

# Verbesserung von Sprachverständlichkeit bei Telefongesprächen für Probanden mit einem Hochtonhörverlust

Jan Rennies<sup>1</sup>, Vanessa Rautenstrauch<sup>1</sup>, Inga Holube<sup>2</sup>

<sup>1</sup>Fraunhofer Institut für Digitale Medientechnologie IDMT, Projektgruppe Hör-, Sprach- und Audiotechnologie, Oldenburg, E-Mail: jan.rennies@idmt.fraunhofer.de

<sup>2</sup>Jade Hochschule, Institut für Hörtechnik und Audiologie, Oldenburg, E-Mail: inga.holube@jade-hs.de

## Einleitung

Das Telefon ist ein unentbehrlicher Bestandteil der Alltagskommunikation geworden. Bei Menschen mit Hörschädigung ist diese wichtige Alltagskompetenz jedoch oft eingeschränkt. Die beeinträchtigten auditorischen Funktionen führen zu einer Abschwächung und/oder einer Verzerrung von Sprache, was sich negativ auf die Sprachverständlichkeit auswirkt. Neben der reduzierten Sprachwahrnehmung führen die Eigenschaften des Telefonnetzes zu weiteren Kommunikationsproblemen. Die Sprache am Telefon besitzt im Vergleich zu Sprache einer normalen Unterhaltung eine mindere Übertragungsqualität. Dies liegt zum einen in der Bandbegrenzung und zum anderen in der Lautstärke des Telefonsignals begründet, da der zur Verfügung stehende Dynamikbereich vom Telefonnetz oft nicht ausgenutzt wird. Aus der Praxis der Hörgeräteakustik ist bekannt, dass das Telefonieren mit Hörgeräten oft nicht reibungslos funktioniert. Häufig treten Rückkopplungen auf, wenn das Hörgerätemikrofon zur Telekommunikation verwendet wird. Darüber hinaus wird der Telefonhörer vom Nutzer häufig an die Ohrmuschel gehalten und nicht an das Mikrofon. Demzufolge existieren viele Gründe, Hörgerätealgorithmen direkt in Telefone zu integrieren. Das Ziel der vorliegenden Arbeit ist es, verschiedene Signalverarbeitungsstrategien zu eruiieren, die das Sprachverstehen von Schwerhörigen am Telefon in Anwesenheit von Störgeräuschen verbessern sollen. Dafür wurden Sprachsignale von den verwendeten Strategien vorverarbeitet, die anschließend für die Messung der Sprachverständlichkeit und Klangqualität verwendet wurden. Verglichen wurde das Peak-Clipping mit zwei weiteren Algorithmenansätzen: einer adaptiven Verstärkungsmaximierung und einer SII-basierten Dynamikkompression.

## Algorithmen

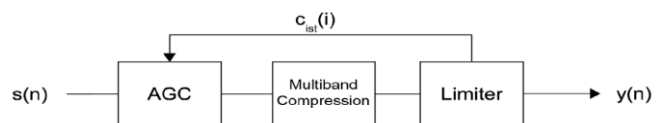
### Peak-Clipping:

Beim „Peak-Clipping“ handelt es sich um eine Funktion, die häufig in Hörgeräten genutzt wird, wenn der Hörer durch eine zu hohe Verstärkung an seine Grenzen gerät. Die Telefonsignale wurden linear verstärkt und anschließend begrenzt, sodass der Grenzwert von 0 dB FS nicht überschritten wird. Insgesamt wurde das in dieser Studie verwendete Sprachmaterial mit zwei Varianten des Peak-Clipping bearbeitet, indem die Verstärkung so gewählt wurde, dass in jedem Satz entweder 5% oder 10% der Samples übersteuern. Die durch diese Methode erzeugten Datensätze werden „Clip-5%“ und „Clip-10%“ genannt.

### Lautheitsmaximierung:

In [1] wurde ein mehrstufiges System vorgestellt, bei dem eine adaptive Lautstärkanhebung des Telefonsignals

durchgeführt wird. Das Ziel dieses Systems ist es maximal ausgesteuerte Signale zu erzeugen, die nicht übersteuern. Das schematisch in Abb. 1 dargestellte System setzt sich aus einer automatischen Verstärkungsregulierung (AGC), einer Filterbank, einem Multibandkompressor und einem sog. Look-ahead-Limiter zusammen. Dabei wird das Telefonsignal frequenzabhängig (vier Kanäle) in seiner Dynamik verändert, ohne dabei Clipping-Artefakte entstehen zu lassen. Durch Variation der Verstärkungen in den einzelnen Frequenzkanälen wurden verschiedene Formen der Frequenzgewichtung erzeugt: flaches Spektrum (abgekürzt LM für Lautheitsmaximierung), Mittenbetonung (LM+MT) bzw. Höhenbetonung (LM+HT).



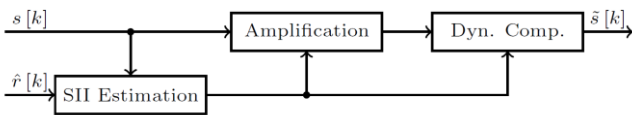
**Abbildung 1:** Schematische Darstellung der Lautheitsmaximierung, die das Eingangssignal  $s(n)$  derart komprimiert und verstärkt, dass ein lauterer Ausgangssignal  $y(n)$  entsteht, ohne dass dabei Peak-Clipping auftritt.

### Adapt-DRC-Verfahren:

In [2] wurde der Vorverarbeitungsalgorithmus AdaptDRC (adaptive dynamic range compression, Schema siehe Abb. 2) zur Verbesserung des Sprachverstehens im Störgeräusch vorgestellt. In Abhängigkeit vom vorliegenden Störgeräusch wird hierbei eine frequenzabhängige Verstärkungsregelung und Dynamikkompression verwendet. Als Steuergröße dient eine vereinfachte Form des Sprachverständlichkeitsindex (SII). Das auszugebene Sprachsignal  $s[k]$  wird bei einem gegebenen Störgeräusch  $r[k]$  adaptiv gemäß Schätzungen des SII modifiziert, so dass in Situationen mit schlechter Verständlichkeit das Sprachsignal am Ausgang zu einem rosa Spektrum geformt und zudem in seiner Dynamik komprimiert wird, wohingegen bei hinreichend guter Verständlichkeit keine Veränderung des Sprachsignals erfolgt. Der Pegel am Eingang und Ausgang bleibt dabei erhalten, d.h. es erfolgt keine Erhöhung des Breitbandpegels.

## Experimente

An den Experimenten nahmen zehn Probanden (drei davon weiblich) mit mittel- bis hochgradiger Hochtonschwerhörigkeit teil. Das Durchschnittsalter lag bei 69 Jahren. Mit allen Probanden wurden sowohl Sprachverständlichkeitsmessungen als auch Sprachqualitätsmessungen durchgeführt. In beiden Fällen wurde das Sprachmaterial des Oldenburger Satztests verwendet. Als Störgeräusche dienten das stationäre, sprachsimulierende Rauschen des Oldenburger Satztests sowie ein aus vielen simultanen Gesprächen bestehendes Cafeteria-Rauschen bei jeweils 80 dB SPL.



**Abbildung 2:** Schematische Darstellung des AdaptDRC-Verfahrens zur störgeräuschadaptiven Vorverarbeitung von Sprachsignalen. Als Steuergröße des Algorithmus dient der Sprachverständlichkeitsindex (SII).

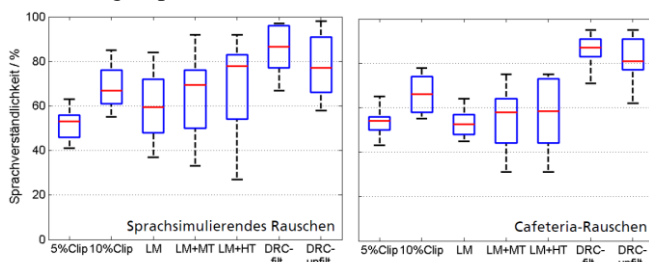
Alle Stimuli wurden monaural (links) über einen Kopfhörer (HD650) dargeboten. Die eingeschränkte Bandbreite der Sprachsignale wurde durch einen Bandpassfilter simuliert. Für die mit AdaptDRC verarbeiteten Signale wurden zusätzlich Messungen mit ungefilterter Sprache durchgeführt.

Die Sprachverständlichkeit wurde als Anteil korrekt verstandener Wörter bei festem SNR erfasst. Der SNR wurde individuell für jeden Probanden und jedes Störgeräusch anhand einer Referenzsituation festgelegt und entsprach dem SNR, bei dem mit einem 5%-geclipten Sprachmaterial die Hälfte der Wörter korrekt verstanden wurde. Dieser Referenz-SNR lagen zwischen ca. -7 und -5 dB (sprachsimulierendes Rauschen) bzw. -3,5 und +3 dB (Cafeteria-Rauschen) und damit ca. 5 dB oberhalb der Referenz für Normalhörende [2].

Mit denselben Probanden wurden Qualitätsmessungen durchgeführt. Hierfür wurde die subjektive Klangpräferenz der mit den unterschiedlichen Verfahren bearbeiteten Sprachsignale in einem vollständigen Paarvergleich ermittelt, wobei die Darbietung der Sprachsignale in Ruhe erfolgte.

## Ergebnisse

Die Boxplots in Abb. 3 repräsentieren die für das Probandenkollektiv ermittelte Sprachverständlichkeit für beide Störgeräusche und alle Algorithmen. Für das sprachsimulierende Rauschen (links) konnten alle untersuchten Verfahren die Verständlichkeit gegenüber der Referenzkondition im Median verbessern. Der größte Gewinn wurde dabei durch AdaptDRC erzielt. Für die Lautheitsmaximierung zeigte sich, dass eine Mitten- bzw. eine Höhenbetonung im Median progressiv einen Gewinn an Verständlichkeit bewirkt, wobei die erreichte Sprachverständlichkeit interindividuell stark schwankte und trotz eines ähnlichen Hörverlustes nicht für alle Probanden zu beobachten war. Für das Cafeteria-Rauschen (rechts) wurde im Median nur durch Erhöhung des Clipping und durch AdaptDRC eine Verbesserung der Verständlichkeit erzielt, nicht aber durch Lautheitsmaximierung. Für beide Störgeräusche war die Verständlichkeit bei AdaptDRC für bandpass-gefilterte Sprache durch die Energiekonzentration im sprachrelevanten Bereich leicht besser als für breitbandige Sprache.



**Abbildung 3:** Boxplot-Darstellung der Sprachverständlichkeit für die untersuchten Algorithmen und beide Störgeräusche.

**Tabelle 1:** Intervallskala der Qualitätsbewertung.

Algorithmus	Skalenwert
AdaptDRC-Cafeteria-ungefiltert	3,0
AdaptDRC-Sprachrauschen-ungefiltert	2,9
AdaptDRC-Sprachrauschen-gefiltert	2,5
AdaptDRC-Cafeteria-gefiltert	2,3
Clipping 10%	1,1
LM+HT	1,1
Clipping 5%	0,8
LM	0,4
LM+MT	0,0

Der Präferenzvergleich ergab, dass die Probanden die dargebotene Sprache in der Regel konsistent und konkordant bewerteten. Die Häufigkeiten der Präferenzurteile wurden nach Thurstone's Law of Comparative Judgement skaliert. Die resultierenden Skalenwerte (Tab. 1) ergaben deutliche Tendenzen zu einer Präferenz von AdaptDRC gegenüber allen anderen Verfahren, wenngleich durch die geringe Anzahl an Probanden nur qualitative Aussagen möglich sind. Wie erwartet wurden die breitbandigen Sprachsignale (*unfiltered*) dabei gegenüber den bandpassgefilterten Signalen bevorzugt.

## Diskussion

Diese Studie zeigte, dass die Verständlichkeit von Telefonsprache im Störgeräusch für Probanden mit Hochtonhörverlust deutlich verbessert werden kann. Der mögliche Gewinn hängt dabei jedoch stark von der Art des Störgeräusches, dem jeweiligen Algorithmus und auch vom individuellen Probanden ab. Zwar zeigte sich, dass eine Anhebung der hohen Frequenzanteile auch mit dem Lautheitsmaximierer im Median zu besseren Ergebnissen führte, jedoch war nur durch AdaptDRC in beiden Störgeräuschen eine Verständlichkeitsverbesserung zu beobachten. Interessanterweise wurde derselbe Algorithmus auch hinsichtlich seiner Klangqualität am besten bewertet. Ein möglicher Grund für die hier beobachtete Überlegenheit von AdaptDRC gegenüber der Lautheitsmaximierung könnte eine zwar wirkungsvolle aber nicht zu stark in das Signal eingreifende Dynamikkompression sein, die einerseits die Verständlichkeit verbessert, andererseits aber das Signal nicht zu stark verzerrt (für Details siehe [3]). Insgesamt verdeutlichen die Ergebnisse das große Potenzial von Algorithmen zur Verständlichkeitsverbesserung in Telefonanwendungen, um auch Schwerhörenden ein leichteres Sprachverstehen zu ermöglichen.

## Literatur

- [1] Asendorf, R., Oetting, D. und Bitzer, J. (2013). Improvement of speech intelligibility in telephone-systems by adaptive gain maximization, Fortschritte der Akustik – DAGA 2013.
- [2] Schepker, H., Rennies, J. und Doclo, S. (2013). Improving speech intelligibility in noise by SII-dependent preprocessing using frequency dependent amplification and dynamic range compression, Proc. Interspeech 2013, 3577-3581.
- [3] Rautenstrauch, V. (2013). Wirkungsweise verschiedener Signalverarbeitungsstrategien zur Verbesserung der Sprachverständlichkeit von Telefongesprächen bei Probanden mit einem Hochton-Hörverlust, Bachelorarbeit, Jade Hochschule, Oldenburg.