

Innovative Hörunterstützung in Kommunikationssystemen

*Jan Rennies, Stefan Goetze, Thomas Rohdenburg &
Jens-E. Appell*

Fraunhofer Institut für Digitale Medientechnologie
Projektgruppe Hör-, Sprach- und Audiotechnologie
Marie-Curie-Str. 2, 26129 Oldenburg

Kontakt: jan.rennies@idmt.fraunhofer.de

ZUSAMMENFASSUNG

Trotz des technischen Fortschritts ist akustische Qualität in modernen Kommunikationssystemen in vielen Situationen nicht optimal. Sowohl äußere Faktoren, wie z.B. Hintergrundlärm oder schlechte Raumakustik, als auch persönliche Faktoren, wie z.B. individuelle Hörstörungen, können die akustische Kommunikation erheblich einschränken. Vor dem Hintergrund einer im Mittel älter werdenden und daher häufiger schwerhörenden Arbeitnehmerschaft und der zunehmenden Bedeutung von akustischer Kommunikation am Büroarbeitsplatz sind innovative Lösungen erforderlich. Diese reichen von Strategien zur Reduktion störender Signalanteile über den Einsatz von Gehörmodellen zur Überwachung der Kommunikationsqualität bis zur Kompensation individueller Hörstörungen.

ABSTRACT

Despite technological advances acoustic quality in modern communication systems is still not perfect in many situations. Both external factors, e.g. background noise or poor acoustics as well as personal factors such as individual hearing loss can considerably impede acoustic communication. In the light of the demographic change which on average leads to older and therefore to an increased number of hearing impaired employees, and the increasing importance of acoustic communication at office working places, innovative solutions are required. They range from strategies for reducing acoustic disturbances and the use of hearing models to monitor communication quality to compensation techniques for individual hearing impairments.

EINLEITUNG

Die Qualität von akustischer Kommunikation macht einen wesentlichen Anteil der ergonomischen Eigenschaften eines modernen Arbeitsplatzes aus. Dies betrifft einerseits die direkte zwischenmenschliche Sprachkommunikation und die Kommunikation mithilfe von Kommunikationssystemen und andererseits auch die Wahrnehmung und Lokalisierung von Alarmsignalen und die akustische Orientierung in der Arbeitsumgebung. Moderne Strukturen der Arbeitsorganisation beruhen immer mehr auf einem intensiven akustischen Informationsaustausch, der zunehmend auch in Produktionsbereichen mit beachtlichem Umgebungslärm vordringlich wird. Auch im Büroumfeld sind wesentliche Strukturen direkt an akustische Kommunikation geknüpft, was durch die wachsende Verbreitung videogestützter Konferenz- oder Chatsysteme weiter verstärkt wird.

Den herausragenden Möglichkeiten der modernen Kommunikationsmittel stehen dabei Risiken gegenüber, die entstehen, wenn die Bedingungen für eine qualitativ hochwertige akustische Kommunikation nicht gegeben sind. Ist dies der Fall, sind häufig erhöhte Höranstrengung der Mitarbeiter und damit reduzierte Konzentrations- und Leistungsfähigkeit die Folge. In gravierenden Fällen können fehlende oder falsch übermittelte Informationen erhebliche Störungen des Arbeitsablaufes verursachen. Eine optimale akustische Kommunikation ist daher eine Voraussetzung für reibungslose und effiziente Abläufe an modernen Arbeitsplätzen.

Faktoren, die die Kommunikationsqualität einschränken können, sind einerseits physikalischer (äußerer) Natur. Dies umfasst störende Nebengeräusche (z.B. Maschinen, Geräte, Gespräche der Mitarbeiter), deren Lautstärke durch schlechte raumakustische Eigenschaften noch verstärkt werden kann, sowie nicht-optimale Systemeigenschaften von Kommunikationsgeräten (z.B. Band- und Dynamikbegrenzung, Störeinflüsse durch Aufnahme, Kodierung und Übertragung). Andererseits können

persönliche (innere) Faktoren eine barrierefreie akustische Kommunikation erschweren, bspw. wenn eine Hörstörung oder eine erhöhte Lärmempfindlichkeit vorliegt. Dadurch kann die Arbeitsfähigkeit derart eingeschränkt sein, dass hörgeschädigte Mitarbeiter gänzlich von der Nutzung der im Betrieb etablierten (Kommunikations-)Systeme ausgeschlossen werden, da Sie diese nur in ruhiger Umgebung nutzen könnten, der Umgebungslärm z.B. in einem Großraumbüro sie jedoch daran hindert. Aktuelle Studien belegen, dass Kommunikationseinschränkungen durch eine Hörstörung sowohl schwerwiegende individuelle Folgen für die betroffenen Mitarbeiter als auch erhebliche betriebs- und volkswirtschaftliche Konsequenzen haben (Bormann, Sust, Heinecke-Schmitt et al., 2005; Shield, 2006). Da etwa 16% der erwachsenen Europäer (d.h. ca. 70 Mio. Menschen) an einem behandlungsbedürftigen Hörverlust leiden und nur 20% dieser Personen ein Hörgerät nutzen, ist die Problematik bereits heute hoch relevant und wird sich vor dem Hintergrund des demografischen Wandels weiter verstärken, da Schwerhörigkeit überwiegend altersbedingt ist (Davis, 2003; Shield, 2006). Dieser Beitrag beschreibt Technologien und aktuelle Forschungsarbeiten, die eine Verbesserung der akustischen Kommunikation zum Ziel haben, welche durch äußere und innere Faktoren eingeschränkt ist. Die Technologien umfassen dabei die Reduktion ungewünschter Signalanteile, die Kompensation von Hörverlusten und den Einsatz von Hörmodellen zur optimalen Planung und Qualitätsüberwachung von Arbeitsplätzen.

REDUKTION UNGEWÜNSCHTER SIGNALANTEILE

Moderne Kommunikationssysteme (z.B. Telefone, Konferenzsysteme, Audiochatsysteme, PA-Systeme) haben das gemeinsame Prinzip, dass das akustische Signal durch ein Mikrofon aufgenommen wird. Anschließend wird es nach mehr oder weniger aufwändiger Verarbeitung (z.B. Kodierung, Langzeitpegelanpassung) an den Empfänger gesendet, dort

verarbeitet (z.B. dekodiert) und schließlich akustisch wiedergegeben. In der Praxis werden diese Kommunikationssysteme häufig in nicht-optimalen akustischen Umgebungen eingesetzt, die Störgeräusche, Raumhall oder Verzerrungen hinzufügen. Die entstehende Problematik ist in Abbildung 1a) schematisch dargestellt: Die Störeinflüsse wie z.B. Rauschen (R) werden zusammen mit dem Sprachsignal (S) von Mikrofon aufgenommen und an den Empfänger übertragen, so dass bei diesem eine reduzierte Signalqualität vorliegt.

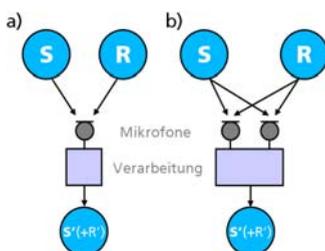


Abbildung 1: Schematische Darstellung von a) einkanaliger und b) mehrkanaliger Störgeräuschbefeuerung.

Um die dadurch entstehenden Qualitätseinbußen zu verringern, zielen Kommunikationssysteme darauf ab die störenden Signalanteile zu reduzieren. Verfügt das System über nur ein Mikrofon, besteht der Ansatz darin, solche Frequenzanteile des gemischten Signals abzuschwächen, die überwiegend Störungen enthalten, und Frequenzanteile mit starken Sprachsignalanteilen nicht abzuschwächen. Dadurch wird insgesamt die Energie der Störung reduziert. Idealerweise ist das verarbeitete Signal (S') gleich dem gewünschten Sprachsignal am Eingang (S). In der Praxis ist eine perfekte Trennung nicht möglich, da in den abgeschwächten Frequenzanteilen auch immer ein Anteil gewünschter Sprache und in den nicht abgeschwächten Anteilen immer ein Anteil der Störung enthalten ist. Dadurch liegt am Ausgang noch eine Reststörung (R') vor und das Sprachsignal (S') stimmt qualitativ nicht mit dem Originalsignal überein. In

besonders störrauschbehafteten Umgebungen ist die Trennung von Sprache und Störungen auch für moderne Systeme sehr schwierig. Typischerweise treten dann derartige Verzerrungen im Sprachsignal auf, dass trotz der Reduktion der Störrausche keine Vorteile bzgl. der Sprachverständlichkeit erreicht werden können. Da diese Verzerrungen nicht vollständig vermeidbar sind, zielen aktuelle Forschungsansätze für Systeme mit einem Mikrofon darauf ab, die Verzerrungen so zu gestalten, dass sie so wenig wie möglich die Hörwahrnehmung beeinträchtigen (Vary, Heute und Hess, 1998; Gustafsson, 1999; Rohdenburg, Hohmann und Kollmeier, 2005; Goetze, Mildner und Kammeyer, 2006).

Ein großer Vorteil ergibt sich, wenn Systeme über mehr als ein Mikrofon verfügen (Abbildung 1b). Räumlich getrennte Sprach- und Störrauschquellen kommen in unterschiedlichen Zeitabständen an den Mikrofonen an. Intelligente Signalverarbeitungsstrategien können dies ausnutzen, um eine bessere Trennung zwischen gewünschten und ungewünschten Signalanteilen zu erreichen, indem räumliche Richtungen, aus denen vornehmlich Störrausche kommen, abgeschwächt werden (Bitzer und Simmer, 2001; Goetze, Mildner und Kammeyer, 2006; Rohdenburg, Goetze, Hohmann et al., 2008).

Die Einsatzmöglichkeiten solcher Technologien im Büroumfeld reichen dabei von der Verwendung von Telefonen und Diktiersystemen in lärmbehafteter Umgebung bis hin zu Videokonferenzsystemen. Insbesondere die Verfahren mit mehrkanaligen Signalverbesserungsstrategien sind hier jedoch noch Gegenstand aktueller Forschung und haben bisher nur vereinzelt Anwendung in marktfähigen Geräten gefunden.

KOMPENSATION VON HÖRVERLUSTEN

Etwa 16% der Bevölkerung in Europa ist von einer Hörstörung betroffen. Obwohl Schwerhörigkeit mit dem Alter zunimmt, geht man davon aus, dass etwa die Hälfte der Schwerhörigen noch im

berufsfähigen Alter ist¹. Jeder Zweite über 65 Jahren gibt an Probleme beim Hören und bei der Kommunikation zu haben. Abbildung 2 verdeutlicht die Verteilung der unterschiedlichen Abstufungen von Schwerhörigkeit in Deutschland. Wie schematisch angedeutet leidet der überwiegende Teil der Betroffenen an leichter Schwerhörigkeit, die sich vor allem dadurch auszeichnet, dass Kommunikationsprobleme auftreten, wenn die äußeren akustischen Bedingungen nicht optimal sind. In dieser Gruppe ist nur ein geringer Anteil mit Hörgeräten versorgt. Zwar steigt die Versorgungsrate mit dem Grad der Schwerhörigkeit, aber erst bei schweren Hörverlusten tragen mehr als die Hälfte der Betroffenen Hörgeräte. Die Gründe hierfür sind neben den hohen Anschaffungskosten, nicht zufrieden stellender Anpassung und Stigmatisierung auch Probleme im Berufsalltag, die durch die Inkompatibilität zwischen Hörgeräten und im Betrieb etablierten Kommunikationssystemen (z.B. Telefon, Headsets, Kopfhörer, Gehörschutz) bedingt sind und die das Tragen von Hörgeräten prinzipiell nicht ermöglichen.

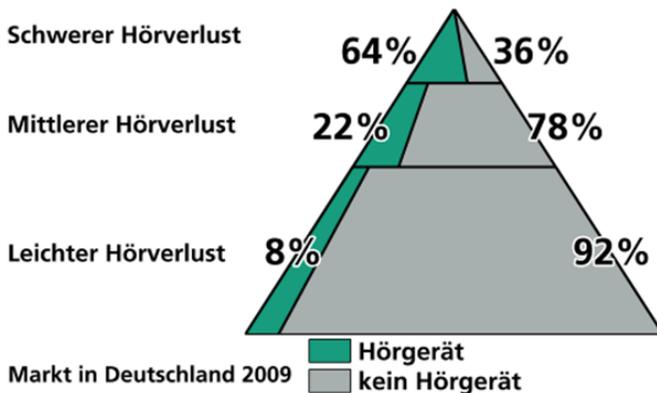


Abbildung 2: Hörgeräteversorgungsrate für leichte, mittlere und schwere Hörverluste in Deutschland.

¹ <http://www.german.medical.hear-it.org/Ursachen-von-Horverlust>

Hörverlustkompensation in Kommunikationssystemen

Die geringe Versorgungsrate und die Inkompatibilität von Hörgeräten und Kommunikationssystemen erfordern innovative Ansätze, wie betroffenen Personen die Kommunikation am Arbeitsplatz erleichtert oder sogar erst wieder zugänglich gemacht werden kann. Die Algorithmen, die eine individuelle Kompensation von Hörverlusten ermöglichen, sind dabei durch die Entwicklung in Hörgeräten gut erforscht und bieten Hörgeräteträgern in vielen Situationen eine gute Unterstützung. Das Prinzip besteht darin, frequenzabhängig leise Signalanteile zu verstärken, die aufgrund der erhöhten Hörschwelle nicht hörbar wären, deutlich hörbare Signalanteile der gleichen Frequenz werden nicht weiter verstärkt, um eine unangenehm hohe Lautstärke zu vermeiden (Dynamikkompression). Die optimalen Parameter der Dynamikkompression sind dabei individuell sehr unterschiedlich. Dies liegt unter anderem daran, dass einerseits die Form der Hörschwelle meist frequenzabhängig und individuell stark unterschiedlich ist, und andererseits auch überschwellige Wahrnehmungsstörungen in sehr unterschiedlichen Ausprägungen auftreten können (z.B. erhöhte Lärmempfindlichkeit oder schlechtere räumliches und zeitliches Auflösungsvermögen). Daher ist die Anpassung von Hörgeräten in der Regel ein sehr individueller und langwieriger Prozess, bei dem in mehreren Anpasssitzungen und Gewöhnungsphasen die Parameter durch einen professionellen Hörgeräteakustiker immer weiter optimiert werden.

Sollen nun ähnliche Signalaufbereitungsverfahren in alltäglichen Kommunikationssystemen für den Arbeitsalltag eingesetzt werden, besteht die technische Herausforderung nicht in der Entwicklung der Algorithmen selbst, sondern vielmehr darin, wie die Parameter individuell eingestellt werden können, da eine professionelle Anpassung nicht praktikabel ist. Aktuelle Forschungsarbeiten untersuchen daher verschiedene Verfahren, die eine schnelle, selbständige und dennoch akkurate Anpassung erlauben und somit den Einsatz von Hörunterstützung in

Kommunikationssystemen erst ermöglichen (z.B. Irandoost, 2011).

Im Folgenden werden Ergebnisse einer Studie beschrieben, in denen hörunterstützende Technologie in Festnetztelefonen implementiert und evaluiert wurde. Die Anpassung für die einzelnen Probanden wurde durch eine selbständige Auswahl aus einem Satz typischer Einstellungen (Presets) von Hörverlustkompensation bei altersbedingter Schwerhörigkeit vorgenommen. Altersbedingte Schwerhörigkeit zeichnet sich in den meisten Fällen durch eine mit der Frequenz ansteigende Hörschwelle aus, d.h., hohe Frequenzen werden schlechter wahrgenommen, während die Wahrnehmung tiefer Frequenzen weitestgehend normal bleibt. Die Presets wurden daher so gestaltet, dass sie zehn verschiedene Grade an Verstärkung von höheren Frequenzen abdeckten. Die Aufgabe der Probanden bestand darin, einem Gespräch über das Telefon zu folgen und dabei das Preset auszuwählen, das für sie individuell am angenehmsten war.

Subject (Gender)	Ear	PTA / dB	Mean Preset	Beneficial or not
1(male)	left	43	1	no
	right	53	4	yes
2(male)	left	43	4	yes
	right	41	2	yes
3(female)	left	23	3	yes
	right	18	3	yes
4(male)	left	55	5	yes
	right	56	5	yes
5(female)	left	38	1	no
	right	30	1	no
6(male)	left	22	3	no
	right	25	1	no
7(male)	left	37	6	yes
	right	33	6	yes
8(female)	left	31	1	no
	right	38	1	no

Tabelle 1: Evaluationsergebnisse der Hörunterstützung im Telefon für alle Probanden.

Die Auswahl erfolgte über das Zifferblatt (0-9) des Telefons. Von den acht teilnehmenden Probanden (25-76 Jahre) waren sechs altersschwerhörig und zwei zeigten eine Hörstörung, die gerade im Sprachbereich ausgeprägt war. Alle Probanden besaßen ein Hörgerät, das während des Auswahlprozesses nicht getragen wurde. In einem Vergleichstest wurde dann das Telefon mit integrierter Hörunterstützung mit dem individuell angepassten Hörgerät in der Telefonsituation verglichen. Die Tests fanden in ruhiger Umgebung statt, um den Einfluss äußerer Störgeräusche zu verhindern. Beide Ohren aller Probanden wurden getrennt getestet. Die Ergebnisse der Messreihe sind in Tabelle 1 dargestellt. Die Hörverluste der Probanden sind als pure-tone average (PTA) angegeben, welcher den mittleren Hörverlust bei 0,5, 1 und 2 kHz widerspiegelt. Die vierte Spalte gibt das im Mittel eingestellte Preset wieder, wobei 1 für das Preset steht, das keine Hochtonverstärkung enthält (normale Telefonübertragung) und die Hochtonverstärkung mit steigender Presetzahl ansteigt. Es ist erkennbar, dass in 10 von 16 Fällen ein von 1 verschiedenes Preset bevorzugt wurde, d.h., dass die Hörunterstützung im Telefon gegenüber der normalen Telefonübertragung ohne Verarbeitung bevorzugt wurde. Im Allgemeinen stellten sich Probanden mit einem stärkeren Hörverlust eine größere Verstärkung (ein höheres Preset) ein. Die letzte Spalte in Tabelle 1 gibt die Ergebnisse des zweiten Tests wieder, in dem die Probanden angaben, ob die integrierte Hörunterstützung ggü. dem eigenen Hörgerät bevorzugt wurde. Dabei gaben die Probanden in 9 der 16 Fälle an, dass die integrierte Hörunterstützung gegenüber dem eigenen Hörgerät bevorzugt wurde. Dieses Ergebnis ist besonders interessant, da es zeigt, dass durch die Auswahl aus einem sehr begrenzten Satz an Presets bereits eine gute Unterstützung erreicht werden kann, die in Telefonsituationen der Unterstützung durch individuell angepasste Hörgeräte gleichwertig oder sogar überlegen ist. Dieses Ergebnis ist auch dadurch begründet, dass einige der Probanden nach eigener Angabe in der Regel ohne Hörgerät

telefonieren, da sie Schwierigkeiten bei der Ankopplung von Telefonhörer und Hörgerät haben. Da diese Inkompatibilität wie oben erwähnt ein Grund für das Nichtverwenden von Hörgeräten am Büroarbeitsplatz ist, deuten die Messergebnisse an, dass ein großes Potenzial in hörunterstützenden Technologien liegt, die direkt in Kommunikationssysteme integriert werden, und dass eine Anpassung an individuelle Hörverluste in bestimmten Anwendungsfällen bereits durch einfache auf den Anwendungsfall optimierte Presets erfolgen kann.

VORHERSAGE VON KOMMUNIKATIONSQUALITÄT

Im Bereich der akustischen Planung von Gebäuden, Einrichtungen und Kommunikationsumgebungen wird vor allem auf präzise messbare physikalische Größen zurückgegriffen, wie z.B. den Schallpegel oder die Nachhallzeit. Diese Messgrößen geben die Hörwahrnehmung des Menschen jedoch nur begrenzt wieder und führen ggf. zu falschen Planungsgrundlagen. So ist es bspw. möglich, dass verschiedene Geräusche mit demselben physikalischen Pegel völlig andere empfundene Lautstärken hervorrufen. Auch die Verständlichkeit von Sprache hängt von vielen Parametern ab, deren Zusammenwirken bei der Sprachwahrnehmung nicht direkt physikalisch messbar ist. Aus diesem Grund werden vermehrt auch auf Hörmodellen basierte Messverfahren in der Planung von Raumakustik eingesetzt. Ein weit verbreitetes Beispiel ist der Sprachübertragungsindex (Speech Transmission Index, STI, siehe Steeneken und Houtgast, 1980; IEC, 2003). Er basiert auf einer Messmethodik, in der festgestellt wird, wie gut zeitliche Fluktuationen, die wichtige Informationsträger von Sprache sind, in einem Übertragungssystem erhalten bleiben, z.B. bei Wiedergabe von Sprache durch einen Büroraum. Dabei wird ein Messsignal abgespielt und an der Stelle des Empfängers aufgezeichnet. Der Vergleich zwischen dem Messsignal und dem aufgezeichneten Signal gibt dann Aufschluss über die Energieverhältnisse von gewünschter

Sprache und Störgeräuschen sowie der Nachhallzeit, so dass die zu erwartende Sprachverständlichkeit abgeschätzt werden kann. Dabei werden wichtige Eigenschaften des menschlichen Gehörs, wie etwa die frequenzabhängige Hörschwelle, die gegenseitige Maskierung benachbarter Frequenzen, die Maskierung von Sprache durch Störgeräusche und der Einfluss der Verzerrung von Sprache durch Raumhall, berücksichtigt. Die Präzision der Vorhersagen des STI und seine Überlegenheit ggü. einfachen physikalischen Messgrößen sind also vor allem darin begründet, dass die wesentlichsten Eigenschaften des Gehörs in der Berechnung berücksichtigt werden. Trotz dieser Vorteile ergeben sich Grenzen in der praktischen Anwendbarkeit des STI. Zum einen ist zur Ermittlung des STI immer das Abspielen und Aufnehmen eines Messsignals notwendig, so dass eine regelmäßige Überprüfung der Raumakustik nur mit hohem Aufwand und mit einer Unterbrechung des Arbeitsalltags möglich ist. Zum anderen sind wichtige Eigenschaften des auditorischen Systems, wie etwa das binaurale (zweiohrige) Gehör des Menschen, nicht berücksichtigt, woraus sich teilweise erhebliche Falschaussagen des STI ergeben. Beide Formen der Limitation des STI sind aktuelle Forschungsfelder, aus denen einige Ergebnisse im Folgenden dargestellt werden.

Referenzlose Abschätzung des STI

Das oben geschilderte Prinzip des STI erfordert die genaue Kenntnis des Testsignals und des aufgenommenen Signals, da nur dann eine genaue Aussage über die Verhältnisse von Sprach- und Störenergie als Grundlage für die weitere Berechnung vorliegt. Das Testsignal dient also als Referenz, deren Störung durch das aufgenommene Signal beschrieben werden kann. Die Tatsache, dass eine Referenz bekannt sein muss, verhindert prinzipiell, dass der klassische STI für eine regelmäßige oder gar kontinuierliche Überwachung der Kommunikationsqualität zum Einsatz kommt. In der Praxis ist das gewünschte Sprachsignal,

das durch ein Übertragungssystem gestört wird nicht bekannt, sondern es liegt lediglich das gestörte Signal am Eingang eines Mikrofons vor (vgl. Abschnitt zu Störgeräuschbefreiung).

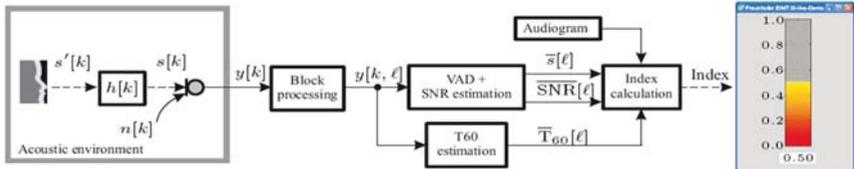


Abbildung 3: Schematische Darstellung des Systems zur referenzlosen STI-Schätzung. Die einzelnen Blöcke symbolisieren die Verarbeitungsschritte vom Mikrofonsignal (Kreis links) bis zum Index (rechts).

Zur Lösung dieses Problems wurde ein System entwickelt, das auf den Berechnungsvorschriften des STI beruht und ebenfalls die zwei wesentlichen Eingangsparameter Signal-Rausch-Abstand (Energieverhältnis zwischen Sprache und Rauschen, engl. signal-to-noise ratio [SNR]) und Nachhallzeit (T60) verwendet. Das System berechnet diese Größen jedoch nicht auf Grundlage einer bekannten Referenz, sondern schätzt diese aus einem mit einem Mikrophon aufgenommenen Signal ab. Die schematische Darstellung in Abbildung 3 zeigt, dass zunächst ein ähnliches Problem besteht wie in der einkanaligen Störgeräuschbefreiung, da aus einem gemischten Signal abgeschätzt werden muss, ob Sprache anwesend ist (engl. voice-activity detection [VAD]) und wenn ja, welche Frequenzanteile Sprache und welche Störgeräusche sind. Aus den geschätzten Parametern SNR und T60 wird schließlich der STI berechnet. Dabei kann optional auch ein Audiogramm berücksichtigt werden, falls dieses als zusätzliche Information vorliegt, um dadurch auch die erhöhte Hörschwelle bei Schwerhörigkeit zu berücksichtigen (Rennies, Albertin, Goetze et al., 2010). Da die Abschätzung in kurzen Zeitblöcken erfolgt, ist eine Abschätzung des STI und damit eine Überwachung der Kommunikationsqualität quasi in Echtzeit möglich, so dass sich ändernde akustische Bedingungen und

eventuelle Einbußen der Sprachqualität detektiert und quantifiziert werden können².

Abbildung 4 zeigt eine exemplarische Antwort des Systems, wenn 1 Minute laufender Sprache bei konstantem Pegel durch ein Störgeräusch überlagert wird, dessen Pegel über die Zeit ändert. Aufgetragen ist der geschätzte Index über die Zeit, der alle 2 s neu berechnet wurde. Das Störgeräusch hatte zunächst einen relativ geringen Pegel, so dass der Sprachpegel 15 dB oberhalb des Rauschens lag (d.h., der SNR war 15 dB). Nach 20 s wurde der Pegel des Rauschens um 5 dB erhöht (SNR = 10 dB) und nach weiteren 20 s um weitere 10 dB (SNR = 0 dB). Nach jeweils weiteren 20 s wurde der Störgeräuschpegel in zwei Schritten wieder auf seinen ursprünglichen Wert verringert. Es ist erkennbar, dass die Antworten des Systems den äußeren Änderungen der akustischen Umgebung (dem Pegel des Rauschens) gut folgen. In Phasen von konstantem Störgeräusch ist die Antwort des Systems relativ konstant, abgesehen von kleineren Abweichungen, welche durch die zeitlichen Fluktuationen des Sprachsignals (und damit der Schätzwerte) begründet sind. Der Störfaktor Rauschen kann vom System also prinzipiell bewertet werden.

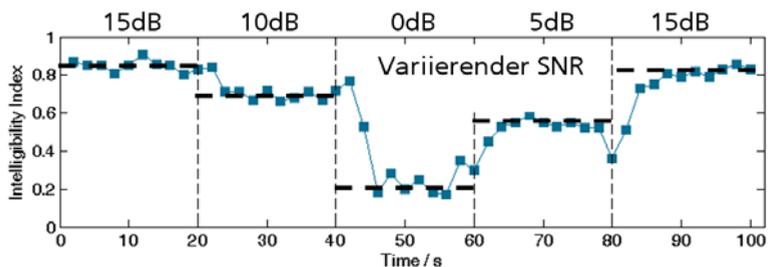


Abbildung 4: Antwort des referenzlosen STI-Schätzers auf laufende Sprache, welche durch ein Störgeräusch überlagert ist, dessen Pegel sich alle 20 s ändert (gestrichelte Linien).

²http://www.idmt.fraunhofer.de/eng/press_media/download/product_information/09_si_live_demo_en.pdf

Um die Anwendbarkeit des Systems weiter zu validieren, wurden in einer weiteren Testreihe (Scheper, 2011; Scheper, Rennies, Holube et al., 2011) die für jeden Zeitblock geschätzten STI-Werte mit denen verglichen, die mit den *berechneten* anstatt den *geschätzten* Werten ermittelt wurden. Um allgemeinere Aussagen treffen zu können, wurden verschiedene Arten von Störgeräuschen und verschiedene Arten der zeitlichen Pegeländerungen der Störgeräusche getestet. Zusätzlich wurden verschiedene SNR verwendet und jede Bedingung mehrmals wiederholt, um die Reproduzierbarkeit der Werte sicher zu stellen. Wie zuvor diente laufende Sprache, die von den verschiedenen Störungen überlagert wurde, als Testsignal.

Die Gegenüberstellung der geschätzten und berechneten (theoretischen) STI-Werte (in Prozent) ist in Abbildung 5a) dargestellt. Die durchgezogene Linie symbolisiert die perfekte Abbildung, d.h. bei optimaler Übereinstimmung sollten alle Messwerte auf dieser Linie liegen. Die gestrichelte Linie gibt die beste Anpassung einer Gerade mit Steigung 1 an die Daten wieder. Wie zu erkennen ist sind die Abweichungen zur idealen Abbildung sehr gering (Abweichung [Bias] 2.6%) und alle Messpunkte liegen in ihrer unmittelbaren Umgebung. Dies drückt sich auch in der hohen Korrelation aus ($R^2 = 0,943$). Die blockweise Abschätzung des STI für Sprachen in verschiedenen Arten von Hintergrundlärm kann also mit hoher Präzision erfolgen. Der letzte Teil der Studie (Scheper, 2011; Scheper, Rennies, Holube et al., 2011) beschäftigte sich mit der Frage, inwieweit das System zur Schätzung des STI in der Lage ist subjektiv wahrgenommene Höranstrengung vorherzusagen. Hierfür wurden 22 normalhörende Probanden aufgefordert, die Höranstrengung eines laufenden Sprachsignals mit variierendem Hintergrundgeräusch zu bewerten.

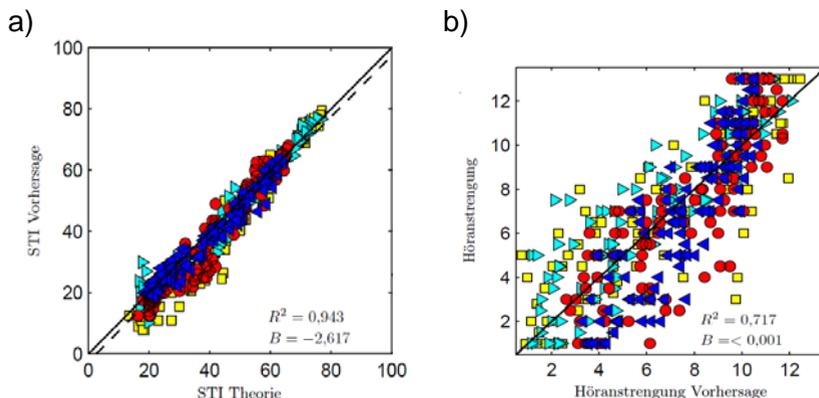


Abbildung 5a: Gegenüberstellung von geschätztem und berechnetem (theoretischem) STI. Unterschiedliche Symbole repräsentieren unterschiedliche Arten von zeitlich veränderlichen Störgeräuschen.

Abbildung 5b: Gemessene und durch das System vorhergesagte Höranstrengung für dieselben Störgeräuschbedingungen wie in Abbildung 5a)..

Die Bewertung wurde anhand einer 13-stufigen Skala von mühelos (1) bis extrem anstrengend (13) durchgeführt und erfolgte kontinuierlich, d.h. die Probanden bewegten einen „Slider“ entlang der Skala und bewerteten jeweils die momentan empfundene Höranstrengung. Änderungen des Störgeräusches hatten in der Regel Änderungen der Höranstrengung zur Folge, worauf die Probanden entsprechend durch Justierung des Sliders reagierten. Dieselben Mischungen von Sprache und Rauschen wurden anschließend durch das System bewertet. Die abgeschätzten Indizes wurden mit den subjektiven Daten korreliert und eine lineare Abbildung der Daten auf die Indizes berechnet. Die mit dieser Abbildung transformierten Werte sind als vorhergesagte Höranstrengung in Abbildung 5b) den tatsächlichen Messwerten gegenüber gestellt. Verschiedene Symbole stellen erneut unterschiedliche Störgeräuschbedingungen dar, während die durchgezogene Linie einer idealen Vorhersage entspräche. Es ist zu erkennen, dass die Streuung insgesamt größer ist als beim Vergleich zwischen theoretischem

und geschätztem STI (Abbildung 5a). Dies ist nicht verwunderlich, da Höranstrengung eine in hohem Maße subjektive Wahrnehmung ist und sich die Probanden teilweise stark unterscheiden. Der Anspruch eines Systems wie dem referenzlosen STI ist es daher stets, die generellen Trends derart komplexer Wahrnehmungsgrößen wiedergeben zu können. Dass dies der Fall ist, zeigt die erneut hohe Korrelation ($R^2 = 0,717$). Das System ist somit in der Lage subjektive Wahrnehmungsdaten vorherzusagen, ohne dabei eine bekannte Referenz vorauszusetzen. Dadurch ergeben sich völlig neue Anwendungsmöglichkeiten von Hörmodellen im Büroumfeld. Diese liegen bspw. in der visualisierten Rückmeldung an den Sprecher in Videokonferenzsystemen bzgl. der eigenen Verständlichkeit beim Empfänger, so dass letzterer nicht mehr aktiv in die Kommunikation eingreifen muss, wenn die Verständlichkeit zu schlecht wird. Weitere Möglichkeiten bestehen in der automatischen Steuerung von Soundmasking-Systemen, die bspw. nur dann aktiviert werden, wenn die Verständlichkeit des Büronachbarn *zu gut* wird und die eigene Konzentrationsfähigkeit mindern würde. Eine solche bedarfsangepasste Steuerung könnte den Lärmpegel und damit die Belastung der Mitarbeiter minimieren.

Vorhersage binauraler Sprachverständlichkeit

Wie im vorherigen Abschnitt erwähnt ist der STI in der Lage den Einfluss wichtiger Faktoren wie SNR und Nachhallzeit auf die Sprachwahrnehmung gut wiederzugeben. Andere Faktoren bleiben jedoch unberücksichtigt. Einer davon ist die Wahrnehmung von räumlich verteilten Schallquellen. Der Mensch ist mit erstaunlicher Genauigkeit in der Lage Informationen aus komplexen akustischen Szenen zu extrahieren (z.B. Gespräch in einer lauten Bar. Dies ist vor allem darauf zurückzuführen, dass mit zwei Ohren Schall empfangen wird. Wie schon in Abbildung 1b) dargestellt erreichen im Raum verteilte

Schallquellen die beiden Ohren (Mikrofone) mit unterschiedlicher Verzögerung. Zusätzlich erfolgt eine Abschattung durch den Kopf, so dass der Pegel auf dem der Quelle abgewandten Ohr in der Regel niedriger ist als auf dem zugewandten Ohr. Diese Effekte führen dazu, dass enorme Gewinne in der Sprachverständlichkeit erreicht werden können, wenn Sprache und Störgeräusch aus unterschiedlichen räumlichen Richtungen präsentiert werden. Daten aus einer Studie von Beutemann und Brand (2006) verdeutlichen dieses Phänomen. Gefüllte Kreise in Abbildung 6 zeigen Sprachverständlichkeitsschwellen, das heißt die SNR, bei denen 50% der Wörter verstanden werden können, die für eine frontale Sprachquelle (0°) und eine Störquelle aus verschiedenen Richtungen in der Horizontalebene gemessen wurden. Messungen wurden in einem schallarmen Raum (links) und einem Büro (rechts) durchgeführt. In beiden Fällen ist zu erkennen, dass eine räumliche Trennung von Sprach- und Störsignal eine erhebliche Absenkung der Schwelle zur Folge hat, d.h., dass der Proband bei einem wesentlich höheren Rauschpegel immer noch die gleiche Verständlichkeit zeigt. Dieser Effekt tritt auch signifikant in Büroräumen auf, wenngleich der Gesamteffekt reduziert ist.

Der STI wurde für frontale Sprachpräsentation entwickelt. Die Vorhersagen des STI (horizontale Linien in Abbildung 6) für räumlich getrennte Schallquellen stimmen daher nur für ebenfalls frontale Störung mit den Messdaten überein, während bei anderen Störgeräuschrichtungen erheblich Abweichungen auftreten. In der Praxis könnte dies bei richtungsabhängigen Schallquellen ungenaue Planungsvorgaben zur Folge haben. Beutemann und Brand (2006) entwickelten daher ein Modell, dass die binaurale Verarbeitung des menschlichen Gehörs nachahmt.

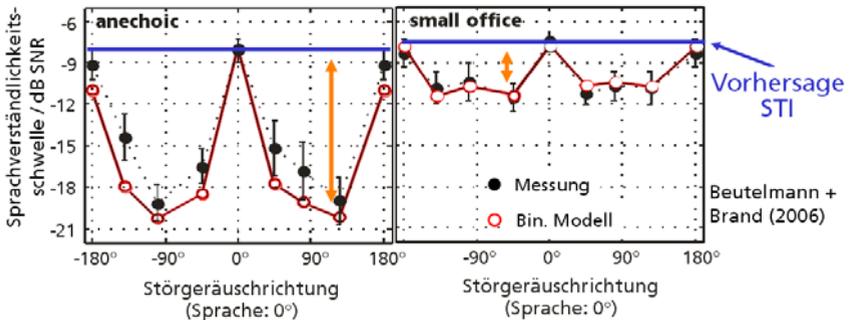


Abbildung 6: Sprachverständlichkeitsdaten (gefüllte Kreise) und –vorhersagen mit dem Modell von Beutelmann und Brand (offene Kreise) und dem STI (horizontale Linie).

Offene Kreise in Abbildung 6 stellen die entsprechenden Vorhersagen dar, die in allen Bedingungen gut mit den Messdaten übereinstimmen. Aufgrund seiner erhöhten Komplexität ist das Modell bisher zwar nicht Echtzeitfähig und somit noch nicht wie das oben beschriebene System zur kontinuierlichen Überwachung von Sprachqualität geeignet. Die erhöhte Präzision und aktuelle Forschungsarbeiten zur noch generellen Anwendbarkeit (Rennies, Brand, Kollmeier, 2011) lassen jedoch auf ein hohes Potenzial der binauralen Modelle in der akkuraten Planung von Büroraumakustik hoffen.

ZUSAMMENFASSUNG UND FAZIT

Trotz des technischen Fortschritts ist akustische Kommunikation in Büroumgebungen noch nicht optimal. Äußere Störfaktoren wie Nebengeräusche und schlechte Raumakustik schränken die Kommunikationsfähigkeit ein, was insbesondere für den beachtlichen Anteil von schwerhörenden Mitarbeitern erhebliche Probleme aufwirft und diese in Extremfällen sogar von der Kommunikation mit im Betrieb etablierten Systemen ausschließt. Technische Lösungen zur Unterstützung von akustischer Kommunikation sind daher unabdingbar. Hierfür wurden

verschiedene Ansätze vorgestellt. Ein Bereich befasst sich mit der Befreiung des zu Sprachsignals von unerwünschten Anteilen wie z.B. Lärm oder Nachhall. Dabei sind besonders gute Ergebnisse zu erzielen, wenn mehr als ein Mikrofon verwendet wird, da dann das System (ähnlich wie das menschliche Gehör) in der Lage ist eine Richtung, aus der vornehmlich Störgeräusch kommt, auszublenden. Weitere Ansätze zielen auf eine individuelle Kompensation von Hörverlusten ab, die direkt in Kommunikationssysteme integriert ist. Dadurch kann nicht nur die Problematik der geringen Hörgeräteversorgungsrate und der Inkompatibilität mit gängigen Geräten umgangen werden. Der konkrete durch das Gerät vorgegebene Nutzungskontext, beispielsweise bei Integration in einem Telefon, erlaubt zudem auf den Anwendungsfall spezifisch zugeschnittene Lösungen. Die Herausforderung besteht hier hauptsächlich in der Anpassung der Kompensationsparameter ohne professionelle Unterstützung. Erste Studien zeigen jedoch viel versprechende Ergebnisse und verdeutlichen damit das Potenzial, die Kommunikation für Schwerhörende am Arbeitsplatz zu verbessern. Als weiterer Bereich wurden Hörmodelle vorgestellt, die die menschliche Sprachwahrnehmung vorhersagen können. Solche Modelle sind in ihrer Aussagekraft einfachen, technischen Maßen deutlich überlegen und finden daher bereits Anwendung in der raumakustischen Planung. Aktuelle Ansätze der angewandten Hörforschung zielen darauf ab, etablierte Verfahren (z.B. STI) in referenzlose Echtzeitanwendungen zu überführen, um damit neue Anwendungsfelder wie die kontinuierliche Überwachung der Kommunikationsqualität zu erschließen. Fortlaufende Verbesserungen der zugrunde liegenden Hörmodelle schließen mittlerweile auch Wahrnehmungen in komplexen akustischen Situationen mit räumlich verteilten Schallquellen ein. Der Einsatz solcher weit entwickelten Modelle wird die Präzision von raumakustischen Planungen und die Evaluation von Kommunikationsumgebungen in Zukunft erheblich verbessern.

LITERATUR

- Bitzer, J. und Simmer, K.U. (2001). Superdirective microphone arrays. In M. S. Brandstein and D. Ward, editors, *Microphone Arrays: Signal Processing Techniques and Applications*, Kapitel 2, pp. 19–38. Springer-Verlag.
- Bormann, V., Sust, C.A., Heinecke-Schmitt, R., Fuder, G. und Lazarus, H. (2005). Schwerhörigkeit und Sprachkommunikation am Arbeitsplatz. Fb 1041 – Schriftenreihe der Bundesanstalt für Arbeitsschutz und Arbeitsmedizin. Wirtschaftsverlag NW, Dortmund-Berlin-Dresden.
- 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. *Int. 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. Dissertation, RWTH Aachen, Wissenschaftsverlag Mainz, Aachener Beiträge zu digitalen Nachrichtensystemen, Band 11.
- IEC (2003). Sound System Equipment - Part 16: Objective rating of speech intelligibility by speech transmission index. International Standard IEC 60268-16 (International Electrotechnical Commission).
- Irandoost, A. (2011). Entwicklung und Evaluation eines Verfahrens zur Personalisierung von Mediendaten für leichte bis mittelgradige Schwerhörigkeit. Bachelorarbeit, Jadehochschule Oldenburg + Fraunhofer IDMT, Oldenburg, 2011.
- Rennies, J., Albertin, E., Goetze, S. und Appell, J.-E. (2010). Automatic live monitoring of communication quality for normal-hearing and hearing-impaired listeners. In Miesenberger, K.; Klaus, J.; Zagler, W. und Karshmer, A. (ed.) *Computers Helping People with Special Needs, ICCHP 2010, Part II, LNCS 6180*, Springer – Berlin/Heidelberg, pp. 568-575.
- Rennies, J., Brand, T. und Kollmeier, B. (2011). Prediction of the influence of reverberation on binaural speech intelligibility in noise and in quiet. *J.Acoust.Soc.Am.*, *in press*.

- Rohdenburg, T., Hohmann, V. und Kollmeier, B. (2005). Objective perceptual quality measures for the evaluation of noise reduction schemes, In 9th International Workshop on Acoustic Echo and Noise Control, 2005, pp. 169-172.
- Rohdenburg, T., Huber, R., van Hengel, P., Bitzer, J. und 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. und Kammeyer, K.-D. (2008). Combined Source Tracking and Noise Reduction for Application in Hearing Aids, 8. ITG-Fachtagung Sprachkommunikation, Aachen
- Schepker, H. (2011). Weiterentwicklung und Evaluation von Verfahren zur subjektiven und objektiven Bestimmung von Höranstrengung. Bachelorarbeit, Jadehochschule Oldenburg + Fraunhofer IDMT, Oldenburg, 2011.
- Schepker, H., Rennies, J., Holube, I. und Kollmeier, B. (2011). Kontinuierliche Messung und Vorhersage von Höranstrengung. Proc. 37th Annual Convention for Acoustics (DAGA), pp. 589-590.
- Shield, B. (2006). Evaluation of the social and economic costs of hearing impairment. Technical report, Hear-It.
- Steeneken, H.J.M. und Houtgast, T. (1980). A physical method for measuring speech-transmission quality. J.Acoust.Soc.Am. 67(1), pp. 318-326.
- Vary, P., Heute, U. und Hess, W. (1998) Digitale Sprachsignalverarbeitung, Teubner-Verlag, Stuttgart
- Xiong, F., Schneider, D., Goetze, S., Ewert, S., Rohdenburg, T., Appell, J.-E. (2011). Hearing-Loss Compensation in a telephone system. Proc. 37th Annual Convention for Acoustics (DAGA), pp. 377-378.