Individualized and noise-adaptive enhancement of speech intelligibility

A. Kubiak[1,2], J. Rennies [1], A. Volgenandt[1], J. Drefs[1,2], B. Kollmeier[1,2]

[1] *Fraunhofer IDMT Project Group Hearing, Speech and Audio Technology and Cluster of Excellence "Hearing4all", Oldenburg, Germany*

[2] *Medizinische Physik and Cluster of Excellence "Hearing4all", Universität Oldenburg, Germany*

Introduction

The intelligibility of speech played back via audio reproduction systems is often impaired in noisy backgrounds. Ideally, algorithms enhancing speech intelligibility should be adaptive to the type and temporal variations of the noise, and also account for differences in individual listening preferences. While noise-adaptive algorithms have been investigated in several studies, individual preferences have not yet been addressed in this context.

Motivation

The current study investigated the inter-individual variability of normal-hearing subjects' preferences with respect to intelligibility enhancement in noise for communication applications using the **AdaptDRC** algorithm [1], which had been shown to be highly effective in various types of background noise. Originally, the algorithm uses estimations of the SII [2] to control spectral shaping and compression characteristics of the speech signal. An extension of the algorithm, which also includes an adaptive level increase - **AdaptDRCPlus** – has shown to increase intelligibility even further [3]. In this study, the generic SII model controlling both versions of the algorithm is replaced with listeners' individual preferences allowing for further enhancements when accounting for individual factors in speech processing - **AdaptDRCPerso**. The data are discussed with respect to the relation between individual listening preferences and generic model-based parameters and the predictability of individually preferred parameter settings, which would allow a complete individualization of the algorithm.

Algorithms

AdaptDRC uses short-time estimates of the SII to adaptively control a linear frequency-shaping stage and a stage introducing dynamic range compression to modify an input speech signal depending on the current environmental noise (**Fig. 1**). The SII is used to progressively increase the degree of speech signal modification with decreasing SII. The rms-level of the speech is not modified.

AdaptDRCPlus additionally contains a processing stage which adaptively amplifies the output of **AdaptDRC** (**Fig. 2**) using the same SII estimation with an equal-peak-level constraint. If amplification is applied, samples beyond the maximum input amplitude are clipped hard.

AdaptDRCPerso replaces the generic SII model with listeners' individual preferences, which allows for an individually tailored speech enhancement in relation to hearing abilities and personal needs of a user, (**Fig.3**).

Methods

In order to investigate the degree of signal modification preferred by normal-hearing listeners as well as the inter-subject variability of listening preferences, subjects were instructed to select the processing parameters using 1- or 2-dimensional, real-time sliders in four different scenarios:

- adjusting AdaptDRC only – using slider, (**Fig. 1**)
- adjusting linear gain/clipping only – using slider, (**Fig. 2**)
- adjusting linear gain and AdaptDRC in parallel – using slider, (**Fig. 3**)
- adjusting linear gain and AdapDRC independently – using two-dimensional matrix (**Fig. 4**)

Effectively, the subjects' choices replaced the generic SII model, which could take values from zero (corresponding to SII estimation set to zero which causes a maximum degree of signal modification) to one (corresponding to SII estimation set to one, resulting in no signal modification). Fourteen normal-hearing subjects were invited to the experiment, 5 female, 9 male, mean age 26.6. Presented audio probes were played as diotic .wav files at a sampling rate of 16 kHz. The starting level of the speech signals was equalized to 60 dB SPL. As speech material, sentences from the Oldenburg sentence test were used [4], mixed with non-stationary cafeteria background noise at fixed SNRs of -14, -10, -6, -2, +2 and +6 dB. Stimuli were presented via Sennheiser HD650 headphones. Subjects could listen to the stimuli for as long as they needed to adjust the sliders to their preferred degree of processing. Changes in slider positions resulted in real-time changes of the processed signals so that an immediate acoustic feedback was available. Subjects were instructed to test different settings by moving the sliders and stop at the position that corresponded to their preferred sound. Each combination of scenario and SNR was repeated twice.

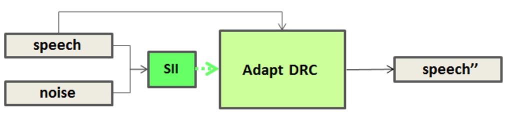


Figure 1: Schematic of AdaptDRC [1]

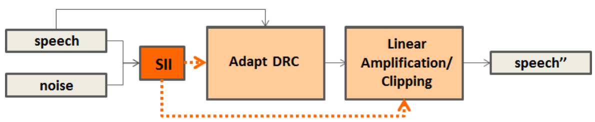


Figure 2: Schematic of AdaptDRCPlus [3].

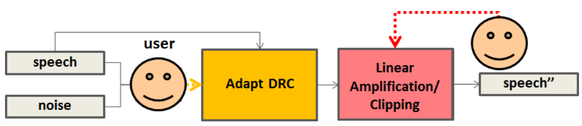


Figure 3: Schematic of AdaptDRC Perso

Results

The data collected in previous studies on AdaptDRC and AdaptDRCPlus are re-plotted in **Fig. 4**. Considerable improvements of speech intelligibility compared to unprocessed reference signals have been shown, although the spread among (normal-hearing) subjects was severe, indicating that the individual benefit depended strongly on listener. Hence this study aimed at investigating between-listener spread of preferences with respect to the degree of signal modification by using the subjects' own choices rather that the generic SII estimation. The results of an exemplary subject's choices in four different scenarios are presented in **Fig. 5**. The abscissa represents the amount of desired AdaptDRC processing on the scale from zero to one, where zero indicates maximum applied processing (equal to SII estimation set to zero), and one indicates no processing (SII equals one). Symbols on the abscissa (triangles, color-coded for different SNRs) indicate the subject's choices for the scenario when only the AdaptDRC processing could be adjusted and no linear amplification was applied (**Fig. 1**). The ordinate represents the amount of desired broad band gain followed by clipping, so, that each additional gain introduces clipping. Axis's scale ranges from zero to one, where one means no gain applied, followed by no clipping, and zero means maximum gain applied and 30% of samples hard clipped; any values in between result in clipping ranging from 0% to 30%. In this condition, the broadband gain (followed by clipping) was the only hearing support that could be applied to the speech signal (**Fig 3, red block only**). Symbols on the axis (circles, color-coded) represent the subject's ratings. Diamonds on the diagonal show the adjustments of AdaptDRC and broadband gain, when both are controlled by the same factor – SII estimation, equal for both, (**Fig. 2**). Squares, distributed freely on the two-dimensional plane indicate subjective choices from the scenario, where AdaptDRC processing and broadband gain could be adjusted separately and independently (**Fig. 3, yellow and red block**). Each symbol in **Fig. 5** appears twice, showing subjective preferences for both the test and the retest measurement. With a few exceptions, test and retest resulted in very similar choices of this subject, which was a general observation also true for the other subjects. Results from the gain adjustment scenario only (ordinate) indicate that subjects tend to increase the amount of support while the SNRs worsen. On the other hand, looking on the AdaptDRC adjustments only (abscissa), it is clear that subjects preferred a lot of processing even for very good SNRs, rating it as very helpful for speech enhancement and listening comfort. Diamonds on the diagonal standing for AdaptDRC and gain processing adjusted in parallel show similar spread as the gain adjustments on the ordinate, but do not align with the squares obtained from independent adjustments of both stages of the algorithm. This may indicate that controlling gain and AdaptDRC processing with only one estimate does not fully match subsects' needs or that, at least, the two processing stages should be weighted differently. Concerning the independent scenario only (squares) it can be observed that subjects highly prefer AdaptDRC support, but additionally tend to increase the gain when SNRs worsen, severely deviating from the generic SII-controlled processing schemes.

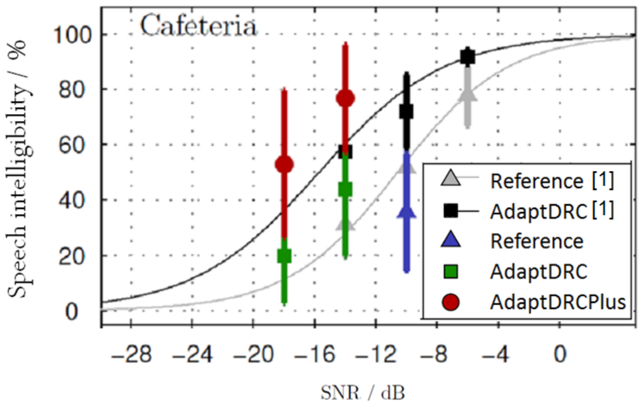


Figure 4: Speech intelligibility as a function of SNR for cafeteria noise and speech from the Oldenburg sentence test [4]. Comparison of unprocessed speech, AdaptDRC and AdaptDRCPlus [3].

Median values of subjective choices from the scenario, where subjects were asked to choose the individually preferred processing schemes of linear gain/clipping stage and AdaptDRC independently (**Fig. 6 top**) indicate a positive trend between AdaptDRC and gain usage in different listening conditions with clear preference towards AdaptDRC. These data and the substitution rate between these two variables (amount of AdaptDRC processing versus amount of preferred gain for a given SNR condition, as a fraction) were used as a basis for an adaptive model that could potentially replace the generic SII estimation that is not aware of any individual features of a particular user. The proposed model has to provide predictions of a desired amount of AdaptDRC processing from gain adjustments only. The model tracks introduced gain changes in different SNRs, providing AdaptDRC estimations and constantly updating the substitution rate between these two variables. Monitoring a trend and a spread of changes of the known variable (gain), and the interpolated one (amount of AdaptDRC processing), based on former subject's choices, provides a knowledge resource for the model. The predicted amount of AdaptDRC processing versus experimental data across all subjects is presented in the bottom panel of **Fig. 6**.

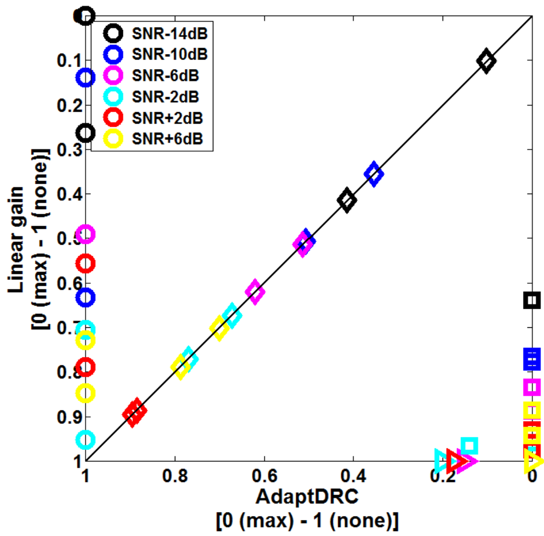
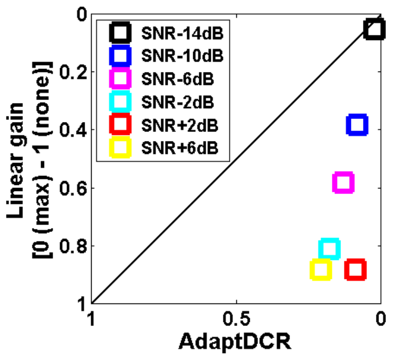


Figure 5: Test/re-test results of an exemplary subject across different scenarios and listening conditions (color-coded)



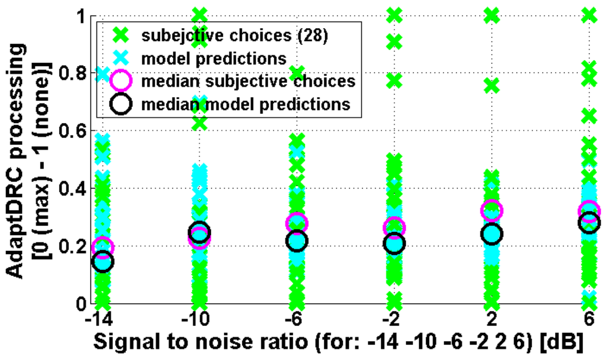


Figure 6 Top: Median preferences across subjects for the condition of independent gain and AdaptDRC processing adjustments as a function of SNR (color-coded). **Bottom:** Model predictions of desired amount of AdaptDRC processing at different SNRs based on subjects' gain adjustments vs. experimental data.

Conclusions

The current study (AdaptDRC Perso) aimed at providing better understanding of individual factors that may lead to individualization of speech enhancing algorithms and account for user's hearing abilities and preferences for further improvement of listening comfort in difficult scenarios. Results obtained from listening tests investigating speech intelligibility in the presence of non-stationary noise for two versions of the algorithm – AdaptDRC and AdaptDRC Plus - revealed significant increase in the amount of correctly understood words in comparison to reference conditions, but still indicated severe spread in gained improvement between subjects [3]. This led to the idea of replacing the generic SII model that controls the processing stages of both versions of the algorithm by subjective choices in order to account for personal abilities and preferences and make use of them in the processing schemes. Subjective tests, on a data set of 14 subjects, revealed high test/retest accuracy in listeners' choices, as well as a clear preference towards AdaptDRC processing over broadband gain in demanding listening conditions. These results, embedded in the scenario of, e.g., telephone conversations, provided a basis for an adaptive model capable of predicting the desired amount of AdaptDRC processing based on linear gain adjustments only. The model tracks introduced gain changes at different SNRs and estimates the preferred AdaptDRC processing. Constant updating of the substitution rate between these two variables and monitoring a trend and a spread of changes of the known variable (gain) and an interpolated one (amount of AdaptDRC processing) results in designation of an optimum function of a consumer. Such a function indicates personalized speech enhancement schemes for given listening conditions, but is also easily accessible, adjusting dynamically to current settings, following user's preferences. Further development of this idea for predicting two-dimensional preferences based on data from one dimension only may lead to the replacement of the generic, non-individualized SII estimation with a highly individualized approach tailored to the listening environment, but also to the subjects' hearing abilities and listening preferences and have a potential to be easily applicable in many everyday scenarios.

Literature

[1] Schepker,, H., Rennies, J. & Doclo, S. (2013). Improving speech intelligibility in noise by SII-dependent preprocessing using frequency-dependent amplification and dynamic range compression. Proc. of Interspeech, Lyon, France, Aug. 2013, pp. 3577-3581.

[2] ANSI (1997). Methods for calculation of the speech intelligibility index. American National Standard ANSI S3.5-1997 (American Nationa Standards Institute, Inc.), New York, USA.

[3] Drefs, J., Rennies, J., Schepker, H. & Doclo, S. (2015). Weiterentwicklung und Evaluation eines Algorithmus zur SII-basierten Sprachverständlichkeitsverbesserung in störgeräuschbehafteter Umgebung. Fortschritte der Akustik - Proc. of the 41th Annual Conference of the German Acoustical Society DAGA 2015, Nürnberg.

[4] Wagener, K., Kühnel, V., and Kollmeier, B. (1999). Entwicklung und Evaluation eines Satztests für die deutsche Sprache I: Design des Oldenburger Satztests. Zeitschrift für Audiologie/Audiological Acoustics, 38:4–15.