

# Intelligente Konferenzsysteme für natürliche Freisprechkommunikation

*Stefan Goetze, Jan Rennies, Jens-E. Appell*

Fraunhofer IDMT, Hör-, Sprach- und Audiotechnologie, Oldenburg

*Kontakt: s.goetze@idmt.fraunhofer.de*

## ZUSAMMENFASSUNG

In modernen Arbeitsprozessen gewinnt der Einsatz von akustischen Kommunikationssystemen zunehmend an Bedeutung. Häufig kommt es jedoch vor, dass die Kommunikation durch äußere Einflüsse wie Störgeräusche oder Nachhall gestört wird. Während dies bereits für Normalhörende eine besondere Herausforderung darstellt, können die daraus resultierenden Beeinträchtigungen der Kommunikation für Menschen mit Schwerhörigkeit gravierend sein und sie gegebenenfalls sogar von der Nutzung des Systems ausschließen. Intelligente technologische Hilfsmittel können zu einer verbesserten Kommunikation beitragen, indem sie Störgeräusche und Nachhall reduzieren. Zudem ist es möglich die akustischen Signale auf den individuellen Nutzer zugeschnitten zu bearbeiten und somit eine Teilkompensation einer Schwerhörigkeit zu erreichen. Durch den Einsatz objektiver Maße zur rechnergestützten Abschätzung der Sprachverständlichkeit kann darüber hinaus die Qualität der akustischen Kommunikation kontinuierlich überprüft werden. Dieser Beitrag diskutiert die Anwendung solcher Technologien als Lösung in Kommunikationssystemen.

## ABSTRACT

In modern work processes, the role of acoustic communication systems becomes more and more important. In practice, however, interfering noises or reverberation can disturb the communication. While this already challenges normal-hearing users, the consequences for people with hearing impairments can be serious and may even prevent them from using the communication systems at all. Intelligent technological methods can improve the communication quality by removing undesired noise and reverberation. A further pre-processing of the acoustic signals tuned to the needs of the individual user can partly compensate for the detrimental effects of a possible hearing loss. Additionally, objective methods to estimate speech intelligibility can be used to continuously monitor the quality of acoustic communication. In this contribution, the potential of such technologies is discussed with respect to enhancement of communication systems.

## EINLEITUNG

Akustische Kommunikation hat durch die Entwicklung moderner Kommunikationstechnologien eine grundlegende Veränderung erfahren. Im privaten Bereich sowie als Basis vieler Arbeitsprozesse ist der Einsatz von mobilen Kommunikationsgeräten mittlerweile ebenso selbstverständlich wie die Nutzung von videogestützten Konferenz- oder Chatsystemen. Vom Sprecher gesendete akustische Signale werden dabei durch ein oder mehrere Mikrofone aufgenommen, durch ein System übertragen und über Lautsprecher oder Kopfhörer für den Empfänger wiedergegeben. Abhängig vom Einsatzgebiet des Kommunikationsgerätes treten dabei unterschiedliche akustische Gegebenheiten auf, die die Kommunikation behindern können. Abbildung 1 veranschaulicht diese Problematik in allgemeiner Form.

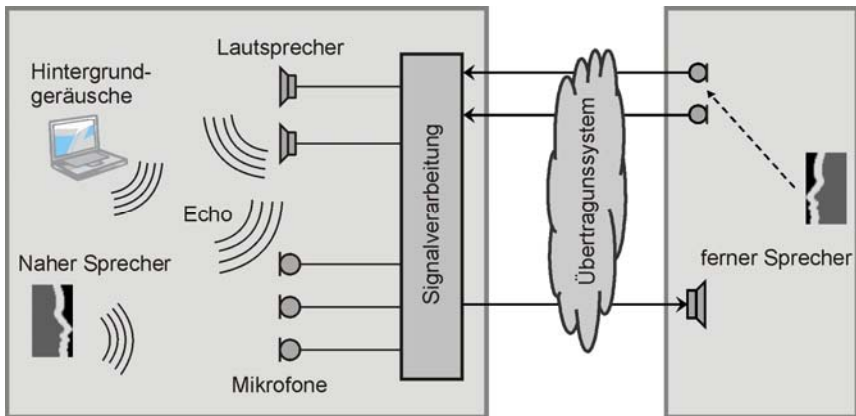


Abbildung 1: Problematik bei akustischen Kommunikationssystemen. Das Sprachsignal des nahen Sprechers wird überlagert von Hintergrundgeräuschen, Echos und Nachhall sowie dem empfangenen Signal des fernen Sprechers.

Das Signal des fernen Sprechers wird aufgenommen, übertragen und über Lautsprecher in den Raum abgegeben, in dem sich der nahe Sprecher befindet. Durch die Raumakustik bedingt bilden sich dort Reflektionen und Echos. Zusätzlich können weitere Geräuschquellen aktiv sein wie zum Beispiel Lüfter elektrischer Geräte oder Gespräche anderer Personen. Diese zusätzlichen akustischen Signale überlagern sich mit dem Sprachsignal, das der nahe Sprecher dem fernen Sprecher übermitteln möchte. Das Signal, das der ferne Sprecher empfängt, ist also ohne weitere Verarbeitung eine Mischung aus gewünschtem Sprachanteil und unerwünschten Signalanteilen (Störgeräusche inkl. der am Ort des fernen Sprechers aufgenommenen Signale, akustische Echos und Hallanteile). Diese unerwünschten Störgeräusch- und Nachhallanteile können die Qualität der Kommunikation erheblich beeinträchtigen, da sie das gewünschte Sprachsignal schwerer verständlich machen.

In klassischen Kommunikationssystemen wie Telefonen besteht die beschriebene Problematik nur bedingt, da sich der Mund des Sprechers meistens sehr nah an den Mikrofonen befindet (close-talk). Die zunehmend eingesetzten Konferenzsysteme erlauben jedoch die gleichzeitige Nutzung durch mehrere Personen. Dies wird zumeist dadurch realisiert, dass das Mikrofon für alle erreichbar platziert wird, zum Beispiel in der Mitte des Konferenztisches. In solchen Fällen ist der Abstand zwischen den Sprechern und dem Mikrofon erheblich größer (non-close talk) und Störgeräusche und Nachhall haben einen entsprechend größeren Einfluss auf die Kommunikationsqualität. Dies stellt bereits für Normalhörende durch die erhöhte Höranstrengung eine besondere Herausforderung dar und kann bei längerer Nutzung zu Müdigkeit oder erhöhtem Stress führen. Für Menschen mit Schwerhörigkeit können die resultierenden Beeinträchtigungen der Kommunikation gravierend sein und sie in Extremfällen sogar von der Nutzung des Systems ausschließen, da sie - abhängig vom individuellen Hörverlust - bereits unter akustisch guten Bedingungen Probleme in der Sprachkommunikation haben können. Da 16% der erwach-

senen Europäer (ca. 70 Mio.) an einem behandlungsbedürftigen Hörverlust leiden und nur 20% dieser Personen ein Hörgerät nutzen, ist die Gefahr einer solchen Ausgrenzung der Betroffenen groß und die Problematik wird sich vor dem Hintergrund des demografischen Wandels verstärken, da Schwerhörigkeit überwiegend altersbedingt ist (Davis, 2003; Shield, 2006). Eine Verbesserung der Kommunikationsqualität in Konferenzsystemen ist also sowohl für Normal- als auch für Schwerhörende sehr wünschenswert.

Raumakustische Maßnahmen können die Einflüsse von Nachhall und Störgeräuschen reduzieren (z.B. durch zusätzliche Dämmung oder räumliche Trennung der Quellen), nicht jedoch die speziellen Probleme von Hörschäden ausgleichen. Außerdem können sie aus architektonischen oder finanziellen Gründen unter Umständen nicht oder nicht in akustisch sinnvollem Umfang realisierbar sein. Dieser Beitrag beschäftigt sich mit technischen Methoden und Verfahren, die über raumakustische Maßnahmen hinaus gehen und legt dabei den Fokus auf die Verbesserung der Signalqualität in Kommunikationssystemen. Die vorgestellten Methoden umfassen zum einen Signalverarbeitungsstrategien für die allgemeine Verbesserung der Signalqualität, von der sowohl Normal- als auch Schwerhörende profitieren können. Zum anderen werden auf den individuellen Hörverlust angepasste Verarbeitungsmethoden zur Kompensation von Hörstörungen betrachtet. Zusätzlich wird ein System vorgestellt, das eine kontinuierliche technische Überwachung der Kommunikationsqualität ermöglicht und somit eine Rückmeldung an den Sprecher über die Verständlichkeit der eigenen Sprache auf Empfängerseite bietet.

## **VERBESSERUNG DER SIGNALQUALITÄT**

Im Folgenden werden Methoden für den Einsatz in Kommunikationssystemen vorgestellt, von denen sowohl für Normalhörende als auch für Schwerhörende profitieren können.

## **Störgeräuschreduktion**

Für die Kompensation von Störgeräuschen können einkanalige sowie mehrkanalige Ansätze Anwendung finden, je nachdem, ob ein oder mehrere Mikrofone zur Signalaufnahme genutzt werden können. Da bei der Verwendung nur eines Mikrofons immer Nutzanteil des Signals und Störung gleichzeitig aufgenommen werden, haben einkanalige Algorithmen die Unterdrückung des Störgeräusches zum Ziel, indem sie das Mikrofonsignal in verschiedenen Frequenzbändern in Abhängigkeit vom Signal-Rausch-Abstand abschwächen. Ist z.B. in einem Frequenzband sehr viel Störung im Vergleich zum Nutzsignalanteil enthalten, wird dieses stark bedämpft. Frequenzbänder, die keine oder geringe Störungen enthalten, werden hingegen nur wenig verändert. Somit kann das Verhältnis von Nutzsignalanteil zu Störung im Mikrofonsignal deutlich verbessert und die Signalqualität erhöht werden. Da jedoch in jedem Band immer beide Signalanteile gleichzeitig bedämpft werden, kommt es bei einkanaligen Störgeräuschreduktionsverfahren in der Regel zu einer Beeinflussung des Nutzsignalanteils und es verbleibt eine Reststörung, die durch ihren unnatürlichen Klang trotz ihrer geringen Leistung deutlich hörbar ist. Diese Reststörung kann durch Ausnutzung von Modellen der menschlichen Hörwahrnehmung so beeinflusst werden, dass sie vom Menschen als weniger störend empfunden wird (Vary, Heute und Hess, 1998; Gustafsson, 1999; Goetze, Mildner und Kammeyer, 2006).

Im Gegensatz zu einer einkanaligen Störgeräuschreduktion können Verfahren, die das Signal mehrerer Mikrophone nutzen, Informationen über die räumliche Verteilung verschiedener Signalquellen ausnutzen. Sie orientieren sich dabei an der Verarbeitung akustischer Signale durch den Menschen. Durch seine zwei Ohren ist der Mensch in der Lage sich auch in äußerst komplexen akustischen Umgebungen mit hohen Störanteilen auf einzelne Signalquellen zu fokussieren. Auch wenn das menschliche auditive System sehr komplex und bis heute noch nicht vollständig verstanden ist, basiert diese Fähigkeit des Menschen im Wesent-

lichen auf der Ausnutzung von Laufzeit- und Lautstärkedifferenzen zwischen den zwei von den Ohren aufgenommenen Signalen. Diese Unterschiede zwischen zwei an unterschiedlichen Orten aufgenommenen Signalen lassen sich auch bei der mehrkanaligen Störgeräuschunterdrückung ausnutzen. Es ist damit einerseits möglich Quellen akustisch im Raum zu lokalisieren (Knapp und Carter, 1976; Doblinger, 2006; Rohdenburg et al., 2008), wie auch andererseits „in eine bestimmte Richtung zu hören“. Dabei werden alle Signalanteile, die nicht aus der gewählten Vorzugsrichtung der Mikrofonanordnung eintreffen, gedämpft, während eine unbeeinflusste Übertragung der Signalanteile aus der Vorzugsrichtung erfolgt (Bitzer und Simmer, 2001; Goetze, Mildner und Kammeyer, 2006). Da die beschriebenen mehrkanaligen Algorithmen den einkanaligen Algorithmen sowohl in Bezug auf erreichbare Signalqualität, als auch bei der zu erzielenden Störgeräuschreduktion überlegen sind, sollten, wo immer es die Randbedingungen wie Gehäusegröße oder Kosten möglich machen, mehrere Mikrofone zur Signalakquisition genutzt werden.

### **Kompensation akustischer Echos**

Akustische Echos werden durch die akustische Kopplung zwischen Lautsprecher und Mikrofon verursacht. Wie in Abbildung 1 dargestellt, wird das Signal des fernen Sprechers, das im Raum des nahen Sprechers über die Lautsprecher wiedergegeben wird, von den Mikrofonen des Konferenzsystems wieder aufgefangen und mit den sich insgesamt ergebenden Verzögerungen zurück zum fernen Sprecher übertragen, wenn keine Gegenmaßnahmen getroffen werden. Der ferne Sprecher hört also seine eigene Stimme, verzögert um die Laufzeit des Gesamtsystems. Dies wird von Nutzern als sehr störend empfunden und kann eine natürliche Kommunikation sogar vollkommen unmöglich machen. Auch akustische Echos, die vom Mikrofon aufgefangen werden, können als Störgeräusche interpretiert werden und somit durch die oben beschriebenen Ansätze unterdrückt werden. Allerdings bietet die Ausnutzung der Kenntnis über das zu unterdrückende Signal –

nämlich das Lautsprechersignal – die Möglichkeit einer deutlich verbesserten Reduktion der Störung *Echo* bei gleichzeitig verminderter Beeinflussung des Nutzsignals. Theoretisch ist es sogar möglich, sämtliche Echoanteile, die im Mikrofonsignal enthalten sind, herauszufiltern (Hänsler und Schmidt, 2004). Jedoch muss dazu die Raumimpulsantwort, die die Übertragung des Signals vom Lautsprecher zum Mikrofon beschreibt, exakt bestimmt werden. In der Praxis ist dies allerdings in der Regel nicht möglich, da Raumimpulsantworten im Allgemeinen sehr lang sind und sich über die Zeit verändern können, so dass man Verfahren einsetzt, die die Raumimpulsantwort ständig neu durch ein Filter zur Echo-kompensation schätzen. Ein durch nicht perfekte Schätzung der Raumimpulsantwort verbleibendes Restecho kann durch sogenannte Nachfilter weiter reduziert werden, die im Frequenzbereich nach einem ähnlichen Prinzip arbeiten wie die beschriebenen Filter zur Störgeräuschunterdrückung (Goetze, Kallinger und Kammeyer, 2005).

### **Reduktion von Hallanteilen**

Neben Störungen durch Umgebungsgeräusche und akustische Echos können auch Anteile des Nutzsignals selbst zu einer verminderten Verständlichkeit der Sprache führen. Wie aus Kirchen oder großen Bahnhofshallen bekannt ist, führt starker Hall zu einer verminderten Sprachverständlichkeit. Auch technische Systeme zur automatischen Spracherkennung oder die bereits erwähnten Algorithmen zur Richtungsschätzung bzw. Ortung von Schallquellen zeigen eine deutlich reduzierte Leistung in halligen Umgebungen. In solchen Situationen ist es sinnvoll das eigentliche Nutzsignal zusätzlich von Hallanteilen zu befreien. Mit diesem Ziel wird ähnlich wie bei den Algorithmen zur Störgeräuschreduktion die Leistung des Hallanteils in verschiedenen Bändern geschätzt und starker Nachhall bedämpft (Habets, 2007). Es sei hier angemerkt, dass durch die starke Zeitveränderlichkeit eines Sprachsignals die Enthaltung eine größere Herausforderung darstellt als die Unterdrückung von Umgebungsstörungen.

## Kompensation von Hörbeeinträchtigungen

Die zuvor beschriebenen Technologien zielen stets auf eine Verbesserung der Qualität oder Verständlichkeit von Sprache bzw. auf eine Verminderung der Höranstrengung ab. Obwohl auch Personen, die an einer Schwerhörigkeit leiden, von solchen Maßnahmen deutlich profitieren, kann die Sprachkommunikation für diese Menschen aufgrund ihrer Hörbeeinträchtigung dennoch schwierig oder unmöglich sein. Anhand eines beispielhaften Hörverlustes soll im Folgenden gezeigt werden, dass eine einfache Verstärkung von Signalen für hörgeschädigte Personen nicht zu einer Kompensation des Hörverlustes führt.

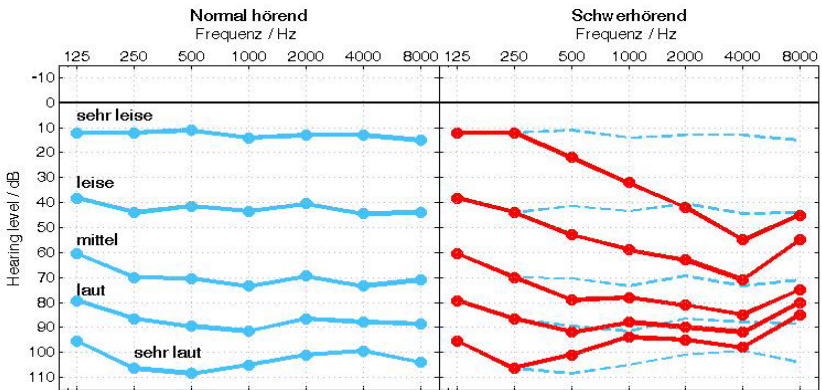


Abbildung 2: Kurven gleicher Lautheit für Normalhörende (links) und für Personen mit einer Hörbeeinträchtigung (rechts) über der Frequenz normiert auf die Hörschwelle Normalhörender.

Abbildung 2 zeigt die Kurven gleicher Lautheit eines Normalhörenden (links) und einer Personen mit einer beeinträchtigten Hörwahrnehmung beispielhaft für einen bestimmten Hörverlust (rechts). Höhere Pegel werden dabei wie in der audiologischen Praxis üblich nach unten dargestellt. Ein Pegel von 0 dB ent-



spricht dabei der sogenannten Hörschwelle Normalhörender, also dem bei der jeweiligen Frequenz gerade noch wahrnehmbaren Stimulus. Der Dynamikbereich des menschlichen Gehörs, also der Bereich zwischen gerade noch hörbaren und unangenehm lauten Tönen umfasst ungefähr 110-130 dB. Während die Lautheitskurve für *sehr leise* Töne insbesondere bei höheren Frequenzen bei der schwerhörigen Person (rechts in Abbildung 2) deutlich angehoben ist, ähneln die Pegel, bei denen Töne als *laut* empfunden werden, denen von Normalhörenden. Ein höherer Pegel wird von der schwerhörigen Person oft sogar eher als *sehr laut* empfunden, als dies bei Normalhörenden der Fall ist. Algorithmen zur Kompensation eines Hörverlustes dürfen also das Signal nicht einfach verstärken, sondern müssen das Signal an den verminderten Dynamikbereich von schwerhörenden Personen anpassen (Appell, Hohmann, Gabriel et al., 2002; Rohdenburg, Huber, van Hengel et al., 2009). Dies führt allgemein zu einer Verstärkung leiser Signalanteile, die ansonsten für die hörgeschädigte Person nicht mehr wahrnehmbar wären, während laute Signalanteile nicht verstärkt oder gar bedämpft werden. Diese bereits in Hörgeräten angewendete Dynamikkompensation kann auch in anderen Kommunikationssystemen wie bspw. Konferenzsystemen eingesetzt werden und so die Sprachverständlichkeit für schwerhörige Personen durch individuelle Signalverarbeitung verbessern.

## **KONTINUIERLICHE ÜBERPRÜFUNG DER KOMMUNIKATIONSQUALITÄT**

Ein wesentlicher Aspekt der Kommunikationsqualität ist die Verständlichkeit von Sprache, da das grundlegende Ziel die Übermittlung der darin enthaltenen Information ist. Neben der bereits diskutierten reduzierten Signalqualität besteht bei modernen Kom-

munikationssystemen wie Videokonferenzsystemen ein Problem darin, dass der Sprecher keinerlei Information darüber hat, wie gut seine Sprache für den Empfänger zu verstehen ist. Sollte der Empfänger Verständlichkeitsprobleme haben, ist die einzige Lösungsmöglichkeit, dass er den Sprecher explizit darauf hinweist. Dies führt zu störenden Unterbrechungen der Kommunikation, so dass der Empfänger bei anhaltend schlechter Verständlichkeit oftmals aufgibt und die schlechte Kommunikationsqualität akzeptiert, um nicht zu oft zu unterbrechen. Eine technische Überwachung der Verständlichkeit könnte für den Sprecher eine wichtige Unterstützung sein, wenn sie ihn automatisch auf die Verständlichkeitsprobleme des Empfängers hinweist, ohne dass dieser persönlich einschreiten muss. Der Sprecher hat dann die Möglichkeit, zum Beispiel durch günstigere Platzierung des Mikrofons die Verständlichkeit zu verbessern, ohne die Kommunikation unterbrechen zu müssen.

Die verlässlichste Methode Sprachverständlichkeit zu ermitteln ist die Durchführung von Sprachtests (siehe z.B. Wagener, Kühnel und Kollmeier, 1999; Wagener, Brand und Kollmeier, 1999a; Wagener, Brand und Kollmeier, 1999b), die allerdings für eine kontinuierliche Verständlichkeitsüberprüfung nicht geeignet sind. Stattdessen können Modelle zur Berechnung der Sprachverständlichkeit verwendet werden, die die wichtigsten Faktoren wie zum Beispiel Störgeräusche oder Nachhall berücksichtigen. Ein solches Modell ist der standardisierte Sprachübertragungsindex (Speech Transmission Index, STI, IEC1998). Auf der Grundlage des Sprachsignals, des Störgeräusches und des Nachhalls kann der STI als Index zwischen 0 (völlig unverständlich) und 1 (vollständig verständlich) berechnet werden. Dazwischen liegende Werte können in bekannte Maße für die Verständlichkeit umgerechnet werden, wie z.B. den Anteil der korrekt verstandenen Worte. Viele Experimente haben gezeigt, dass der STI den messbaren Einfluss von Störgeräuschen und Nachhall gut vorhersagen kann (siehe z.B. Houtgast und Steeneken, 1973; Steeneken und Houtgast,

1980), so dass er prinzipiell für den Einsatz in Kommunikationssystemen geeignet ist.

In der Praxis tritt bei der Verwendung von Sprachverständlichkeitsmodellen jedoch das Problem auf, dass die zur Berechnung nötigen Eingangsgrößen nicht direkt vorliegen. So beinhaltet das von einem Mikrofon aufgenommene Signal eine Mischung aus verhallter Sprache und Störgeräusch, während der STI beide Signalanteile getrennt als Eingabe benötigt. Ebenso ist die raumakustische Umgebung zunächst unbekannt, so dass der Einfluss von Nachhall nicht direkt berechnet werden kann. Zur Verwendung des STI ist es daher notwendig, diese Parameter auf der Grundlage des Mikrofonsignals zu schätzen, um anschließend die Sprachverständlichkeit berechnen zu können. Das Schaltbild eines solchen Systems ist in Abbildung 4 dargestellt.

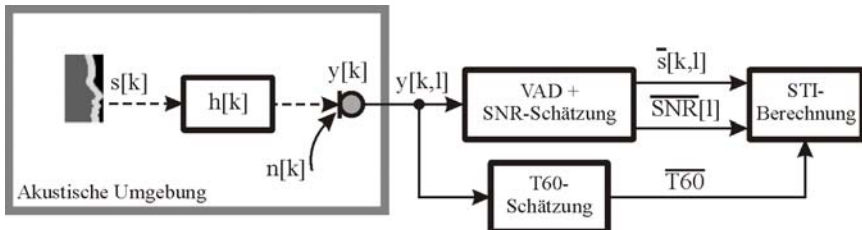


Abbildung 4: Schematische Darstellung des Verfahrens zur kontinuierlichen Berechnung des STI aus einem Mikrofonsignal, das Sprache, Rauschen und Verhallung enthält. Die zur Berechnung des STI notwendigen Parameter Signal-zu-Rauschabstand (signal-to-noise ratio, SNR) und Nachhallzeit des Raumes (T60) werden dazu blockweise geschätzt.

Das vom Mikrofon aufgenommene Signal  $y[k]$  enthält das gewünschte Sprachsignal  $s[k]$  sowie auch Störgeräusche  $n[k]$  und den Einfluss von Nachhall, der durch die Impulsantwort  $h[k]$  beschrieben werden kann. Das Mikrofonsignal wird in zeitlichen Blöcken verarbeitet. Für jeden Block wird der Sprach- und Rauschanteil geschätzt, so dass das Signal-zu-Rausch-Verhältnis (signal-to-noise ratio, SNR) berechnet werden kann. Hierzu wird ein Verfah-

ren angewendet, das berechnet, zu welchen Zeitpunkten Sprache vorhanden ist (voice activity detection, VAD). Zusätzlich wird die Nachhallzeit des Raumes  $T_{60}$  geschätzt. Aus den geschätzten Parametern kann anschließend für jeden Block der STI berechnet und visualisiert werden, der eine Aussage über die Verständlichkeit erlaubt.

In Abbildung 5 ist eine beispielhafte Anwendung des Systems dargestellt, wobei der STI alle 2 s neu berechnet wurde. Als Signale wurden Sätze des Oldenburger Satztests aneinandergereiht, so dass eine Gesamtdauer von 80 s erreicht wurde. Die Sprache wurde von einem Rauschen überlagert, das die gleichen spektralen Eigenschaften wie die verwendeten Sätze hatte. Zusätzlich wurde die Raumakustik eines typischen Büros simuliert, indem dem Signal eine Nachhallzeit von etwa 300 ms aufgeprägt wurde. Nach jeweils 20 s wurde das SNR von ursprünglich +15 dB zu 0 dB, +5 dB und wieder +15 dB verändert. Die Änderung des SNR wurde durch Absenken bzw. Anheben der Rauschenergie erreicht. Für den Zeitraum von 38,5 bis 41,5 s ist dies im oberen Teil von Abbildung 5 dargestellt. Die Energie des Rauschens sinkt nach 40 s, während die Sprachenergie konstant bleibt, der SNR erhöht sich also. Eine solche sprunghafte Änderung des SNR tritt in der Praxis bspw. auf, wenn zusätzliche Störgeräuschquellen an- bzw. abgeschaltet werden. Um den Einfluss des Nachhalls auf die geschätzte Sprachverständlichkeit zu untersuchen, wurde die Berechnung einmal mit und einmal ohne die Schätzung der Nachhallzeit durchgeführt, was im unteren Teil von Abbildung 5 durch Quadrate bzw. Kreise gekennzeichnet ist.

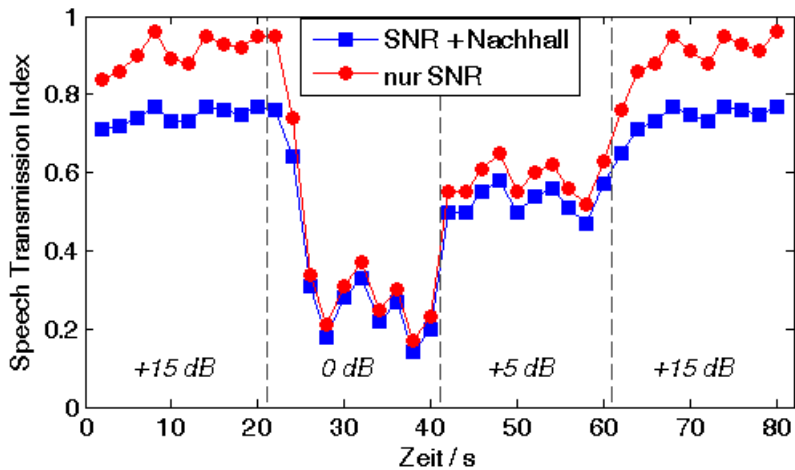
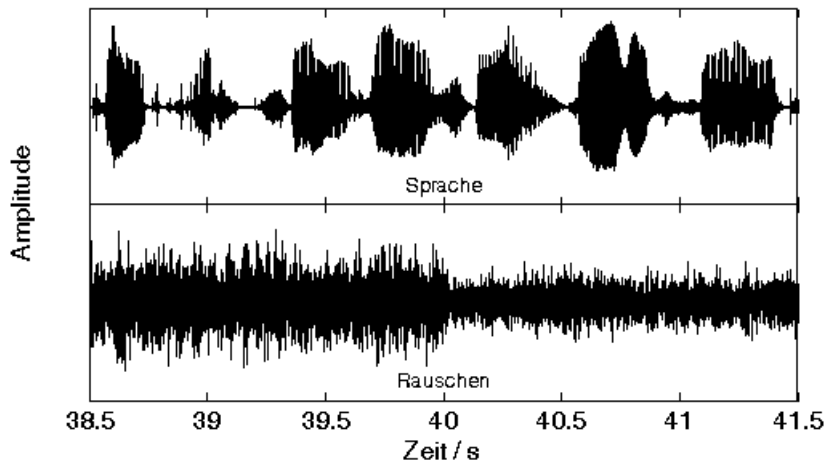


Abbildung 5: Oben: Ausschnitte des Sprach- und Rauschsignals. Nach einer Dauer von 40 s wurde die Energie des Rauschens abgesenkt und damit der SNR erhöht. Unten: Kontinuierliche Schätzung der Sprachverständlichkeit durch den Speech Transmission Index. Die Schätzung wurde auf Grundlage des SNR (Kreise) bzw. auf Grundlage von SNR und Nachhall (Quadrate) durchgeführt. Vertikale Linien kennzeichnen Zeitpunkte, an denen sich der SNR ändert (siehe unterer Bildrand).

Die Berechnungen zeigen, dass der geschätzte STI auf Änderungen des SNR reagiert und wie erwartet eine schlechtere Verständlichkeit bei schlechterem SNR voraussagt. Bei konstantem SNR in den einzelnen Segmenten sind leichte Variationen des geschätzten STI zu erkennen. Dies liegt zum einen daran, dass die Energie der Sprache zeitlich fluktuiert. Zum anderen verwenden die Verfahren zur Schätzung von Nachhallzeit und SNR statische Methoden, die intrinsisch mit Schätzungenauigkeiten verbunden sind.

Generell gilt, dass die Schätzungen umso genauer sind, je mehr Information zur Verfügung steht, d.h. in diesem Fall, je länger die Blockdauer gewählt wird. Andererseits ist eine schnelle Anpassung der Schätzung an Änderungen der akustischen Umgebung wünschenswert, wenn das System zur kontinuierlichen Überwachung benutzt werden soll. Dies erfordert eine eher kurze Blockdauer. Die hier im Beispiel gewählte Blockdauer von 2 s kann als praktikabler Kompromiss zwischen Schätzgenauigkeit und Anpassgeschwindigkeit angesehen werden. Der Vergleich der Schätzungen mit und ohne Berücksichtigung der Nachhallzeit zeigt, dass die Nachhallzeit die Sprachverständlichkeit beeinflusst. Wird sie in die Berechnungen mit einbezogen, führt eine größere Nachhallzeit zu einem geringeren STI. Insbesondere in Umgebungen mit viel Hall ist dies wichtig, da ansonsten die Sprachverständlichkeit überschätzt wird und Verständnisprobleme übersehen werden können.

## **FAZIT**

Störgeräusche, Nachhall, Echos und individuelle Hörverluste hindern Menschen daran, moderne Kommunikationssysteme wie Videokonferenzenanlagen in optimaler Weise zu nutzen. Nichtsdestotrotz wächst das Einsatzfeld komfortabler Kommunikationssysteme in modernen Arbeitsprozessen ebenso wie im privaten Umfeld ständig. Es besteht also großer Bedarf an Verbesserungen in der akustischen Kommunikation. Eine Lösungsmöglichkeit besteht

in Signalverarbeitungsstrategien, die die Audioqualität verbessern, indem sie Störgeräusche, Echos und Nachhall reduzieren. Zusätzlicher Gewinn kann erreicht werden, wenn die Kenntnis über den individuellen Hörverlust dazu eingesetzt wird, die Signalverarbeitung zu personalisieren und damit die Schallpräsentation individuell zu optimieren. Während alle genannten Verfahren zur Signalverbesserung noch Gegenstand aktueller Forschung sind, bieten die derzeit bestehenden Methoden bereits große Möglichkeiten zur Verbesserung von Kommunikationssystemen. Gleiches gilt für den Einsatz von Modellen zur Vorhersage von Sprachverständlichkeit, die verwendet werden können, um die Kommunikationsqualität zu überwachen. Solche Systeme ermöglichen ein direktes Feedback an den Sprecher, wie gut die eigene Sprache am Ort des Empfängers verstanden werden kann. Der Sprecher kann so auf schlechte Verständlichkeit reagieren (z.B. durch bessere Mikrofonpositionierung), ohne dass die Kommunikation unterbrochen werden muss.

Als Alternative oder idealerweise als Ergänzung zu raumakustischen Maßnahmen können die Verfahren die Kommunikationsqualität verbessern. Generell gilt dabei, dass das Verbesserungspotential umso größer ist, je genauer die Eigenschaften der Raumakustik und/oder des Hörverlustes bekannt sind.

Abbildung 6 zeigt ein System, das die oben beschriebenen Signalverbesserungsstrategien in einer Anwendung für Videotelefonie beispielhaft kombiniert. Es können Text-, Video- und Sprachsignale übertragen werden. Störgeräusche und akustische Echos werden unterdrückt und ein individueller Hörverlust kann kompensiert werden. Weiterhin kann die Sprachverständlichkeit im Empfangsraum geschätzt und angezeigt werden, um für den Fall einer zu geringen Sprachverständlichkeit beispielsweise das Mikrofon besser zu platzieren und somit die Sprachverständlichkeit zu erhöhen.

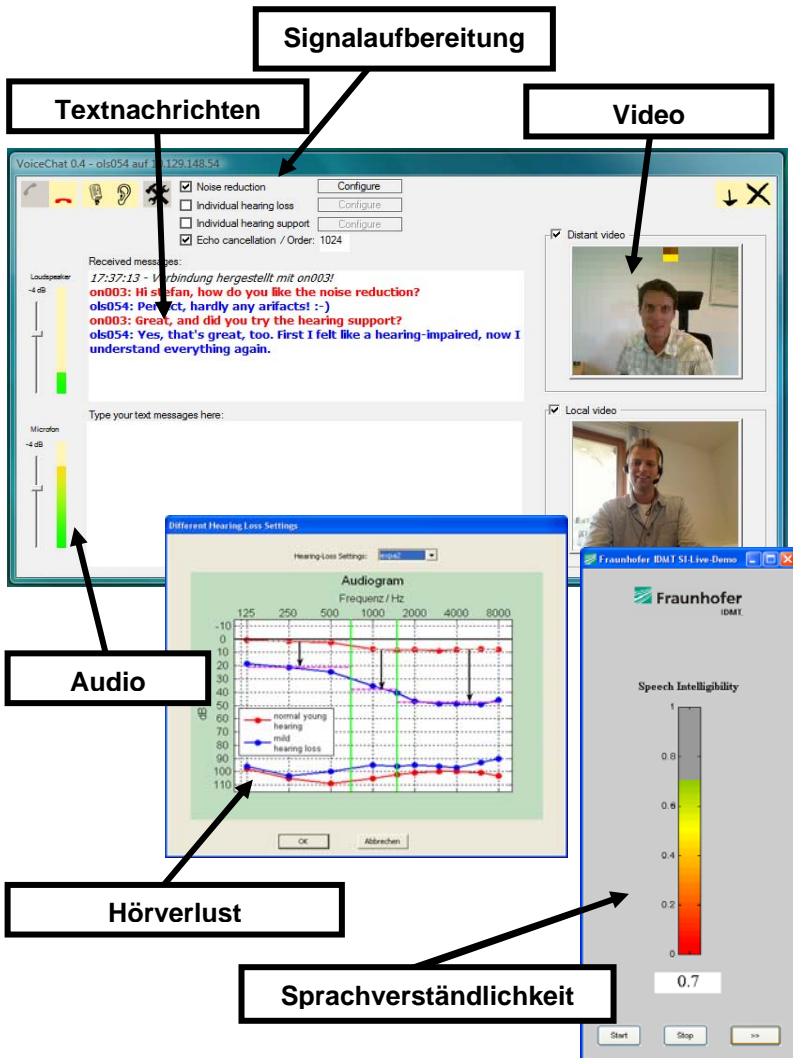


Abbildung 6: Das entwickelte Kommunikationssystem beinhaltet Signalaufbereitung, personalisierte Signalverstärkung sowie eine automatische Überwachung der Sprachverständlichkeit.



## LITERATUR

- Appell, J.-E., Hohmann, V., Gabriel, B. & Kollmeier, B. (2002). Evaluation of different 3-channel dynamic compression schemes in a field-test with a wearable DSP prototype hearing aid. *Zeitschrift für Audiologie / Audiological Acoustics* 41(1), pp. 22-46.
- Bitzer, J. & Simmer, K.U. (2001). Superdirective microphone arrays. In: M. S. Brandstein & D. Ward (Eds.), *Microphone Arrays: Signal Processing Techniques and Applications*. Berlin: Springer; pp. 19-38.
- Davis, A. (2003). Population study of the ability to benefit from amplification and the provision of a hearing aid in 55-74 year old first time hearing aid users. *International Journal of Audiology* 42(2), pp. S39-S52.
- Doblinger, G. (2006). Localization and Tracking of Acoustical Sources. In: *Topics in Acoustic Echo and Noise Control*. Berlin: Springer, pp. 91-122.
- Goetze, S., Kallinger, M. und Kammeyer, K.-D. (2005). Residual echo power spectral density estimation based on an optimal smoothed misalignment for acoustic echo cancelation. In: *International Workshop on Acoustic Echo and Noise Control (IWAENC-2005)*, Eindhoven, Niederlande.
- Goetze, S., Mildner, V. und Kammeyer, K.-D. (2006). A psychoacoustic noise reduction approach for stereo hands-free systems. In: *Audio Engineering Society (AES), 120. Convention, Paris, Frankreich*
- Gustafsson, S. (1999). *Enhancement of Audio Signals by Combined Acoustic Echo Cancellation and Noise Reduction*. Aachener Beiträge zu digitalen Nachrichtensystemen, Band 11. Mainz: Wissenschaftsverlag. Zugleich: Dissertation, RWTH Aachen.
- Habets, E.A.P. (2007). *Single and Multi-Microphone Speech Dereverberation using Spectral Enhancement*. Dissertation, Universität Eindhoven, Niederlande
- Hänsler, E., & Schmidt, G. (2004). *Acoustic echo and noise control: a practical approach*. Hoboken: Wiley.
- Houtgast, T., & Steeneken, H.J.M. (1973). The modulation transfer function in room acoustics as a predictor of speech intelligibility, *Acustica* 28, pp. 66-73.
- IEC1998 (1998). *Sound System Equipment - Part 16: Objective rating of speech intelligibility by speech transmission index*.

- Knapp, C.H., & Carter, G.C. (1976) The Generalized Correlation Method for Estimation of Time Delay. *IEEE Transactions on Acoustics, Speech and Signal Processing* 24(4), pp. 320–327.
- Rohdenburg, T., Huber, R., van Hengel, P., Bitzer, J., & Appell, J.-E. (2009). Hearing aid technology and multi-media devices. In: ITG Fachtagung für Elektronische Medien. 13. Dortmunder Fernsehseminar.
- Rohdenburg, T., Goetze, S., Hohmann, V., Kollmeier, B. & Kammeyer, K.-D. (2008). Combined Source Tracking and Noise Reduction for Application in Hearing Aids. In: 8. ITG-Fachtagung Sprachkommunikation, Aachen.
- Shield, B. (2006). Evaluation of the social and economic costs of hearing impairment. Technical report, Hear-It.
- Steeneken, H.J.M. & Houtgast, T. (1980). A physical method for measuring speech-transmission quality. *Journal of the Acoustical Society of America* 67(1), pp. 318-326.
- Vary, P., Heute, U., & Hess, W. (1998) *Digitale Sprachsignalverarbeitung*. Stuttgart: Teubner-Verlag.
- Wagener, K., Kühnel, V., & Kollmeier, B. (1990). Entwicklung und Evaluation eines Satztests für die deutsche Sprache I: Design des Oldenburger Satztests. *Zeitschrift für Audiologie / Audiological Acoustics* 38(1), pp. 4-15.
- Wagener, K., Brand, T., & Kollmeier, B. (1999a). Entwicklung und Evaluation eines Satztests für die deutsche Sprache II: Optimierung des Oldenburger Satztests. *Zeitschrift für Audiologie / Audiological Acoustics* 38(2), pp. 44-56.
- Wagener, K., Brand, T. & Kollmeier, B. (1999b). Entwicklung und Evaluation eines Satztests für die deutsche Sprache III: Evaluation des Oldenburger Satztests. *Zeitschrift für Audiologie / Audiological Acoustics* 38(3), pp. 86-95.
- Zwicker, E., & Fastl, H. (1990). *Psychoacoustics. Facts and models*. Berlin: Springer.