

Signalverarbeitung für Hörhilfen und Audiologie

Norbert Dillier / Zürich

Einführung

Ausgehend von der Feststellung und diagnostischen Eingrenzung einer Hörstörung schliesst sich wenn möglich eine Behandlung dieser Hörstörung durch medizinische und technische Mittel an. Medizinische Behandlungsmöglichkeiten beinhalten medikamentöse Therapien und chirurgische Eingriffe. Technische Lösungen bestehen aus der Anpassung von Hörinstrumenten bzw. implantierbaren Systemen (Mittelohr-, Cochlea- und Hirnstammimplantate). In diesem Beitrag beschränken wir uns auf die technisch-apparativen Rehabilitationsmöglichkeiten und dabei insbesondere auf Aspekte der Signalverarbeitung.

Ziele der Signalverarbeitung für Hörinstrumente, Hörhilfen, Sprachprozessoren oder Kommunikationsgeräte ist es, die zur sprachlichen Verständigung erforderlichen Schallsignale für die hörgeschädigte Person wahrnehmbar („Hören“) und unterscheidbar („Verstehen“) zu machen und wenn möglich einem natürlichen Klangempfinden („Hörqualität“) anzunähern.

Traditionelle Hörgeräte mussten sich aufgrund der begrenzten Möglichkeiten der Analogschaltungstechnik auf einfache Verstärkung, Filterung und Pegelbegrenzung beschränken. Moderne Systeme mit digitaler Signalverarbeitung erlauben vielfältigere und neuartige Transformationen der Schallsignale, sodass vielleicht in nicht allzu ferner Zukunft die Vision eines „Verstehgerätes“ im Unterschied zu den bisherigen „Hörgeräten“ erfüllt sein könnte.

Die Algorithmen und Verfahren der digitalen Signalverarbeitung können in zwei grosse Klassen unterteilt werden. Die erste Gruppe ist relativ eng an die konkreten Ausprägungen eines Hörschadens gekoppelt und versucht, aufgrund möglichst genauer individueller Angaben über Art und Ausmass dieses Hörschadens Signalverarbeitungsparameter so zu wählen, dass Grundfunktionen der gestörten Hörwahrnehmung wie Lautheits- und Tonhöhenempfindung, Detektion zeitlicher Modulationen und weitere Klangeigenschaften möglichst gut kompensiert werden. Die zweite Gruppe betrifft globalere Funktionen, welche als Vorverarbeitung für praktisch beliebige nachgeschaltete Systeme benutzt werden können. Reduzierung von Störlärmeinfluss durch gesteuerte Richtmikrofone, automatische Klassifikation von Klangsituationen zur Wahl unterschiedlicher Verarbeitungsprogramme, Integration von Kommunikationssystemen wie Telefon, Raumbeschallungsanlagen, Internetanbindung usw. Im folgenden werden zu diesen beiden Signalverarbeitungsklassen einige Beispiele vorgestellt.

Signalverarbeitung für Innenohrschwerhörigkeit

Die Signalverarbeitung im Gehör und die bei einer Innenohr-Schwerhörigkeit möglicherweise auftretenden Störungen können vereinfacht wie folgt beschrieben werden. Das Schallsignal gelangt über Außen- und Mittelohr in das Innenohr, dessen Wirkung grob durch eine Filterbank mit anschließender einhüllender Extraktion und Dynamikkompression (kompressive Nichtlinearität z.B. durch aktive Prozesse im Innenohr) modelliert werden kann. Die weitere Signalanalyse im Hirnstamm kann durch eine Zerlegung nach Modulationsfrequenzen (Modulationsfilterbank) und einen binauralen Vergleich mit Störschallunterdrückung charakterisiert werden. Am Ausgang dieser Vorverarbeitungsstufen stehen die akustischen Merkmale des Eingangsschallsignals in Form einer „internen Repräsentation“ der kognitiven Ver-

arbeitung im Hörkortex zur Verfügung, wo eine optimale Verarbeitung und Mustererkennung der eingehenden, transformierten Schallsignale vorgenommen wird. Limitiert wird dabei die „Schärfe“ der internen Repräsentation durch das „interne Rauschen“. Als Störungen ist an erster Stelle die Abschwächungswirkung zu nennen, die durch Schalleitungsschwerhörigkeit und Ausfall der inneren sowie teilweisen Ausfall der äußeren Haarzellen bedingt ist, sowie als zweiter Faktor der Kompressionsverlust (Ausfall der äußeren Haarzellen), an dritter Stelle der binaurale Verlust (Reduktion der binauralen Störschall-Unterdrückung) und an vierter Stelle eine Erhöhung des internen Rauschens als „zentraler Hörverlust“.

Ausgehend von diesem grob-schematischen Modell sind folgende Möglichkeiten zur Kompensation der Verarbeitungsstörung denkbar:

1. Die Abschwächungswirkung lässt sich durch eine lineare Verstärkung kompensieren.
2. Der Kompressionsverlust lässt sich durch eine Dynamikkompensation kompensieren, die modellbasiert erfolgen sollte, um gemäß einem möglichst validierten Verarbeitungsmodell zu einer optimalen Kompensation dieses Funktionsausfalls zu gelangen.
3. Der binaurale Verlust sollte durch eine „echt“ binaurale Signalverarbeitung (z. B. Verstärkung der Unterschiede zwischen beiden Ohren) erfolgen, die normalerweise im Gehirn durch binaurale Signalverarbeitung durchgeführt wird.
4. Der zentrale Hörverlust kann durch eine Stör-Reduktion und eine Anpassung der jeweiligen Signalverarbeitung an die akute aktuelle akustische Situation erfolgen.

Die Wiederherstellung der Dynamik- und Lautheitsempfindung in unterschiedlichen Frequenzbereichen erfolgt sinnvollerweise durch einen Lautheitsmodell-gesteuerten Ansatz, bei dem das jeweilige Eingangssignal einerseits mit einem Lautheitsmodell für Normalhörende bewertet wird und zugleich mit einem Lautheitsmodell für den individuellen Schwerhörigen. Aus dem Vergleich kann die Modifikation des Eingangssignals im Hörgerät so durchgeführt werden, daß am Ausgang beider Lautheitsmodelle möglichst der gleiche (frequenzabhängige) Wert resultiert (vgl. Hohmann, 1993, Launer, 1995, Appell, 2002). Da dieser Ansatz wesentlich von der Güte und Validierung des verwendeten Lautheitsmodells abhängt, gilt es, hier eine an die Gegebenheiten von Schwerhörenden und an praktische Messungen angepasste und validierte Version des Lautheitsmodells zu verwenden.

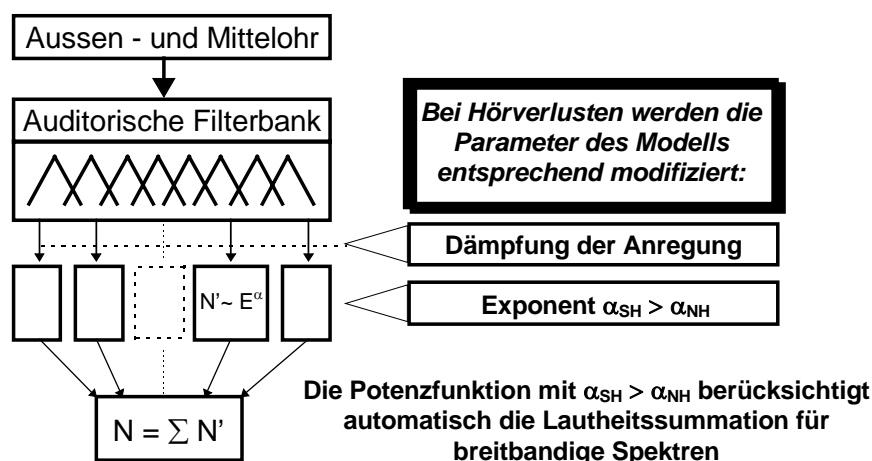


Abb. 1: Modell zur Berechnung der Lautheit für Normal- und Schwerhörige (modifiziert nach Zwicker).

Das in Abb.1 gezeigte modifizierte Modell berücksichtigt die cochleären Defizite Schwerhöriger. Die erhöhte Hörschwelle dämpft direkt die Anregung der Cochlea, die Steilheit der Lautheitsfunktion beeinflusst die Potenzfunktion und die Lautheitsbildung wird damit im Vergleich zu Normalhörenden stark verändert. Die optimalen Parameter für das Lautheitsmodell liefern das Audiogramm und die Lautheitsskalierung.

Die Zielverstärkung in einem solchen Signalverarbeitungsalgorithmus ist abhängig vom Signal. In jedem einzelnen Filterband wird die Verstärkung so gewählt, dass die entsprechende Bandlautheit (spezifische Lautheit) derjenigen eines Normalhörenden bei gleichem Signal entspricht; die Lautheit wird normalisiert.

Die Umwandlung des Erregungspegel-Musters in eine spezifische Lautheit geschieht durch eine Lautheitstransformation, die die Form der Isophonen von Normal- und Schwerhörenden insbesondere bei tiefen Frequenzen und dabei auch neuere Messungen zum Verlauf dieser Isophonen berücksichtigt (Gabriel et al., 1997). Nach der sich anschließenden spektralen Aufintegration der spezifischen Lautheit und Bestimmung der Gesamt-Lautheit in sone erfolgt eine Transformation in kategoriale Einheiten, so daß die Kategorial-Lautheit (KU) vorhergesagt werden kann, die sich unmittelbar mit Hilfe der kategorialen Lautheitsskalierung experimentell nachmessen lässt

Herkömmliche Hörgeräte modellieren die Cochlea als lineare, pegelunabhängige Filterbank und berechnen die Zielverstärkung als Funktion des Schalldruckpegels. Mit einem solchen Multiband-Konzept können jedoch maskierte Signale nicht erkannt werden, daher werden auch maskierte Komponenten verstärkt. Diese Fehlsignale äussern sich als Rauschen und Verzerrungen und beeinträchtigen die Klangqualität des Hörgerätes. Klassische Multiband-Systeme sind auch nicht in der Lage, schmal- und breitbandige Signale zu differenzieren und können damit auch die verminderte Lautheitssumation Schwerhöriger nicht kompensieren.

Eine gehörgerechte Signalverarbeitung hingegen bezieht die elementaren Funktionen der normalen und geschädigten Cochlea wie Maskierung, Lautheit und Lautheitssumation mit ein.

Möglichkeiten der Störgeräusch-Reduktion

Die derzeitige Lage von Störgeräusch-Unterdrückungs-Algorithmen bei Hörsystemen stellt sich wie folgt dar:

- Die Annahmen verschiedener Störunterdrückungs-Algorithmen sind in der Realität nie vollständig erfüllt, so daß in modernen Hörgeräten verschiedene Algorithmen kombiniert bzw. je nach vorliegender akustischer Umgebungssituation aktiviert und deaktiviert werden müssen.
- Akustisch „einfache“ Situationen sind mit derzeitiger Technologie beherrschbar (dazu zählt stationäres Störgeräusch mit einer großen spektralen Differenz zwischen Sprache und Störgeräusch und stabilen räumlichen Differenzen).
- Probleme bereiten dagegen nach wie vor „schwierige Situationen“, die sich durch instationäre Störschallquellen, durch Nachhall und durch mehrere gleichzeitig aktive Sprecher von verschiedenen Richtungen auszeichnen. Hier besteht noch ein großer Forschungsbedarf.

Die neueste Entwicklung in diesem Bereich stellen adaptive Mehr-Mikrofon-Systeme dar (Spriet et al., 2005). Bei ihnen wird die Richtcharakteristik in Abhängigkeit von der Einfallsrichtung des Störgeräusches variiert (siehe Abb. 2)

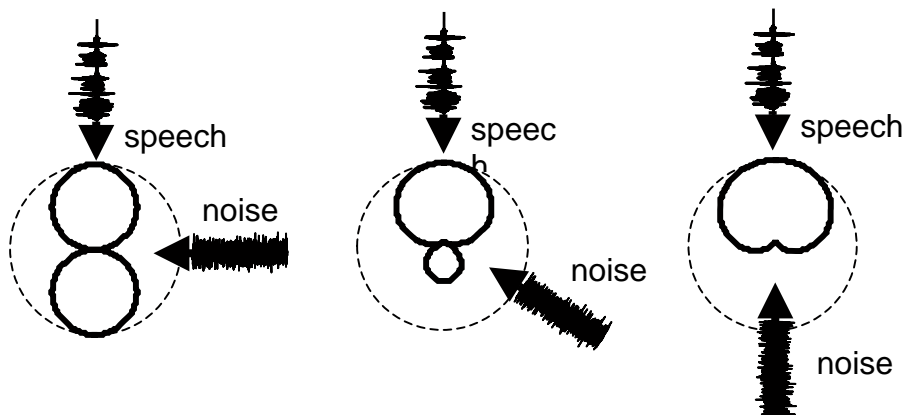


Abb. 2: Adaptives Richtmikrofonsystem. Die Richtcharakteristik wird in Abhängigkeit von der Einfallsrichtung des Störgeräusches variiert.

Zur Reduzierung von Störgeräuschen können in digitalen Hörgeräten die Eingangssignale analysiert und dadurch Nutzsignale von Rauschen unterschieden werden. Die frequenzabhängige Verstärkung des Hörgerätes kann dann in Abhängigkeit von der Umgebungssituation geregelt werden. Um Sprache von Störgeräuschen zu unterscheiden, müssen spezielle Kenntnisse über die Eigenschaften von Sprache herangezogen werden. Mit deren Hilfe können z. B. Sprachpausen erkannt werden. In solchen Pausen kann das Rauschsignal näherungsweise berechnet werden und vom Eingangssignal subtrahiert werden. Auf diese Weise wird das Rauschsignal reduziert. Da die Subtraktion im Frequenzbereich erfolgt, nennt man dieses Verfahren spektrale Subtraktion (s. z.B. Marcincik und Kollmeier, 1999).

Bei einem anderen Algorithmus werden z. B. die langsamen Schwankungen der Einhüllenden der Signale berechnet. Die Änderungsgeschwindigkeit der Einhüllenden wird als Modulationsfrequenz bezeichnet. Die Einhüllende zeigt im Frequenzbereich einen für Sprache typischen Verlauf mit einem Maximum im Bereich von 2-8 Hz. Diese Frequenzen sind durch den Rhythmus der Sprache (die Dauer der Wörter, Silben und Phoneme) bestimmt und sind sowohl vom Klang der Stimme als auch von der gesprochenen Sprache unabhängig. Störgeräusche dagegen beinhalten zumeist andere Modulationsfrequenzen. Ein Algorithmus kann nun die Einhüllenden der Eingangssignale analysieren. Wenn die typischen Modulationsfrequenzen von Sprache im Bereich von bis zu 10 Hz nicht vorhanden sind, wird die Verstärkung in dem entsprechenden Frequenzbereich und damit das Störgeräusch reduziert.

Feedbackreduktion

Ein wichtiges Qualitätsmerkmal von Hörgeräten ist die Unterdrückung der akustischen Rückkopplungen (Feedback). Sie wird meist durch eine Reduzierung der Verstärkung bei der Anpassung des Hörgerätes vermieden. Eine andere Möglichkeit sind schmalbandige Lückenfilter (Notch-Filter), die bei den kritischen Feedbackfrequenzen positioniert werden und die Verstärkung bei der betreffenden Frequenz damit selektiv reduzieren. Solche statischen Anpassungen sind jedoch wirkungslos, falls die Rückkopplungen als Reaktion auf veränderte äußere akustische Bedingungen plötzlich bei einer anderen Frequenz auftreten. Dieses Problem können adaptive Algorithmen lösen. Sie analysieren ständig das Eingangssignal und erkennen Rückkopplungen. Ein Notch-Filter oder ein Kompensationsfilter eliminieren den Feedbackanteil im Signal und werden ständig an die aktuellen Bedingungen an-

gepaßt. Der Kompensationsfilters-Ansatz vermeidet Rückkopplungen durch die Einstellung eines adaptiven Filters, so dass sein Ausgangssignal die rückgekoppelten Anteile im Eingangssignal gerade auslöscht.

Auditorische Objekterkennung und Klangklassifizierung

Hörgeräte ermöglichen es, für verschiedene akustische Situationen unterschiedliche Programme zu wählen, um den Frequenzgang und Kompressionsparameter zu ändern, oder Richtmikrofon, Störgeräusch-Reduktion oder Feedback-Unterdrückung zu aktivieren. Der Hörgeräträger hat dabei die nicht immer leichte Aufgabe, die akustische Hörsituation zu beurteilen und dann per Schalter am Hörgerät oder über eine Fernbedienung das entsprechende Programm zu wählen. Ein automatisches Erkennen der aktuellen akustischen Situation und automatisches Umschalten in das geeignetste Programm könnten den Hörgerätekomfort verbessern. Ein System zur Geräuschklassifizierung wird im folgenden vorgestellt, welches die vier Klassen 'Sprache', 'Sprache im Störgeräusch', 'Störgeräusch' und 'Musik' robust erkennen soll.

Es sind bereits einige Hörgeräte im Handel, die eine Geräuschklassifizierung vornehmen. In einem Gerät von Widex wird die Amplitudenstatistik des Signals ausgewertet, um zwischen impulshaften und kontinuierlichen Klängen zu unterscheiden und damit ein Störgeräusch-Unterdrückungssystem zu steuern (Ludvigsen, 1993). Im Zusammenhang mit Störgeräusch-Unterdrückung wird auch von ReSound eine Geräuschklassifizierung vorgenommen (Edwards et al., 1998). Anhand einer Analyse der Amplitudenmodulationen wird Sprache im Signal detektiert. Von Phonak ist ein Gerät im Handel, das abhängig von der akustischen Umgebung zwischen zwei Programmen umschaltet (Phonak, 1999). Die Klassifizierung in die Klassen 'Sprache im Störgeräusch' und 'Andere' erfolgt aufgrund einer Analyse der zeitlichen Fluktuationen und der Form des Spektrums, wie es von Kates (1995) vorgeschlagen wurde. Ostendorf et al. (1998) unterscheiden basierend auf einer Modulationsfrequenz-Analyse die drei Klassen 'Sprache', 'Sprache im Störgeräusch', und 'Störgeräusch'. Von Nordqvist (2000) wird ein Verfahren vorgeschlagen, mit dem reine Sprache und verschiedene Hintergrund-Geräusche mittels einer LPC-Analyse und HMMs klassifiziert werden. Ein Ansatz von Feldbusch (1998) erkennt reine Sprache, Stimmengewirr und Verkehrslärm, indem verschiedene zeitliche und spektrale Merkmale mittels neuronalen Netzen ausgewertet werden.

Alle genannten Ansätze erlauben ein robustes Erkennen von reiner Sprache; Musik hingegen kann bis anhin nicht identifiziert werden, und es ist nur teilweise möglich, Störgeräusch von Sprache im Störgeräusch zu trennen.

Auditorische Szenenanalyse

Auditorische Szenenanalyse beschreibt die Mechanismen und Strategien, die das auditorische System verwendet, um die akustische Umgebung zu analysieren. Obwohl dieser Prozess noch nicht vollständig verstanden wird, ist bekannt, dass das auditorische System charakteristische Merkmale aus dem akustischen Signal extrahiert. Diese auditorischen Merkmale umfassen spektrale Trennung, spektrales Profil, Harmonizität, Onsets und Offsets, kohärente Amplituden- und Frequenz-Modulationen sowie räumliche und zeitliche Trennung.

Auditorische Merkmale

Im folgenden Ansatz für die Geräuschklassifizierung werden bis anhin vier Merkmalsgruppen verwendet: Amplituden-Modulationen, spektrales Profil, Harmonizität und Onsets. Für die Modellierung der Amplituden-Modulationen werden die genannten bisherigen Ansätze von Ludvigsen (1993), Ostendorf et al. (1998) und Kates (1995) verwendet. Fig. 3 zeigt

als Beispiel, wie das Amplituden-Histogramm mittels Perzentilen modelliert wird, und der Abstand zwischen den Perzentilen Aufschluss über die Modulationstiefe gibt.

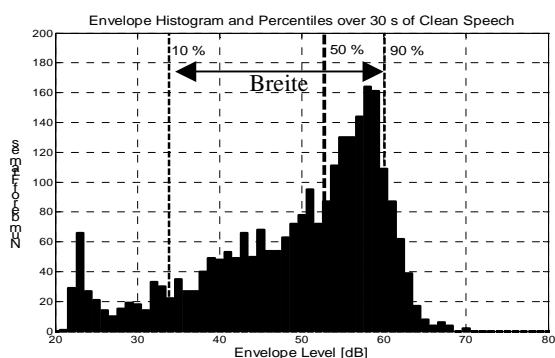


Fig. 3: Amplitudenhistogramm und Perzentile eines Sprachsignals. Das Merkmal 'Breite' ist definiert als Differenz zwischen der 90 % und der 10 % Perzentile.

Das spektrale Profil wird auf eine rudimentäre Weise durch zwei Merkmale modelliert, den spektralen Schwerpunkt und die zeitlichen Schwankungen des spektralen Schwerpunktes (Fig. 4). Diese beiden Merkmale wurden bereits im Klassifizierer von Phonak (1999) verwendet.

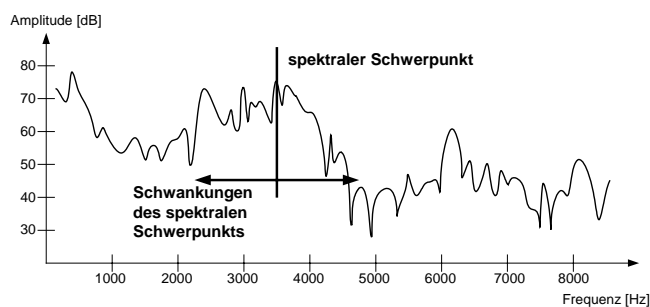


Fig. 4: Zur Beschreibung des spektralen Profils werden der spektrale Schwerpunkt und dessen zeitliche Schwankungen berechnet.

Für die Beschreibung der Harmonizität wird normalerweise der Pitch eines Klanges benutzt (Fig. 5). In unserem Ansatz wird die Harmonizität durch zwei Merkmale charakterisiert: die Tonalität und die Pitch-Varianz. Die Tonalität ist definiert durch das Verhältnis der harmonischen zu den unharmonischen Anteilen im Zeitsignal.

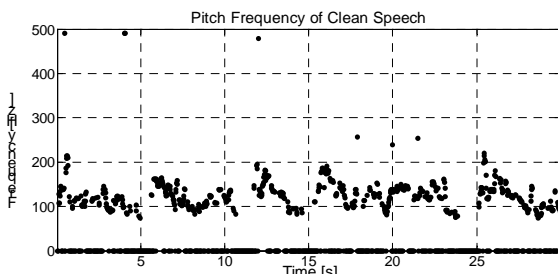


Fig. 5: Typischer zeitlicher Pitch-Verlauf eines Sprachsignals. Das Merkmal Tonalität ist das Verhältnis von harmonischen zu unharmonischen (hier durch 0 Hz angezeigten) Anteilen.

Um Amplituden-Onsets zu modellieren, wird das Auftreten hoher und schneller Amplitudenanstiege im Bark-Spektrum analysiert. Fig. 6 zeigt ein typisches Onsetmuster für Popmusik.

Da im Onsetmuster auch sehr schön der Rhythmus eines Signals sichtbar wird, wird zusätzlich ein Merkmal für den Takt berechnet.

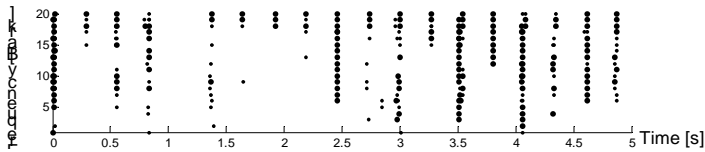


Fig. 6: Onsetmuster von Pop-Musik. Starke Onsets sind durch Punkte dargestellt. Der Rhythmus der Musik ist deutlich erkennbar.

Die genannten auditorischen Merkmale wurden zusammen mit Klassifizierern unterschiedlicher Komplexität evaluiert. Fünf verschiedene Klassifizierer-Typen kamen zur Anwendung: Regelbasierter Ansatz, Minimum-Distance Klassifizierer, Bayes Klassifizierer, neuronales Netz, und Hidden Markov Modell (HMM).

Der Vorteil der ersten zwei Klassifizierer ist, dass diese etwa viermal weniger Rechenzeit und Speicherplatz benötigen als die übrigen Ansätze; dafür ermöglichen die letzteren eine genauere Unterteilung des Merkmalsraums.

Die Performance der einzelnen Klassifizierer mit unterschiedlichen Merkmalssets wurde mit etwa 300 verschiedenen Alltagsklängen getestet. Die Klang-Datenbank enthält unterschiedlichste Beispiele zu den vier Klassen; für die Klasse 'Sprache' zum Beispiel verschiedene Sprecher, die unterschiedliche Sprachen mit unterschiedlichen Geschwindigkeiten und Tonlagen in unterschiedlich halligen Räumen sprechen, oder Signale der Klasse 'Störgeräusch' aus Haushalt, Verkehr, Restaurant, Industrie usw., 'Sprache im Störgeräusch' bei verschiedenen SNR, und 'Musik' von Klassik über Pop und Rock zu einzelnen Instrumenten und Gesang.

Von jeder Klasse wurden 80 % der Klänge fürs Training und 20 % fürs Testen der Klassifizierer verwendet. Viele Klänge wurden sehr robust erkannt, insbesondere reine und schwach verhallte Sprache, Verkehrs- und Stimmen-Geräusche, klassische Musik, Instrumente und Gesang. Probleme bereiten vor allem stark verhallte Sprache, fluktuierende Störgeräusche und komprimierte Pop-Musik. Dafür sollten weitere Merkmale aus dem Signal extrahiert werden, wie zum Beispiel ein Mass für die Stärke des Nachhalls im Signal.

Merkmale zur auditorischen Objekterkennung

Einige charakteristischen spezifischen Merkmale (Yost & Sheft, 1993), welche für die Bildung akustischer Objekte wichtig sind, werden im folgenden aufgelistet.

- Spektrale Trennung

Die bemerkenswerte Fähigkeit des auditorischen Systems, den spektralen Inhalt eines Signals zu bestimmen, ist eine wichtige Basis für die weitere Trennung in die nachfolgenden Merkmale.

- Spektrales Profil

Die meisten Klangquellen produzieren ein bestimmtes Amplitudenspektrum, dessen Profil relativ konstant bleibt wenn der Gesamtpegel verändert wird. Verschiedene Klangklassen, wie z.B. Sprache, stationäres Störgeräusch und Musik, haben oft unterschiedliche spektrale Profile, was zur Unterscheidung von Quellen beitragen kann. Das spektrale Profil ist auch für die Klangfarbe (Timbre) von Musikinstrumenten und Stimmen wesentlich. Das Timbre wird jedoch ebenso durch temporale Faktoren bestimmt, wie Anstiegs- und Abfallzeiten von gespielten Noten eines Instruments.

- Harmonische Struktur

Die Teiltöne vieler natürlicher Klangquellen haben eine harmonische oder annähernd harmonische Struktur, die ihren Ursprung in einem schwingenden Medium hat. Die Harmonischen formen ein einzelnes auditorisches Bild, beschrieben als Pitch. Dies ist ein hervorragendes Beispiel von Fusion, denn es ist in der Regel unmöglich, einzelne Teiltöne herauszuhören. Da unterschiedliche Quellen meist verschiedene harmonische Serien bilden, ist mit dem Pitch eine Trennung von Quellen möglich.

- Gemeinsame Ein / Ausschwingzeit

Eines der stärksten Merkmale für die Gruppierung von Teiltönen ist der gemeinsame Anstieg der Amplitude. Gleichzeitig einsetzende Teiltöne werden in der Regel zu einem auditorischen Objekt fusioniert. Kleinere Asynchronitäten bestimmen zu einem wesentlichen Teil das Timbre der Quelle, grössere asynchrone Amplitudenanstiege einzelner Teiltöne tragen andererseits stark zur Trennung in separate Quellen bei, auch wenn die Teiltöne untereinander harmonische Verhältnisse aufweisen. Messungen haben ergeben, dass Musiker in einem Orchester um sich selbst zu hören bis zu 50 ms versetzt spielen (Mellinger & Mont-Reynaud, 1996). Auch das Publikum kann dadurch einzelne Instrumente heraushören, ohne aber den Eindruck eines gleichzeitigen Einsatzes zu verlieren.

- Kohärente Amplituden- und Frequenzmodulation

Amplitudenmodulation ist charakteristisch für viele natürliche Klangquellen wie z.B. Sprache. Unter kohärenter Amplitudenmodulation versteht man die gleichzeitige Amplitudenänderung von mehreren Teiltönen. Das auditorische System vermag spektrale Komponenten zu gruppieren, die mit demselben zeitlichen Muster moduliert sind, oder zu trennen, wenn die Muster unterschiedlich sind. Dies funktioniert für Modulationsfrequenzen unter 50 bis 100 Hz.

Weit weniger stark scheint die gemeinsame Frequenzmodulation mehrerer Partialtöne zur Fusion oder Trennung beizutragen. Trotzdem kann sie gerade bei Musik zur Trennung von Instrumenten beitragen (Vibrato).

- Lokalisation

Das auditorische System verrichtet eine schwierige Aufgabe, wenn es entscheidet, woher im Raum ein Signal kommt. Interaurale Zeit- und Pegeldifferenzen und Aspekte der Übertragungsfunktion von Kopf / Ohrmuschel sind entscheidend für die Lokalisation. Für die Trennung von Klangquellen ist die räumliche Lokalisation zwar signifikant aber nicht unerlässlich.

Gruppierungsverfahren

Unter Gruppierung versteht man die auf den vorgängig beschriebenen physikalischen Merkmalen basierende Zuordnung der Ereignisse in einem akustischen Signal zu verschiedenen akustischen Objekten (Fusion oder Trennung).

Primitive Gruppierung und Gruppierung nach Schema

Bregman (1990) unterscheidet zwischen zwei Mechanismen um zu entscheiden, welche Komponenten zu einer bestimmten Klangquelle gehören.

- Primitive Gruppierungsmechanismen teilen das Eingangssignal auf der Basis einfacher physikalischer Signalmerkmale auf, wie dies im vorhergehenden Abschnitt beschrieben wurde. Diese Mechanismen funktionieren ohne vorheriges Wissen.
- Bei der Gruppierung nach Schema geschieht die Auswahl und Kombination von Komponenten des komplexen Signals aufgrund von Erfahrungen und gelernten Mustern, also aufgrund der Eigenschaften von Klängen oder Klangklassen, die einem schon früher begegnet sind.

Die Erfahrung zeigt, dass die kontextunabhängige, primitive Gruppierung viel einfacher modellierbar und anzuwenden ist als die Gruppierung nach Schema. Allerdings ist nicht zu erwarten, dass eine vergleichbare Klassifizierungs- und Quellentrennleistung wie beim audito-

rischen System erreicht werden kann, wenn die höher liegenden Gruppierungsmechanismen nicht berücksichtigt werden.

Gestaltprinzipien

Am Anfang des zwanzigsten Jahrhunderts hat die Forschung der visuellen Wahrnehmung die sogenannten Gestaltprinzipien formuliert, welche nun auch für auditorische Empfindungen Anwendung finden. Bregman (1990) nennt dazu zwei Grundregeln (allocation und accounting):

- Jedes Element eines komplexen Signals kann nur einer Quelle zugeordnet werden.
- Alle Elemente des akustischen Ereignisses müssen einer Quelle zugeordnet werden. Kann ein Element keiner existierenden Quelle zugeordnet werden, so wird es selbst zu einer Quelle.

Für die Bildung von auditorischen Objekten werden diese Gestaltprinzipien auf verschiedenen Ebenen angewendet.

- **Kontinuität:** Amplitudenmoduliertes Rauschen wird als einzelnes Objekt wahrgenommen, falls die Modulation glatt (zum Beispiel sinusförmig) ist. Bei Rechteckmodulation hingegen hört man zwei Objekte. Pfeifen, zusammenhängend oder abgesetzt, erzeugt eine oder zwei Quellen.
- **Nähe:** Aufeinanderfolgende Noten, die spektral nahe beieinander liegen (Pitch), werden zur gleichen Quelle gezählt. Es braucht viel, bis eine Melodie durch Störsignale zerfällt.
- **Ähnlichkeit:** Ähnliche Klangkomponenten werden gruppiert; Noten mit demselben Timbre werden meistens derselben Quelle zugeordnet.
- **Gemeinsames Schicksal:** Ändern sich Teiltöne eines Klangs auf gleiche Weise, werden sie gruppiert. Gemeinsamer Amplitudenanstieg und gemeinsame Modulationen sind starke Gruppierungsprinzipien. Werden in einem harmonischen Klang einige Teiltöne moduliert, heben sie sich von den anderen ab, und man hört zwei Klänge.
- **Geschlossenheit:** Geschlossenheit hilft, Kontinuität trotz abrupter Änderungen zu erhalten. Lücken in einem Signal trennen Ereignisse, während Rauschen in den Lücken hilft, sie wieder zu fusionieren. Ein Beispiel dazu wäre Klatschen beim Sprechen. Auch die Trennung von mehreren gleichzeitigen Sprechern wird dadurch unterstützt.
- **Gute Fortsetzung:** Die meisten Klänge ändern nicht plötzlich den Charakter; ein Klavier wird nicht plötzlich wie eine Violine tönen. Das heisst, dass feine, glatte Änderungen einer einzigen Quelle zugeordnet werden, plötzliche Wechsel erzeugen eine neue Quelle.

Wenn wir uns die Prozesse zur akustischen Szenenanalyse nochmals vor Augen führen und mit den Methoden aktueller Hörgeräte-Automaten vergleichen so erkennen wir, dass die heutigen Lösungsansätze sehr technisch inspiriert sind und noch wenig mit unserem eigenen Wahrnehmungssystem zu tun haben. Wir haben aber die Grenzen der heutigen Lösungen erkannt und neue Wege entdeckt, die noch fehlenden Klassifizierungsleistungen einzubringen. Die primitive Gruppierung auditorisch basierter Merkmale mit Hilfe der Gestaltprinzipien ist ein vielversprechender Lösungsansatz zur besseren Klassifizierung der akustischen Umgebung. Es dürfte möglich sein, einige Prinzipien zu formulieren und in Hörsystemen einzusetzen.

Schlussfolgerungen

Die eingangs genannten Ziele der rehabilitativen Audiologie (Hören, Verstehen, Klangqualität) sind durch digitale Signalverarbeitung heute nur unzureichend erfüllbar, obwohl in den letzten Jahren ein deutlicher Fortschritt erzielt worden ist und auch mit weiteren Fortschritten

in nächster Zeit zu rechnen ist. Aus audiologischer Sicht ist dem Versuch eines Ausgleichens des Hörschadens eine besondere Bedeutung beizumessen, wobei es um eine Verbesserung der Diagnostik und des Verständnisses von Verarbeitungsdefiziten bei Innenohr-Schwerhörigkeit primär geht. Dabei erscheint der Modell-basierte Ansatz erfolgversprechend, da durch Verarbeitungsmodelle unser derzeitiges Wissen über die Funktion und etwaige Fehlfunktion des Hörsystems in für die Hörgeräte-Signalverarbeitung nutzbarer Form charakterisiert werden kann. Komplexe Computermodelle zur auditorischen Szenenanalyse sind allerdings derzeit zur Implementation in Hörgeräten noch zu aufwendig. Wenn nicht nur rein technisch inspirierte Konzepte zur Klassifizierung akustischer Situationen und Objekte verwendet werden sollen, lassen sich den zusammengefassten Theorien folgend aber Modelle und Regeln zur Verwendung in Hörinstrumenten ableiten.

Literatur

- Appell, Jens-Ekkehardt, Loudness models for rehabilitative Audiology, Dissertation, Universität Oldenburg, 2002 (<http://docserver.bis.uni-oldenburg.de/publikationen/dissertation/2002/applou02/applou02.html>)
- Büchler, M. et al. (2000). „Klassifizierung der akustischen Umgebung für Hörgeräte-Anwendungen“, DAGA 2000.
- Dau, T. et al. (1998). „Psychophysics, Physiology and Models of Hearing“. World Scientific Publishing.
- Edwards, B. W., Hou, Z., Struck, C. J., Dharan, P. (1999). "Signal-processing algorithms for a new software-based, digital hearing device," *The Hearing Journal*, Vol. 51, No. 9, 44-52.
- Feldbusch, F. (1998). "Geräuscherkennung mittels Neuronaler Netze," *Zeitschrift für Audiologie* 1/1998, 30-36.
- Gabriel, B., B. Kollmeier and V. Mellert (1997). "Influence of individual listener, measurement room and choice of test tone levels on the shape of equal-loudness level contours." *Acustica united with acta acustica* 83(4): 670-683.
- Hohmann, V. (1993). *Dynamikkompensation für Hörgeräte- Psychoakustische Grundlagen und Algorithmen*. Düsseldorf, VDI-Verlag.
- Kates, J. M. (1995). „Classification of background noises for hearing-aid applications“, *Journal of the Acoustical Society of America*, 97 (1), 461-470.
- Launer, S. (1995). Loudness perception in listeners with sensorineural hearing loss. Fachbereich Physik. Oldenburg, Carl-von-Ossietzky Universität Oldenburg. (http://www.physik.uni-oldenburg.de/Docs/medi/pub_down.html)
- Ludvigsen, C. (1997). „Schaltungsanordnung für die automatische Regelung von Hörhilfsgeräten“, *Europäische Patentschrift*, EP 0 732 036 B1.
- Marzinzik, M. and Kollmeier, B. (1999). "Development and Evaluation of Single-Microphone Noise Reduction Algorithms for Digital Hearing Aids," in *Psychophysics, Physiology and Models of Hearing*, edited by T. Dau, V. Hohmann, and B. Kollmeier (World Scientific, Singapore), pp. 279-282.
- Mellinger, D. K., Mont-Reynaud, B. M. (1996). „Scene Analysis,“ in *Auditory Computation*, edited by H. L. Hawkins et al., Springer, New York.
- Nordqvist, P. (2000). "Automatic Classification of Different Listening Environments in a Generalized Adaptive Hearing Aid," *International Hearing Aid Research Conference, IHCON 2000*, Lake Tahoe, CA.
- Ostendorf, M. et al. (1997). „Empirische Klassifizierung verschiedener akustischer Signale und Sprache mittels einer Modulationsfrequenzanalyse“, DAGA 97, 608-609.
- Phonak Hearing Systems (1999). "Claro AutoSelect", company brochure no. 028-0148-02.
- Spriet, A., Moonen, M., and Wouters, J. (2005). "Stochastic gradient-based implementation of Spatially Preprocessed Speech Distortion Weighted Multichannel Wiener filtering for noise reduction in hearing aids," *IEEE Trans. Signal Process.* **53**, 911-925.
- Yost, A. W., Sheft, S. (1993). „Auditory Perception,“ in *Human Psychophysics*, edited by W. A. Yost et al., Springer, Berlin.
- Zwicker, E., Fastl, H. (1990). „Psychoacoustics, Facts and Models“, Springer, Berlin.