

7. Ein Signalverarbeitungs-Rahmen für binaurale Modelle

7.1. Verarbeitung der Eingangssignale

Die Vorverarbeitungsstufen des Verfahrens haben die Aufgabe, die aufgenommenen reellen Zeitsignale für die Analyse in einem Cocktail-Party-Processor aufzubereiten. Da die Prozessoren analytische Zeitsignale in Frequenzgruppen benötigen, besteht die Vorverarbeitung aus folgenden Stufen: Filterung der Signale, Erzeugung von analytischen Zeitsignalen, evtl. Datenreduktion. Das Prinzipschaltbild der Vorverarbeitungsstufen ist in Bild 7.1 dargestellt.

7.1.1 Frequenzgruppen-Filter

Die Signale sollen, analog zum menschlichen Gehör, innerhalb von Frequenzgruppen verarbeitet werden. Die Verarbeitung in Frequenzgruppen hat folgende Eigenschaften:

- Es können relativ schnelle Änderungen der Schallsignal-Parameter erfaßt werden. Gemäß dem Zeit-Bandbreite-Produkt liegt die mögliche Zeitauflösung bei Frequenzgruppen-Filtern zwischen 10 ms bei tiefen Frequenzen (Bandbreite 100 Hz) und einigen 100 µs bei hohen Frequenzen (Bandbreite bis zu 3 kHz).
- Die Anzahl der Frequenzbereiche ist hoch genug, um unterschiedliches Signalverhalten in unterschiedlichen Frequenzbereichen auswerten zu können.

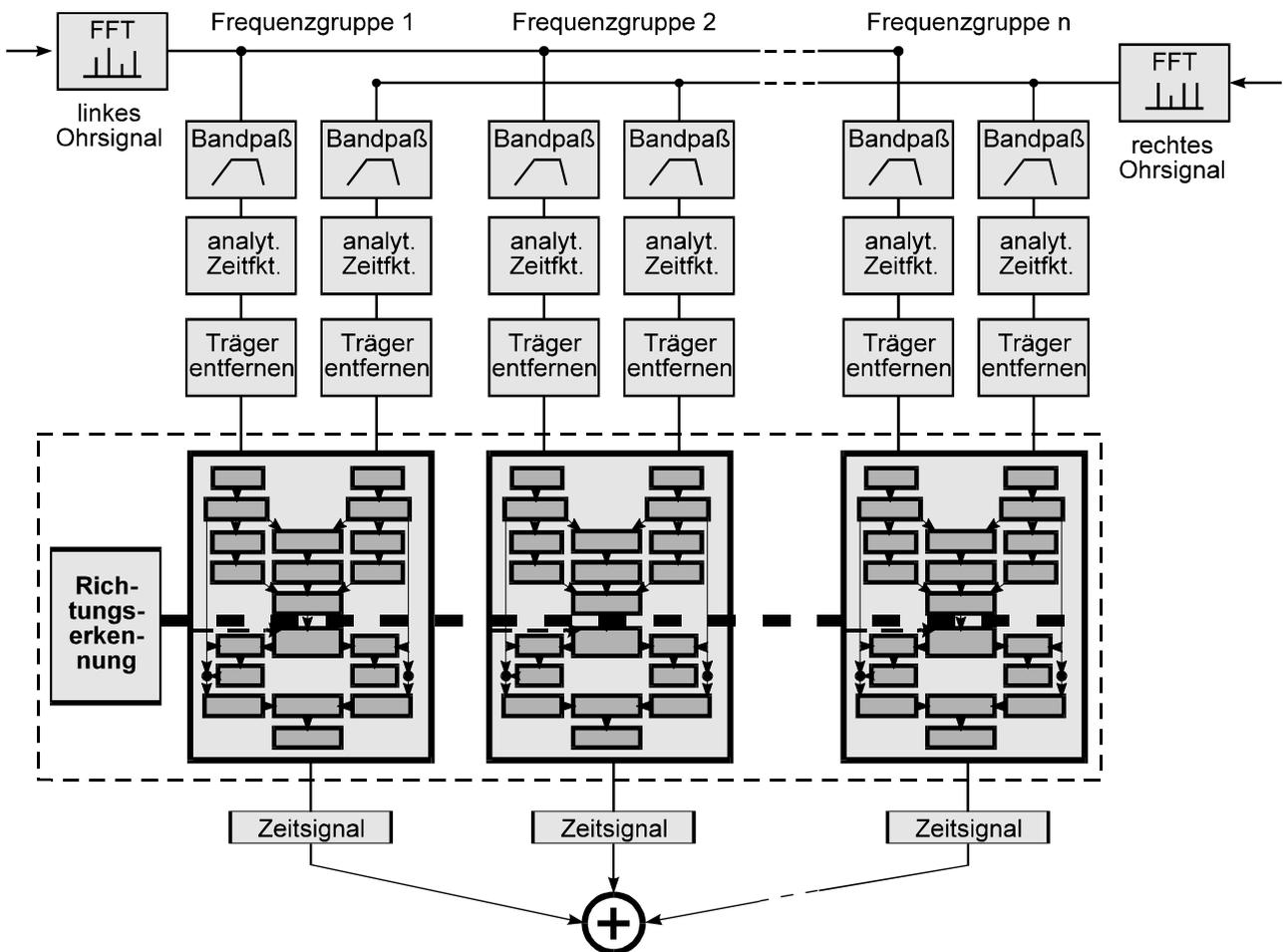


Bild 7.1: Signalverarbeitungs-Rahmen für das binaurale Modell
Der Modellaufbau innerhalb der Frequenzgruppen entspricht Bild 7.5

- Analyse-Bandbreite und Zeitauflösung sind an die Eigenschaften des Gehörs angepaßt.

An ein Filterverfahren zur Frequenzgruppen-Filterung werden folgende Anforderungen gestellt:

- Es soll möglichst nur der gewünschte Frequenzbereich herausgefiltert werden.
- Innerhalb des Durchlaßbereichs sollen die Signale nicht verändert werden.
- Um Querbezüge zwischen Signalen unterschiedlicher Frequenzgruppen herstellen zu können, sollen Laufzeitverzerrungen möglichst gering bleiben.
- Um lange Signalstücke rationell verarbeiten zu können, sollte die Impulsantwort eine bestimmte Länge nicht überschreiten (Anwendung des Overlapp-Add-Verfahrens).
- Um bei festgelegter Bandbreite eine möglichst hohe Zeitauflösung zu erhalten, sollten die Impulsantworten der Bandpaßfilter möglichst kurz sein.
- Das Filterverfahren sollte auf unterschiedliche Signaltypen angewandt werden können (reelle Zeitfunktion, analytisches Zeitsignal, Modulations-Funktionen) und sowohl im Zeit- als auch im Frequenzbereich arbeiten können.

Gewählt wurde ein laufzeitfreies, nichtkausales Filterverfahren im Frequenzbereich (Phase der Übertragungsfunktion=0) mit zum Zeitpunkt $t=0$ symmetrischer Impulsantwort. Die Übertragungsfunktionen werden hierzu so konstruiert, daß sie möglichst keine Unstetigkeiten in Ableitungen niedriger Ordnung enthalten (siehe Anhang D), um so eine möglichst steil abfallende Impulsantwort zu erhalten und Fehler durch das Abschneiden der Impulsantwort auf eine vorgegebene Länge möglichst gering zu halten. Hiermit lassen sich trotz relativ kurzer Impulsantworten steile Filterflanken erzielen.

Folgendes Verfahren wird zur Konstruktion einer optimierten Übertragungsfunktion angewendet (vergl. Slatky [37] bzw. Anhang D):

- Vorgabe der Eckfrequenzen, Zusammensetzen der Roh-Übertragungsfunktion aus einem "Funktions-Baukasten" (Geradenstücke, Cosinus-Funktionen, Exponential-Funktionen, Gauß-Kurven u.a.),
- nötigenfalls Glätten der Roh-Übertragungsfunktion (gleitende Mittelung mit wählbarem Mittelungsfenster),
- Fourier-Transformation zur Bestimmung der Impulsantwort,
- Beschränken der Impulsantwort auf eine vorzugebende Länge mittels Fensterfunktion (wählbar aus "Funktions-Baukasten"),
- Rücktransformation in den Frequenzbereich mit erhöhter Frequenzauflösung und Kontrolle der Ergebnisse.

Eine Filterung erfolgt mit Hilfe eines angepaßten Overlapp-Add-Verfahrens. Die Art der Frequenz-Transformationen ist hierbei abhängig von der Art der gewünschten Signale: reelle FFT zur Erzeugung reeller Zeitsignale, komplexe FFT zur Erzeugung des analytischen Zeitsignals, komplexe FFT des frequenzmäßig verschobenen Signals zur Bestimmung von Modulations-Funktionen oder unterabgetasteten analytischen Zeitsignalen.

Folgende Vorgaben für die Übertragungsfunktionen der Frequenzgruppen-Filter sollten hierbei eingehalten werden:

- Eckfrequenzen und Bandbreiten gemäß Zwicker [52] (Anhang B),
- Filterflanken entsprechend psychoakustisch gemessenen Verdeckungsfunktionen (niederfrequente Flanke 30..100 dB/Okt, hochfrequente Flanke bis zu 300 dB/Okt), simuliert durch entsprechende Exponentialfunktionen,
- kurze Impulsantworten (realisiert mit Längen unter 30 ms).

7.1.2 Erzeugung des analytischen Zeitsignals

Die Cocktail-Party-Prozessor-Verfahren (Kap.5 und 6) bauen auf der Analyse des analytischen Zeitsignals auf:

- Analytisches und reelles Zeitsignal sind zwei Beschreibungsformen für den gleichen Signalzusammenhang (das reelle Zeitsignal ist der Realteil des analytischen Zeitsignals).
- Für abgetastete analytische Zeitsignale sind die Amplituden für Frequenzen $\frac{1}{2}f_{\text{abt}} < f < f_{\text{abt}}$ gleich Null. Datenreduktion bei Bandpaßsignalen ist durch Frequenz-Transformationen und Reduktion der Abtastrate möglich.
- Betrag und Phase harmonischer Schwingungen sind als Betrag und Phase des analytischen Zeitsignals getrennt verfügbar. Eine Signalanalyse, die vor allem die Hüllkurven der Quellsignale auswertet, wird hierdurch vereinfacht (Betragsbildung bzw. konjugiert komplexe Multiplikation zur Phasen- und Träger-Elimination).

Da die Frequenz-Transformierte des analytischen Zeitsignals im Bereich $0 < f < \frac{1}{2}f_{\text{abt}}$ mit dem entsprechenden reellen Zeitsignal übereinstimmt und im Bereich $\frac{1}{2}f_{\text{abt}} < f < f_{\text{abt}}$ zu Null wird, läßt sich durch inverse komplexe Fourier-Transformation des Frequenzbereichs $0 < f < \frac{1}{2}f_{\text{abt}}$ das analytische Zeitsignal leicht aus der Frequenztransformierten des reellen Zeitsignals bestimmen. Das verarbeitete reelle Zeitsignal erhält man dann durch Extraktion des Realteils (bei analytischen Zeitsignalen mit gleicher Abtastrate) oder mit Hilfe von Fourier-Transformations-Verfahren (siehe Kapitel 7.2).

7.1.3 Datenreduktion

Für Bandpaßsignale, die auf den Frequenzbereich $f_{\text{min}}..f_{\text{max}}$ begrenzt sind ($\underline{A}(f)=0$ für $f < f_{\text{min}}, f > f_{\text{max}}$), gilt für das analytische Zeitsignal $\underline{a}(t)$ und die Fourier-Transformierte $\underline{A}(f)$:

$$\underline{a}(t) = \int_{f_{\text{min}}}^{f_{\text{max}}} \underline{A}(f) e^{j2\pi ft} df$$

$$\underline{a}(t) = e^{j2\pi f_{\text{min}} t} \int_0^{f_{\text{max}}-f_{\text{min}}} \underline{G}(f) e^{j2\pi ft} df \quad \text{mit} \quad \underline{G}(f) = \underline{A}(f+f_{\text{min}})$$

$$\underline{a}(t) = e^{j2\pi f_{\text{min}} t} \underline{g}(t) \tag{7.1.3/1}$$

$\underline{a}(t)$ entspricht der Modulation eines komplexen Trägers $e^{j2\pi f_{\text{min}} t}$ mit einem analytischen Zeitsignal $\underline{g}(t)$, das auf den Frequenzbereich $0..f_{\text{max}}-f_{\text{min}}$ beschränkt ist. Ist die Frequenz f_{min} bekannt, kann ohne Informationsverlust die Funktion $\underline{a}(t)$ durch die Funktion $\underline{g}(t)$ beschrieben werden.

Für eine abgetastete Funktion $\underline{a}(it_{\text{abt}})$ ergibt sich folgender Summenausdruck (f_{abt} =Abtastfrequenz, t_{abt} =Abtastperiode, N =Länge der Fourier-Transformation):

$$\underline{a}(it_{\text{abt}}) = \sum_{n=n_{\text{min}}}^{n_{\text{max}}} \underline{A}(nf_0) e^{j2\pi i n/N}$$

Wobei gilt: $n_{\text{max}}f_0 \geq f_{\text{max}}; n_{\text{min}}f_0 \leq f_{\text{min}}; f_0 = f_{\text{abt}}/N;$

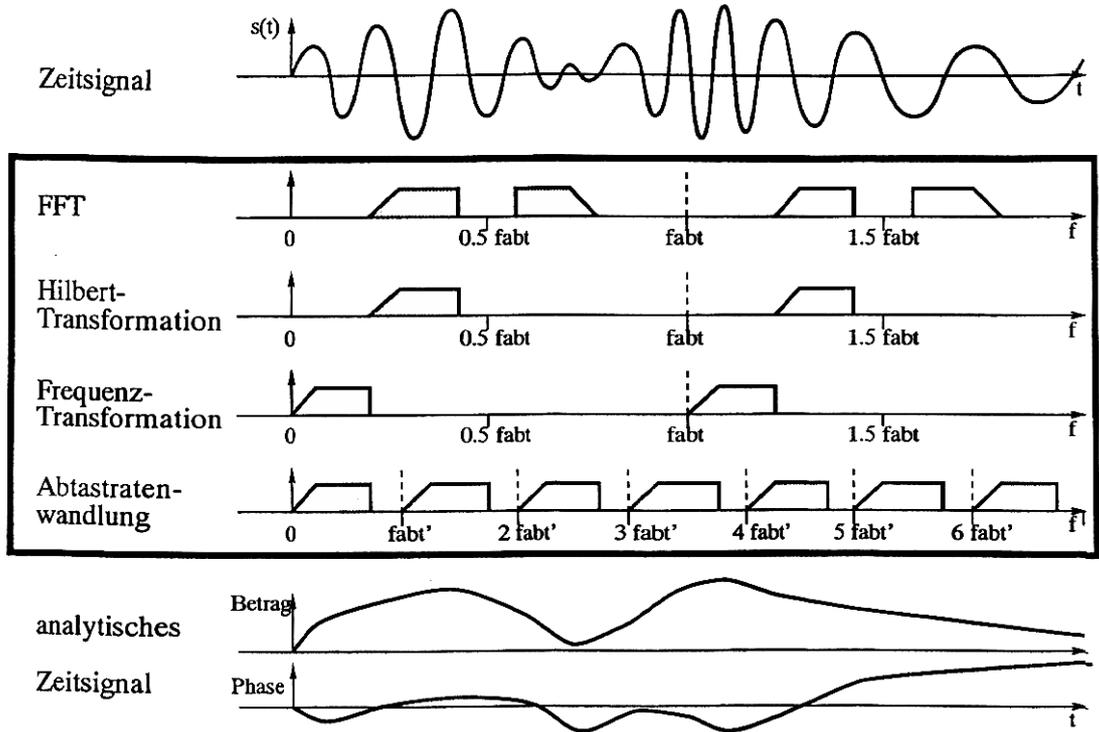


Bild 7.2: Transformation der Eingangsdaten

$$\underline{a}(it_{abt}) = e^{j2\pi n_{min}i/N} \frac{1}{\pi/2} \sum_{n=0}^{n_{max}-n_{min}} \underline{G}(nf_0) e^{j2\pi i n/N} \quad \text{mit} \quad \underline{G}(nf_0) = \underline{\Delta}((n+n_{min})f_0)$$

$$\underline{a}(it_{abt}) = e^{j2\pi n_{min}i/N} \underline{g}(it_{abt}) \quad (7.1.3/2)$$

Zur vollständigen Beschreibung von bandpaßgefilterten Signalen ist somit nur die Kenntnis der niedrigsten Frequenzlinie und der komplexen Modulatorfunktion $\underline{g}(it_{abt})$ erforderlich. Die Modulatorfunktion $\underline{g}(it_{abt})$ erhält man durch Transformation des gefilterten Zeitsignals $\underline{a}(it_{abt})$

$$\underline{g}(it_{abt}) = e^{-j2\pi n_{min}i/N} \underline{a}(it_{abt})$$

Die Funktion $\underline{g}(it_{abt})$ ist auf den Frequenzbereich $0..(n_{max}-n_{min})f_0$ beschränkt. Laut Abtasttheorem ist zur Wahrung des Informationsgehaltes nur eine Abtastung erforderlich mit

$$f_{abt} = 2(n_{max}-n_{min})f_0 = N_g/N f_{abt}$$

$$t_{abt} = N/N_g t_{abt} \quad ; \quad N_g = 2(n_{max}-n_{min})$$

Ist die Länge der Fourier-Transformation N ein ganzzahliges Vielfaches von N_g , läßt sich aus $\underline{g}(it_{abt})$ eine niedrig abgetastete Version $\underline{g}'(kt_{abt})$ durch Verwenden jedes (N_g/N) -ten Abtastwertes erstellen und die Funktion \underline{a} vollständig durch \underline{g}' , n_{min} und t beschreiben.

Bei Verwendung einer Bandpaßfilterbank zur Frequenzgruppen-Filterung würde sich bei Verwendung dieser Methode und Bandpässen mit unendlicher Flankensteilheit die Datenrate durch die Filterung nicht ändern. Bei Verwendung der Frequenzgruppen-Filter aus Anhang D und einem erlaubten Störpegel von -80 dB erhält man etwa die doppelte bis dreifache Datenrate des Originalsignals. Bei konventioneller Digitalfilterung in 24 Frequenzgruppen, einer Abtastfrequenz von 40 kHz und 16 Bit Auflösung (80 kB/s) beträgt die Datenrate ca. 2000 kB/s. Wird hierbei die

Abtastrate der Frequenzgruppen-Signale entsprechend der jeweils höchsten vorkommenden Frequenz reduziert, sinkt die Datenrate auf ca. 600 kB/s. Bei Verwendung des oben vorgestellten Verfahrens sind bei gleichen Filtercharakteristiken Datenraten von ca. 200 kB/s erreichbar.

Im Anschluß an die Verarbeitung des Signals $g'(k t_{abt})$ muß die Abtastrate wiederum gewandelt werden und das Signal durch Multiplikation mit dem Träger $e^{j2\pi n_{min}i/N}$ in den Original-Frequenzbereich transformiert werden. Das verarbeitete Gesamtsignal ergibt sich dann durch Addition der jeweiligen Bandpaßsignale (siehe Bild 7.2).

Im übrigen ähnelt dieses Verfahren zur Datenreduktion durch Elimination des Trägers in der Methodik der Fourier-Transformation. Um Betrag und Phase einer Frequenz f zu erhalten, wird bei der Fourier-Transformation das Signal so in der Frequenz verschoben, daß die Frequenz f zur Frequenz 0 Hz wird (Demodulation). Betrag und Phase der Frequenz ergeben sich dann durch Integration über die so erhaltenen Gleichwerte, wobei Wechselanteile, die von anderen Frequenzen herrühren, herausgemittelt werden.

$$\underline{A}(f) = \int_t a(t) e^{j2\pi ft} dt$$

Mittlung bei $f'=0$ Signal Demodulation: Transformation zu $f'=0$

7.2. Verarbeitung der Ausgangssignale

7.2.1 Anforderungen an ein Resynthese-Verfahren

Aufgabe einer Resynthese-Einheit ist:

- das Erzeugen von verarbeiteten Ohrsignalen aufbauend auf den Schätzern der Cocktail-Party-Prozessoren,
- die Verteilung der verarbeiteten Ohrsignale auf die gewünschte Anzahl von Ausgangskanälen,
- die Kombination der Ohrsignale einzelner Frequenzgruppen zu breitbandigen verarbeiteten Signalen.

Bild 7.3 zeigt das Prinzipschaltbild der Resynthese-Einheit

Folgende Anforderungen sollten hierbei erfüllt werden:

- Das Verfahren soll möglichst schnell sein.
- Die Signale sollen möglichst wenig verzerrt werden, insbesondere sollen keine nichtlinearen Verzerrungen auftreten.
- Binaurale Informationen der Eingangssignale sollen erhalten bleiben.
- Je nach gewünschter Fortführung der Analyse sollen die resultierenden Signale als reelles Zeitsignal, analytisches Zeitsignal oder Frequenz-Transformierte vorliegen können.

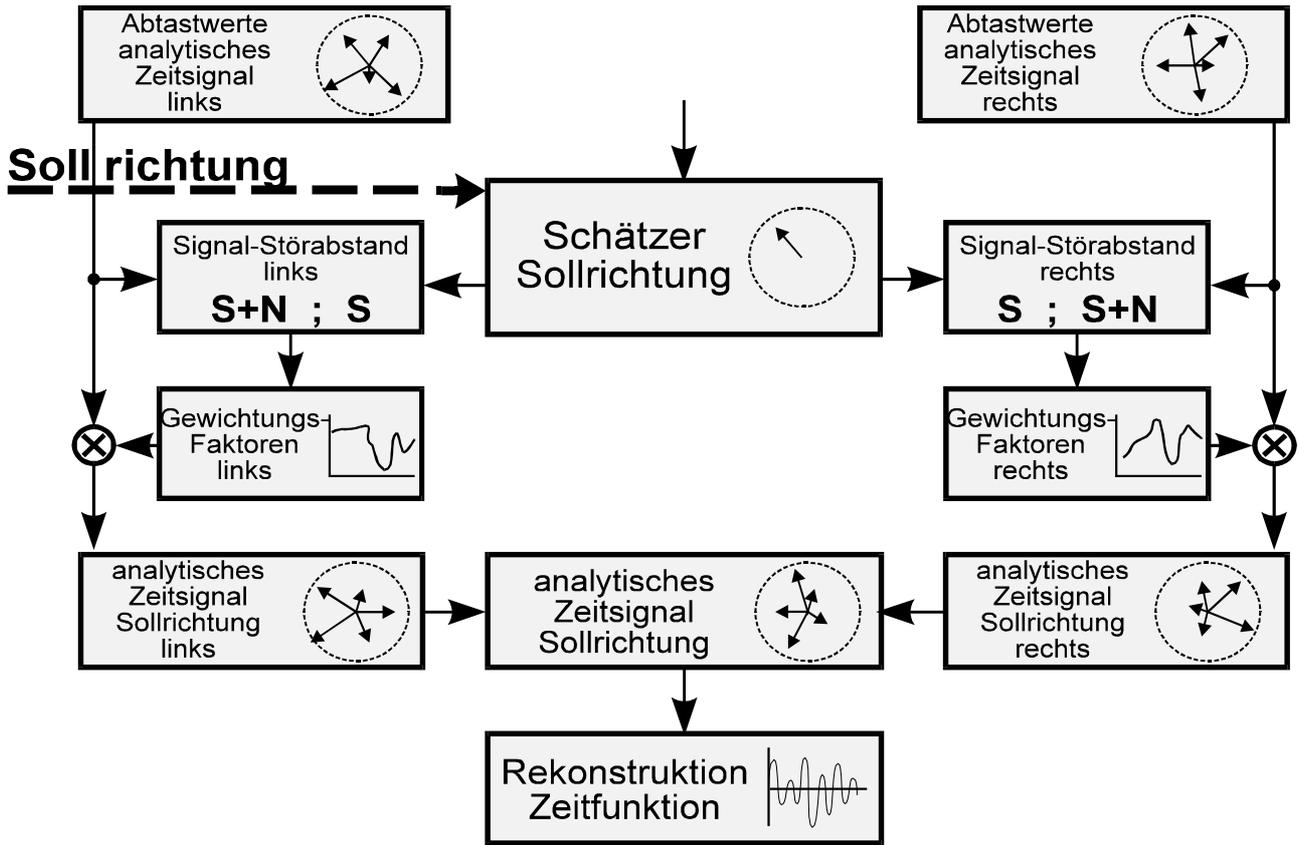


Bild 7.3: Erzeugen von Zeitsignalen aus Schätzern eines Cocktail-Party-Prozessors

7.2.2 Anpassung der Eingangssignale an Signalschätzer

Erzeugen von Ohrsignal-Schätzern

Ergebnis der Cocktail-Party-Prozessoren sind Schätzer eines auf "Kopfmittle" (Kap. 4.1) bezogenen Signals der Sollrichtung. Die benötigten Schätzer für die Ohrsignale können hieraus mit Hilfe der entsprechenden Freifeld-Außenohr-Übertragungsfunktionen bestimmt werden. Diese Ohrsignal-Schätzer $a_l'(t), a_r'(t)$ beschreiben jeweils für ein bestimmtes Zeitfenster die Amplitude des Sollrichtungs-Signals in den Ohren.

Das Resynthese-Verfahren soll die analytischen Zeitsignale der Ohrsignale, die am Eingang des Modells vorliegen, so verändern, daß innerhalb des Zeitfensters $2T_{\mu}$, in dem die Ohrsignal-Schätzer gültig sind, die Leistung der Ohrsignale der geschätzten Leistung der Sollrichtung entspricht.

Erzeugen von Gewichtungsfaktoren

Hierzu werden für jedes Zeitfenster Gewichtungsfaktoren $g_l(t), g_r(t)$ bestimmt, die dem Quotienten aus geschätzter und vorhandener Ohrsignal-Leistung entsprechen. Durch Multiplikation der Eingangssignale mit den Gewichtungsfaktoren sollen diese dann an die Ohrsignal-Schätzer angepaßt werden (Bild 7.4). Aufbau und Eigenschaften dieses Verfahrens entsprechen dem von Bodden/ Gaik [9] vorgestellten Wiener-Filter-Verfahren.

$$g_l(t)^2 = \frac{a_l'(t)^2}{\int_{T_G} |l(t_G)|^2 dt_G}$$

$$g_r(t)^2 = \frac{a_r'(t)^2}{\int_{T_G} |r(t_G)|^2 dt_G}$$

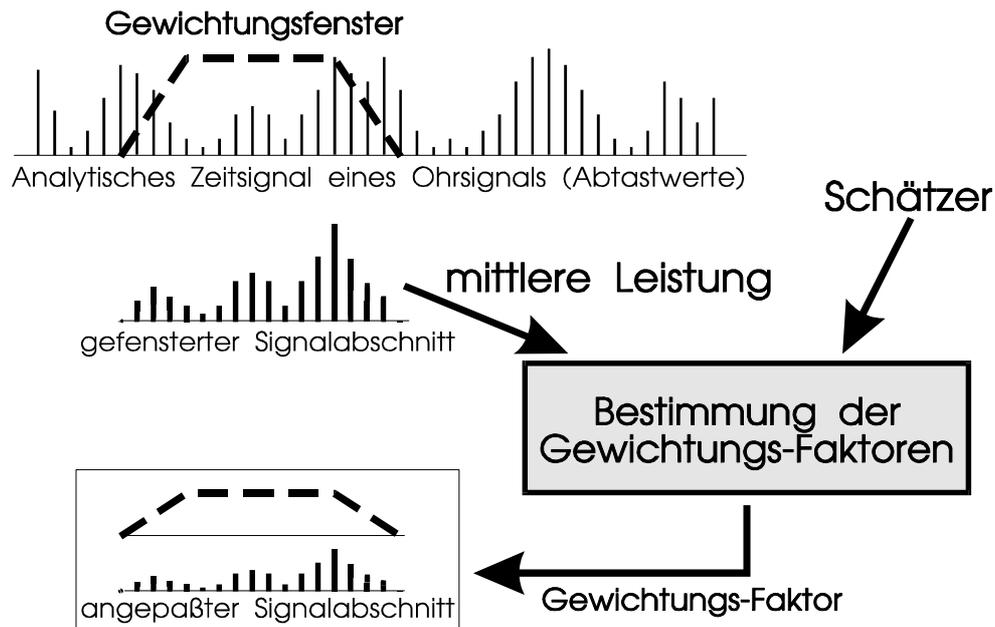


Bild 7.4: Bestimmung von Gewichtungsfaktoren aus der mittleren Leistung eines Signalabschnitts und den Schätzwerten

Für diese zeitabhängigen Gewichtungsfaktoren gelten folgende Anforderungen:

- Sie dürfen nicht größer als 1 werden. Eine geschätzte Signalleistung, die größer als die Leistung der Ohrsignale ist, deutet auf Schätzfehler hin.
- Zur Vermeidung von nichtlinearen Verzerrungen muß die Änderungsgeschwindigkeit der Gewichtungsfaktoren begrenzt sein. Bei bandbegrenzten Signalen darf die maximale Änderungsgeschwindigkeit der Hüllkurve die maximale Steigung einer Schwingung der Bandbreiten-Frequenz nicht überschreiten.
- Um innerhalb der Frequenzgruppe die Signale vollständig zu übertragen, muß das Zeitfenster zur Erzeugung der Gewichtungsfaktoren größer als die Periode der kleinsten in dieser Frequenzgruppe enthaltenen Frequenz sein.

Für die Gewichtungsfaktoren ist also eine Nachbearbeitung erforderlich: Begrenzen auf eine maximale Größe von 1 und Mittelung über mehrere aufeinanderfolgende Schätzintervalle. Damit einmalige Schätzfehler sich nicht auf die verarbeiteten Daten auswirken, sollten die Schätzungen und Gewichtungen mit überlappendem Zeitfenster erfolgen.

Das Verfahren wird wie folgt angewandt (Bild 7.4): Jedem Schätzer wird eine trapez- bzw. dreieckförmige Fensterfunktion zugewiesen mit dem geschätzten Gewichtungsfaktor als Maximalwert. Die Länge des Trapez-Fensters (Meßpunkt: 50% des Maximums) entspricht dem Schätzintervall $2T_{\mu}$. Anstieg und Abfall sind so bemessen, daß das Überblenden zu den vorhergehenden bzw. folgenden Schätzern ausreichend sanft erfolgt. Nach Möglichkeit werden die Schätzungen mit überlappenden Zeitintervallen durchgeführt. Es entsteht so eine Gewichtungsfunktion aus der Überlagerung und Mittelung der Gewichtungsfenster.

Die resultierenden Gewichtungsfaktoren passen sich einerseits schnellen Signaländerungen gut an (max. Verzögerung = 2 Fensterlängen), beschreiben aber andererseits durch eine hohe Zahl von separaten Berechnungen die Signale sehr sicher (bei 90% Überlappung Mittelung der Gewichtungsfaktoren aus 10 Schätzungen) und vermeiden zusätzlich Störungen der Signale (knackfreies Überblenden durch den sanften Anstieg der Trapezflanken).

Durch Multiplikation mit diesen Gewichtungsfaktoren werden die Eingangssignale an die Schätzwerte angepaßt.

Eigenschaften des Resynthese-Verfahrens

Dieses Verfahren zur Anpassung von gefilterten Eingangsdaten entspricht im Prinzip einer Weiterentwicklung eines Verfahrens, das von Bodden/Gaik [9] zur Unterdrückung von Störsprechern bei bekannten Signal/Stör-Abständen angewendet wurde.

Bei Verwendung von Datenreduktionsverfahren (Kap. 7.1.3) liegen die gewichteten Signale mit verringerter Abtastrate vor. Durch Hochtransformieren im Frequenzbereich (zeitsynchrone komplexe Multiplikation mit dem eingangs entfernten Trägersignal) können hieraus Signale der ursprünglichen Abtastrate hergestellt werden, aus denen die breitbandigen Zeitsignale durch Zusammenmischen der Frequenzgruppen-Signale erzeugt werden können.

Dieses Resynthese-Verfahren hat folgende Eigenschaften:

- Durch die Verwendung der Eingangssignale bleibt die Feinstruktur der Ohrsignale erhalten.
- Die Verarbeitung von analytischen Zeitsignalen erlaubt eine relativ schnelle Synthese bei Verarbeitung datenreduzierter Signale.
- Der eigentliche Signalverarbeitungsschritt ist das Einprägen von geschätzten Signalleistungen auf die Eingangsdaten (Wiener-Filter-Verfahren, s.o.). Dieses Verfahren ist sehr flexibel. Bei Anwendung entsprechender Schätzverfahren könnten auch Phasen- oder Frequenzschätzer den analytischen Zeitsignalen durch eine einfache komplexe Multiplikation eingeprägt werden.
- Durch Begrenzung der Änderungsgeschwindigkeit der Gewichtungsfaktoren und Wahl eines ausreichend langen Gewichtungs-Zeitraums können nichtlineare Verzerrungen vermieden und der Bandbereich der Frequenzgruppe vollständig übertragen werden.
- Durch die relativ langsame Modulation der Eingangssignale mit den Gewichtungsfaktoren bleiben die Feinstruktur der Signale, deren spektrale Zusammensetzung und die Signalphasen erhalten. Dies ist besonders bei binauralen bzw. Sprachsignalen wichtig.

Bei diesem Verfahren wird das ansonsten unveränderte Signalgemisch mit der Leistung eines erwünschten Signals übertragen. Das Vorgehen ähnelt der Vocoder-Technik, wo zur verständlichen Übertragung von Sprache in Frequenzgruppen nur die Übertragung der Hüllkurven erforderlich ist, während die Trägersignale nur in der Grundstruktur (tonales Signal mit entsprechender Grundfrequenz/Rauschen) übereinstimmen müssen. Werden, wie hier, die unverarbeiteten Signale als Trägersignale gewählt, stimmen bei dominanten Sollquellen (positives Signal/Stör-Verhältnis) die übertragenen Träger mit dem erwünschten Signal überein. Bei relativ leisem Nutzsignal und negativem Signal/Stör-Verhältnis wird das Trägersignal konkurrierender Quellen verwendet, wodurch weniger die Sprachverständlichkeit leidet als der Klang des Signals.

Durch die unveränderte Übertragung der Signal-Feinstruktur läßt sich das Cocktail-Party-Prozessor-System als Vorverarbeitungsstufe für weitere Analyseverfahren einsetzen. So eignet sich das System im Prinzip für die Vorverarbeitung von Mikrofon-Array-Signalen. Da die Signalphasen erhalten bleiben und bei der obigen Frequenzgruppen-Filterung keine Laufzeitverzerrungen auftreten, können die Ausgangssignale des Systems mit Array-Techniken weiter verarbeitet werden und u.U. ein zusätzlicher Störabstands-Gewinn für die Nutzrichtung erzielt werden. Cocktail-Party-Prozessor-Technik und Array-Technik würden sich hier sinnvoll ergänzen, da im

Niederfrequenzbereich, wo der Phasendifferenz-Prozessor allein zu guten Ergebnissen führt, lineare Arrays nur bei sehr großen Abmessungen eine gute Richtwirkung erreichen. Ebenso wäre auch die Möglichkeit gegeben, Kunstkopf-Aufnahmen durch die Cocktail-Party-Prozessoren vorzuverarbeiten, um so auch bei binauraler Beschallung einen zusätzlichen Störabstands-Gewinn zu erzielen und die binaurale Übertragung auf die Signale bestimmter Richtungen zu beschränken.

7.2.3 Verringern der Anzahl der Ausgangskanäle

Werden weniger Ausgangskanäle als Eingangskanäle benötigt, kann nach folgenden Verfahren vorgegangen werden

- *Zusammenfassen von Ausgangskanälen:*
Hierzu müssen die Signale der Ausgangskanäle invers mit den jeweiligen Freifeld-Außenohr-Übertragungsfunktionen gefiltert werden und so Zeitverzögerungen und Pegelunterschiede für die Sollrichtung kompensiert werden. Durch die Aufsummierung ohne Phasen- und Pegelunterschied werden die Signale der Nutzrichtung verstärkt und Signale anderer Richtungen geringer verstärkt oder abgeschwächt. Dieser zusätzliche Gewinn für die Nutzrichtung ist bei Vorliegen interauraler Pegeldifferenzen nicht allzu groß, so daß dieses Verfahren eher bei Stereo-Mikrofon-Anordnungen bzw. Mikrofon-Arrays von Vorteil ist. Erforderlich ist hierzu eine relativ genaue Schätzung der Einfallrichtung, da ansonsten besonders bei höheren Frequenzen das Nutzsignal sogar abgeschwächt werden könnte (Kammfilter-Effekt).
- *Übertragen von Kanälen mit maximalem Störabstand*
Die Gewichtungsfaktoren der Kanäle geben die geschätzten Signal/Stör-Abstände wieder. Hieraus können die Kanäle mit maximalem Signal-Stör-Abstand ausgewählt werden. Da in unterschiedlichen Frequenzgruppen das Signal/Stör-Verhältnis für unterschiedliche Eingangskanäle maximal werden kann, müssen auch hier die interauralen Differenzen kompensiert werden, bevor die Kanäle unterschiedlicher Frequenzgruppen zu einem Signal zusammengemischt werden. Dieses Verfahren ist sehr robust; Fehlschätzer für die Einfallrichtung und die Signalleistungen wirken sich nur wenig auf das Nutzsignal aus.

7.2.4 Bildung der Zeitfunktion aus Abtastwerten des analytischen Zeitsignals

In den einzelnen Frequenzgruppen liegen somit analytische Zeitsignale mit verminderter Datenrate vor, die an die Leistung des Nutzsignals angepaßt worden sind. Ist eine Phasen- oder Zeitkorrektur vorgenommen worden, so muß dieses bei den folgenden Schritten durch die Korrektur der Transformationszeit berücksichtigt werden.

Das reelle Zeitsignal kann mit einem inversen Verfahren zu dem aus Bild 7.2 erzeugt werden: Bei datenreduzierten Frequenzgruppen-Signalen $g(i t_{abt})$ liegen nur komplexe Hüllkurven-Informationen im Frequenzbereich $0..(f_{max}-f_{min})$ vor. Bei der Transformation in den Frequenzbereich entsteht ein periodisches Spektrum mit der Periode $1/(f_{max}-f_{min})$. Die Datenrate wird erhöht durch Definition einer neuen Frequenzfunktion, die im Bereich $0..(f_{max}-f_{min})$ mit der obigen übereinstimmt und im Bereich $(f_{max}-f_{min})..f_{abt}$ zu Null wird. Durch Verschieben des Frequenzbereichs $0..(f_{max}-f_{min})$ um f_{min} werden die Daten in den Original-Frequenzbereich transformiert, wodurch das Spektrum des analytischen Zeitsignals entsteht.

Die reelle Zeitfunktion läßt sich durch reelle inverse Fourier-Transformation aus dem Spektrum des analytischen Zeitsignals bestimmen (Spiegelung des Spektrums an $1/2f_{abt}$). Die reellen Zeitsignale der einzelnen Frequenzgruppen lassen sich durch einfache Addition zu einem breitbandigen verarbeiteten Signal zusammenfassen.

7.3 Gesamtdarstellung des Cocktail-Party-Prozessor-Modells

7.3.1 Modellstruktur

In **Bild 7.5** zeigt für eine Frequenzgruppe eine Übersicht über die Verarbeitungsstufen des hier vorgestellten Cocktail-Party-Prozessor-Modells (Gesamtaufbau des Modells in Bild 7.1). Der Kern

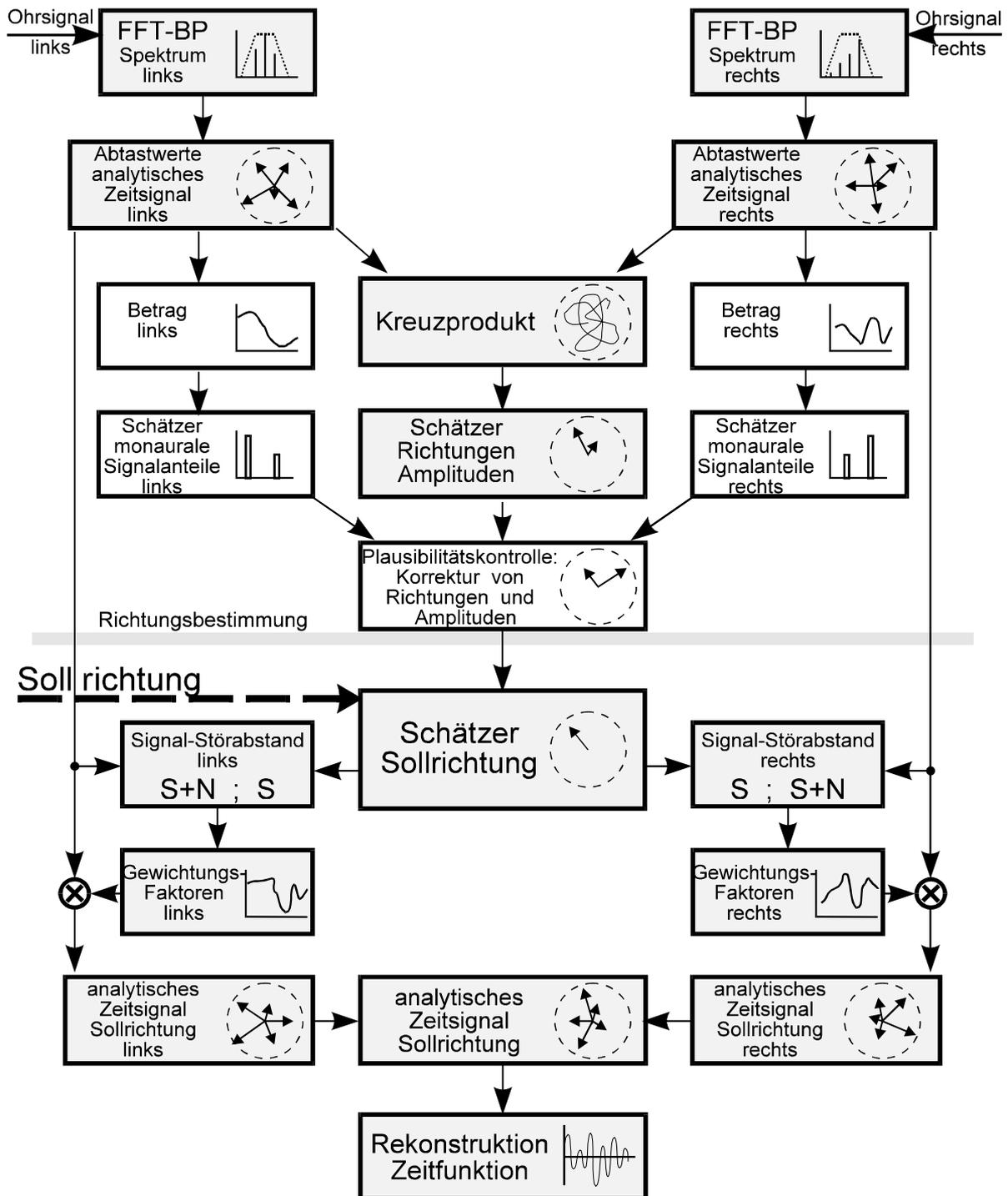


Bild 7.5: Gesamtaufbau des binauralen Prozessors.
 grau unterlegt: Module des Phasendifferenz-Prozessors
 weiß: zusätzliche Module für den Pegeldifferenz-Prozessor

des Cocktail-Party-Prozessors besteht aus dem Phasendifferenz-Cocktail-Party-Prozessor (Kap.5) mit der Analyse des interauralen Kreuzprodukts und dem Pegeldifferenz-Cocktail-Party-Prozessor (Kap.6), der die Beträge der Ohrsignale analysiert. Die Kombination der Schätzer erfolgt in einer verbindenden Stufe gemäß Kapitel 6.4. Das Verfahren ist zwar in der Lage, Einfallsrichtungen von Schallquellen zu schätzen (Siehe Bild 5.7), es muß aber von außen bzw. von entsprechenden Modellstufen (Kap. 8) vorgegeben werden, welche der möglichen Richtungen als Sollrichtung der Verarbeitung dienen soll. Ist eine Sollrichtung vorgegeben, können die Schätzer auf diese Richtung abgebildet und korrigiert werden (Kap. 5.6 und 6.3). Die Vor- und Nachverarbeitung der Signale erfolgt analog zu diesem Kapitel.

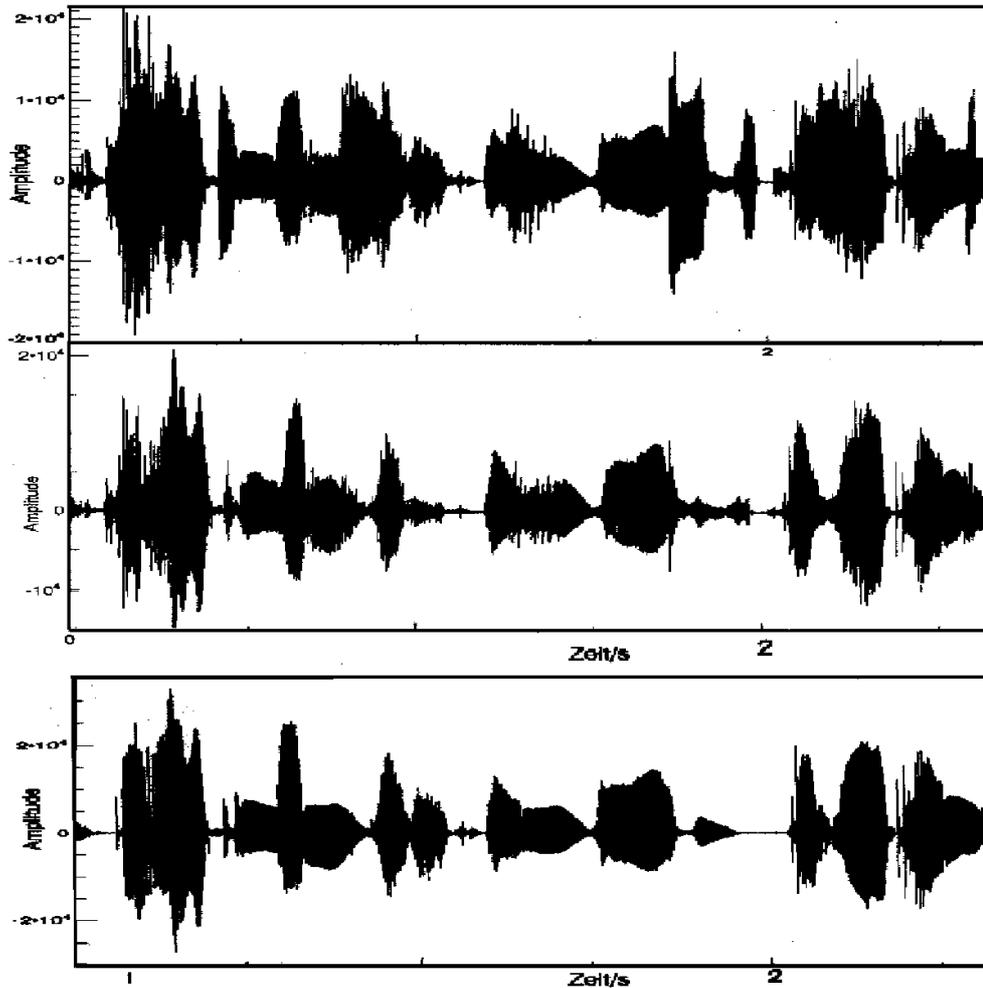
7.3.2 Leistungsfähigkeit der Cocktail-Party-Prozessoren

Testbedingungen

Die Möglichkeiten des Modells sollen an einem Beispiel demonstriert werden. Hierzu wurden im Freifeld (reflexionsarmer Raum) Sprachsignale eines Sprechers und einer Sprecherin aufgenommen (2.7 s fortlaufender Text) und unter Verwendung vereinfachter Freifeld-Außenohr-Übertragungsfunktionen (Anhang C) mit unterschiedlichen interauralen Parametern zu Ohrsignalen zusammengesetzt: In dieser Simulation wurden die Signale der Sprecherin ohne interaurale Differenzen dargeboten, die Signale des Sprechers von rechts mit einer interauralen Laufzeitdifferenz von 400 μ s und den dazugehörigen interauralen Pegeldifferenzen (nach Anhang C). Sollrichtung war die Richtung der Sprecherin. Der Test umfaßte 3 unterschiedliche Signal-Stör-Abstände: 0 dB, -10 dB und -20 dB. Die Angaben der Signal-Stör-Abstände beziehen sich jeweils auf die Gesamtenergie der 2.7 Sekunden langen Sprachsignale.

Am Beispiel des Phasendifferenz-Cocktail-Party-Prozessor soll die Leistungsfähigkeit der entwickelten Algorithmen dargestellt werden. Folgende Systemparameter wurden für den Test benutzt:

- Leichte Datenreduktion bei der Erzeugung der analytischen Zeitsignale der Ohrsignale (ein Abtastwert pro Periode)
- Zeitkonstante für die Bestimmung der statistischen Parameter des interauralen Kreuzprodukts: 20 ms,
- Bestimmungen neuer Schätzer: einmal pro 1 ms,
- Mittelungszeitraum der Schätzer: 20 ms,
- Breite der Richtkeule (Fangbereich): $\pm 0.1\pi$ (entspricht $\pm 18^\circ$),
- Korrekturmethode für von der Sollrichtung abweichende Schätzer: "Gültiger Schätzerbereich" (Kap.5.6),
- Fenster zum Anpassen der analytischen Zeitsignale der Ohrsignale an Gewichtungsfaktoren: Dreieck,
- Methode zur Reduktion der Ausgangskanäle: Auswahl des Kanals mit größtem Signal-Stör-Abstand (hier: linkes Ohrsignal).



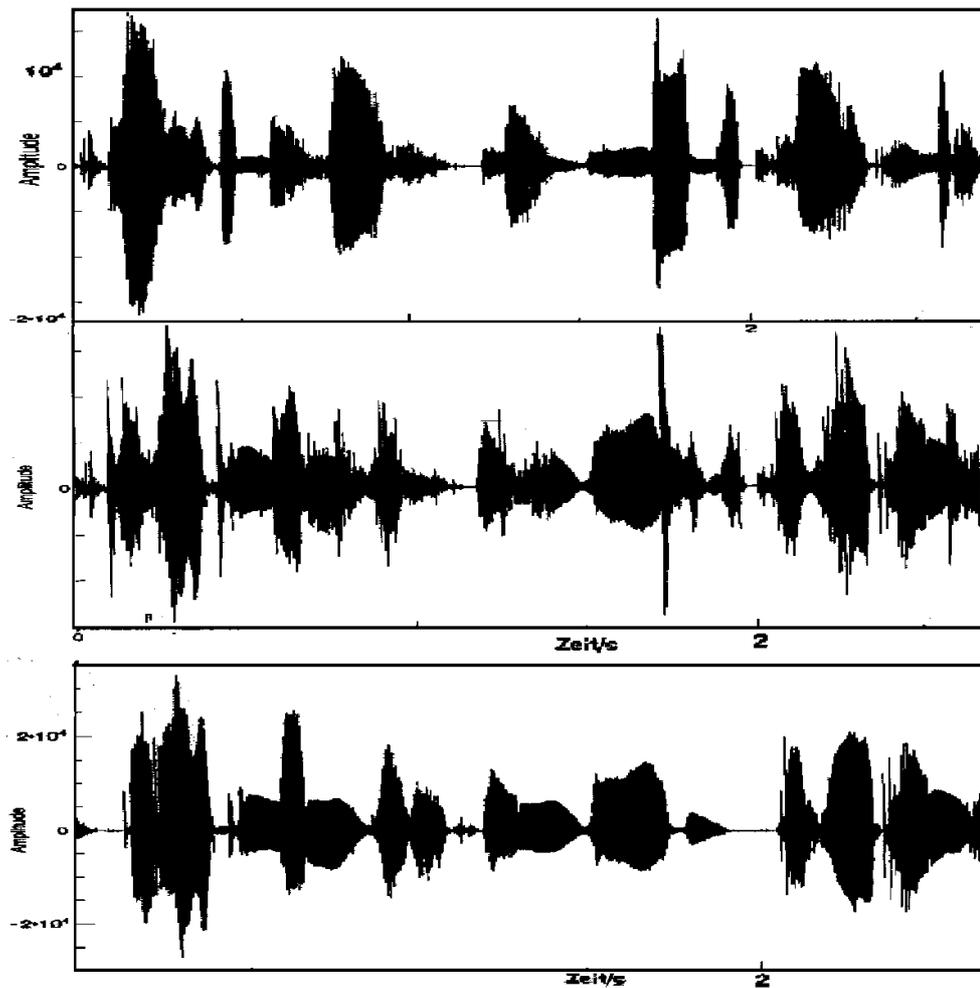
*Bild 7.6: Verarbeitungs-Ergebnisse des Phasendifferenz-Cocktail-Party-Prozessors bei 2 Sprechern im Freifeld; 2.7 s fortlaufender Text
interaurale Laufzeit der Richtung des Störsprechers: 400 μ s
interaurale Laufzeit der Richtung der Sprecherin (Nutzsignal): 0 μ s
Signal-Stör-Abstand des Nutzsignals: 0 dB
oben: linkes Ohrsignal
Mitte: verarbeitetes Signal
unten: ungestörtes Nutzsignal*

Ergebnisse

In Bild 7.6, Bild 7.7 und Bild 7.8 werden die Ergebnisse der richtungsselektiven Verarbeitung (verarbeitete Signale) verglichen mit dem linken Ohrsignal, aus dem die Resynthese-Einheit das verarbeitete Signal erzeugt, und dem unbeeinflussten Nutzsignal. Sollrichtung war die Richtung der Sprecherin ($\tau=0$).

In Bild 7.6 sind die Ergebnisse der richtungsselektiven Verarbeitung bei gleichstarkem Nutz- und Störsignal dargestellt. Während in den Ohrsignalen starke Einflüsse des Störsprechers erkennbar sind (oberes Bild), sind diese Einflüsse im verarbeiteten Signal fast vollständig eliminiert (mittleres Bild). Das verarbeitete Signal entspricht nahezu dem Nutzsignal (unteres Bild).

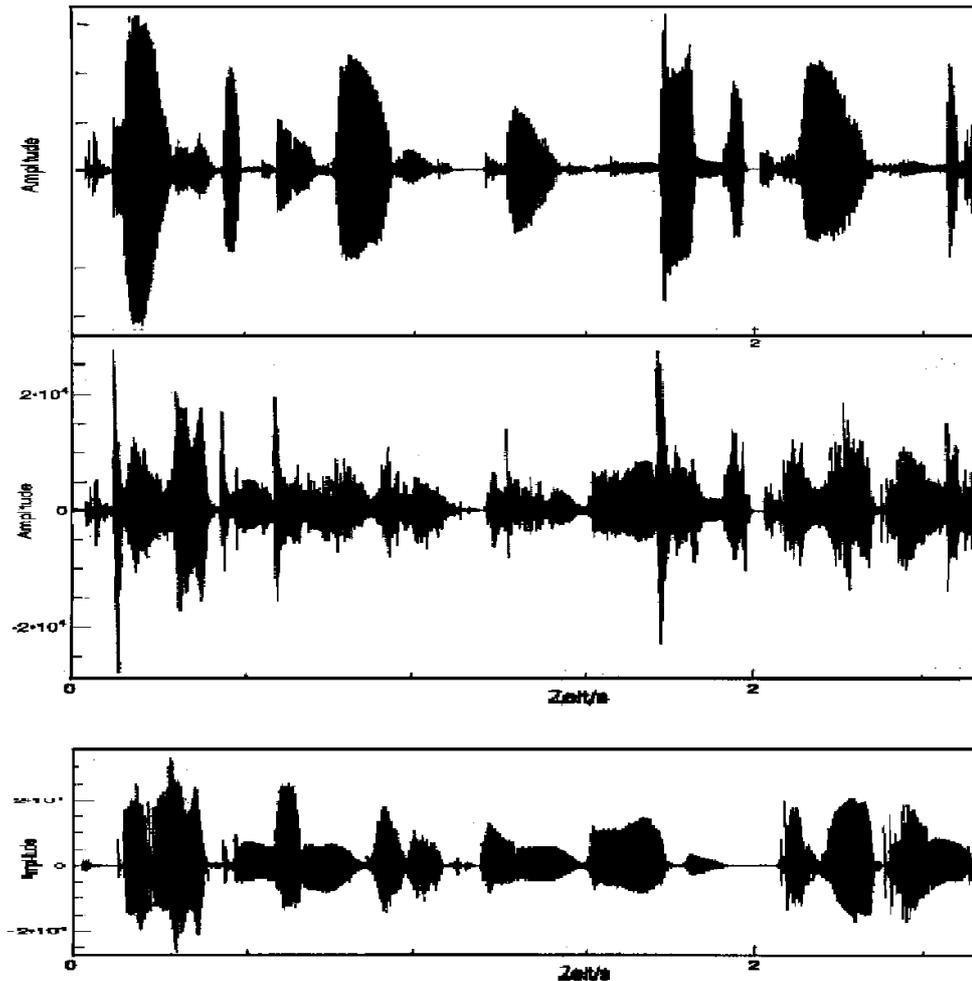
Das verarbeitete Nutzsignal weist bei akustischer Präsentation eine recht hohe Qualität auf. Die Lautstärke des Störsprechers ist stark vermindert.



*Bild 7.7: Verarbeitungs-Ergebnisse des Phasendifferenz-Cocktail-Party-Prozessors bei 2 Sprechern im Freifeld; Signal-Stör-Abstand des Nutzsignals: -10 dB sonstige Bedingungen: siehe Bild 7.6 und Text
oben: linkes Ohrsignal
Mitte: verarbeitetes Signal
unten: ungestörtes Nutzsignal*

Wird der Signal-Stör-Abstand des Nutzsignals negativ, wie in der in [Bild 7.7](#) dargestellten Situation, überwiegt in den Ohrsignalen das Signal der Störquelle (oberes Bild). Durch die richtungsselektive Verarbeitung im Cocktail-Party-Prozessor kann das Störsignal mit 10 dB höherem Pegel noch ausgeblendet werden. Wie ein Vergleich zwischen verarbeitetem Ohrsignal (mittleres Bild) und dem entsprechenden Ohrsignal bei ungestörtem Nutzsignal (unteres Bild) zeigt, entspricht der Verlauf des verarbeiteten Signals im wesentlichen dem Verlauf des Nutzsignals, wenn auch an einigen Stellen schon Fehler auftauchen.

Die akustische Qualität der verarbeiteten Signale ist noch recht hoch, die Sprecherin aus der gewünschten Richtung wird lauter wahrgenommen als der Störsprecher.



*Bild 7.8: Verarbeitungs-Ergebnisse des Phasendifferenz-Cocktail-Party-Prozessors bei 2 Sprechern im Freifeld; Signal-Stör-Abstand des Nutzsignals: -20 dB sonstige Bedingungen: siehe Bild 7.6 und Text
oben: linkes Ohrsignal
Mitte: verarbeitetes Signal
unten: ungestörtes Nutzsignal*

Bei sehr niedrigen Signal-Störabständen von -20 dB, wie in Bild 7.8, sind die Anteile des Nutzsignals in den Ohrsignalen kaum noch erkennbar (oberes Bild). Der Cocktail-Party-Prozessor schwächt aber dennoch das Störsignal so weit ab, daß im verarbeiteten Signal (mittleres Bild) die Struktur des Nutzsignals (unteres Bild) wieder erkennbar wird. Die Struktur des verarbeiteten Signals ähnelt ein wenig der Struktur des linken Ohrsignals bei einem Signal-Stör-Abstand von 0 dB.

Während bei akustischer Präsentation des unverarbeiteten Ohrsignals die Nutzquelle als relativ leise erschien, erscheint sie in den verarbeiteten Signalen mit nahezu der gleichen Lautstärke wie die Störquelle.

Für noch geringere Signal-Störabstände überwiegt auch im verarbeiteten Signal die Störquelle. Während bei Signal-Störabständen von -30 dB die Nutzquelle bei akustischer Präsentation der Ohrsignale nicht mehr erkennbar ist, wird durch die Verarbeitung im Cocktail-Party-Prozessor das Signal der Nutzquelle wieder wahrnehmbar.

Bewertung

Die geschilderten Beispiele zeigen, daß die verwendeten Algorithmen auch unter relativ ungünstigen Bedingungen in der Lage sind, Verbesserungen des Signal-Stör-Abstands für eine Nutzrichtung zu erzielen.

Der Grund für die Möglichkeit, auch stark negative Signal-Stör-Abstände zu verarbeiten, liegt vor allem darin begründet, daß die Algorithmen auf einem 2-Quellen-Modell aufbauen, alle Schallfeld-Parameter also als Ergebnis der Überlagerung zweier Schallquellen bzw. Frequenzlinien interpretiert werden. Somit können leichte Schwankungen von Schallfeld-Parametern einer dominierenden Quelle, die durch den Einfluß einer schwachen Quelle hervorgerufen werden, noch ausgewertet werden.

Zur Verarbeitung dieser ca. 2.7 Sekunden dauernden Signale in 24 Frequenzgruppen benötigte ein Vektorrechner (Stardent Titan) in einer nicht rechenzeit-optimierten Testversion ca. 7 Minuten Rechenzeit (Echtzeitfaktor \approx 150), wobei die Verarbeitung der Signale in den unteren 5 Frequenzgruppen jeweils in Echtzeit erfolgte. Durch Optimierung des Verfahrens (Herausnahme von Testroutinen und grafischen Ausgaben, Funktionstabellen, externe Frequenzgruppen-Filterung, stärkere Datenreduktion besonders in den oberen Frequenzgruppen) wären Echtzeit-Realisierungen, zumindest für jede Frequenzgruppe, möglich. Allein die Anwendung der in Kapitel 7.1.3 diskutierten Möglichkeiten zur Datenreduktion innerhalb von Frequenzgruppen (Abtastung der komplexen Modulationsfunktion statt des Originalsignals) kann ohne Informationsverlust eine Rechenzeiterparnis um den Faktor 3 ergeben. Das Modell ist so konzipiert, daß jede Frequenzgruppe unabhängig verarbeitet werden kann und in jeder Frequenzgruppe die Verarbeitung auf maximal 14 unabhängige Prozesse, die auch auf unabhängigen Prozessoren ablaufen können, aufgeteilt werden kann (siehe Anhang F). Somit ist, zumindest bei entsprechendem Hardware-Aufwand, ein Echtzeit-Ausbau des Modells mit Hilfe von Signalprozessor- oder Transputer-Boards bei einem rechenzeitoptimierten Modellaufbau durchaus möglich.

Die Sollrichtung der Cocktail-Party-Prozessoren muß bisher noch von außen vorgegeben werden. Eine Weiterentwicklung des Modells könnte darin bestehen, Verfahren einzubinden, die die Sollrichtung der Cocktail-Party-Prozessoren automatisch auswählen. Im folgenden Kapitel sollen deshalb Möglichkeiten diskutiert werden, eine Prozessorstufe zur Richtungssteuerung der Prozessoren, angelehnt an die Richtungssteuerung des menschlichen Gehörs, zu konstruieren.

8. Steuerung des Cocktail-Party-Prozessors

Grundlage für die Steuerung der Sollrichtung der Cocktail-Party-Prozessoren sollen die Fähigkeiten des menschlichen Gehörs zur Wahrnehmung von Richtungen sein. Von besonderem Interesse sind in diesem Zusammenhang Richtungs-Wahrnehmungen in komplexen Schallfeldern und dynamische Effekte. Die Auswahlkriterien für die Aufmerksamkeits-Richtung des Gehörs können so als Vorbild für eine Richtungs-Steuerungseinheit für Cocktail-Party-Prozessoren dienen, nach der die Sollrichtung automatisch gesucht und weiterverfolgt wird.

8.1. Detektionskriterien für Richtungsinformationen:

Das Hören in Räumen (Franssen-Effekt)

In geschlossenen Räumen stellen die Lokalisation und das richtungsselektive Verarbeiten von Schallsignalen ein besonderes Problem für das binaurale System des Menschen dar. Befindet sich der Hörer genügend weit außerhalb des Hallradius aller Schallquellen, so ist gemäß den Versuchen von Franssen [16] das binaurale System des Menschen nicht mehr in der Lage, stationäre Signale zu lokalisieren und richtungsmäßig zu selektieren (Franssen-Effekt):

In einem Raum (Hörsaal) sind 2 Lautsprecher an unterschiedlichen Positionen aufgestellt. Zu Beginn der Darbietung strahlt Lautsprecher 1 ein Rauschsignal mit steil ansteigender Hüllkurve ab, das dann in der Leistung konstant bleibt. Die Zuhörer können diesen Lautsprecher lokalisieren. Im stationären Teil der Hüllkurve wird das Signal sehr sanft von Lautsprecher 1 zu Lautsprecher 2 überblendet. Obwohl Lautsprecher 2 nun den Schall abstrahlt, verbleibt für die Zuhörer das Hörereignis an der Position von Lautsprecher 1. An dieser (Fehl-)Lokalisation ändert sich selbst dann nichts, wenn vom Versuchsleiter demonstrativ die Zuleitungskabel zu Lautsprecher 1 unterbrochen werden.

Für die Signalverarbeitung des menschlichen Gehörs in reflexionsbehafteter Umgebung mit Quellen außerhalb des Hallradius bedeutet dies:

- Das menschliche Gehör ist hier nicht in der Lage, stationäre Signale zu lokalisieren (sonst wäre der Lautsprecher 2 lokalisiert worden).
- Das menschliche Gehör ist hier nicht in der Lage, stationäre Signale richtungsselektiv zu verarbeiten (sonst wäre nach dem Überblenden die Lautstärke des Hörereignisses gering geworden).
- Das menschliche Gehör ist aber sehr wohl in der Lage, während schneller Signaländerungen oder bei Signaleinsätzen die zugehörige Schallquelle zu lokalisieren (korrekte Lokalisation von Lautsprecher 1 zu Beginn des Versuchs).

In komplexen Schallfeldern wie in großen geschlossenen Räumen stehen also nur zu wenigen Zeitpunkten Richtungsinformationen zur Verfügung, die vom Gehör auswertbar sind. Ist die Leistung der Reflexionen in allen Teilbändern größer als die Leistung des Direktschalls einer Schallquelle, kann das Gehör weder die Richtung und die Signaleigenschaften einer Quelle bestimmen, noch den Direktschall dieser Quelle selektiv verarbeiten.

Das Gehör ist zur Lokalisation und richtungsmäßigen Verarbeitung also darauf angewiesen, daß zumindest in kurzen Zeitabschnitten der Direktschall einer Quelle überwiegt und auswertbar ist. Möglicherweise müssen diese Zeiträume auch dazu genutzt werden, Signal-Störabstände, bzw. die Leistung des verbleibenden reflexionsbehafteten Störfeldes zu bestimmen. Solche Zeiträume sind

durch ansteigende Flanken der Hüllkurve - zumindest in einzelnen Teilbändern - gekennzeichnet. Wolf [49] hat diese Eigenschaften ausgenutzt, um ein Lokalisationssystem für Schallquellen in geschlossenen Räumen aufzubauen.

Überwiegt der Direktschall, wird die Varianz (bezogen auf die Verschiebung) der interauralen Kreuzkorrelationsfunktion klein. Die Signale werden lokalisierbar. Überwiegt der Nachhall, steigt diese Varianz. Hierauf aufbauend, wurde von Allen/Berkley/Blauert [1] ein Enthüllungssystem aufgebaut, das über varianz-gesteuerte Gewichtungsfaktoren Signalabschnitte abschwächt bzw. ausblendet, in denen die interaurale Varianz und damit der Nachhall-Anteil groß wird.

Die Cocktail-Party-Prozessoren des oben beschriebenen Modells liefern während ansteigender Signalflanken mit überwiegendem Direktschallanteil Schätzer mit den interauralen Parametern des Direktschalls mit geringer Richtungsvarianz. (Schätzer bei dominanten Quellen, Kap. 5.5 und 6.2.5). Als Detektionskriterium für die auswertbare Richtungsinformationen kann somit die Varianz der Richtungsschätzung dienen.

Dieses Detektionskriterium entspricht einer Anwendung der Flankenmethode nach Wolf [49] auf Cocktail-Party-Prozessor-Modelle. Es wären möglicherweise Verbesserungen der Detektion zu erwarten, da ansteigende Flanken, die durch die Überlagerung mehrerer Quellen unterschiedlicher Momentanfrequenz verursacht werden (z.B. Spiegelschallquellen mit unterschiedlicher Laufzeit bei Signalen mit veränderlichem Kurzzeit-Spektrum), nicht detektiert würden.

Auch gegenüber der reinen Varianz-Methode könnten Vorteile erwartet werden, da eine Reihe von Signalen mit varianter interauraler Korrelationsfunktionen nun auswertbar würden. Z.B. würden interaurale Schwebungen, die bei Überlagerung zweier Signale unterschiedlicher Momentanfrequenz entstehen, als Signale zweier Quellen mit invarianten Richtungsschätzern ausgewertet. Die Methoden der Varianz-Analyse werden jedoch bei der Bewertung der Schätzer übernommen: Variante Schätzer mit breiter Bewertungsfunktion (vergl. Kapitel 5.6 und 6.3.) lassen darauf schließen, daß diese Schätzer das Ergebnis der Überlagerung von Reflexionen und Nachhall sind, die bei der Richtungsanalyse ignoriert werden sollten.

Mit der Cocktail-Party-Prozessor-Methode lassen sich Signaleinsätze (ansteigende Flanken) des Direktschalls noch in den Reflexionsbereich hinein verfolgen. Solange nur eine frühe Reflexion auftritt, handelt es sich um ein reines 2-Quellen-Problem, das mit Hilfe der Cocktail-Party-Prozessor-Algorithmen für beide Quellen exakt lösbar ist. Ist die Gesamtleistung aller Reflexionen geringer als die Leistung des Direktschalls, kann dieser als dominante Quelle in den Reflexionsberg hinein ausgewertet werden.

8.2. Dynamische Effekte der Richtungserkennung: Der Präzedenz-Effekt

Der Präzedenz-Effekt beschreibt die Richtungs-Wahrnehmungen bei Präsentation von Signalen aus unterschiedlichen Richtungen in enger zeitlicher Abfolge, z.B. Direktschall aus einer Richtung und eine Reflexion aus einer anderen.

Die Wahrnehmungen sind von der zeitlichen Abfolge von "Direktschall" und "Reflexion" abhängig. Ist der Zeitraum zwischen zwei Richtungsinformationen geringer als die maximale interaurale Laufzeit ($<1\text{ms}$), tritt Summenlokalisierung auf und eine "gemittelte" Richtung wird wahrgenommen (vergl. Kap. 4.2). Ist der Zeitabstand zwischen Direktschall und Reflexion größer als die Echoschwelle, werden die Richtungen beider Signale wahrgenommen. Die Echoschwelle ist vom

Spektrum und dem dynamischen Verhalten der Signale abhängig. Im Bereich zwischen Summenlokalisation und Echschwelle wird "normalerweise" nur die Direktschall-Richtung wahrgenommen ("Gesetz der ersten Wellenfront"). Die spektralen Informationen des "Reflexions"-Signals werden dann der Direktschall-Richtung zugerechnet.

Untersuchungen von Clifton [11], Wolf [49], Blauert/Col [8] ergaben, daß der Präzedenz-Effekt nicht immer dem "Gesetz der ersten Wellenfront" folgt, sondern von den verwendeten Signalen, der zeitlichen Folge und der Vorgeschichte abhängig ist. So kann eine Signalkonfiguration aus Direktschall und Reflexion einmal zum Wahrnehmen nur einer Einfallrichtung führen ("Gesetz der ersten Wellenfront"), bei einer bestimmten Vorgeschichte (Reflexionsrichtung als vorherige Direktschall-Richtung) aber zum Wahrnehmen beider Richtungen führen. Dies bedeutet:

- Das menschliche Gehör ist prinzipiell in der Lage, bei Signalkonfigurationen zum Präzedenz-Effekt (schnelle Folge zweier unterschiedlicher Richtungsinformationen) sowohl die Direktschall- als auch die Reflexionsrichtung wahrzunehmen und die zugehörigen Signale richtungsselektiv zu verarbeiten.

Da die Ausprägung des Präzedenz-Effekts von der Vorgeschichte abhängt und die entscheidenden Signale mehrere Sekunden zurückliegen können, scheint der Präzedenz-Effekt die Auswirkung von Eingriffen "höherer Ebenen" der Wahrnehmung auf die bevorzugte Richtung zu sein und somit die Folge einer Steuerung der Aufmerksamkeits-Richtung. Der Präzedenz-Effekt kann somit Auskunft über Steuerungs-Mechanismen des Gehörs geben und als Vorbild für die Steuerung eines Cocktail-Party-Prozessors dienen.

Eine Deutung des Präzedenz-Effekts könnte folgendermaßen aussehen: Das Gehör bestimmt eine Richtung, der die Aufmerksamkeit zugewandt wird und auf die der gehörinterne Cocktail-Party-Prozessor ausgerichtet wird. Liefert das binaurale System keine sichere Richtungsinformation, wird die bisherige Sollrichtung beibehalten. Bei sicheren Schätzungen der Einfallrichtung (z.B. ansteigende Flanken bei Signaleinsätzen) muß die Steuereinheit entscheiden, ob die erkannte Richtung als neue Sollrichtung übernommen wird oder nicht. Experimente zum Präzedenz-Effekt können so Aufschluß geben über die Entscheidungskriterien des binauralen Systems zur Aufnahme einer Richtung als neue Aufmerksamkeits-Richtung.

8.3. Beschreibung des Präzedenz-Effekts durch ein binaurales Cocktail-Party-Prozessor-Modell

"Gesetz der ersten Wellenfront"

Bei Experimenten zum Präzedenz-Effekt werden z.B. Clicks mit einigen Millisekunden Dauer aus unterschiedlichen Richtungen mit unterschiedlichen Laufzeiten dargeboten, um die Auswirkungen von Direktschall und Reflexionen auf die Wahrnehmungen zu untersuchen.

Vor dem Eintreffen der ersten Reflexion wird das Schallfeld nur vom Direktschall bestimmt. Einfallrichtung und Signaleigenschaften des Schalls können leicht bestimmt werden. Das interaurale Kreuzprodukt (oder eine interaurale Kreuzkorrelationsfunktion) weist nur eine geringe Richtungsvarianz auf. Die Direktschall-Richtung kann als sicher lokalisierte Richtung im Gedächtnis verbleiben.

Bei Eintreffen der ersten Reflexion überlagern sich innerhalb der einzelnen Frequenzgruppen die Signale von Direktschall und Reflexion. Die Varianz des Kreuzprodukts steigt. Ein oben

beschriebener Cocktail-Party-Prozessor-Algorithmus ist aber in der Lage, die Leistungen und Einfallsrichtungen von Direktschall und Reflexion zu bestimmen.

War zu Beginn der Darbietung keine sichere Richtungsinformation vorhanden, wird von der Richtungs-Steuereinheit während der Direktschall-Phase die analysierte Einfallsrichtung als Sollrichtung übernommen. Da das System nun auf die Direktschall-Richtung ausgerichtet ist, wird die Reflexionsrichtung als unerwünschte Richtung unterdrückt. Wegen der höheren Varianz des Kreuzprodukts wird die Richtungsinformation der Reflexion als unsicherer eingestuft, und eine Neuausrichtung des Systems unterbleibt. Das Ergebnis ist das "Gesetz der ersten Wellenfront".

Ausnahmen des "Gesetzes der ersten Wellenfront"

Während der Direktschall-Darbietung zu Beginn der Präzedenz-Effekt-Signale ergibt sich immer ein Zeitraum niedriger interauraler Varianz und damit ein sicher lokalisiertes Hörereignis, welches im Prinzip zu einer Neuausrichtung des Systems führen könnte.

Werden gleichartige Folgen von Direktschall und einer Reflexion dargeboten, wie bei Experimenten zu Ausnahmen vom "Gesetz der ersten Wellenfront", scheint sich die Information über die Sollrichtung zu verfestigen. Wird dann ein Signalpaar dargeboten, bei dem die Richtungen von Direktschall und Reflexion vertauscht sind, werden beide Richtungen wahrgenommen, die Reflexionsrichtung als bisherige Sollrichtung und die Direktschall-Richtung als sicher lokalisierte Richtung.

Werden die neuen Direktschall-Reflexions-Kombinationen wiederholt, werden so lange 2 Hörereignis-Richtungen wahrgenommen, bis die Steuerung des Systems die Direktschall-Richtung als neue Sollrichtung übernommen hat. Danach werden die Richtungsinformationen der Reflexion wieder ausgeblendet und das Ergebnis entspricht dem "Gesetz der ersten Wellenfront".

Blauert/Col [8] wiederholten Experimente zu diesem Effekt mit laufend vertauschten Direktschall- und Reflexionsrichtungen. Auch hier wurden bei Tausch von Direktschall- und Reflexionsrichtung beide Richtungen wahrgenommen. Nach einigen Vertauschungen ergab sich aber wieder das "Gesetz der ersten Wellenfront".

Die Zeitkonstanten der Richtungssteuerung scheinen sich an die Schallfeld-Situation anzupassen. Während zu Beginn des obigen Experiments jede sicher lokalisierte Direktschall-Richtung als neue Sollrichtung übernommen wird und nach dem Vertauschen der Richtungen die Reflexionsrichtung als alte Sollrichtung wahrgenommen wird, scheint bei sich laufend ändernden Direktschall-Richtungen das auditorische System die "Gedächtnis-Zeiten" zu verkürzen und sich auf "laufende Neu-Ausrichtung" umzustellen, so daß die Reflexionsrichtungen nicht mehr wahrgenommen werden.

Echoschwelle

Überlagern sich Direktschall und Reflexionen in den einzelnen Frequenzgruppen nicht mehr, so wird auch die Richtung der Reflexion varianzarm und kann als sichere Einfallsrichtung registriert werden.

Querbeziehungen zwischen Frequenzgruppen

Nach Untersuchungen von Blauert/Divenyi [6] wirkt der Präzedenz-Effekt auch über Frequenzgruppen-Grenzen hinweg. Nach Blauert u.a.[7] kann Direktschall im mittleren Frequenzbereich um

1 kHz die Richtungsinformationen von Reflexionen in anderen Frequenzbereichen unterdrücken, wogegen der umgekehrte Einfluß nicht beobachtet wurde. Für andere Frequenzbereiche ergeben sich abgestufte Wirkungen über Frequenzgruppen-Grenzen hinweg.

Die Frequenzbereiche für Direktschall und Reflexionen überlappten sich bei diesen Untersuchungen meist nicht, so daß in den Frequenzbereichen von Direktschall und Reflexion invariante Richtungsinformationen vorlagen, die im Prinzip beide zu einer Neu-Ausrichtung der Aufmerksamkeits-Richtung hätten führen können.

Sichere Richtungsinformationen werden demnach frequenzabhängig bewertet. Sie führen nur dann zur Neu-Ausrichtung des Systems, wenn dem keine Informationen aus wichtigeren Frequenzbereichen entgegenstehen.

8.4. Konsequenzen für die Steuerung eines Cocktail-Party-Prozessors

Eingangssignale einer Steuereinheit für Cocktail-Party-Prozessoren sind die per Wahrscheinlichkeitsfunktion gewichteten Schätzer für Einfallsrichtungen und Signalleistungen.

Wichtige Aufgabe einer Steuereinheit ist die Detektion varianzarmer Zeiträume des interauralen Kreuzprodukts (Direktschall-Detektion). Während dieser Zeiträume sollten die erkannten Richtung und Signalleistungen an übergeordnete Modellstufen weitergegeben werden.

Die Steuereinheit muß bei Vorliegen mehrerer varianzarmer Richtungs-Schätzer (z.B. aus mehreren Frequenzgruppen) die Entscheidung fällen, welche Richtung als Sollrichtung des Systems übernommen wird. Kriterien für diese Entscheidung können sein: die Sicherheit der Richtungserkennung, die Wertigkeit des Frequenzbereichs, in dem die Information vorliegt, die Vorgeschichte (oft erkannte Richtungen gelten als sehr sichere Information) sowie das Unterstützen bestimmter Richtungen durch andere Informationsquellen (z.B. optische Informationen, bewußte Entscheidung).

Wird eine Einfallsrichtung als neue Sollrichtung übernommen, wird aus der Varianz dieser Einfallsrichtung ein Schwellenkriterium für neue Einfallsrichtungen gebildet. Eine Sollrichtung wird beibehalten bzw. verstärkt, wenn sie durch neue sichere Richtungsinformationen, z.B. durch varianzarme Schätzer einzelner Frequenzgruppen, bestätigt wird. Fehlt diese Bestätigung über eine gewisse Zeit, sinkt die Varianz-Schwelle, und andere sichere Richtungsinformationen, z.B. varianzarme Schätzer anderer Einfallsrichtungen, können als neue Sollrichtung übernommen werden.

Der Zeitraum, für den die Sollrichtungs-Entscheidung gilt, kann signalabhängig variiert werden. Bei schneller Änderung der als wichtig eingestuften erkannten Einfallsrichtungen kann das System auch auf "Permanente Neu-Ausrichtung" umgestellt werden.

Ist eine Entscheidung für eine Sollrichtung gefallen, werden erkannte Richtungen, die nicht der Sollrichtung entsprechen, nicht weitergeleitet und nicht eindeutige Analyse-Ergebnisse auf mögliche Sollrichtungs-Anteile untersucht (vergl. Kap. 5.6 und 6.4). Starke varianzarme Signale können aber, auch wenn sie nicht der Sollrichtung entsprechen, an höhere Stufen weitergeleitet werden.

Die Abhängigkeit des Präzedenz-Effekts von der Frequenz kann durch eine frequenzabhängige Bewertung der Schätzer unterschiedlicher Frequenzgruppen berücksichtigt werden.

8.5. Von der Prozessor-Steuerung zum Präzedenz-Prozessor

Informationen über die Leistung des Direktschalls bzw. des Hintergrundgeräusches können bei der Detektion ansteigender Flanken bzw. während invarianter Zeitabschnitte des interauralen Kreuzprodukts bestimmt werden. Diese Schätzer könnten während der Zeiträume, in denen aufgrund von Reflexionen und Nachhall keine verlässliche Schätzung möglich ist, als Referenz zur Bestimmung von Signal- bzw. Störpegeln angesehen werden.

Als sogenannter Präzedenz-Prozessor, der während ansteigender Flanken Schallquellen- und Störsignal-Parameter aufzeichnet und diese Informationen im Nachhallfeld fortschreibt (sozusagen als Voreinstellung für Zeiträume unsicherer Richtungsschätzung), könnte ein solcher Algorithmus als dritter Cocktail-Party-Prozessor-Typ neben Phasendifferenz- und Pegeldifferenz-Prozessor eingesetzt werden.

9. Zusammenfassung und Ausblick

Zentraler Punkt der vorliegenden Arbeit ist die Untersuchung von Algorithmen zur richtungsselektiven Verarbeitung von Schallsignalen.

Grundlage für den Entwurf von Signalverarbeitungs-Algorithmen waren psychoakustische Versuche, die Aufschluß über die Signalverarbeitung des menschlichen Gehörs bei Anwesenheit mehrerer Schallquellen lieferten. Fragestellung dieser Versuche war, wie ähnlich Signale sein dürfen, damit diese noch richtungsselektiv verarbeitet werden können. Die Untersuchungen fanden einmal im Frequenzbereich unter 800 Hz statt, wo das Gehör zur Richtungsbestimmung interaurale Phasen auswertet, und dann im Bereich oberhalb von 1,6 kHz, wo die Richtungserkennung auf der Analyse interauraler Gruppenlaufzeit- und Pegeldifferenzen fußt.

Im Frequenzbereich unter 800 Hz konnten Versuchspersonen schon bei relativ geringen Signalunterschieden die Einfallsrichtungen zweier Schallsignale korrekt bestimmen, so z.B. bei Sinussignalen von 500 Hz und 530 Hz bzw. bei zwei unabhängigen Rauschsignalen mit 7% relativer Bandbreite von 500 Hz Mittenfrequenz. Bei diesen Signalen wird ein Großteil der Signalleistung in einer Frequenzgruppe zusammengefaßt. Die Versuchspersonen waren in der Lage, die relativen Tonhöhen der lokalisierten Quellen zu bestimmen (höher/tiefer als andere Quellen) sowie den lokalisierten Signalen entsprechende Lautheiten zuzuweisen. Die Klänge der Schallsignale konnten allerdings nicht korrekt angegeben werden. Dies war erst möglich, wenn die (Mitten-)Frequenzdifferenz eine Frequenzgruppen-Breite überschritt.

Im Frequenzbereich oberhalb von 2 kHz konnten zwei Schallquellen unterschiedlicher Richtung erst ab Frequenzdifferenzen oberhalb einer Frequenzgruppen-Breite korrekt lokalisiert werden. Dann war auch eine korrekte Zuordnung von Tonhöhen, Lautheiten und Klängen zur lokalisierten Richtung möglich.

Geht man davon aus, daß für die Bestimmung von Richtungen das schmalste vom Gehör verwendete Analyseband die Frequenzgruppe ist, wäre ohne "Cocktail-Party-Prozessor-Mechanismen" eine nach Richtungen getrennte Analyse nur möglich, wenn Signale unterschiedlicher Richtungen in unterschiedlichen Frequenzgruppen vorliegen. Im Bereich niedriger Frequenzen, wo das Gehör interaurale Phasenlaufzeiten auswertet, würde demnach das Gehör "Cocktail-Party-Prozessor"-Mechanismen anwenden, im Bereich hoher Frequenzen, wo interaurale Gruppenlaufzeiten ausgewertet werden, dagegen nicht. Die Eigenschaften dieses gehörinternen Prozessors wären demnach: Auswertung interauraler Phasen, Bestimmung von Einfallsrichtungen und Lautheiten von (mindestens) 2 Schallquellen, keine Analyse der Signalklänge, d.h. keine richtungsselektive Trennung von Phasen- bzw. spektralen Informationen.

Technische Modelle, die die Fähigkeiten des Gehörs nachvollziehen sollen, müssen ähnliche Eigenschaften wie der vermutete gehörinterne Cocktail-Party-Prozessor besitzen. Funktionen, diese Forderungen erfüllen, sind z.B. die interaurale Kreuzkorrelationsfunktion (Analyse reeller Signale) und das interaurale Kreuzprodukt (Analyse komplexer analytischer Zeitsignale). Für beide Funktionen werden Cocktail-Party-Prozessor-Algorithmen vorgestellt, die Richtungen und Signalleistungen zweier sich überlagernder Schallquellen aus Ohrsignalen bestimmen können. Bei der Kreuzkorrelationsfunktion ist dies eine Methode im Frequenzbereich, beim Kreuzprodukt im Zeitbereich.

Wesentlicher Punkt der vorliegenden Arbeit ist die Beschreibung eines Cocktail-Party-Prozessors, der auf der Analyse des interauralen Kreuzprodukts aufbaut. Dieses Verfahren läßt sich mathematisch einfach beschreiben, schnell berechnen und reagiert schnell auf Signaländerungen.

Mit Hilfe des interauralen Kreuzprodukts werden interaurale Schwebungen analysiert, die bei der Überlagerung von Signalen mit unterschiedlichen Spektren entstehen. Die statistischen Parameter des interauralen Kreuzprodukts werden dazu benutzt, um die Leistungen und Einfallrichtungen zweier Schallquellen zu bestimmen, die solche Schwebungen erzeugen können.

Die Besonderheit dieses Verfahrens zur binauralen Signalverarbeitung, dem Phasendifferenz-Cocktail-Party-Prozessor, besteht darin, daß hier ein 2-Quellen-Ansatz zur Beschreibung der binauralen Interaktionen verwendet wird. Hierdurch können Einfallrichtungen und Leistungen zweier Quellen simultan analysiert werden und die Parameter einer Quelle auch dann noch bestimmt werden können, wenn eine Störquelle mit wesentlich höherer Leistung anwesend ist. Es lassen sich Verbesserungen des Signal-Stör-Abstands um bis 20 dB erreichen und bei Sprachsignalen selbst bei Signal-Stör-Abständen von -30 dB noch hörbare Verbesserungen des Signal-Stör-Verhältnisses erreichen.

Bei komplexeren Schallfeldern (mehr als 3 Schallquellen, Reflexionen und Nachhall) sind Verbesserungen durch das Verfahren dann erzielbar, wenn die Leistung der Nutzquelle zu bestimmten Zeiten bzw. in bestimmten Frequenzbereichen die Leistungen anderer Quellen überwiegt oder die Nutzquelle zumindest die zweitstärkste Einzelquelle ist (dominante Quelle). Mit Hilfe zusätzlicher Schätzalgorithmen ist es jedoch möglich, auch bei schwachen Nutzquellen und komplexen Schallfeldern andere dominante Einzelquellen auszublenden bzw. die mögliche Leistung der Nutzquelle abzuschätzen und fortzuschreiben. Allerdings steigen hier die Schätzfehler an. Bei solch komplexen Schallfeldern ist allerdings auch das menschliche Gehör nicht mehr in der Lage, schwächere Schallquellen korrekt zu lokalisieren und richtungsselektiv zu verarbeiten (vergl. verringerte BILD in halliger Umgebung, Franssen-Effekt).

Der verwendete Cocktail-Party-Prozessor-Algorithmus ist auch in der Lage, einkanalige Signale zu verarbeiten und als Ergebnis der Überlagerung zweier Signale unterschiedlichen Spektrums darzustellen. Diese Eigenschaft wird zur Konstruktion des "Pegeldifferenz-Cocktail-Party-Prozessors", ausgenutzt. Hier werden die Hüllkurven beider Ohrsignale getrennt ausgewertet und Schätzer für die Signalleistungen und Einfallrichtungen zweier Schallquellen durch passende interaurale Kombination der monauralen Schätzer bestimmt.

Die Eigenschaften dieses Prozessors ähneln dem Phasendifferenz-Verfahren: Leistungen und Richtungen zweier stationärer Schallquellen können selbst bei negativen Signal-Stör-Verhältnissen noch abgeschätzt werden. Die Parameter dominanter Quellen in komplexen Schallfeldern können bestimmt werden. Auch hier lassen sich Nachbearbeitungs-Verfahren finden, die auch für schwache Nutzquellen in komplexen Schallfeldern eine Abschätzung der möglichen Leistung vornehmen können.

Für breitbandige Analysen bei natürlichem Ohrabstand müssen beide Cocktail-Party-Prozessoren kombiniert werden. Bei niedrigen Frequenzen, wo die interaurale Phase eindeutig ist, liefert der Phasendifferenz-Prozessor sehr genaue Ergebnisse, wogegen der Pegeldifferenz-Prozessor besonders zur Analyse hoher Frequenzen, wo ausreichend hohe Pegeldifferenzen auftreten, geeignet ist. Im mittleren Frequenzbereich können die Fehler des Phasendifferenz-Prozessors in Form mehrdeutiger Richtungsschätzungen durch Hinzuziehung des Pegeldifferenz-Prozessors korrigiert werden. Die Ergebnisse beider Verfahren werden in einer entsprechenden Prozessor-Stufe kombiniert, die aus dem mehrdeutigen Schätzer-Angebot die wahrscheinlichsten ausgewählt. An dieser Stelle könnten auch die Ergebnisse andere Informationsquellen in die Richtungsanalyse und Signalverarbeitung einbezogen werden. (z.B. optische Informationen, Vorwissen, Präzedenz-Prozessor).

Psychoakustische Erkenntnisse könnten dazu verwendet werden, eine gehörgemäße Steuerungsstrategie für die Cocktail-Party-Prozessoren zu entwerfen. Analog zu den Hörversuchen von Gaik [21] könnte die Verknüpfung von Phasendifferenz- und Pegeldifferenz-Prozessor so gestaltet werden, daß bei Widersprüchen zwischen interauralen Zeit- und Pegeldifferenzen monaurale Schätzer (z.B. durch Aufspaltung widersprüchlicher Schätzer) gebildet werden.

Erkenntnisse über den Präzedenz-Effekt, der das dynamische Verhalten der Richtungserkennung des Gehörs beschreibt, könnten zur Steuerung der Aufmerksamkeits-Richtung des Cocktail-Party-Prozessors herangezogen werden und so die Steuerungsstrategie des Gehörs technisch nachvollzogen werden.

Für die Verarbeitung komplexer Schallfelder mit Reflexionen und Nachhall wäre ein sog. Präzedenz-Effekt-Prozessor denkbar, der bei Vorhandensein sicherer Richtungsinformationen (Direktschall-Zeiträume) Informationen über die entsprechende Schallquelle bzw. das Störfeld sammelt und diese Informationen in Zeiträumen ohne sichere Richtungsinformation als Grundlage der Verarbeitung übernimmt.

Denkbar wäre ein Ergänzen der akustischen Analyse durch weitere Informationsquellen, wie optische Informationen (Erkennen von Richtungen und Mundbewegungen), Einbeziehen von Wissen (z.B. über die möglichen Signalleistungen und Spektren bestimmter Signale <Männer/Frauenstimme/ Geräusche>).

Die vorgestellten Cocktail-Party-Prozessor-Verfahren verarbeiten analytische Zeitsignale, die auch datenreduziert sein dürfen, relativ schnell. Bei einer Weiterentwicklung des Verfahrens wären Echtzeitrealisierungen möglich. Bei einer entsprechenden Vor- und Nachbearbeitung der Signale (nichtkausale Frequenzgruppen-Filter ohne Laufzeit-Verzerrungen; sanftes Modulieren der Signale durch die Schätzer der Prozessoren) lassen sich verarbeitete Signale relativ hoher Qualität erzeugen, die für weitere Signalverarbeitungs-Verfahren als Eingangssignale dienen können.

Der Signalverarbeitungs-Rahmen der Cocktail-Party-Prozessoren entspricht einem adaptiven richtungsselektiven Filter, dessen Übertragungsfunktion laufend an das Spektrum der Signale einer bestimmten Einfallsrichtung angepaßt wird.

Da keine Laufzeit-Verzerrungen auftreten und die Signalphasen nicht verändert werden, könnte das Verfahren z.B. zur Unterstützung von Mikrofon-Arrays eingesetzt werden. Hierzu müßten alle Ausgangskanäle des Cocktail-Party-Prozessors die gleiche (richtungsselektive) Dämpfung erhalten, so daß die Pegelverhältnisse zwischen den Empfängern des Arrays nicht verändert werden. Die Richtwirkung von Mikrofon-Arrays ließe sich hiermit besonders im tieffrequenten Bereich verbessern. Bei einem Einsatz in Verbindung mit zwei Richtmikrofonen könnte hier besonders im tieffrequenten Bereich die Richtwirkung verbessert werden.

Ebenso wäre dieses Verfahren als Vorverarbeitungsstufe für Kunstkopf-Systeme einsetzbar. Die Spektralbereiche störender Quellen könnten durch den Cocktail-Party-Prozessor eliminiert werden, ohne die binauralen Analysemöglichkeiten einzuschränken. Hiermit bestünde die Möglichkeit, in die verbliebenen Signale hineinhören zu können.

Als Mikrofonträger wären beliebige Gegenstände denkbar, z.B. Telefongehäuse. Voraussetzung für die Anwendung der Verfahren wäre die Kenntnis der Übertragungsfunktionen Mikrofon - freies Schallfeld, analog zu den Freifeld-Außenohr-Übertragungsfunktionen des Kopfes. Mit Mehrmikrofon-Anordnungen bzw. passend konstruierten Gehäusen ließen sich optimale Bedingungen für die

Prozessoren schaffen, z.B. eindeutige Phasenbeziehungen für den Phasendifferenz-Prozessor, glatte Übertragungsfunktionen für den Pegeldifferenz-Prozessor.

Der Phasendifferenz-Cocktail-Party-Prozessor ließe sich mit einer entsprechenden Mikrofonanordnung auch breitbandig einsetzen. Hierbei müßte die Mikrofonanordnung eindeutige Phasenbeziehungen im gesamten Frequenzbereich gewährleisten (z.B. Mikrofonabstände von 20 cm, 5 cm und 1 cm). Mit diesem Mini-Array wären über nahezu den gesamten Hörfrequenzbereich scharfe Richtkeulen zu erreichen.

Einsatzmöglichkeiten für ein solches Cocktail-Party-Prozessor-Verfahren wären Anwendungsfälle, in denen richtungsselektive Signalverarbeitung erforderlich oder erwünscht ist, d.h. wo ein erwünschtes Signal von Signalen anderer Richtungen gestört wird, z.B.

- als Eingabeeinheit für Sprachsysteme (Spracherkenner, Freisprechttelefone, Aufnahmesysteme in Räumen), z.B. auf mikrofonbesetzten Gegenständen, um andere störende Sprecher und Geräusche aus der Übertragung auszuschließen,
- als Hörgeräte-Algorithmus zur Störsignal-Unterdrückung für einohrig Ertaubte ("künstliches binaurales System"), z.B. Hörbrille mit Mini-Array und Phasendifferenz-Prozessor, um die ansonsten stark verminderten Kommunikationsmöglichkeiten in gestörter Umgebung zu verbessern,
- als akustisches Signalanalyse-System (Lärmquellen-Ortung, richtungsselektive Verarbeitung von Lärmquellen, akustische Qualitätsbeurteilung von Räumen/Konzertsälen), z.B. als Vorverarbeitungs-System für Mikrofon-Arrays oder Kunstköpfe, um eine richtungsselektive Lärmquellen-Diagnose und eine subjektive Beurteilung bestimmter Schallquellen durchführen zu können,
- zur Simulation der Wahrnehmungen des Menschen (Untersuchung der Beeinflussung/Lästigkeit bestimmter räumlicher Quellen, Simulation binauraler Übertragungssysteme z.B. für Audio-Anwendungen), um die Auswirkung richtungsabhängiger Informationen auf den Menschen erfassen und beurteilen zu können,
- als binaurales Modell zum Nachvollziehen der Signalverarbeitung des Menschen, zur Erklärung psychoakustischer Phänomene, um so für Mehr-Quellen-Phänomene neue Modellierungsmöglichkeiten zu erhalten und so weitere Grundlagen zu schaffen, Gehör-Prozesse für technische Anwendungen nutzbar zu machen.