 2015, 41$^{st}$ Annual Conference of German Acoustical Society, Nuremberg, Germany, 2015.*

[7] H. Schepker, J. Rennies, and S. Doclo, "Speech-in-noise enhancement using amplification and dynamic range compression controlled by the speech intelligibility index," *J. Acoust. Soc. Am., under review.*