

# Entwicklung und Evaluation eines auditorisch motivierten Algorithmus zur Verbesserung der Sprachqualität in Telefonsystemen

Tobias Bruns<sup>1</sup>, Dirk Oetting<sup>1</sup>, Jan Rennies<sup>1</sup>, Volker Hohmann<sup>2</sup>

<sup>1</sup> Fraunhofer IDMT Hör-, Sprach- und Audiotechnologie, 26129 Oldenburg,  
E-Mail: tobias.bruns@idmt.fraunhofer.de

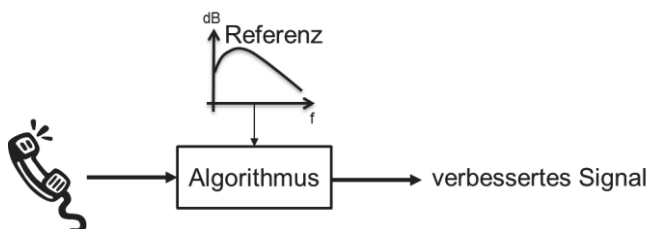
<sup>2</sup> Carl von Ossietzky Universität Oldenburg, Medizinische Physik, 26129 Oldenburg

## Einleitung

Bei der Übertragung von Telefonsignalen kann es zu unterschiedlichsten Qualitätseinbußen kommen [1]. Neben statischen Einschränkungen wie z.B. dem begrenzten Übertragungsbereich, haben dynamische Änderungen wie die Gesprächs- und Übertragungssituation Auswirkungen auf den wahrgenommenen Klang. Insbesondere dynamische Unterschiede sind nur durch eine adaptive Anpassung auszugleichen, da unterschiedliche Sprecher und Situationen im Vorhinein nicht feststehen. Der hier vorgestellte Algorithmus hat zum Ziel Telefonsignale adaptiv in Richtung eines optimalen Sprachsignals einzustellen.

## Anforderungen

Zur Verbesserung von Sprachsignalen ergeben sich einige grundlegende Anforderungen an den Algorithmus. Große Abweichungen von Signaleigenschaften wie die spektrale Verteilung der Sprachenergie im Vergleich zu einem optimalen Sprachsignal sollten von dem Algorithmus kompensiert werden. Dabei steht der subjektive Eindruck bei der Beurteilung einer Qualitätsoptimierung im Vordergrund. Der Algorithmus sollte daher einer grundlegenden auditorischen Motivation folgen. Dazu gehört die Verarbeitung in Bändern mit kritischer Bandbreite um eine gezielten Lautheits- und Schärferegelung zu erreichen. Das Verarbeitungsschema ist in Abb. 1 dargestellt.



**Abbildung 1:** Verarbeitungsschema des Algorithmus. Die spektralen Signalparameter des Eingangssignals werden mit einem Referenzsignal verglichen. Abweichungen von der Referenz werden vom Algorithmus kompensiert.

## Referenzwerte

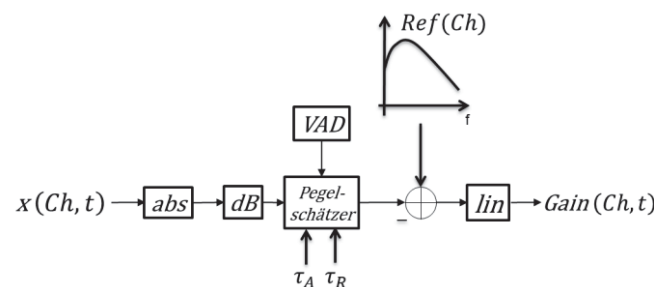
Da die Verarbeitung des Algorithmus auf vorher definierte Referenzwerte beruht, müssen die Eigenschaften üblicher guter Sprachsignale vorher bekannt sein. Zur Festlegung der Referenzwerte wurde der Frequenzgang verschiedener Telefonhörer gemessen und das Spektrum eines Sprachreferenzsignals ermittelt.

Als Basis für die vom Algorithmus angestrebten Zielwerte diente das „International Speech Test Signal“ (ISTS) [2].

Das ISTS basiert auf echten Sprachsignalen und vereinigt typische Signaleigenschaften, wie ein übliches Langzeit-spektrum und einen Dynamikumfang von etwa 30 dB. Die Übertragungsfunktionen der gemessenen Telefonhörer zeigen alle einen ähnlich begrenzten Übertragungsbereich von 750 Hz bis 4 kHz mit einem betonten Mittenbereich um 2 kHz. Zwischen den einzelnen Hörern zeigen sich allerdings auch starke Abweichungen von über 10 dB. Zur Linearisierung des Übertragungsverhaltens wurde daher ein Filter entworfen, der nur den grundlegenden Verlauf ausgleicht ohne tiefere Frequenzanteile über die Kapazität der Telefonhörer hinweg zu verstärken.

## Algorithmus

Die Funktionsweise des Algorithmus basiert auf einer automatischen Pegelsteuerung in Bark-skalierten Frequenzbändern bis 4kHz. Dabei wird in jedem Frequenzband eine Zielverstärkung in Abhängigkeit zum Referenzsignal ermittelt. Das Signalspektrum wird in jedem Frequenzband zum Referenzsignalpegel gezogen. Abbildung 2 verdeutlicht die Verstärkungsberechnung pro Frequenzband.



**Abbildung 2:** Blockschaltbild eines Kanals zur Berechnung der Zielverstärkung. Die Differenz des Pegelschätzers pro Kanal zum Referenzsignal wird als Zielverstärkung verwendet.

Ein logarithmischer Pegelschätzer analysiert das Eingangssignal in jedem Kanal und vergleicht die Werte mit abgespeicherten Referenzwerten. Entscheidend für die Verarbeitung ist die Verwendung einer kurzen Attackzeit und einer langen Releasezeiten. Bei sehr starken Pegelanstiegen reagiert der Algorithmus mit einer schnellen Reduktion der Verstärkung. Die lange Releasezeit soll einer qualitätsmindernden Dynamikreduktion vorbeugen [3].

## Validierung im Hörversuch

Zunächst wurden 28 Telefonsprachsignale (Kurzgeschichten) von verschiedenen Telefonen mit unterschiedlichen Sprechern auf einem digitalen VoIP-Anrufbeantworter aufgenommen. Die Aufnahmen umfassen ein ausgewogenes Abbild für verschiedene Endgeräte und unterschiedliche

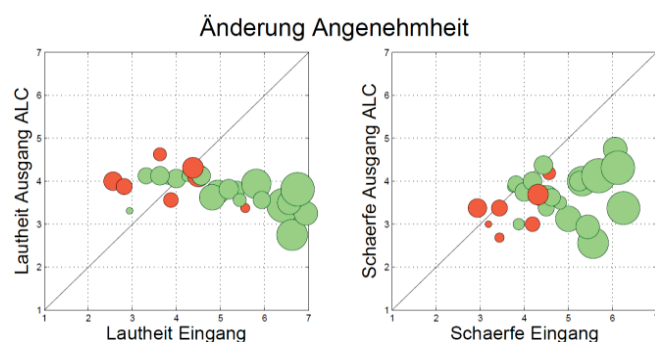
Sprecher. Aus jeder Aufnahme wurde jeweils ein Satz für die Experimente entnommen.

Im ersten Experiment wurde eine kategoriale Qualitätsbeurteilung für die Merkmale Lautheit, Schärfe, Höranstrengung, Angenehmheit, Natürlichkeit und Gesamtqualität durchgeführt. Die Beurteilung folgte auf einer diskreten 7-stufigen Skala für das unverarbeitete und das verarbeitete Signal. In diesem Experiment wurden 16 normalhörende Probanden befragt.

Zur Beurteilung der individuellen Präferenz wurde in einem zweiten Experiment ein direkter Präferenzvergleich zwischen unbearbeiteten und bearbeiteten Signalen durchgeführt. Dieser Vergleich zeigt auch Einflüsse von geringen Signalunterschieden, die durch die einzelnen Qualitätsmerkmale aus dem ersten Experiment nicht vollständig erfasst wurden. Es wurden weiteren 10 Probanden jeweils randomisierte Signalpaare im 2-AFC Verfahren dargeboten.

## Ergebnisse

Die Verarbeitung des Algorithmus hatte zum einen das Ziel Lautheit und Schärfe gezielt zu steuern, um dadurch einen positiven Einfluss auf die Qualität zu erzielen. Aus diesem Grund sollen diese beiden Merkmale in der Auswertung als Basis der Ergebnisdarstellung dienen.



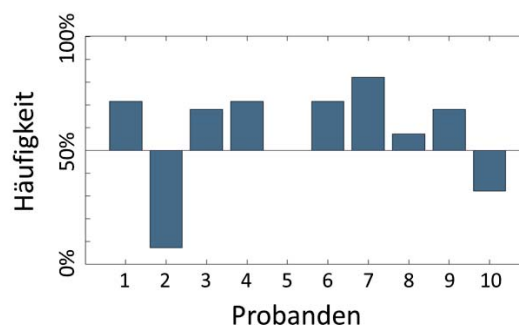
**Abbildung 3:** Lautheits- (links) und Schärfebewertungen (rechts) und Änderung der Dimension „Angenehmheit“ als farblich markierte Kreise.

In Abb. 3 wird die Lautheits- bzw. Schärfebewertung vor der Verarbeitung entlang der Abszisse und nach der Verarbeitung entlang der Ordinate abgebildet. Wenn keine Veränderung dieser Merkmale wahrgenommen wurde, befindet sich der dazugehörige Datenpunkt auf der Winkelhalbierenden. Wurde die Lautheit/Schärfe nach der Verarbeitung als deutlich geringer empfunden, so befindet sich der dazugehörige Datenpunkt rechts von der Winkelhalbierenden. Der Umfang der Kreise zeigt die Bewertungsänderung des dritten Merkmals. Ein roter Kreis signalisiert eine Verringerung ein grüner Kreis eine Erhöhung auf der Skala.

Die Daten zeigen, dass Signale, die zunächst als sehr laut bzw. sehr scharf beurteilt wurden, von der Verarbeitung profitierten und am Ausgang des Algorithmus ausgewogener bewertet wurden. Jedoch wurden bereits vorher als mittelgradig scharf bewertete Signale nach der Verarbeitung als eher zu dumpf empfunden, was sich negativ auf die

anderen Merkmale auswirkte. Für die überwiegende Anzahl der Stimuli wurde eine Verbesserung der Angenehmheit erzielt. Alle Qualitätsmerkmale korrelieren signifikant mit Lautheit und Schärfe. Die Gesamtqualität wurde schlechter bewertet und korrelierte nicht signifikant mit einer Schärfenänderung.

In Abb. 4 sind die Ergebnisse des Präferenzvergleichs der einzelnen Probanden, gemittelt über alle Testsignale, dargestellt. Auf der y-Achse ist aufgetragen, wie häufig das verarbeitete Signal gegenüber dem unverarbeiteten präferiert wurde. Außer einzelnen individuellen Abweichungen, zeigt sich eine überwiegende Präferenz für den Algorithmus.



**Abbildung 4:** Ergebnisse des Präferenzvergleichs. Aufgetragen ist die Häufigkeit, mit der die Verarbeitung gegenüber dem Originalsignal bevorzugt wurde.

## Zusammenfassung

Gerade besonders laute bzw. scharfe Signale wurden nach der Verarbeitung überwiegend besser bewertet. Es ließ sich außerdem eine deutliche Steigerung der Angenehmheit feststellen. Bereits vorher als ausgewogen scharf bewertete Signale wurden nach der Verarbeitung eher schlechter beurteilt. Diese Signale sind als zu dumpf bewertet worden. Dieses Defizit ließe sich durch Anpassen des statischen Ausgleichsfilters verhindern. Eine gleichbleibend schlechte Bewertung für Gesamtqualität könnte auf Verzerrungen oder auf eine Dynamikreduktion durch die Verarbeitung hindeuten. Im Sinne dieser Problematik sollten weitere Untersuchungen und Weiterentwicklungen des Algorithmus folgen.

## Literatur

- [1] ITU-T: Effect of transmission impairments. International Telecommunication Union, Recommendation P.11 (1994)
- [2] Holube, I; Fredelake, S; Vlaming, M; Kollmeier, B.: Development and analysis of an International Speech Test Signal (ISTS). International Journal of Audiology 49 (2010), 891-903
- [3] Hansen, M.: Effects of multi-channel compression time constants on subjectively perceived sound quality and speech intelligibility. Ear and Hearing 23(4) (2002), 369-380.