Weiterentwicklung und Evaluation eines Algorithmus zur SII-basierten Sprachverständlichkeitsverbesserung in störgeräuschbehafteter Umgebung

Jakob Drefs[1], Jan Rennies[1], Henning Schepker[2], Simon Doclo[1,2]

[1] *Fraunhofer IDMT/Hör-, Sprach- und Audiotechnologie, Oldenburg, Deutschland,*
Email: {jakob.drefs, jan.rennies}@idmt.fraunhofer.de

[2] *CvO Universität Oldenburg, Dept. of Medical Physics and Acoustics and Cluster of Excellence Hearing4All, Germany,*
Email: {henning.schepker, simon.doclo}@uni-oldenburg.de

Einleitung

Häufig ist die Sprachkommunikation durch Störgeräusche und/oder Nachhall gestört. Dies kann zu einer verringerten Sprachverständlichkeit und einer erhöhten Höranstrengung führen [1]. Um eine hohe Kommunikationsqualität zu gewährleisten, wurden daher in der Vergangenheit verschiedene Verfahren vorgestellt, welche zum Ziel haben die Sprachverständlichkeit in störgeräuschbehafteten Umgebungen durch Vorverarbeitung des (ungestörten) Sprachsignals zu verbessern [2].
Eine häufige Annahme bei der Entwicklung dieser Verfahren ist, dass der Sprachpegel vor und nach der Verarbeitung identisch sein muss, um eine Übersteuerung der Verstärkungskette (Verstärker, Lautsprecher, etc.) und zu hohe Darbietungspegel zu vermeiden. In [3] wurde der sogenannte AdaptDRC-Algorithmus vorgestellt, welcher auf der Kombination einer zeit- und frequenzabhängigen Verstärkung und einer zeit- und frequenzabhängigen Dynamikkompression basiert, wobei beide Stufen mit Hilfe des Sprachverständlichkeitsindex (SII) [4] gesteuert werden. Es konnte für normalhörende Probanden gezeigt werden, dass mit Hilfe des AdaptDRC-Algorithmus eine signifikante Verbesserung der Sprachverständlichkeit bei gleichbleibendem Sprachpegel sowohl in stationären als auch in instationären Störgeräuschen erzielt werden kann [3, 5]. In praktikschen Anwendungen, z.B. bei Bahnhofsdurchsagen oder in Mobiltelefonen, ist es jedoch nicht zwingend notwendig, den Sprachpegel unverändert zu lassen. Häufig ist bei diesen Wiedergabesystemen der Maximalpegel, oberhalb dessen ungewünschte Verzerrungen des Sprachsignals auftreten, limitiert. Ziel dieser Studie ist daher die Erweiterung des AdaptDRC-Algorithmus um eine adaptive Verstärkungsstufe, welche das Sprachsignal bei einer schlechten vorhergesagten Sprachverständlichkeit gezielt innerhalb der Grenzen des Wiedergabesystems verstärkt. Ähnlich wie die Verarbeitungsstufen des AdaptDRC-Algorithmus wird auch die neu eingeführte, adaptive Verstärkungsstufe über den SII gesteuert. Die subjektive Evaluation des neuen AdaptDRCplus-Algorithmus zeigt eine signifikante Verbesserung der Sprachverständlichkeit für normalhörende Probanden gegenüber dem AdaptDRC-Algorithmus.

AdaptDRC-Algorithmus

Der AdaptDRC-Algorithmus kombiniert eine zeit- und frequenzabhängige Verstärkung, welche bei einer niedrigen vorhergesagten Sprachverständlichkeit eine Verstärkung hochfrequenter Signalanteile bewirkt, und eine zeit- und frequenzabhängige Dynamikkompression, welche nicht hörbare Signalanteile verstärkt und laute Signalanteile abschwächt. Im Folgenden wird kurz auf die Wirkungsweise des AdaptDRC-Algorithmus eingegangen, für eine detaillierte Beschreibung sei der Leser an [3] verwiesen.

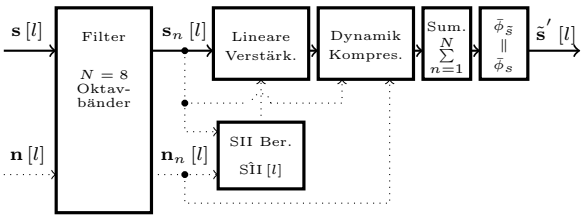

Abbildung 1: Signalflussplan AdaptDRC-Algorithmus

In Abbildung 1 ist der Signalflussplan des AdaptDRC-Algorithmus dargestellt. Hierbei ist die Annahme, dass sowohl das ungestörte Sprachsignal als auch eine Schätzung des Störgeräusches vorliegen. Das Sprachsignal $\mathbf{s}[l]$ und das Störgeräusch $\mathbf{n}[l]$ werden in nichtüberlappende Blöcke der Länge M unterteilt mit dem Blockindex l und anschließend mit Hilfe einer nichtdezimierenden Filterbank in $N = 8$ Oktavbänder mit Mittenfrequenzen von 125 Hz bis 16000 kHz gefiltert. Basierend auf den Subbandsignalen $\mathbf{s}_n[l]$ und $\mathbf{n}_n[l]$, $n = 1, \ldots, N$ wird anschließend eine Schätzung des SII $\hat{SII}[l]$ im l-ten Block berechnet. Diese steuert die zeit- und frequenzabhängige Verstärkung, sowie die zeit- und frequenzabhängige Dynamikkompression. Nach Summierung der Subbandsignale erfolgt die Angleichung des Ausgangspegels $\bar{\phi}_{\tilde{s}}$ auf den Eingangspegel $\bar{\phi}_s$.

AdaptDRCplus-Algorithmus

Das Ziel des AdaptDRCplus-Algorithmus ist die zusätzliche Verstärkung des Sprachsignals am Ausgang des AdaptDRC-Algorithmus, wenn die vorhergesagte Sprachverständlichkeit gering ist. Hierbei wird somit die Randbedingung des identischen Ein- und Ausgangspegels verworfen. Stattdessen wird angenommen, dass der Maximalpegel des Sprachsignals vor und nach der Verarbeitung identisch sein muss. Somit wird in praktischen Anwendungen des Algorithmus (z.B. Bahnhofsdurchsage, Mobiltelefon) trotz Anhebung des Langzeitpegels vermieden, den limitierten Maximalpegel

des Verstärkersystems zu überschreiten.
Die neu eingeführte Verarbeitungsstufe des AdaptDRCplus-Algorithmus umfasst eine Verstärkungsfunktion, die adaptiv über den SII gesteuert wird. Um die Randbedigung des konstanten Maximalpegels zu erfüllen, werden Signalanteile außerhalb einer festgelegten Amplitudengrenze abgeschnitten (Hard-Clipping). Je größer der Anteil der geclippten Samples, desto stärker treten hörbare Verzerrungen des Sprachsignals auf. Daher wird der sogenannte Clipping-Prozentsatz adaptiv über den SII gesteuert, sodass nur bei niedrigem Signal-Rausch-Abstand (SNR) deutliche Übersteuerungen des Sprachsignals durchgeführt werden. In diesem Fall werden Verzerrungen zum einen teilweise durch das Störgeräusch maskiert und zum anderen zugunsten einer guten Verständlichkeit vom Hörer akzeptiert.
Abbildung 2 stellt den Signalflussplan des AdaptDRCplus-Algorithmus dar. Zunächst wird das Sprachsignal $\mathbf{s}[l]$ wie bereits erläutert durch den AdaptDRC-Algorithmus verarbeitet. Gemäß (1) erfolgt anschließend die Anwendung eines blockweise berechneten Verstärkungsfaktors $g\,[l]$ auf das Ausgangssignal $\tilde{\mathbf{s}}^{'}\,[l]$ des AdaptDRC-Algorithmus, wodurch der Langzeitpegel des Sprachsignals erhöht wird:

$$\tilde{\mathbf{s}}^{''}\,[l] = g\,[l]\,\tilde{\mathbf{s}}^{'}\,[l] \tag{1}$$

Um die Verstärkung von unerwünschten Geräuschen in Sprachpausen zu vermeiden, erfolgt die Anwendung von (1) nur in Blöcken, die Sprache enthalten. Zur Sprachaktivitätserkennung (VAD) wird die ITU-T Empfehlung P.56 verwendet [6]. Der zeitabhängige Verstärkungsfaktor $g\,[l]$ berechnet sich gemäß (2). Darin beschreibt $s_{max}\,[l]$ die maximale Amplitude eines Blocks und $s_{q[l]}$ denjenigen Wert, der oberhalb des $(1 - q\,[l])$-Perzentils aller Samples des l-ten Blocks liegt:

$$g\,[l] = \frac{s_{max}\,[l]}{s_{q[l]}} \tag{2}$$

Der Term $q\,[l]$ bezeichnet den sogenannten adaptiven Clipping-Prozentsatz und gibt an, zu welchem Anteil das Signal übersteuert wird bzw. wie viele Samples des l-ten Blocks in Folge der Anwendung von (1) außerhalb der Amplitudengrenzen $\pm s_{max}\,[l]$ liegen. Der adaptive Clipping-Prozentsatz wird gemäß (3) über einen linearen Zusammenhang mit dem SII gesteuert:

$$q\,[l] = q_{max} \cdot \left(1 - \hat{SII}\,[l]\right) \tag{3}$$

Dabei wird auf die Schätzung des SII $\hat{SII}[l]$ aus der Berechnung des AdaptDRC-Algorithmus zurückgegriffen. q_{max} in (3) bezeichnet den maximalen Clipping-Prozentsatz. Dieser Wert wird extern festgelegt und bestimmt, zu welchem Anteil das Signal im Falle einer minimalen vorhergesagten Sprachverständlichkeit ($\hat{SII}\,[l] = 0$) übersteuert werden soll. Der tatsächlich verwendete, adaptive Clipping-Prozentsatz $q\,[l]$ ist stets kleinergleich q_{max}. Um Verzerrungen zwischen Blöcken zu vermeiden, wird der Verstärkungsfaktor aus (2) gemäß (4) mit einer

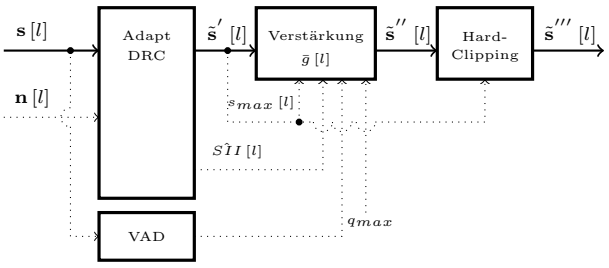

Abbildung 2: Signalflussplan AdaptDRCplus-Algorithmus

Konstanten α_g rekursiv geglättet:

$$\bar{g}\,[l] = \alpha_g \bar{g}\,[l-1] + (1 - \alpha_g)\, g\,[l] \tag{4}$$

α_g wird gemäß (5) durch eine Glättungszeitkonstante τ_g bestimmt, f_s bezeichnet dabei die Sampling-Frequenz:

$$\alpha_g = exp\left((-\tau_g f_s)^{-1}\right) \tag{5}$$

Tatsächlich berechnet sich das verstärkte Signal $\tilde{\mathbf{s}}^{''}\,[l]$ in (1) durch Anwendung von (4) anstelle von (2) auf das Ausgangssignal des AdaptDRC-Algorithmus $\tilde{\mathbf{s}}^{'}\,[l]$. Das Ausgangssignal des AdaptDRCplus-Algorithmus $\tilde{\mathbf{s}}^{'''}\,[l]$ ergibt sich schließlich, indem alle Samples von $\tilde{\mathbf{s}}^{''}\,[l]$, die außerhalb der Amplitudengrenzen $\pm s_{max}\,[l]$ liegen, abgeschnitten werden. In Abbildung 3 sind die durch den AdaptDRCplus-Algorithmus erzielten Langzeitpegel-Erhöhungen der Sprachsignale für verschiedene Werte von q_{max} und für verschiedene Eingangs-SNR dargestellt. Die Pegelerhöhungen entsprechen einer Verbesserung des Eingangs-SNR. Als Sprachmaterial wurden zehn zufällig gewählte Sätze des Oldenburger Satztests [7] verwendet und als Störgeräusch ein sprachgefärbtes, stationäres Rauschen. Die Glättungszeitkonstante betrug $\tau_g = 150\,\text{ms}$ und die Sampling-Frequenz 44,1 kHz. Es wird deutlich, dass bei niedrigen Eingangs-SNR und 20 % maximalem Clipping der Sprachpegel um bis zu ca. 6,5 dB durch den AdaptDRCplus-Algorithmus erhöht wird.

Subjektive Evaluation

Für die subjektive Evaluation des AdaptDRCplus-Algorithmus wurden Hörtests mit insgesamt elf normalhörenden Probanden durchgeführt. Dabei wurden dem Probanden Sätze bei gleichzeitigem Störgeräusch dargeboten. Die Aufgabe des Probanden bestand darin, die von ihm verstandenen Wörter mündlich wiederzugeben. Die Sprachverständlichkeit wurde anhand der korrekt verstandenen Wörter pro Satz gemessen. Die verwendeten Stimuli wurden erzeugt, indem einzelne Sätze durch den AdaptDRC- und den AdaptDRCplus-Algorithmus vorverarbeitet und mit verschiedenen Störgeräuschen bei unterschiedlichen Eingangs-SNR überlagert wurden. Zusätzlich wurden unverarbeitete Sätze als Referenz dargeboten. Als Sprachmaterial diente der Korpus des Oldenburger Satztests (OLSA) [7], dessen Sätze einer festen syntaktischen Struktur (Name, Verb, Zahl, Adjektiv, Objekt) unterliegen, z.B. Peter hat

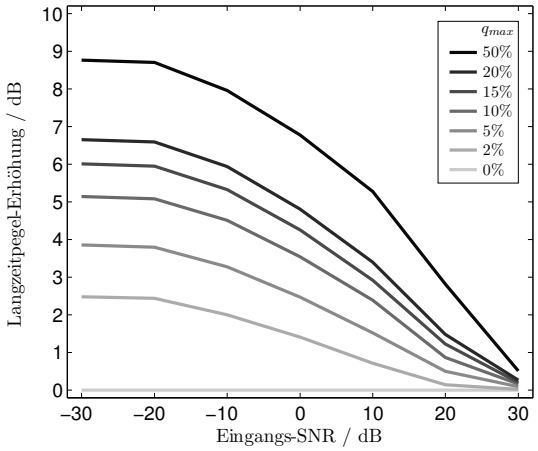

Abbildung 3: Erhöhung des Langzeit-Sprachpegels durch den AdaptDRCplus-Algorithmus für verschiedene Werte des maximalen Clipping-Prozentsatzes q_{max}. Als Sprachmaterial wurden zehn zufällig gewählte Sätze des Oldenburger Satztests [7] verwendet und als Störgeräusch ein sprachgefärbtes, stationäres Rauschen. Die Glättungszeitkonstante betrug $\tau_g = 150\,\text{ms}$ und die Sampling-Frequenz 44,1 kHz.

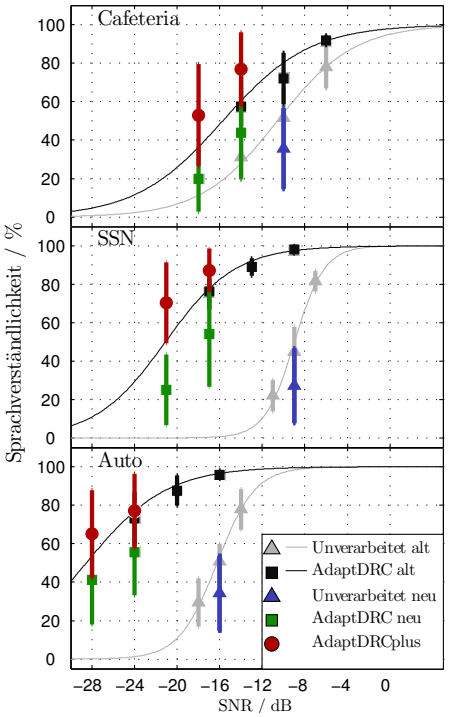

Abbildung 4: Ergebnisse der subjektiven Evaluation des AdaptDRCplus-Algorithmus für das Cafeteria-Störgeräusch (oben), das sprachgefärbte Rauschen (SSN, Mitte) sowie das Auto-Störgeräusch (unten). Farbige Symbole geben die über alle Probanden gemittelten Messwerte an, vertikale Linien die zugehörige Standardabweichung. Als *neu* gekennzeichnete Messwerte beschreiben die Daten dieser Studie. Zusätzlich sind für die Konditionen unverarbeitet und AdaptDRC die Daten aus [3, 5], als *alt* gekennzeichnet, abgebildet, wobei graue und schwarze Symbole die über alle Probanden gemittelten Messwerte beschreiben und gleichfarbige durchgezogene Linien daran angepasste psychometrische Funktionen.

drei grüne Messer. Trotz hoher perzeptiver Ähnlichkeit sind die Sätze semantisch nicht vorhersagbar. Für jede Kondition (Kombination aus Vorverarbeitungsstufe, Störgeräusch und Eingangs-SNR) wurde eine Testliste, bestehend aus je 20 Sätzen verwendet. Als Störgeräusche wurden eine Cafeteriaaufnahme (Cafeteria), ein sprachgefärbtes Rauschen (speech-shaped noise, SSN) und eine Innenraumaufnahme eines PKW (Auto) verwendet. Das SSN- und das Auto-Geräusch sind beide stationär, das Cafeteria-Geräusch hingegen instationär.

Für jedes Störgeräusch wurden Messungen bei drei verschiedenen Eingangs-SNR durchgeführt. Um eine Vergleichbarkeit zu den in [3, 5] erhobenen Daten zu ermöglichen, wurden zwei der drei Eingangs-SNR so wie in [3, 5] gewählt. Für das Cafeteria-Störgeräusch betrugen diese −10 dB für die Kondition unverarbeitet, sowie −14 dB und −18 dB jeweils für die Konditionen AdaptDRC und AdaptDRCplus. Beim SSN-Störgeräusch wurden als Eingangs-SNR −9 dB (unverarbeitet), sowie −17 dB und −21 dB (jeweils AdaptDRC und AdaptDRCplus) gewählt. Für das Auto-Geräusch lagen die Werte bei -16 dB (unverarbeitet), sowie -24 dB und -28 dB (jeweils AdaptDRC und AdaptDRCplus). Der Sprachpegel betrug konstant 60 dB SPL, der Störgeräuschpegel wurde für jeden Satz entsprechend des SNR angepasst. In einzelnen Konditionen wurde der Gesamtpegel um 4 dB abgesenkt. Dies war in jedem Störgeräusch jeweils beim kleinsten SNR der Fall, da ansonsten unangenehm laute Störgeräuschpegel dargeboten worden wären. Die Sampling-Frequenz betrug 44,1 kHz. Als maximaler Clipping-Prozentsatz des AdaptDRCplus-Algorithmus wurde $q_{max} = 20\,\%$ gewählt und als Glättungszeitkonstante $\tau_g = 100\,\text{ms}$.

In Abbildung 4 sind die Ergebnisse der subjektiven Evaluation des AdaptDRCplus-Algorithmus in Form von Prozent korrekt verstandener Wörter für die drei verwendeten Störgeräusche Cafeteria (oben), SSN (Mitte) und Auto (unten) bei den entsprechend gewählten Eingangs-SNR dargestellt. Farbige Symbole geben die über alle Probanden gemittelten Messwerte für die Konditionen unverarbeitet (blaue Dreiecke), AdaptDRC (grüne Quadrate) und AdaptDRCplus (rote Kreise) an, vertikale Linien beschreiben die Standardabweichung. Um die Daten dieser Studie (in Abbildung 4 mit *neu* bezeichnet) mit den bisherigen Evaluationen des AdaptDRC-Algorithmus vergleichen zu können, sind zusätzlich die über alle Probanden gemittelten Messwerte, sowie daran angepasste psychometrische Funktionen aus [3, 5] für die Konditionen unverarbeitet (graue Dreiecke und graue Linien) und AdaptDRC (schwarze Quadrate und schwarze Linien) abgebildet (in Abbildung 4 mit *alt* bezeichnet). In [3, 5] konnten bereits signifikante Sprachverständlichkeits-Gewinne für den AdaptDRC-Algorithmus gegenüber den unverarbeiteten Konditionen gemessen werden. Die in dieser Studie gemessene Sprachverständlichkeit für die Konditionen unverarbeitet und AdaptDRC liegt leicht unterhalb der entsprechenden Werte aus [3, 5], jedoch zeigt sich ein vergleichbarer Sprachverständlichkeits-Gewinn durch AdaptDRC gegenüber den unver-

Tabelle 1: Subjektiv gemessene Sprachverständlichkeit für die Konditionen AdaptDRC und AdaptDRCplus (auf ganze Zahlen gerundet)

	SNR	Adapt DRC	Adapt DRCplus	Gewinn
Cafete	-14 dB	$44 \pm 24\,\%$	$77 \pm 19\,\%$	33 %
	-18 dB	$20 \pm 17\,\%$	$53 \pm 27\,\%$	33 %
SSN	-17 dB	$54 \pm 28\,\%$	$87 \pm 11\,\%$	33 %
	-21 dB	$25 \pm 18\,\%$	$70 \pm 21\,\%$	45 %
Auto	-24 dB	$56 \pm 23\,\%$	$77 \pm 19\,\%$	21 %
	-28 dB	$41 \pm 23\,\%$	$65 \pm 23\,\%$	24 %

arbeiteten Konditionen. Durch den AdaptDRCplus-Algorithmus konnte in allen Störgeräuschen eine weitere Sprachverständlichkeitsverbesserung gegenüber dem AdaptDRC-Algorithmus erzielt werden, in einzelnen Konditionen über 40 % (siehe Tabelle 1). Um diese Ergebnisse statistisch belegen zu können, wurden zweiseitige t-Tests durchgeführt. Das Signifikanzniveau wurde dafür auf $\alpha = 0,05$ festgelegt. Dabei zeigte sich, dass die Sprachverständlichkeitsverbesserung durch den AdaptDRCplus-Algorithmus gegenüber dem AdaptDRC-Algorithmus in allen Störgeräuschen signifikant ist: $p < 0,001$ (Cafeteria), $p \leq 0,002$ (SSN) und $p < 0,03$ (Auto).

Zusammenfassung

In dieser Studie wurde der AdaptDRC-Algorithmus [3, 5] mit einer Verstärkungsfunktion zur adaptiven Pegelerhöhung des Sprachsignals kombiniert. Dabei wurden die bisher angenommenen physikalischen Randbedingungen geändert, sodass nicht mehr der Langzeit-, sondern der Maximalpegel am Ein- und Ausgang des Algorithmus konstant bleibt. In subjektiven Untersuchungen mit normalhörenden Probanden konnten zum einen die Evaluationsergebnisse aus [3, 5] bzgl. des AdaptDRC-Algorithmus reproduziert und zum anderen sowohl in stationären als auch in instationären Störgeräuschen weitere Sprachverständlichkeits-Gewinne durch den neuen AdaptDRCplus-Algorithmus im Vergleich zum AdaptDRC-Algorithmus gemessen werden.

Literatur

[1] Rennies, J., Schepker, H., Holube, I. und Kollmeier, B.: *Listening effort and speech intelligibility in listening situations affected by noise and reverberation.* The Journal of the Acoustical Society of America, 136(5):2642–2653, Nov. 2014.

[2] Cooke, M., Mayo, C. und Valentini-Botinhao, C.: *Intelligibility-enhancing speech modifications: the Hurricane challenge.* In: *Proc. Interspeech*, Seiten 3552–3556, Lyon, Frankreich, Aug. 2013.

[3] Schepker, H., Rennies, J. und Doclo, S.: *Improving speech intelligibility in noise by SII-dependent preprocessing using frequency-dependent amplification and dynamic range compression.* In: *Proc. Interspeech*, Seiten 3577–3581, Lyon, Frankreich, Aug. 2013.

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

[6] ITU-T: *Recommendation P.56, Objective Measurement of Active Speech Level.* International Telecommunication Union, Geneva, 1993.

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