

Evaluation eines Algorithmus zur SII-basierten Sprachverständlichkeitsverbesserung in störrauschbehafteter Umgebung mit schwerhörenden Probanden

David Hülsmeier¹, Jan Rennies¹, Jakob Drefs¹, Henning Schepker², Simon Doclo^{1,2}

¹ Fraunhofer IDMT / Hör-, Sprach- und Audiotechnologie und Exzellenzcluster Hearing4All, Oldenburg,

² Universität Oldenburg, Signal Processing Group und Exzellenzcluster Hearing4All, Oldenburg

{david.huelsmeier, jan.rennies, jakob.drefs, simon.doclo}@idmt.fraunhofer.de,

{henning.schepker, simon.doclo}@uni-oldenburg.de

Einleitung

Die Sprachverständlichkeit (SV) ist in Umgebungen mit lauten Hintergrundgeräuschen häufig eingeschränkt. Dieses ist zum Beispiel bei Durchsagen in Bahnhöfen oder Flughäfen der Fall. Daraus resultiert besonders für schwerhörende Personen häufig ein Informationsverlust. Das Gleiche gilt für das Telefonieren in lauten Umgebungen. In diesen Situationen liegt das Sprachsignal getrennt vom Störrausch vor und kann so gezielt verändert werden, um die Sprachverständlichkeit zu erhöhen. Eine Vielzahl von Algorithmen existiert, die sich diesem Problem annehmen [1]. Diese Algorithmen wurden jedoch ausschließlich mit normalhörenden Personen evaluiert.

In dieser Studie wurden zwei verschiedene Varianten eines Störrausch-adaptiven Algorithmus zur modellbasierten Sprachverständlichkeitsverbesserung hinsichtlich ihrer Wirksamkeit für schwerhörende Probanden untersucht:

1. AdaptDRC [2]
2. AdaptDRCplus [3]

AdaptDRC ändert die spektrale Form und Signaldynamik bei unveränderter Signalleistung. Die AdaptDRCplus Variante erlaubt eine adaptive Erhöhung der Signalleistung.

Für die Evaluation wurde der Oldenburger Satztest mit instationärem Cafeteria-Störrausch und unterschiedlichen SNRs verwendet. Subjektive Hörversuche wurden mit zehn leicht bis mittelgradig schwerhörenden, unverstärkten Probanden durchgeführt.

Methoden

Akustisches Szenario

Sowohl für Durchsagen in Bahnhöfen, Flugzeugen oder Bussen, als auch während eines Telefongesprächs kann das in Abb. 1 dargestellte Szenario angenommen werden. Das Sprachsignal $s[k]$, mit diskretem Zeitindex k , ist bekannt und kann mit einem Vorverarbeitungsalgorithmus $W\{\cdot\}$ bearbeitet werden. Das verarbeitete Signal wird in einem Raum mit Echopfad $h[k]$ wiedergegeben und von einem Störrausch $n[k]$ gestört. Durch Subtraktion des

geschätzten Echopfads $\hat{h}[k]$ vom Mikrofonsignal $y[k]$ kann das Störrausch $\hat{n}[k]$ geschätzt werden. Zusätzlich wird ein Sprachaktivitätserkennung (VAD) zur Adaption von $\hat{h}[k]$ und $W\{\cdot\}$ verwendet. Das verarbeitete Sprachsignal $\tilde{s}'[k]$ (AdaptDRC) bzw. $\tilde{s}'''[k]$ (AdaptDRCplus), wird wie folgt berechnet:

$$\tilde{s}'[k] = W \left\{ s[k], \hat{n}[k], \hat{h}[k] \right\} \cdot s[k] \quad (1)$$

$$\tilde{s}'''[k] = W \left\{ s[k], \hat{n}[k], \hat{h}[k] \right\} \cdot s[k] \quad (2)$$

Im Folgenden wird eine nachhallfreie Umgebung angenommen ($h[k] = \delta[k]$), so dass eine genaue Schätzung des Störrausches möglich ist, somit entspricht $\hat{n}[k] = n[k]$.

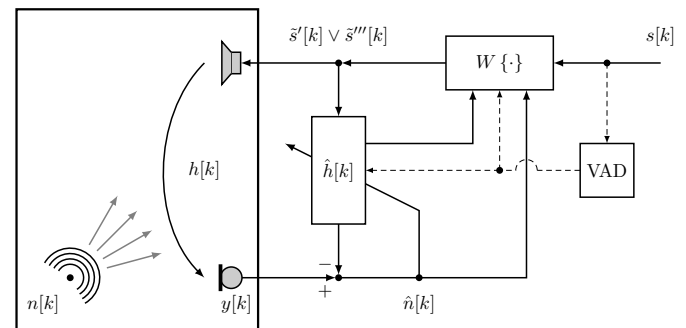


Abbildung 1: Darstellung des betrachteten akustischen Szenarios.

AdaptDRC

Der AdaptDRC Algorithmus [2] (Abb. 2, grau) kann zur Vorverarbeitung eines Signals verwendet werden. Das Ziel des Algorithmus ist, die Sprachverständlichkeit eines Sprachsignals $s[l]$ im Störrausch $n[l]$ zu erhöhen. Die Verarbeitung erfolgt in Blöcken l der Länge M und unterliegt der folgenden Annahme:

- (a) Die Breitbandleistung $\phi_{\tilde{s}}[l]$ des verarbeiteten Signals $\tilde{s}[l]$ entspricht der Breitbandleistung $\phi_s[l]$ des unverarbeiteten Signals.

Das Sprachsignal $s[l]$ und das Störsignal $\hat{n}[l]$ werden jeweils in $N = 8$ Oktavbänder gefiltert. Das Sprachsignal im n -ten Oktavband des l -ten Blocks wird mit

$s_n^l[m] = s_n[lM - m]$, $m = 0, \dots, M - 1$ bezeichnet. Für jeden Block l wird eine Schätzung $\hat{SII}[l]$ des Speech Intelligibility Index [4] berechnet. Diese wird verwendet, um den Grad der linearen Verstärkung $w_n[l]$ pro Oktavband zu bestimmen:

$$w_n[l] = \sqrt{\frac{\phi_n^{\hat{SII}[l]}[l] \phi_s[l]}{\sum_{\lambda=1}^N \phi_{\lambda}^{\hat{SII}[l]}[l] \phi_n[l]}} \quad (3)$$

mit $\phi_n[l]$ als Schmalbandleistung des Sprachsignals im n -ten Oktavbands und im l -ten Blocks. Sofern $\hat{SII}[l] = 1$ wird $s[l]$ nicht verändert. Falls $\hat{SII}[l] = 0$ wird die Leistung gleichmäßig über alle Oktavbänder verteilt.

Die Dynamik-Kompression (DRC) wird dafür verwendet, schlecht hörbare Signalanteile zu verstärken und gut hörbare abzuschwächen. Im Allgemeinen werden für DRC feste Kompressionsverhältnisse verwendet. Dieses kann jedoch zu einer Verschlechterung der Sprachverständlichkeit oder -qualität führen. Aus diesem Grund wird hier eine adaptive Berechnung der anzuwendenden Kompression verwendet.

Zunächst wird die geglättete Einhüllende jedes Frequenzbandes $\bar{s}_n[k]$ berechnet [2]. Das verarbeitete Sprachsignal wird als

$$\tilde{s}_n^l[m] = s_n^l[m] \cdot p(\lambda_n[l], (\bar{s}_n^l[m])^2) \quad (4)$$

berechnet. Die Funktion $p(\lambda_n[l], (\bar{s}_n^l[m])^2)$ gibt die lineare Inter- und Extrapolation der Input-Output-Charakteristik (IOC) im dB-Bereich an. Hierbei entspricht $(\bar{s}_n^l[m])^2$ der Eingangssignalleistung. Die IOC wird durch $\lambda_n[l]$ bestimmt [2].

Artefakte und Verzerrungen werden durch eine rekursive Glättung der kombinierten Verstärkung $w_n[l] \cdot p_n^l[m]$ (Gl. 3 und 4) vermieden. Abschließend werden die Schmalbandsignale summiert und die Breitbandleistung des Ausgangssignals $\bar{\phi}_{\tilde{s}}[l]$ wird auf die Leistung des Eingangssignals $\bar{\phi}_s[l]$ zurückgesetzt.

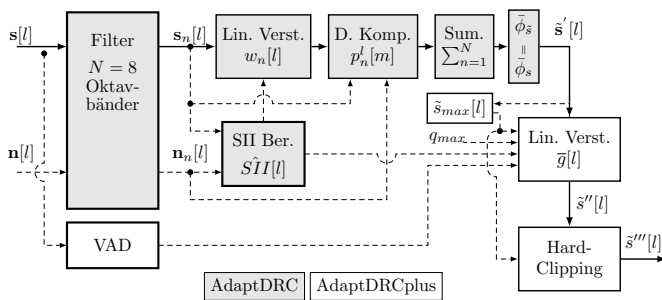


Abbildung 2: Blockschaltbild der AdaptDRC (grau) und AdaptDRCplus (grau und weiß) Algorithmen. Das Sprachsignal $s[l]$ sowie das Störgeräusch $n[l]$ werden blockweise verarbeitet und zu $\tilde{s}'[l]$ bzw. $\tilde{s}''[l]$ verstärkt.

AdaptDRCplus

AdaptDRCplus [3] ist eine Weiterentwicklung von AdaptDRC (siehe Abb. 2, grau und weiß). Es werden die gleichen Annahmen getroffen wie bei AdaptDRC.

Die Randbedingung, die Breitbandleistung nicht zu erhöhen, wurde aufgehoben, da in realen Bedingungen eine Erhöhung der Leistung möglich und der Maximalpegel begrenzt ist. Sollte das Signal bereits ausgesteuert sein, so ist keine rein lineare Verstärkung möglich. Die Randbedingung, die Amplitude des Signals nicht zu verändern, wird daher eingeführt. Diese wird durch eine adaptive Verstärkung $g[l]$ des Sprachsignals mit anschließendem Clipping umgesetzt. Der Grad der Übersteuerung hängt von $\hat{SII}[l]$ ab: für $\hat{SII}[l] = 1$ wird $\tilde{s}'[l]$ nicht verändert, für $\hat{SII}[l] = 0$ wird $\tilde{s}'[l]$ so verstärkt, dass $q_{max}\%$ der Samples clippen. Zudem wird $g[l]$ zu $\bar{g}[l]$ geglättet, so dass keine abrupten Verstärkungen auftreten.

Um Hard-Clipping einzuführen, werden alle Samples bei $\pm \tilde{s}_{max}$, mit $\tilde{s}_{max} = \max(|\tilde{s}[l]|)$, abgeschnitten. Zusätzlich werden nur Blöcke verarbeitet, in denen vom VAD Sprache detektiert wird.

Subjektive Evaluation

Für die erstmalige Evaluation der Algorithmen mit schwerhörenden Probanden wurden sowohl fünf geringgradig, als auch fünf mittelgradig schwerhörende Personen untersucht [Pure Tone Average (0.5|1|2|4) kHz 25-40 dB HL für gering- bzw. 40-60 dB HL für mittelgradig] (siehe Abb. 3 oben rechts). Ihr Alter lag bei 69.3 ± 10.0 Jahren. Sie haben ohne Hörhilfe den Oldenburger Satztest [5] in verschiedenen Vorverarbeitungsbedingungen durchgeführt:

1. unverarbeitet ($s[k] + n[k]$)
2. mit AdaptDRC bearbeitet ($\tilde{s}'[k] + n[k]$)
3. mit AdaptDRCplus bearbeitet ($\tilde{s}''[k] + n[k]$)

Der Sprachpegel lag für Kondition 1 und 2 fest bei $65dB SPL$. In Kondition 3 war je nach SNR eine adaptive Pegelerhöhung möglich. $n[k]$ entsprach einem Cafeteria-Störgeräusch, dessen Pegel je nach Eingangs-SNR relativ zur Sprache skaliert wurde. Für die diotische Signaldarbietung wurde ein Sennheiser HD650 Kopfhörer verwendet. Für den AdaptDRCplus Algorithmus wurde $q_{max} = 20\%$ gewählt.

Vor der Messung wurde ein Training über 60 Sätze mit unverarbeitetem Signal durchgeführt. Anschließend wurde, aufgrund der individuell unterschiedlichen Schwerhörigkeit, der SNR für 50% Sprachverständlichkeit ($SNR_{50\%}$) adaptiv mittels AFC-Framework [6] bestimmt. Dieses wurde für jede Vorverarbeitung durchgeführt.

Um die Sprachverständlichkeitsfunktionen der Probanden zu ermitteln, wurde bei den adaptiv gemessenen SNR sowie bei den $SNR_{50\%} \pm 4dB$ die Sprachverständlichkeit bestimmt. Diese Werte wurden gewählt, um die individuelle Sprachverständlichkeitsfunktion mit 3 Stürzpunkten bestmöglich abzutasten.

Ergebnisse

Die parametrisch gemittelte Sprachverständlichkeitsfunktion ist im oberen linken Panel

von Abb. 3 dargestellt. Für die mittlere Sprachverständlichkeitsfunktion wurden der $SNR_{50\%}$ und die Steigung bei 50% Sprachverständlichkeit über alle Probanden gemittelt. Das mittlere Panel zeigt den Sprachverständlichkeitsgewinn bei 50% Sprachverständlichkeit zwischen Kondition 1 und 2, 1 und 3 sowie 2 und 3. Im unteren Panel ist der SNR Gewinn für die selben Konditionen dargestellt. Individuelle und gemittelte Ergebnisse zeigten, dass sowohl AdaptDRC als auch AdaptDRCplus jedem Probanden eine erhöhte Sprachverständlichkeit ermöglichten. Hierbei zeigte der AdaptDRCplus Algorithmus den größten Gewinn. Im Mittel wurde ein Sprachverständlichkeitsgewinn von 15% für AdaptDRC und 42% für AdaptDRCplus bei 50% Sprachverständlichkeit erzielt. Zudem wurde ebenfalls bei 50% Sprachverständlichkeit ein SNR Gewinn von 1.5dB für AdaptDRC und 7.2dB für AdaptDRCplus ermittelt.

Individuell zeigte VP01 den geringsten Sprachverständlichkeits- und SNR-Gewinn für AdaptDRC gegenüber unverbesselter Kondition von etwa 2% Sprachverstehen bzw. 0.2 dB SNR. VP03 erzielte mit ca. 35% Sprachverständlichkeits- und 3.5 dB SNR-Gewinn für AdaptDRC die größte Verbesserung. Für AdaptDRCplus erzielte jeder Proband deutliche Gewinne mit fast 50% Sprachverstehen für VP05 und etwa 9 dB SNR für VP03.

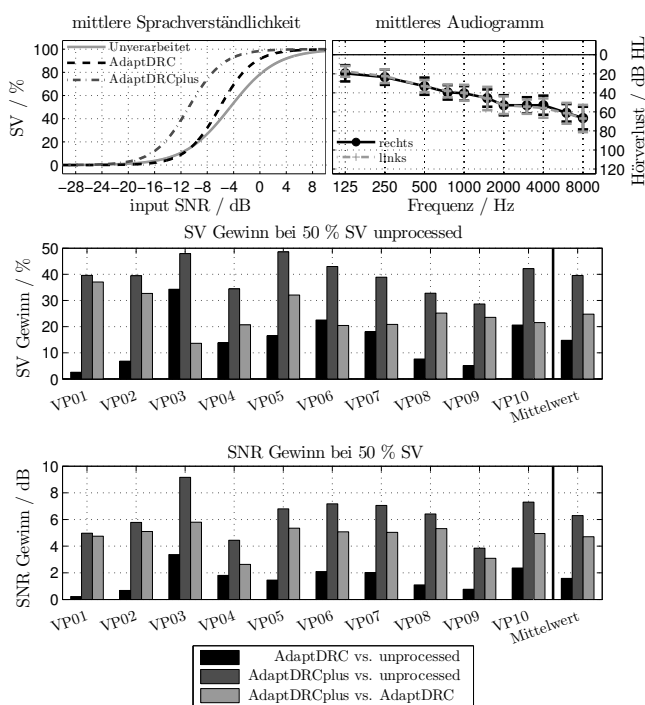


Abbildung 3: Ergebnisse der subjektiven Evaluation. Oben links: Sprachverständlichkeitsfunktion gemittelt über alle Probanden für unverbesserte Kondition, AdaptDRC und AdaptDRCplus. Oben rechts: Mittleres Audiogramm der Probanden. Mitte: Sprachverständlichkeitsgewinn bei 50% Sprachverstehen des unverbesserten Signals (horizontale Differenz). Unten: SNR-Gewinn bei 50% Sprachverstehen des unverbesserten Signals (vertikale Differenz).

Fazit

Das Ziel dieser Studie war es, die Vorverarbeitungsalgorithmen im Störgeräusch mit Schwerhörigen zu evaluieren. Zusätzlich sollte ein Vergleich zu Normalhörenden gezogen werden, um zu sehen, ob gleich hohe Gewinne erzielt werden konnten. Hierfür wurden zwei Algorithmen vorgestellt und mit schwerhörenden Probanden evaluiert. Im Vergleich mit den Ergebnissen Normalhörender [2, 3] erzielten Schwerhörende einen geringeren Gewinn zwischen unverbesselter und mit AdaptDRC bearbeiteter Kondition. Normalhörende erzielten ebenfalls einen SNR-Gewinn von etwa 5 dB bei 50% Sprachverstehen [7]. Im Gegensatz erzielten Schwerhörende nur etwa 2 dB SNR-Gewinn. Ein Vergleich mit AdaptDRCplus ist aufgrund der Datenlage nicht möglich, da in [3] nur zwei SNR für AdaptDRCplus gemessen wurden. Somit konnte keine Sprachverständlichkeitsfunktion an die Daten angepasst werden. Im Allgemeinen unterstützen beide Algorithmen erfolgreich unversorgte, schwerhörige Personen beim Verstehen von Sprache.

Die Verbesserung zwischen AdaptDRC und AdaptDRCplus ist auf den effektiven SNR-Gewinn zurück zu führen, der bis zu 7 dB für Eingangs-SNR von -30 dB beträgt [3]. Der Grund für den unterschiedlichen Sprachverständlichkeitsgewinn zwischen normalhörenden und schwerhörenden Probanden ist Gegenstand weiterführender Untersuchungen. Zusätzlich sollen die weiteren Möglichkeiten einer Individualisierung des Algorithmus überprüft werden.

Literatur

- [1] M. Cooke, C. Mayo, C. Valentini-Botinhao: Intelligibility-enhancing speech modifications: the hurricane challenge, in Proc. of Interspeech, Lyon, Frankreich, Aug. 2013, S. 3552-3556.
- [2] H. Schepker, J. Rennie, S. Doclo: Improving speech intelligibility in noise by SII-dependent preprocessing using frequency dependent amplification and dynamic range compression, in Proc. of Interspeech, Lyon, Frankreich, Aug. 2013, S. 3577-3581.
- [3] J. Drefs, J. Rennie, H. Schepker, S. Doclo: Weiterentwicklung und Evaluation eines Algorithmus zur SII-basierten Sprachverständlichkeitsverbesserung in störrauschbehafteter Umgebung, DAGA 2015, Nürnberg, Deutschland, März 2015.
- [4] ANSI. Methods for calculation of the speech intelligibility index. American National Standard ANSI S3.5-1997 (American National Standards Institute, Inc.), New York, USA, 1997.
- [5] K. Wagener, V. Kühnel, B. Kollmeier: Entwicklung und Evaluation eines Satztests für die deutsche Sprache, Teil 1: Design des Oldenburger Satztests, Z. für Audiologie 1 (2001): 4-15.
- [6] S. D. Ewert: AFC—A modular framework for running psychoacoustic experiments and computational per-

ception models, Proc Conf Acoust AIA-DAGA 2013, Meran, Italien, März 2013.

- [7] H. Schepker: Entwicklung und Evaluation von Vorverarbeitungsalgorithmen zur Verbesserung der Sprachverständlichkeit im Störgeräusch, Masterarbeit, Carl-von-Ossietzky Universität Oldenburg, Oldenburg, Deutschland, Oktober 2012.