Extension and evaluation of a near-end listening enhancement algorithm for listeners with normal and impaired hearing

Jan Rennies,[a)] Jakob Drefs, and David Hülsmeier[b)]
Project Group Hearing, Speech and Audio Technology, Fraunhofer Institute for Digital Media Technology IDMT and Cluster of Excellence Hearing4All, D-26129 Oldenburg, Germany

Henning Schepker and Simon Doclo[c)]
Signal Processing Group, Department of Medical Physics and Acoustics and Cluster of Excellence Hearing4All, University of Oldenburg, D-26111 Oldenburg, Germany

(Received 8 March 2016; revised 25 January 2017; accepted 20 March 2017; published online 10 April 2017)

In many applications in which speech is played back via a sound reinforcement system such as public address systems and mobile phones, speech intelligibility is degraded by additive environmental noise. A possible solution to maintain high intelligibility in noise is to pre-process the speech signal based on the estimated noise power at the position of the listener. The previously proposed *AdaptDRC* algorithm [Schepker, Rennies, and Doclo (2015). J. Acoust. Soc. Am. **138**, 2692–2706] applies both frequency shaping and dynamic range compression under an equal-power constraint, where the processing is adaptively controlled by short-term estimates of the speech intelligibility index. Previous evaluations of the algorithm have focused on normal-hearing listeners. In this study, the algorithm was extended with an adaptive gain stage under an equal-peak-power constraint, and evaluated with eleven normal-hearing and ten mildly to moderately hearing-impaired listeners. For normal-hearing listeners, average improvements in speech reception thresholds of about 4 and 8 dB compared to the unprocessed reference condition were measured for the original algorithm and its extension, respectively. For hearing-impaired listeners, the average improvements were about 2 and 6 dB, indicating that the relative improvement due to the proposed adaptive gain stage was larger for these listeners than the benefit of the original processing stages.

[http://dx.doi.org/10.1121/1.4979591]

[MAH] Pages: 2526–2537

I. INTRODUCTION

In many situations in which speech is played back via sound reinforcement systems (e.g., announcements of public-address systems in train stations or mobile telephony in crowded rooms), speech intelligibility is degraded due to ambient noise at the position of the listener. In such conditions, intelligibility may be improved by so-called near-end listening enhancement (NELE) algorithms, which pre-process the speech signal aiming to remain intelligible in noise. Most previous studies, however, have exclusively focused on normal-hearing listeners. The goal of the present study is (1) to extend a NELE algorithm previously proposed by Schepker *et al*. (2013, 2015) and (2) to evaluate the original and the extended algorithm with both normal-hearing and unaided hearing-impaired listeners.

Several recent studies have proposed NELE algorithms and have investigated their effectiveness in different background noise conditions (e.g., Kleijn *et al*., 2015; Taal *et al.*, 2014; Zorila and Stylianou, 2014; Taal and Jensen, 2013; Sauert and Vary, 2012; Zorila *et al.*, 2012; Tang and Cooke, 2011; Sauert and Vary, 2010). A comprehensive comparison of different NELE algorithms was conducted in the Hurricane Challenge 2013 (Cooke *et al*., 2013a,b), which included 14 different algorithms and two different masker types [stationary speech-shaped noise (SSN) and an interfering talker], each at three different signal-to-noise ratios (SNRs). The tested algorithms employed different types and combinations of time-frequency energy reallocation, frequency-shaping, dynamic range compression (DRC), and time-stretching. As a general result, Cooke *et al*. (2013b) concluded that the speech intelligibility benefit compared to the unprocessed reference condition was generally largest for algorithms applying DRC, especially for the interfering talker. One of the algorithms applying DRC was the *AdaptDRC* algorithm proposed by Schepker *et al.* (2013, 2015), which combines a time-dependent frequency-shaping stage with a time- and frequency-dependent DRC stage. Both the frequency-shaping and the DRC stage are adaptively controlled based on the estimated speech intelligibility. In conditions with high intelligibility the algorithm applies no changes to the original speech signal, while in conditions with low intelligibility the speech signal is modified. The degree of signal modification depends on a short-term estimate of the speech intelligibility index (SII) (ANSI, 1997).

Like all algorithms included in the study of Cooke *et al.* (2013b), the *AdaptDRC* algorithm works under the constraint that the root-mean square (rms) power of the modified

[a)] Electronic mail: jan.rennies@idmt.fraunhofer.de

[b)] Present address: Medical Physics Group, Department of Medical Physics and Acoustics and Cluster of Excellence Hearing4All, University of Oldenburg, D-26111 Oldenburg, Germany.

[c)] Also at: Project Group Hearing, Speech and Audio Technology, Fraunhofer Institute for Digital Media Technology (IDMT), D-26129 Oldenburg, Germany.

0001-4966/2017/141(4)/2526/12/$30.00

speech signal should not exceed the rms power of the original speech signal. This is a very reasonable constraint when comparing different algorithms, since simply increasing the speech power in a constant-noise environment would immediately improve speech intelligibility. For practical applications, however, this constraint may not always be realistic since many technical systems are limited with respect to their peak power rather than their rms power. Consequently, an increase in speech power may be applicable in adverse listening conditions if the playback system allows for a certain headroom. Even if no headroom is available, increasing the speech power may still be beneficial, namely, when distortions introduced by amplification (e.g., due to peak clipping) are outweighed by the increased speech power.

To which extent this compromise between speech distortions and an increased speech power affects speech intelligibility for NELE algorithms has not been investigated for the *AdaptDRC* algorithm or other NELE algorithms before. The first goal of the present study was therefore to extend the *AdaptDRC* algorithm to include an adaptive gain stage working under an equal-peak-power constraint. The underlying research hypothesis was that normal-hearing listeners benefit from additional amplification of the speech signal even at the cost of distortions caused by peak clipping.

The extended algorithm developed in this study comprises DRC (as the original *AdaptDRC* algorithm) and SNR enhancement (due to the additional adaptive gain stage), which are also two main processing stages applied in hearing aids to restore intelligibility. The second goal of this study was to investigate if a benefit for speech intelligibility in noise using NELE algorithms incorporating these stages can also be achieved for hearing-impaired listeners, even when the processing stages are not tuned to the individual hearing loss. This is particularly relevant because the large majority of the hearing impaired are not treated with hearing aids, especially in the groups of mild and moderate hearing loss. For example, Kochkin (2012) summarized that 60% of people with moderate to severe hearing loss and 81% of people with mild hearing loss do not own hearing aids. To the best of our knowledge, NELE algorithms for improving speech intelligibility in noise have not been evaluated with hearing-impaired listeners before. One exception is the study of Ben Jemaa *et al.* (2015), who tested different versions of non-adaptive, linear frequency-shaping derived from hearing-aid fitting rules, but did not find remarkable improvements (or even reduced intelligibility). Arai *et al.* (2010) tested a NELE approach for speech in reverberation, but not in noise. They showed that intelligibility could be improved for both normal-hearing and hearing-impaired listeners by attenuating high-energy vowel components to reduce overlap masking caused by reverberation. Azarov *et al.* (2015) proposed a NELE algorithm which may, in principle, be applicable for hearing-impaired listeners. However, their algorithm requires tuning to the individual hearing loss, and no validation using speech intelligibility tests was conducted. In the present study, we therefore measured speech intelligibility in both normal-hearing and hearing-impaired listeners. The research hypothesis was that the original and extended *AdaptDRC* algorithm are also effective for listeners with sensorineural hearing impairment, even without individualized tuning. Especially the SNR increase enabled by the new equal-peak-power constraint was expected to benefit hearing-impaired listeners.

II. EXTENSION OF THE *ADAPTDRC* ALGORITHM

This section first briefly reviews the *AdaptDRC* algorithm of Schepker *et al.* (2015), followed by a description of the algorithm extensions proposed in this study. Finally, the role of several algorithm parameters is analyzed.

A. *AdaptDRC* algorithm

The *AdaptDRC* algorithm was described in detail by Schepker *et al.* (2015), and in the present study the same notation is used to describe the proposed extensions. Figure 1 depicts the considered acoustic scenario. The general concept of the *AdaptDRC* algorithm is to process a clean speech signal $s[k]$ at discrete time index k using a processing stage $W\{\cdot\}$ to obtain the modified speech signal $\tilde{s}[k]$. For a given additive background noise $r[k]$ at the position of the listener, the goal is to increase the intelligibility of $\tilde{s}[k] + r[k]$ compared to $s[k] + r[k]$ under an equal-power constraint, i.e., the power of $\tilde{s}[k]$ should be equal to the power of $s[k]$. For the algorithm to adapt to time-varying background noises, an estimate of the noise $\hat{r}[k]$ has to be obtained from the microphone signal $y[k] = \tilde{s}[k] * h[k] + r[k]$, where the asterisk denotes convolution and h is the room impulse response between the loudspeaker and the microphone. Several methods exist to obtain a noise estimate in such conditions by using, e.g., adaptive filtering techniques to model the room impulse response $\hat{h}[k]$ (e.g., Haensler and Schmidt, 2008). The modified speech signal is computed as

$$\tilde{s}[k] = W\{s[k], \hat{r}[k], \hat{h}[k]\}. \quad (1)$$

Schepker *et al.* (2015) assumed perfect knowledge of $r[k]$ (i.e., $\hat{r}[k] = r[k]$) and no reverberation to be present (i.e., $\hat{h}[k] = h[k] = \delta[k]$). It was further assumed that the subband powers of both the speech and the noise recorded by the microphone are the same as perceived by the listener, which will generally not be fulfilled in real systems. The same assumptions are made in the present study. The processing $W\{\cdot\}$ consists of two stages: a frequency-shaping stage and a DRC stage, both of which depend on a short-term SII estimate.

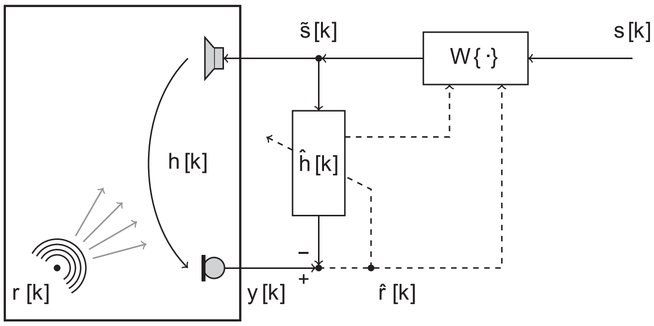

FIG. 1. Considered acoustic scenario (taken from Schepker *et al.*, 2015).

The signal $s[k]$ is split into $N = 8$ octave-band signals $s_n[k]$, $n = 1, \ldots, N$ using a real-valued non-decimated filterbank, with center frequencies ranging from 125 Hz to 16 kHz. Each subband signal $s_n[k]$ is framed into non-overlapping blocks of length M with the lth block denoted as $s_n^l[m] = s_n[lM + m]$, $m = 0, \ldots, M-1$. Correspondingly, the lth block of the time-domain broadband signal is denoted $s^l[m] = s[lM + m]$, $m = 0, \ldots, M-1$. For adaptively controlling the subsequent processing stages, an estimate of the short-term SII in each block, $\hat{SII}[l]$, is calculated from $s_n^l[m]$ and $\hat{r}_n^l[m]$, where $\hat{r}_n^l[m]$ has been defined similarly as $s_n^l[m]$.

(1) In the linear frequency-shaping stage, the subband signals $s_n^l[m]$ are weighted depending on the value of $\hat{SII}[l]$. For $\hat{SII}[l]$ close to 1, the subband weighting factors approach unity, i.e., the spectral shape of the speech signal is not modified. For $\hat{SII}[l]$ close to 0, Schepker *et al.* (2015) designed the subband weighting factors to result in the same power in each octave band, which effectively corresponds to a high-frequency amplification since speech signals generally contain more energy in lower frequency subbands. For intermediate values of $\hat{SII}[l]$, a continuous transition between an unmodified and a flat spectral shape is applied (for details, see Schepker *et al.*, 2015).

(2) In the subsequent DRC stage a transition between no DRC (compression ratio 1:1) and a maximum degree of DRC (compression ratio 1:8) is realized for each subband signal depending on the subband SNR used for the short-term SII estimate. The compressive gain according to the current input-output-characteristics of each band is derived from a smoothed short-term input level estimation (see Table I). To avoid noticeable artifacts due to large gain changes over time, especially at block boundaries, the gain for each subband derived from the frequency-shaping and DRC stages is smoothed recursively. The processed subband signal $\tilde{s}_n[k]$ is then obtained by applying the smoothed gain to the input subband signal $s_n[k]$. All subband signals are then recombined using an inverse filter bank to yield the time-domain broadband signal $\tilde{s}[k]$. To meet the equal-power constraint, a broadband normalization gain is applied to the time-domain signal to yield approximately equal rms powers (see Schepker *et al.*, 2015, for details).

TABLE I. Comparison of *AdaptDRC* time constants used by Schepker *et al.* (2015) and used in the present study. τ_a and τ_r denote the attack and release time constants of the short-term level estimation, respectively [see Eq. (19) of Schepker *et al.*, 2015]. τ_b denotes the time constant used to smooth input-output-characteristics of the DRC stage [see Eq. (25) of Schepker *et al.*, 2015]. τ_p denotes the time constant for smoothing broadband level changes of successive blocks [see Eq. (26) of Schepker *et al.*, 2015], and τ_L denotes the time constant used in the final normalization stage of the *AdaptDRC* algorithm.

Parameter	Schepker *et al.* (2015)	Present study
τ_a/ms	5	25
τ_r/ms	1	500
τ_b/ms	250	500
τ_p/ms	250	500
τ_L/ms	250	500

In the present study, the parameters of the *AdaptDRC* algorithm were the same as used by Schepker *et al.* (2015), except for the smoothing time constants. Schepker *et al.* (2015) applied temporal smoothing at several stages of the *AdaptDRC* algorithm to reduce audible artifacts. In general, rather small time constants were used. In the present study, larger time constants were used because informal listening tests indicated that larger time constants lead to slightly improved sound quality of the processed speech signal. Table I provides an overview of the time constants used by Schepker *et al.* (2015) and those used in the present study. The possible influence of these parameter changes is discussed in Sec. IV A.

B. *AdaptDRCplus* algorithm

In the extended algorithm, referred to as *AdaptDRCplus*, the *AdaptDRC* algorithm was extended with an adaptive gain stage. The equal-rms-power constraint was replaced by an equal-peak-power constraint, i.e., the rms power of each block $s^l[m]$ was allowed to be increased, while keeping the maximum amplitude constant. This boundary condition was selected in order to avoid the need for any assumptions about the incoming speech signal and the application scenario (such as technical headroom of the playback system). Practical implications and alternative formulations of the gain stage are discussed in Sec. IV.

A schematic block diagram of the proposed *AdaptDRCplus* algorithm is shown in Fig. 2. The input speech signal $s^l[m]$ is first processed by the *AdaptDRC* algorithm, including the blockwise broadband normalization to meet the equal-rms-power constraint. The processed block $\tilde{s}^l[m]$ at the output of the *AdaptDRC* algorithm is then linearly amplified according to

$$\tilde{s}'^l[m] = g[l]\tilde{s}^l[m],\ m = 0,\ \ldots,\ M-1, \tag{2}$$

where the gain function $g[l]$ is derived as follows: First, the maximum amplitude of the lth block is calculated as

$$\tilde{s}_{max}[l] = \max_m(|\tilde{s}^l[m]|). \tag{3}$$

The gain function $g[l]$ is then calculated as

$$g[l] = \frac{\tilde{s}_{max}[l]}{\tilde{s}_{q[l]}}, \tag{4}$$

where $\tilde{s}_{q[l]}$ is the amplitude of the sample which just exceeds the amplitude of $(100 - q[l])\%$ of the samples in the lth block $\tilde{s}^l[m]$. Applying the gain in Eq. (4) to $\tilde{s}^l[m]$ leads to a linearly amplified block signal of which $q[l]\%$ of the samples exceed $\tilde{s}_{max}[l]$. The degree of amplification is controlled by the estimated short-term SII calculated in the *AdaptDRC* stage according to

$$q[l] = (1 - \hat{SII}[l])q_{max}, \tag{5}$$

where q_{max} is a constant with $0\% \leq q_{max} \leq 100\%$. This corresponds to a linear mapping of the estimated short-term SII to the variable $q[l]$, resulting in $q[l] \rightarrow 0$ for $\hat{SII}[l] \rightarrow 1$, and

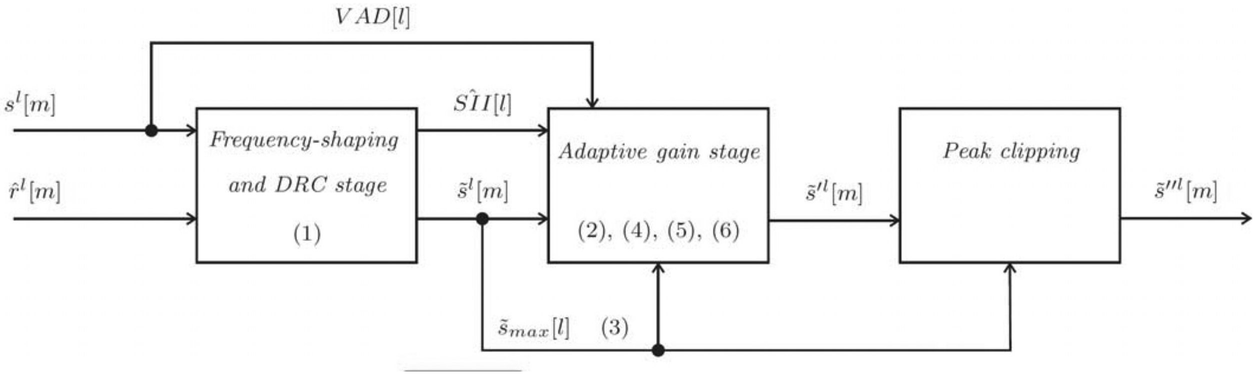

FIG. 2. Schematic diagram of the processing stages of the *AdaptDRCplus* algorithm. Equation numbers are indicated in parentheses.

$q[l] \to q_{max}$ for $\hat{SII}[l] \to 0$, i.e., Eq. (5) provides a transition between no gain and a maximum gain depending on the estimated SII in the lth block. This processing scheme results in an amplification for all blocks with $\hat{SII}[l] < 1$.

To avoid noticeable artifacts at block boundaries, the gain function $g[l]$ is smoothed recursively to obtain

$$\bar{g}[l] = \alpha_g \bar{g}[l-1] + (1-\alpha_g) g[l], \tag{6}$$

where α_g is a smoothing constant. The smoothed gain function $\bar{g}[l]$ is then applied in Eq. (2) instead of $g[l]$.

To avoid amplification in blocks with extremely low energy such as speech pauses, a voice activity detection (VAD) has been implemented based on the active speech level as defined in ITU (2011) and proposed by Kabal (1999). The VAD accounts for the fact that speech contains embedded pauses, i.e., so-called structural pauses shorter than 100 ms are considered to be part of the active speech segment, whereas pauses longer than 350 ms, e.g., grammatical pauses between phrases or before emphasis of specific words, are not counted as active speech. The samplewise VAD decisions are concatenated for the length of one block, where the block is defined to contain speech (i.e., $VAD[l] = 1$) if more than half of the samples in this block have been detected as active speech and otherwise $VAD[l] = 0$. In the *AdaptDRCplus* algorithm the gain for the lth block is set to unity for all blocks with $VAD[l] = 0$, otherwise it is calculated according to Eq. (6).

Since the energy of each block $\tilde{s}'^l[m]$ is larger than or equal to the energy of $\tilde{s}^l[m]$, it is natural to assume that, in a given noise condition, the intelligibility of $\tilde{s}'[k] + r[k]$ is enhanced compared to $\tilde{s}[k] + r[k]$. However, since practical sound playback systems may not allow for an increase in amplitude due to technical limitations, the peak amplitude of each block was again reduced to its original range. In the present study, this was realized by peak clipping, i.e., all samples with amplitudes exceeding $\tilde{s}_{max}[l]$ were set equal to $\tilde{s}_{max}[l]$, i.e., the output block $\tilde{s}''^l[m]$ was computed as

$$\tilde{s}''^l[m] = \max(\min(\tilde{s}'^l[m], \tilde{s}_{max}[l]), -\tilde{s}_{max}[l]), \quad m = 0, \ldots, M-1. \tag{7}$$

Effectively, this means that whenever additional gain is applied, the rms power of a block is increased and peak clipping is applied to fulfill the equal-peak-power constraint. Although peak clipping introduces nonlinear distortions of the speech signal, it was applied in this study to investigate the compromise between speech distortions and increased rms power on speech intelligibility. Alternative ways of introducing peak-power constraints and boundary conditions are discussed in Sec. IV.

C. Role of algorithm parameters

The *AdaptDRCplus* algorithm described above introduces two additional parameters: the constant q_{max}, determining the maximum percentage of samples in each block which are amplified beyond the peak amplitude (and which are then clipped), and the smoothing constant α_g. For $q_{max} = 0$, the gain stage of the *AdaptDRCplus* algorithm is disabled and the processing reduces to the *AdaptDRC* algorithm (i.e., $q[l] = 0$ and hence $\tilde{s}_{q[l]} = \tilde{s}_{max}[l]$, resulting in $g[l] = 1$). The smoothing constant α_g can be considered more intuitively in terms of a time constant τ_g, related to the smoothing constant by $\alpha_g = \exp(-\tau_m/\tau_g)$, where τ_m is the block length (20 ms in this case, i.e., the same block length as used by Schepker *et al.*, 2015).

For all analyses described in the following, speech and noise stimuli were taken from the Oldenburg sentence test (Wagener *et al.*, 1999a,b; Wagener *et al.*, 1999c), which was also used in the listening tests of the present study (see Sec. III B). A concatenation of ten randomly selected sentences from the test was used as speech signal. As described above, for $q_{max} = 0\%$, the *AdaptDRC* and *AdaptDRCplus* algorithms are equivalent. For $q_{max} > 0\%$, the gain stage of the *AdaptDRCplus* algorithm results in an increased speech rms power when $\hat{SII} < 1$, corresponding to an increased SNR at the algorithm output compared to the algorithm input. Figure 3 illustrates the SNR increase as a function of input SNR for the three noise types used in this study (cafeteria noise, speech-shaped noise, and car noise). For $q_{max} = 0\%$ (dashed black line) the SNR increase was always equal to 0 dB due to the equal-power constraint of the *AdaptDRC* algorithm. Increasing q_{max} to values between 2% and 50% (solid lines in different gray scales) resulted in a monotonous increase in SNR. At the lowest input SNR, the SNR increase was between about 2.5 dB ($q_{max} = 2\%$) and 9 dB ($q_{max} = 50\%$) for each noise type. The change in SNR decreased with increasing input SNRs due to the adaptive control of the gain stage. This decrease was similar for speech-shaped noise and car noise, and somewhat less steep for cafeteria noise.

To estimate how much the signal modifications introduced by the proposed algorithm affect speech intelligibility, model predictions were computed using the SII (ANSI, 1997), the extended SII (ESII) (Rehbergen and Versfeld, 2005), and the short-time objective intelligibility measure (STOI) (Taal *et al.*, 2011). Speech-shaped noise was used as interferer. Each of the models computes an index between 0 and 1, where values of 0 and 1 indicate lowest and highest possible extraction of speech information, respectively. For the present evaluation, the predicted improvements of speech intelligibility relative to unprocessed speech were calculated for both the *AdaptDRC* and the *AdaptDRCplus* algorithm. Figure 4 depicts the differences in model index as a function of (input) SNR, where each panel shows predictions of one model. An index difference of 0 indicates no predicted improvement in speech intelligibility. Solid black lines in each panel represent data of the *AdaptDRC* algorithm. In agreement with the results of Schepker *et al.* (2015), an increase in speech intelligibility was predicted by all models. The increase was largest at SNRs between about −15 and −5 dB, and decreased towards lower and higher SNRs. Solid lines in different gray scales represent predictions for the *AdaptDRCplus* algorithm with values of q_{max} ranging between 2% and 50% and a fixed value of $\tau_g = 150$ ms. For all models, increasing q_{max} resulted in an increased speech intelligibility. The SII and the ESII did not show any noticeable saturation effect, indicating that a further increase of q_{max} may have resulted in a further increase of predicted speech intelligibility. The STOI model indicated that improvements were largest when q_{max} was increased from 0% to 2% and from 2% to 5%, while a further increase in q_{max} resulted only in smaller improvements. Even though the model predictions showed reasonable correlation with subjective data in the study of Schepker *et al.* (2015), they cannot be considered reliable enough to fully guide a selection of a good value of q_{max}, especially since SII and ESII are largely spectral measures which tend to consider high-frequency speech distortions as useful speech energy irrespective of their detrimental nature, because frequency bands between 1 and 4 kHz are given

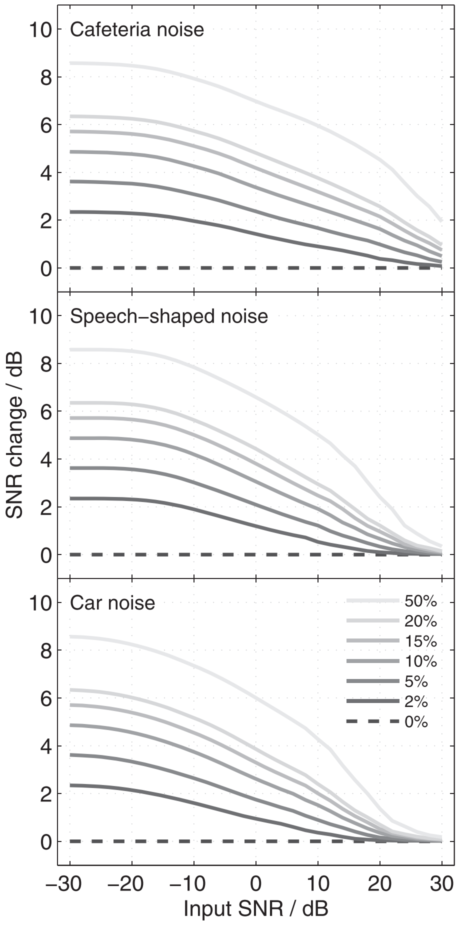

FIG. 3. SNR increase as a function of input SNR for different values of q_{max}. The dashed black curve represents the *AdaptDRC* algorithm ($q_{max} = 0\%$), the other curves represent data for the *AdaptDRCplus* algorithm for different values of $q_{max} > 0\%$.

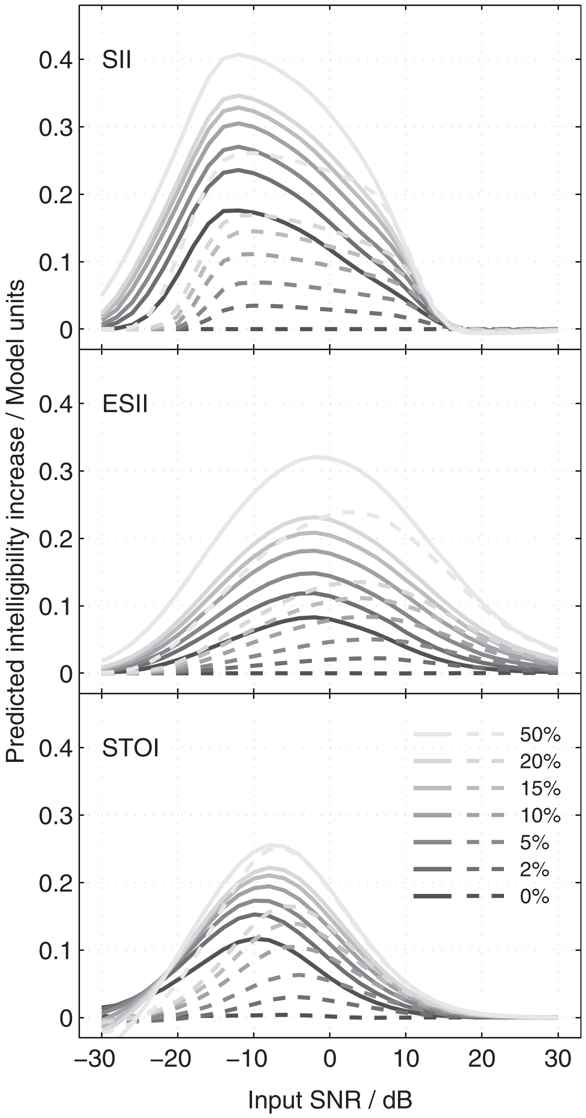

FIG. 4. Increase in speech intelligibility compared to unprocessed speech in speech-shaped noise as predicted by the SII (top panel), the ESII (mid panel), and the STOI measure (bottom panel) as a function of input SNR. Solid black curves represent predictions for the *AdaptDRC* algorithm ($q_{max} = 0\%$), the other solid curves represent predictions for the *AdaptDRCplus* algorithm for different values of $q_{max} > 0\%$. Dashed lines show predictions when only using the adaptive gain stage, i.e., without the *AdaptDRC* algorithm. τ_g was 150 ms.

more weight than lower frequency bands (ANSI, 1997). To set q_{max} in this study, informal listening tests were carried out by three experienced listeners using the same speech material and noise type as used for the model predictions. The listeners listened to speech in noise mixed at various SNRs for the same values of q_{max} as used for the model predictions, and ranked the benefit of q_{max} with respect to speech intelligibility and speech quality. The informal listening confirmed that a benefit from the gain stage may be expected, but that the benefit is not likely to increase further for very large values of q_{max} due to the increased amount of distortion. As a compromise, a value of $q_{max} = 20\%$ was selected and used in the formal listening tests.

Using the same models to evaluate the influence of the time constant τ_g showed that model predictions were largely insensitive to changes in τ_g between 0 and 500 ms for a fixed value of $q_{max} = 20\%$ (not shown). Only predictions with the STOI model tended to favor larger time constants over very small ones. In informal listening tests the preference for larger time constants was confirmed, since too small values of τ_g resulted in an increased audibility of artifacts. Therefore, a value of $\tau_g = 100$ ms was used in the following formal listening tests.

III. EXPERIMENTAL ASSESSMENT OF SPEECH INTELLIGIBILITY

Three experiments were conducted in this study. In experiment 1, speech intelligibility was measured in normal-hearing listeners. In experiments 2 and 3, the possible benefit of the *AdaptDRC* and *AdaptDRCplus* algorithms was evaluated in hearing-impaired listeners.

A. Subjects

The speech intelligibility measurements were conducted with eleven normal-hearing listeners (three female, eight male) and ten listeners with mild to moderate hearing impairment (six female, four male). The normal-hearing subjects were between 20 and 58 years old (median age 26.0 years). The hearing-impaired listeners were between 46 and 78 years old (median age 73.5 years) and had typical age-related high-frequency hearing loss with better-ear pure-tone-averages between 34 and 51 dB hearing level (HL). Their average audiogram is shown as black lines in Fig. 5. Individual audiograms are shown in gray. Subjects were paid an hourly compensation for their participation.

B. Stimuli and equipment

As in the previous evaluation of the *AdaptDRC* algorithm (Schepker *et al.*, 2015), the speech material of the Oldenburg sentence test (Wagener *et al.*, 1999a,b; Wagener *et al.*, 1999c) was used in this study. This material consists of five-word sentences with the fixed syntactical structure *name verb numeral adjective object*. For each of the five words ten alternatives are available, which can be randomly combined to result in grammatically correct, but semantically unpredictable sentences.

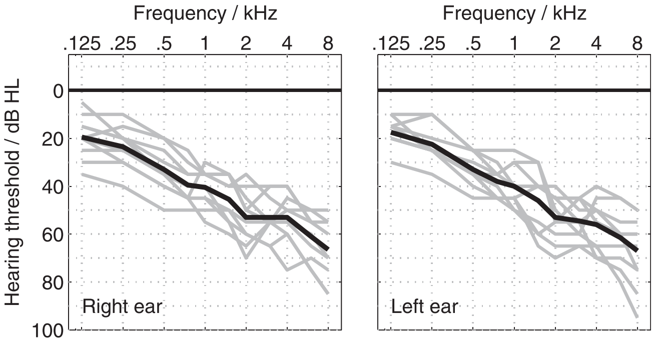

FIG. 5. Average (black) and individual (gray) audiograms of the hearing-impaired listeners.

For the group of normal-hearing listeners (experiment 1) the speech material was scaled to 60 dB sound pressure level (SPL). Three processing conditions of the speech material were used in this study, i.e., unprocessed speech (reference) as well as speech processed by *AdaptDRC* and its proposed extension *AdaptDRCplus*. For speech processed by *AdaptDRCplus*, the speech level typically differs from the input level of 60 dB SPL, depending on the SNR and the noise type. Different types of interfering noise were mixed with the speech material, and the level of the noise was varied to achieve the desired SNRs (see below). For experiment 1 the same noise types as used by Schepker *et al.* (2015) were used, i.e., a nonstationary noise mimicking a cafeteria environment, a stationary speech-shaped noise with a long-term spectrum matching that of the speech material, and a stationary car noise. For each noise three different SNRs were selected based on the results of Schepker *et al.* (2015) (see Table II for an overview). The SNR corresponding to a word recognition rate of about 50% in the unprocessed reference condition was included and measured without processing (top row of each noise) to facilitate the comparison of the present data to the data of the previous study. The selection of the other SNRs was motivated by the expected speech intelligibility for the speech material processed by

TABLE II. SNRs used for normal-hearing listeners in the different conditions (experiment 1).

Noise type	Input SNR/dB	Algorithm
Cafeteria noise	−10	Unprocessed
	−14	*AdaptDRC*
	−14	*AdaptDRCplus*
	−18[a]	*AdaptDRC*
	−18[a]	*AdaptDRCplus*
Speech-shaped noise	−9	Unprocessed
	−17	*AdaptDRC*
	−17	*AdaptDRCplus*
	−21[a]	*AdaptDRC*
	−21[a]	*AdaptDRCplus*
Car noise	−16	Unprocessed
	−24	*AdaptDRC*
	−24	*AdaptDRCplus*
	−28[a]	*AdaptDRC*
	−28[a]	*AdaptDRCplus*

[a]Conditions in which the overall level was reduced by 4 dB.

the algorithms. Schepker *et al.* (2015) found that intelligibility of speech processed by the *AdaptDRC* algorithm was quite high for SNRs corresponding to 50% recognition rate in the unprocessed condition. To allow for a differentiation between processing conditions (i.e., to avoid ceiling effects), one of the lower SNRs used by Schepker *et al.* (2015) as well as another SNR 4 dB lower was included in the present study. The goal was to allow for a direct comparison to the previous data as well as to test the novel *AdaptDRCplus* algorithm at even lower SNRs. However, since this would have resulted in noise levels of up to 88 dB SPL (e.g., for car noise), the overall level of the mixed speech and noise stimuli was reduced by 4 dB for the lowest SNRs for each noise type.

For the hearing-impaired listeners (experiments 2 and 3), the speech material was presented at a fixed level of 65 dB SPL. Only cafeteria noise was used to avoid an overly long measurement duration. This noise type was selected, because the benefit of the *AdaptDRC* algorithm as measured by Schepker *et al.* (2015) was smaller than for the other types of noise, i.e., this interferer provided the most challenging background noise for this algorithm (and also for another NELE algorithm included in the study of Schepker *et al.*, 2015). Since no reference data existed to estimate intelligibility of speech processed by NELE algorithms at different SNRs for this group of listeners, and because a larger interindividual variability was expected, the measurements were divided into two experiments. In experiment 2 the speech recognition threshold (SRT), i.e., the SNR corresponding to a word recognition rate of 50%, was measured for each subject and algorithm (see below for procedural details). These results were then used in experiment 3, in which speech intelligibility was measured at three fixed SNRs corresponding to the individual SRT and the SRT ± 4 dB, respectively. The difference of ±4 dB was chosen because it roughly corresponded to intelligibility scores of 80% and 20%, respectively, in the study of Schepker *et al.* (2015). This way a reasonably accurate sampling of the individual psychometric functions was expected, allowing for a quantitative comparison of the three processing schemes for hearing-impaired listeners.

Speech and noise were digitally processed using MATLAB. The stimuli were preprocessed at the desired SNRs and stored on a hard-drive prior to the listening test. An RME Fireface UC soundcard was used and the stimuli were presented diotically to the subjects via Sennheiser HD650 headphones in a sound-attenuated booth.

C. Procedure

For each experiment and condition, the speech material was presented to the subjects in lists of 20 sentences. After the presentation of each sentence, the task of the subjects was to repeat the words they had recognized and an instructor marked the correctly recognized words on a touch screen. After confirming the input, the presentation of the next sentence started automatically. In experiments 1 and 3, during which the SNR was constant for each list, speech intelligibility was quantified as the percentage of correctly understood words. In experiment 2 to determine individual SRTs for hearing-impaired listeners, an adaptive procedure was applied, i.e., the speech level was fixed at 65 dB SPL and the noise level was adjusted after each sentence depending on the response of the subject to converge to the threshold of 50% word intelligibility. The initial SNR was 0 dB, and the step size of each level change depended on the number of correctly repeated words of the previous sentence and on a convergence factor that decreased exponentially after each reversal of presentation level. The smallest level change was 1 dB. The intelligibility function was represented by a logistic function which was fitted to the data using a maximum likelihood method resulting in an estimated SRT (for details, see Brand and Kollmeier, 2002). These SRTs were then rounded to full dB values and the signals in experiment 3 were selected from stored preprocessed stimuli accordingly. Prior to the measurements all subjects received at least three lists of training using the unprocessed speech material. The data of these lists were discarded to minimize training effects, which are typical to occur for this kind of speech material (see Wagener *et al.*, 1999b).

D. Results

1. Experiment 1: Normal-hearing listeners

Mean speech intelligibility data as a function of input SNR measured with normal-hearing listeners are shown as black symbols in Fig. 6. Error bars represent interindividual standard deviations. Each panel contains data for one noise type. Comparing data for unprocessed speech (triangles) and speech processed by the *AdaptDRC* algorithm (squares) indicated a considerable benefit in speech intelligibility due to the processing. This benefit depended on the noise type: for cafeteria noise, speech intelligibility in the unprocessed reference condition was between that measured for the *AdaptDRC* algorithm at 4 and 8 dB lower SNRs. For speech-shaped noise and car noise, intelligibility of unprocessed speech was similar to or even lower than intelligibility of speech processed by the *AdaptDRC* algorithm at 12 dB lower SNRs.

Comparing speech intelligibility for the *AdaptDRC* algorithm and its extension *AdaptDRCplus* at the same input SNR (vertical distance between squares and circles) showed a benefit due to the extension proposed in this study. For a given noise type, the increase in speech intelligibility was similar for both measured SNRs, and was about 35% (cafeteria noise), 39% (speech-shaped noise), and 23% (car noise). T-tests conducted for each SNR and noise type showed that the differences between the two algorithms were always significant at a confidence level of 5% ($p < 0.001$ for cafeteria noise, $p \leq 0.002$ for speech-shaped noise, and $p \leq 0.01$ for car noise).

2. Experiment 2: SRTs of hearing-impaired listeners

Table III shows the results of experiment 2. In the unprocessed condition, mean SRTs (last column) were −4.0 dB, but a large intersubject variability was observed with SRTs ranging from −8.5 dB (subject #9) to −0.8 dB (subject #4). For the *AdaptDRC* and *AdaptDRCplus*

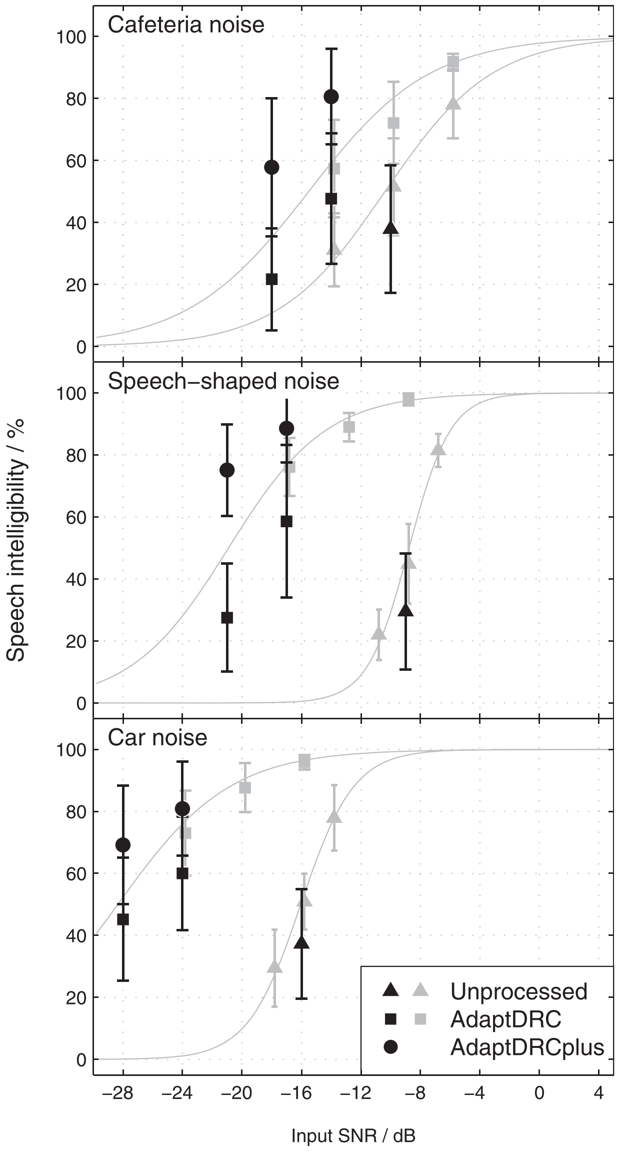

FIG. 6. Results of the speech intelligibility measurements with normal-hearing listeners for cafeteria noise (top panel), speech-shaped noise (mid panel), and car noise (bottom panel). Speech intelligibility is shown as a function of input SNR for the reference condition (black triangles), *AdaptDRC* (black squares), and *AdaptDRCplus* (black circles). Gray curves and symbols represent speech intelligibility data and psychometric functions as measured by Schepker *et al.* (2015) for the reference condition and *AdaptDRC*. Note that these data have been shifted to the right by 0.2 dB.

algorithm, mean SRTs were 1.5 and 6.8 dB lower, respectively. The difference between SRTs for processed and unprocessed speech was termed equivalent intensity changes (EIC) by Cooke *et al.* (2013b) and represents the change in SNR that can be applied to maintain a constant intelligibility score (i.e., 50% speech intelligibility in this experiment). With a single exception (subject #1), all subjects showed a benefit due to processing by the *AdaptDRC* algorithm (i.e., negative EICs), but again EICs varied substantially across subjects ranging from less than −1 dB (subjects #4, #8, and #9) to more than −2 dB (subjects #2, #3, #7, and #10). The largest EIC of −3.6 dB was observed for subject #2. The EICs for the *AdaptDRCplus* algorithm were much larger for all subjects, ranging from −4.2 dB (subject #1) to −9.6 dB (subject #2).

These observations were supported by an analysis of variance (ANOVA) with the factor *processing* (three levels: unprocessed, *AdaptDRC*, and *AdaptDRCplus*) after verifying that the data were normally distributed (according to Shapiro-Wilk tests). The ANOVA showed that the processing condition had a significant effect on measured SRTs [F(2,18) = 169.37, $p < 0.001$]. A *post hoc* analysis using Bonferroni corrections for multiple comparisons showed that both the *AdaptDRC* algorithm ($p = 0.021$) and the *AdaptDRCplus* algorithm ($p < 0.001$) significantly improved SRTs compared to unprocessed speech, and that the difference between *AdaptDRC* and *AdaptDRCplus* was significant ($p < 0.001$).

3. Experiment 3: Psychometric functions of hearing-impaired listeners

Individual results of experiment 3 are shown in Fig. 7. Symbols represent intelligibility measured at the individual SRTs (rounded to full dB values) and at SRTs ± 4 dB. Lines represent individual psychometric functions, which were obtained by fitting the model function (Brand and Kollmeier, 2002)

$$SI(SNR) = \frac{100}{1+e^{-4s_{50}\cdot(SNR-SRT_{50})}} \qquad (8)$$

to the data. The two degrees of freedom, SRT_{50} and s_{50}, represent the SNR at an intelligibility score of 50% and the slope of the function at this point, respectively. The observations from experiment 2 were generally also valid for this experiment. In particular, all subjects showed a benefit due to processing by *AdaptDRCplus* (corresponding to a leftward shift of the psychometric function), while the benefit due to processing by *AdaptDRC* was smaller or even absent (subject #1).

This is also reflected in the average psychometric functions shown as lines in Fig. 8, which were obtained by parametrically averaging the individual psychometric functions. To obtain a visual illustration of the interindividual spread of these average functions, Eq. (8) was calculated four times, employing all combinations of the mean SRT_{50} ± 1 standard error and the mean s_{50} ± 1 standard error. The minimum and

TABLE III. Adaptively measured SRTs in dB SNR for the group of hearing-impaired subjects. The last column contains mean SRTs ± 1 standard deviation (Std).

Subject	#1	#2	#3	#4	#5	#6	#7	#8	#9	#10	Mean ± Std
Unprocessed	−6.3	−1.5	−3.1	−0.8	−7.0	−4.3	−3.0	−1.9	−8.5	−3.9	−4.0 ± 2.4
AdaptDRC	−5.4	−5.1	−5.4	−1.6	−8.3	−5.8	−5.3	−2.4	−8.9	−6.5	−5.5 ± 2.1
AdaptDRCplus	−10.5	−11.1	−10.1	−6.6	−15.2	−11.1	−9.4	−7.7	−15.0	−11.2	−10.8 ± 2.6

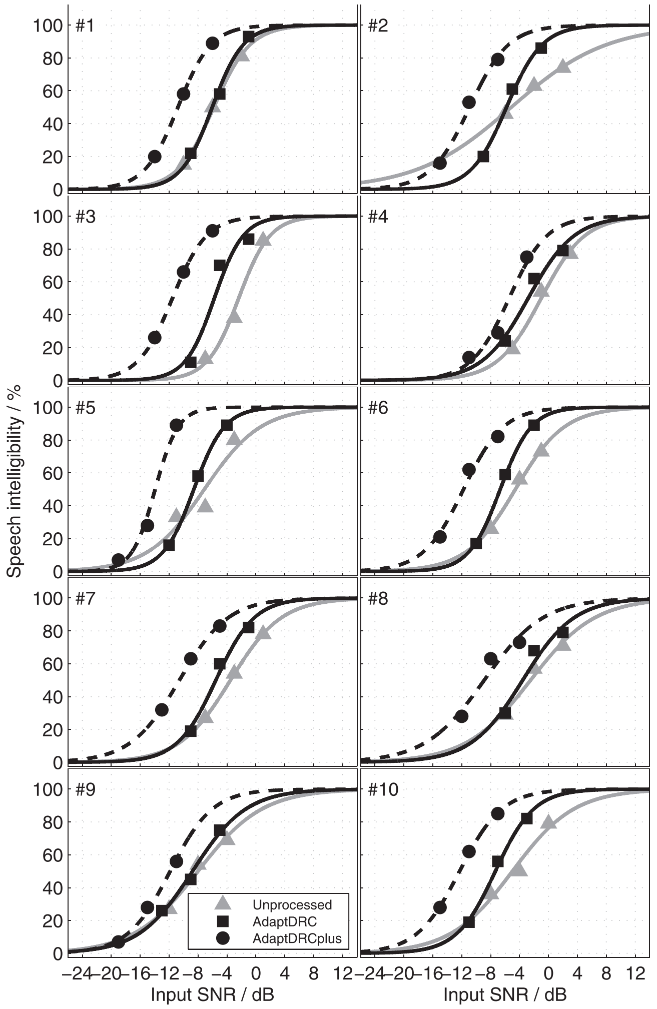

FIG. 7. Individual speech intelligibility scores for unprocessed speech (gray triangles) and speech processed by the *AdaptDRC* algorithm (black squares) and the *AdaptDRCplus* algorithm (black circles). Curves represent psychometric functions fitted to the data. Subject numbers are indicated in each panel.

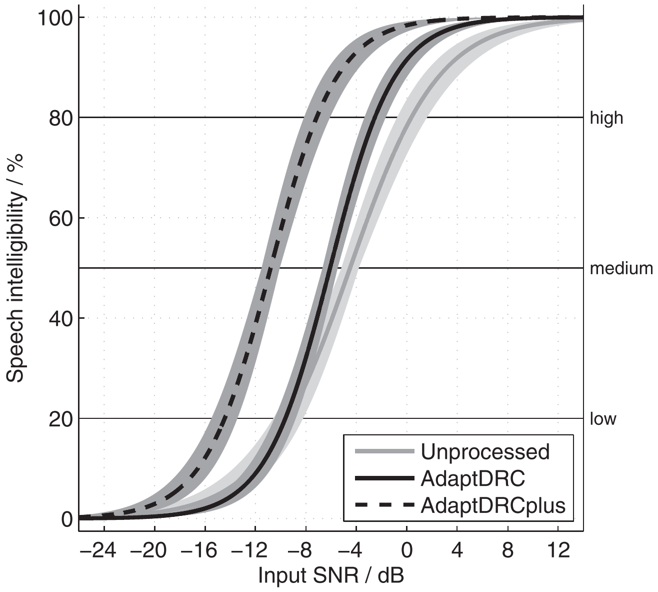

FIG. 8. Average psychometric functions showing speech intelligibility as a function of input SNR for the group of hearing-impaired subjects. Shaded areas provide a visual reference for the standard errors resulting from the parametric averaging of individual psychometric functions.

maximum values of these four functions at each SNR are highlighted as shaded areas in Fig. 8. The curves indicate a considerable benefit due to the *AdaptDRCplus* algorithm, while the benefit due to the *AdaptDRC* algorithm is smaller or even absent at lower SNRs. To explore the statistical significance of these observations, three intelligibility scores were defined as representatives for low (20%), medium (50%), and high speech intelligibility scores (80%). The SNRs required to achieve these scores were derived from the individual psychometric functions and subjected as independent variable to a two-way ANOVA with factors *processing condition* (three levels: unprocessed, *AdaptDRC*, *AdaptDRCplus*) and *intelligibility score* (three levels: 20%, 50%, and 80%). The analysis revealed that both main effects of *processing condition* [$F(2,18) = 146.458$, $p < 0.001$] and *intelligibility score* [$F(2,18) = 224.163$, $p < 0.001$] as well as their interaction [$F(4,36) = 7.519$, $p < 0.001$] were significant. To further explore the sources of significance, a separate one-way ANOVA was conducted for each intelligibility score. In each case, the influence of *processing condition* was significant [$p < 0.001$, $F(2,18) = 33.482$, $F(2,18) = 146.443$, and $F(2,18) = 121.914$ for the low, medium, and high intelligibility score, respectively]. *Post hoc* comparisons showed that, at an intelligibility score of 20%, the difference between *AdaptDRCplus* and both *AdaptDRC* and unprocessed speech was significant ($p \leq 0.001$ for both cases), but the difference between *AdaptDRC* and unprocessed was not significant ($p = 1.000$). At medium and high intelligibility scores (i.e., 50% and 80%), all differences were significant with $p \leq 0.001$, i.e., *AdaptDRCplus* increased speech intelligibility significantly over both *AdaptDRC* and unprocessed, and *AdaptDRC* increased speech intelligibility significantly over unprocessed speech. An overview of the EICs determined at the considered intelligibility scores is provided in Table IV.

IV. DISCUSSION

A. Near-end listening enhancement for normal-hearing subjects

The present study confirmed the results of Schepker *et al.* (2015) in that a considerable benefit in speech intelligibility could be achieved by the *AdaptDRC* algorithm compared to unprocessed speech for normal-hearing subjects. Gray curves and symbols in Fig. 6 represent data measured by Schepker *et al.* (2015). The EICs (i.e., the leftward shifts of the data points in Fig. 6 compared to unprocessed speech) were very similar to the previous study. In contrast, the absolute intelligibility scores showed a consistently lower performance compared to the study of Schepker *et al.* (2015) for the same speech and noise material and SNRs. This intelligibility offset was very similar for all comparable conditions. In particular, it was also similar for unprocessed speech and speech processed by the *AdaptDRC* algorithm (between 14% and 22%). It is therefore likely that the observed differences represent an effect of subject group rather than an impact of the modified smoothing time constants, which could only

TABLE IV. EICs in dB for processed speech compared to unprocessed speech derived from the individual psychometric functions shown in Fig. 7 for speech intelligibility scores of 20% (SI$_{20}$), 50% (SI$_{50}$), and 80% (SI$_{80}$). The last column contains mean EICs ± 1 standard deviation.

Subject		#1	#2	#3	#4	#5	#6	#7	#8	#9	#10	Mean ± Std
AdaptDRC	SI$_{20}$	+0.3	+5.0	−3.1	−2.0	+1.0	−0.4	−0.8	+0.1	+0.3	−0.1	0.0 ± 2.0
	SI$_{50}$	−0.2	−0.7	−3.4	−1.8	−1.5	−2.1	−2.0	−1.1	−0.8	−2.4	−1.6 ± 0.9
	SI$_{80}$	−0.7	−6.4	−3.6	−1.6	−3.9	−3.8	−3.2	−2.3	−1.8	−4.6	−3.2 ± 1.6
AdaptDRCplus	SI$_{20}$	−4.7	−0.4	−9.4	−3.9	−3.6	−6.3	−6.9	−5.9	−1.8	−5.5	−4.8 ± 2.5
	SI$_{50}$	−5.0	−5.8	−9.2	−4.4	−6.8	−7.2	−7.1	−6.4	−3.9	−7.3	−6.3 ± 1.5
	SI$_{80}$	−5.3	−11.2	−8.9	−5.0	−10.0	−8.0	−7.3	−6.9	−6.0	−9.2	−7.8 ± 2.0

have affected processed speech. However, an influence of the time constants cannot be excluded based on the present data since no direct comparison was made within the same subjects.

Compared to the *AdaptDRC* algorithm, a further increase in intelligibility by between 22% and 45% was observed for speech processed by the *AdaptDRCplus* algorithm at the same input SNRs. This shows that the increase in rms power at the cost of introducing distortions caused by peak clipping was beneficial in terms of speech intelligibility. In other words, the speech degradation was outweighed by the increase in SNR. This is not surprising since the auditory system is known to be highly robust against peak clipping in terms of speech intelligibility (e.g., Young *et al.*, 1978).

The observed benefit is generally in line with the trends predicted by the speech intelligibility models presented in Sec. II C. In addition to predictions for the *AdaptDRCplus* algorithm (shown as solid lines in Fig. 4), dashed lines show the predicted speech intelligibility increase if the adaptive gain stage of the *AdaptDRCplus* algorithm is considered in isolation, i.e., without prior processing with the *AdaptDRC* algorithm. Different gray scales represent different values of q_{max}. One general observation is that model predictions for the *AdaptDRCplus* algorithm are always above those for the isolated gain stage. Although no subjective data were collected for speech processed by the isolated gain stage, these model predictions suggest that the combination of *AdaptDRC* preprocessing and gain stage would always be superior to only using the gain stage. Even if one assumes that the introduced distortions due to the peak clipping do not negatively affect speech intelligibility, the maximum theoretically achievable EIC of the isolated gain stage at the selected value of $q_{max} = 20\%$ is about −7 dB at very low SNRs (Fig. 3). In contrast, Schepker *et al.* (2015) reported EICs of up to −12 dB for the *AdaptDRC* algorithm alone for speech-shaped noise and car noise. The present study showed that further improvements (which would correspond to even larger EICs) could be achieved by the *AdaptDRCplus* algorithm if they were measured directly. In conclusion, the combination of *AdaptDRC* processing and the proposed adaptive gain stage is considerably more beneficial than an adaptive gain stage alone.

The relative benefit of gain stage and *AdaptDRC* or *AdaptDRCplus* processing is not predicted correctly by the models considered in this study. Although the predicted relative benefit differs somewhat between models, all models predict that the maximum speech intelligibility improvements should be similar for the *AdaptDRC* algorithm (black solid line, $q_{max} = 0\%$) and an isolated gain stage with q_{max} values of about 10% (STOI), 15% (ESII), or 20% (SII). For the gain stage alone, this would correspond to SNR increases between about 4.5 and 6 dB at an input SNR of −10 dB for speech-shaped noise (Fig. 3). As discussed above, these SNR increases correspond to the maximum theoretically achievable EICs for the isolated gain stage. This is considerably smaller than the EIC of −12 dB reported by Schepker *et al.* (2015) for the *AdaptDRC* algorithm, indicating that the relative benefit of the gain stage is overestimated by the models. This also illustrates that predictions of the tested models are not reliable enough serve as the only source for parameter optimizations, even though their predictions were in reasonable agreement with measured speech intelligibility in the previous study (Schepker *et al.*, 2015).

B. Near-end listening enhancement for hearing-impaired subjects

To the best of our knowledge, the benefit of NELE algorithms including DRC has not been evaluated with unaided hearing-impaired listeners before. The present study showed that this group of subjects could also benefit from preprocessing of speech. In experiment 2, a significant decrease in SRT was found for both the *AdaptDRC* and the *AdaptDRCplus* algorithm. The mean SRT decrease of 1.5 dB (*AdaptDRC*, Table III) was considerably smaller than the SRT decrease of about 5 dB reported for normal-hearing listeners using the same background noise. For the group of hearing-impaired listeners, the average EIC measured for the *AdaptDRCplus* algorithm was −6.8 dB, which is slightly more than the theoretically achievable benefit from the gain stage alone at the employed input SNRs [even for the subject with the lowest SRT for unprocessed speech of −8.5 dB (see Table III), the SNR increase was below 6 dB, see Fig. 3]. This indicates that there was also a benefit of the combined processing of the *AdaptDRC* algorithm and the gain stage. While not all subjects benefited from *AdaptDRC* processing, all showed considerable benefit from the additional gain stage of the *AdaptDRCplus* algorithm. The reasons underlying the large interindividual variations remain unclear. It is not evident from the data of experiment 2 that the baseline performance of the hearing-impaired listeners was related to the benefit due to *AdaptDRC* processing. In fact, the largest EIC was measured for subject #2, who had a comparatively poor SRT for unprocessed speech (Table III). Vice versa, subject #1 had a relatively good SRT for unprocessed speech, but did not benefit from *AdaptDRC* processing at all. Across all listeners, the baseline performance for unprocessed speech was not significantly correlated with the benefit from *AdaptDRC* processing (Pearson's $R = 0.46$, $p = 0.18$), nor with the benefit from *AdaptDRCplus* processing ($R = 0.14$, $p = 0.79$). Similarly, the degree of hearing loss as expressed by the PTA was not significantly correlated with any of the measured SRTs or any of the EICs (all p-values > 0.10).

C. Possible extensions of the proposed algorithm

Since the focus of the present study was on speech intelligibility, possible effects of the proposed processing on perceived speech quality cannot be assessed based on the current data. Although the SNR-dependent nature of the proposed algorithms prevents signal modifications (and hence distortions) for conditions with high SNR, it is likely that the introduced distortions negatively impact perceived quality at lower SNRs, even if some of the distortions are masked by the environmental noise. The optimum trade-off between increased speech intelligibility and reduced speech quality remains a topic of further investigation.

It should be noted that the boundary condition of a maximum peak level per block was set as an artificial limitation of the algorithm. In practical applications, such as public-address systems, knowledge about the employed playback system and speech material may be used to set the boundary conditions in a more accurate way. For example, if the entire speech material (e.g., recorded announcement signals) and the headroom of the playback system are known, the peak level not to be exceeded can be set globally without introducing distortions in each block. This would further increase the output level of speech processed by the *AdaptDRCplus* algorithm, since many samples exceeding the per-block peak clipping limit applied in the present algorithm would not be clipped any more. Thus, one may expect a beneficial effect of a more global peak limiting approach both in terms of speech intelligibility (higher output level) as well as speech quality (less distortions). In the present study, we decided not to make use of such knowledge about the system. This way, the same principle can be repeated with other speech materials, systems and algorithms, which may facilitate comparability in the future. In this light, the present evaluation results can be interpreted as a lower limit of the possible benefit because the applied constraint introduces more distortions than necessary in real systems. As an alternative or in combination, other methods of limiting the peak level which introduce less audible distortions may be applied, e.g., soft clipping with smoother input-output characteristics rather than hard clipping (e.g., Birkett and Goubran, 1996) or even less perceptually invasive clipping techniques (e.g., Defraene *et al.*, 2012). For evaluating such approaches, it would be useful to not only focus on speech intelligibility, but also determine the perceived speech quality or the preference for different processing schemes. Similarly, it would be interesting to evaluate listening effort rather than speech intelligibility. Listening effort has been shown to be a reliable measure for more favorable listening conditions, i.e., for conditions with moderately negative and positive SNRs (e.g., Sato *et al.*, 2005; Klink *et al.*, 2012; Rennies *et al.*, 2014). Although this SNR range is typically more realistic, speech intelligibility is often optimal in terms of correctly recognized words (see Fig. 6), such that speech intelligibility cannot be used to quantify a possible benefit of NELE algorithms at higher SNR. In contrast, a useful distinction between different listening conditions could still be made when listening effort was evaluated, enabling to evaluate the effectiveness of NELE algorithms over much wider SNR ranges.

V. CONCLUSIONS

The following conclusions can be drawn from the present study.

- For normal-hearing listeners the combination of the *AdaptDRC* algorithm proposed by Schepker *et al.* (2015) and the newly integrated adaptive gain stage can be highly beneficial, i.e., both the original *AdaptDRC* processing as well as the adaptive gain stage amplifying the speech signal under an equal-peak-power constraint at the cost of distortions provide considerable improvements.
- In contrast, the benefit from the *AdaptDRC* algorithm for hearing-impaired listeners varies substantially across subjects and is—on average—considerably smaller than for normal-hearing listeners. A larger benefit was only observed for the additional adaptive gain stage.
- For hearing-impaired listeners the benefit from the proposed algorithms is smaller in very difficult listening conditions (i.e., at lower ranges of the psychometric function) than at higher SNRs.
- Speech intelligibility models cannot predict all trends observed in the data. In particular, the role of amplification is overestimated compared to the processing of the *AdaptDRC* algorithm (which works under an equal-rms-power constraint).

ACKNOWLEDGMENTS

This work was supported in part by the Research Unit FOR132 "Individualized Hearing Acoustics" and the Cluster of Excellence 1077 "Hearing4All," funded by the German Research Foundation (DFG). This work was supported in part by the German Federal Ministry of Education and Research (FK 16SV7194).

ANSI (**1997**). ANSI S3.5-1997, *Methods for Calculation of the Speech Intelligibility Index* (American National Standards Institute, New York).

Arai, T., Hodoshima, N., and Yasu, K. (**2010**). "Using steady-state suppression to improve speech intelligibility in reverberant environments for elderly listeners," IEEE Trans. Audio Speech Lang. Process. **18**(7), 1775–1780.

Azarov, E., Vashkevich, M., Herasimovich, V., and Petrovsky, A. (**2015**). "General-purpose listening enhancement based on subband non-linear amplification with psychoacoustic criterion," in *Proceedings of the AES Convention* 138, Warsaw, Poland, May 2015, Paper No. 9265.

Ben Jemaa, A., Mechergui, N., Courtois, G., Mudry, A., Djaziri-Larbi, S., Turki, M., Lissek, H., and Jaidane, M. (**2015**). "Intelligibility enhancement of vocal announcements for public address systems: A design for all through a presbycusis pre-compensation filter," in *Proceedings of Interspeech*, Dresden, Germany, Sep. 2015, pp. 70–74.

Birkett, A. N., and Goubran, R. A. (**1996**). "Nonlinear loudspeaker compensation for hands free acoustic echo cancellation," Electron. Lett. **32**, 1063–1064.

Brand, T., and Kollmeier, B. (**2002**). "Efficient adaptive procedures for threshold and concurrent slope estimate for psychophysics and speech intelligibility tests," J. Acoust. Soc. Am. **111**(6), 2801–2810.

Cooke, M., Mayo, C., and Valentini-Botinhao, C. (**2013a**). "Intelligibility-enhancing speech modifications: The Hurricane challenge," in *Proceedings of Interspeech*, Lyon, France, August 2013, pp. 3552–3556.

Cooke, M., Mayo, C., Valentini-Botinhao, C., Stylianou, Y., Sauert, B., and Tang, Y. (**2013b**). "Evaluating the intelligibility benefit of speech modifications in known noise conditions," Speech Commun. **55**(4), 572–585.

Defraene, B., van Waterschoot, T., Ferreau, H. J., Diehl, M., and Moonen, M. (**2012**). "Real-time perception-based clipping of audio signals using convex optimization," IEEE Trans. Audio Speech Lang. Process. **20**(10), 2657–2671.

Haensler, E., and Schmidt, G. (**2008**). *Speech and Audio Processing in Adverse Environments* (Springer Science & Business Media, Berlin).

ITU (**2011**). "ITU-T P. 56—Objective measurement of active speech level," International Telecommunication Union Recommendation ITU-T P.56, Geneva, Switzerland.

Kabal, P. (**1999**). "Measuring speech activity," technical report, McGill University, Montreal, Canada.

Kleijn, W. B., Crespo, J. B., Hendriks, R. C., Petkov, P., Sauert, B., and Vary, P. (**2015**). "Optimizing speech intelligibility in a noisy environment: A unified view," IEEE Signal Process. Mag. **32**, 43–54.

Klink, K., Schulte, M., and Meis, M. (**2012**). "Measuring listening effort in the field of audiology—A literature review of methods (part 2)," Z. Audiol. **51**, 96–105.

Kochkin, S. (**2012**). "MarkeTrak VIII: The key influencing factors in hearing aid purchase intent," Hear. Rev. **19**, 12–25.

Rennies, J., Schepker, H., Holube, I., and Kollmeier, B. (**2014**). "Listening effort and speech intelligibility in listening situations affected by noise and reverberation," J. Acoust. Soc. Am. **136**, 2642–2653.

Rhebergen, K. S., and Versfeld, N. J. (**2005**). "A speech intelligibility index-based approach to predict the speech reception threshold for sentences in fluctuating noise for normal-hearing listeners," J. Acoust. Soc. Am. **117**, 2181–2192.

Sato, H., Bradley, J., and Morimoto, M. (**2005**). "Using listening difficulty ratings of conditions for speech communication in rooms," J. Acoust. Soc. Am. **117**, 1157–1167.

Sauert, B., and Vary, P. (**2010**). "Recursive closed-form optimization of spectral audio power allocation for near end listening enhancement," in *Proceedings of the ITG Conference on Speech Communication*, Bochum, Germany, Oct. 2010, Paper No. 8.

Sauert, B., and Vary, P. (**2012**). "Near-end listening enhancement in the presence of bandpass noises," in *Proceedings of the ITG Conference on Speech Communication*, Braunschweig, Germany, September 2012, pp. 195–198.

Schepker, H., Rennies, J., and Doclo, S. (**2013**). "Improving speech intelligibility in noise by SII-dependent preprocessing using frequency-dependent amplification and dynamic range compression," in *Proceedings of Interspeech*, Lyon, France, August 2013, pp. 3577–3581.

Schepker, H., Rennies, J., and Doclo, S. (**2015**). "Speech-in-noise enhancement using a two-stage amplification and dynamic range compression algorithm controlled by the speech intelligibility index," J. Acoust. Soc. Am. **138**, 2692–2706.

Taal, C. H., Hendriks, R. C., and Heusdens, R. (**2014**). "Speech energy redistribution for intelligibility improvement in noise based on a perceptual distortion measure," Comput. Speech Lang. **28**(4), 858–872.

Taal, C. H., Hendriks, R. C., Heusdens, R., and Jensen, J. (**2011**). "An algorithm for intelligibility prediction of time frequency weighted noisy speech," IEEE Trans. Audio Speech Lang. Process. **19**(7), 2125–2136.

Taal, C. H., and Jensen, J. (**2013**). "SII-based speech preprocessing for intelligibility improvement in noise," in *Proceedings of Interspeech*, Lyon, France, August 2013, pp. 3582–3586.

Tang, Y., and Cooke, M. (**2011**). "Subjective and objective evaluation of speech intelligibility enhancement under constant energy and duration constraints," in *Proceedings of Interspeech*, Florence, Italy, August 2011, pp. 345–348.

Wagener, K., Brand, T., and Kollmeier, B. (**1999a**). "Entwicklung und Evaluation eines Satztests für die deutsche Sprache II: Optimierung des Oldenburger Satztests" ("Development and evaluation of a German sentence test II: Optimization of the Oldenburg sentence test"), Z. Audiol. **38**, 44–56.

Wagener, K., Brand, T., and Kollmeier, B. (**1999b**). "Entwicklung und Evaluation eines Satztests für die deutsche Sprache III: Evaluation des Oldenburger Satztests" ("Development and evaluation of a German sentence test III: Evaluation of the Oldenburg sentence test"), Z. Audiol. **38**, 86–95.

Wagener, K., Kühnel, V., and Kollmeier, B. (**1999c**). "Entwicklung und Evaluation eines Satztests für die deutsche Sprache I: Design des Oldenburger Satztests" ("Development and evaluation of a German sentence test I: Design of the Oldenburg sentence test"), Z. Audiol. **38**, 4–15.

Young, L. L., Jr., Goodman, J. T., and Carhart, R. (**1978**). "The intelligibility of whitened and peak clipped speech," J. Am. Audiol. Soc. **3**, 167–171.

Zorila, T.-C., Kandia, V., and Stylianou, Y. (**2012**). "Speech-in-noise intelligibility improvement based on spectral shaping and dynamic range compression," in *Proceedings of Interspeech*, Portland, OR, September 2012, pp. 635–638.

Zorila, T.-C., and Stylianou, Y. (**2014**). "On spectral and time domain energy reallocation for speech-in-noise intelligibility enhancement," in *Proceedings of Interspeech*, Singapore, September 2014, pp. 2050–2054.