Modelle zur binauralen Trennung mehrerer Schallquellen: Ein Beitrag zur Entwicklung eines Cocktail-Party-Prozessors[4]

H.SLATKY
(Lehrstuhl für allgemeine Elektrotechnik und Akustik, Ruhr-Universität Bochum, D-4630 Bochum, Germany)

Das menschliche Gehör kann sich auf eine Schallquelle fokussieren, selbst wenn Störungen, Echos, Nachhall oder anderen Schallquellen anwesend sind. Dies ist zum Beispiel der Fall, wenn in einem Raum mehr als ein Sprecher anwesend ist ("Cocktail-Party-Effekt"). In meiner Arbeit versuche ich Algorithmen zu finden, die diese binauralen Phänomene modellieren und für technische Anwendungen nutzbar sind.

Lokalisation mehrerer Schallquellen durch das Gehör

Um die Frage zu beantworten, wie das menschliche Gehör reagiert, wenn mehr als eine Schallquelle gleichzeitig aktiv ist, wurden Hörversuche durchgeführt, bei denen zwei Sinus- oder Schmalband-Rausch-Signale gleichzeitig in einem reflexionsarmen Raum dargeboten wurden.

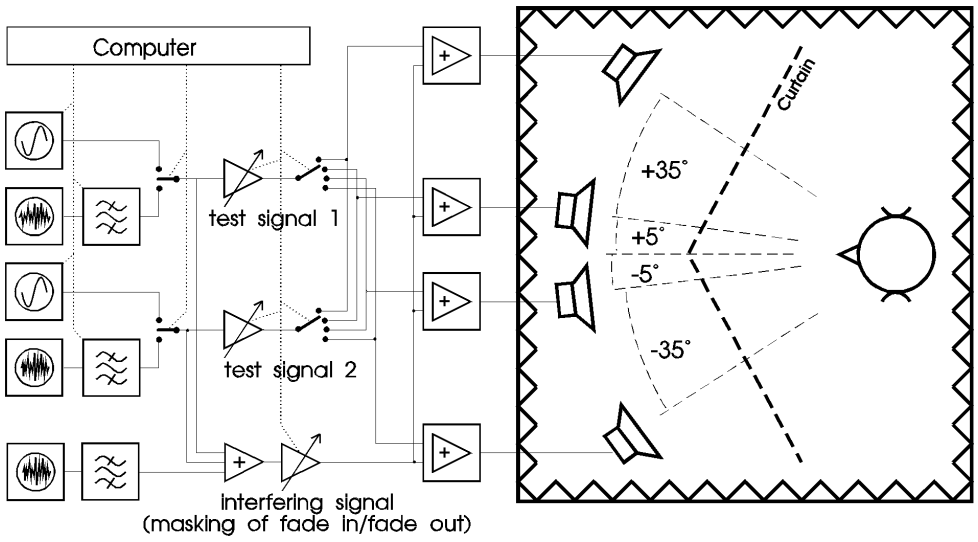


Abb. 1:
Aufbau der Hörversuche zur Lokalisation mehrerer Schallquellen. Dargebotene Signale
1. *Sinus 500 Hz + Sinus (500 Hz + x), x=10.. 160 Hz*
2. *Sinus 2000 Hz + Sinus (2000 Hz + x), x=10..1200 Hz*
3. *Schmalband-Rauschen (7% rel. Bandbreite): Rauschen 500 Hz + Rauschen (500 Hz + x), x=10.. 160 Hz*
4. *Schmalband-Rauschen (7% rel. Bandbreite): Rauschen 2000 Hz + Rauschen (2000 Hz+x) x=10..1200 Hz*

Werden gleichzeitig zwei schmalbandige Schallquellen dargeboten, deren Bandbreiten wesentlich kleiner als eine Frequenzgruppenbreite sind (z.B. Sinussignale 500 Hz + 530 Hz oder Rauschen mit 7% relativer Bandbreite 500 Hz + 510 Hz), kann das Gehör beide Schallquellen korrekt lokalisieren und die Schallquellen an Hand ihrer Tonhöhe identifizieren (höher, tiefer) [3]

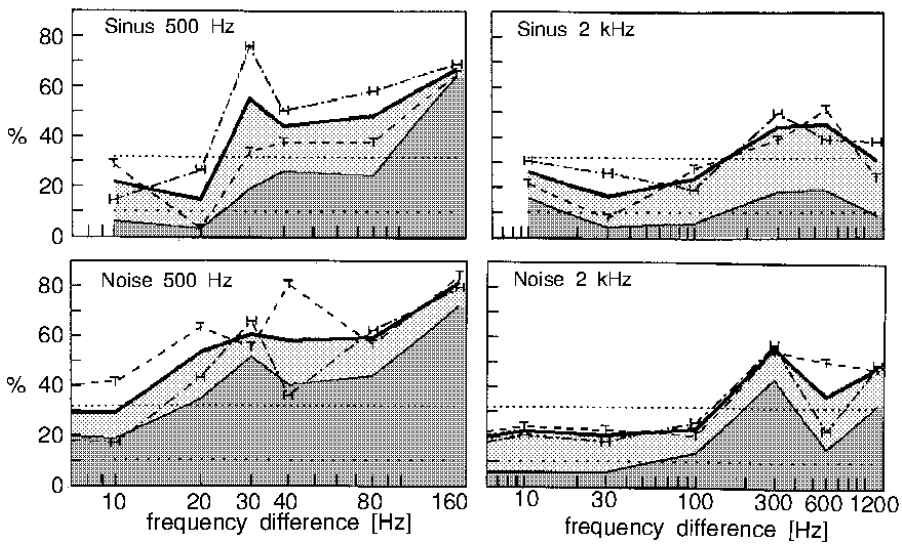


Abb.. 2:
Prozentsatz korrekt lokalisierter Schallquellen.
--H-- höhere Schallquelle
--T--- tiefere Schallquelle
eine Schallquelle korrekt lokalisiert
beide Schallquellen korrekt lokalisiert
..... Ratewahrscheinlichkeit für eine Schallquelle
. . . . Ratewahrscheinlichkeit für beide Schallquelle
um 500 Hz wird die Ratewahrscheinlichkeit überschritten für Δf>10 Hz (Rauschen) bzw. Δf>30 Hz (Sinus)

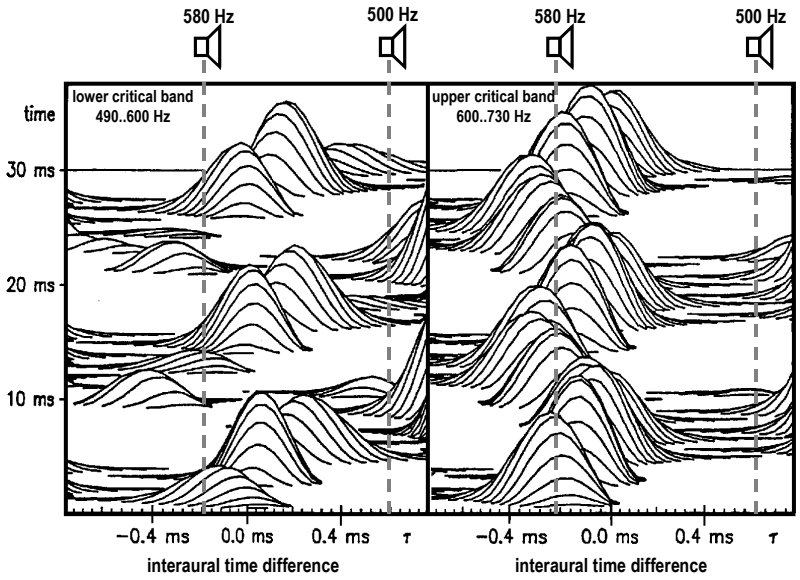


Abb..3:
Interaurale Kreuzkorrelationfunktion[2] der Signale der Hörversuche innerhalb der betroffenen Frequenzgruppen

Dargebotene Signale: :
Sinus 500 Hz τ= 0.6 ms
Sinus 580 Hz τ=-0.2 ms

Gestrichelte Linien: interaurale Zeitdifferenzen der Testsignale

In der unteren Frequenzgruppe gibt es keine Übereinstimmungen zwischen den Positionen der Maxima der Kreuzkorrelationsfunktion und den Richtungen der Schallquellen.
In der oberen Frequenzgruppe stimmen die Positionen der Maxima der Kreuzkorrelationsfunktion mit der Richtung der höheren Schallquelle überein.

Binaurale Modelle

Binaurale Modelle, die auf interauralen Kreuzkorrelationsfunktionen innerhalb von Frequenzgruppen basieren, und die die Richtung von Schallquellen direkt aus der Position der Maxima der Kreuzkorrelationsfunktion bestimmen (z.B. LINKDEMANN[2], GAIK[1]), können bei diesen Signalen nur eine Einfallsrichtung korrekt bestimmen, denn die Position der Maxima ist hierbei nur in einer (von zwei) Frequenzgruppen stabil. Die Kreuzkorrelationsmuster der anderen Frequenzgruppe sind sehr variant, so dass hieraus keine Einfallsrichtung direkt bestimmt werden kann.[3]

Geht man davon aus, dass das Gehör eine Richtungsanalyse in Frequenzgruppen durchführt und dass eine Beschreibung über Kreuzkorrelationsfunktionen gehörgemäß ist, so muss eine Möglichkeit existieren, aus diesen Mustern Informationen über die beteiligten Schallquellen zu extrahieren. ("Rückrechenmechanismus"[3])

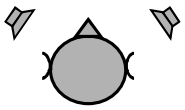
		Lokalisation	**Klang**	**Lautstärke**
Signale in den betroffenen Frequenzgruppen				
Signale in der oberen Frequenzgruppe		Lokalisation des höheren Signals erwartet	Klang des höheren Signals erwartet	Höheres Signal mit verminderter Lautstärke erwartet
Signale in der unteren Frequenzgruppe		keine Lokalisation erwartet	Gemisch aus beiden Signalen erwartet	Sum of both signals exspected
Ergebnisse der Hörversuche		Beide Signale wurden korrekt lokalisiert	Original-Klang für das höhere Signal Signal-Gemisch aus der Richtung des tieferen Signals	Höheres Signal mit 140% der Lautstärke des tieferen Signals
Konsequenz für binaurale Modelle		Eweiterung binauraler Modelle erforderlich	Modell und Experimente stimmen überein	Eweiterung binauraler Modelle erforderlich

Abb.4:
Vergleich zwischen Kreuzkorrelations-Modellen und den Ergebnissen der Hörversuche

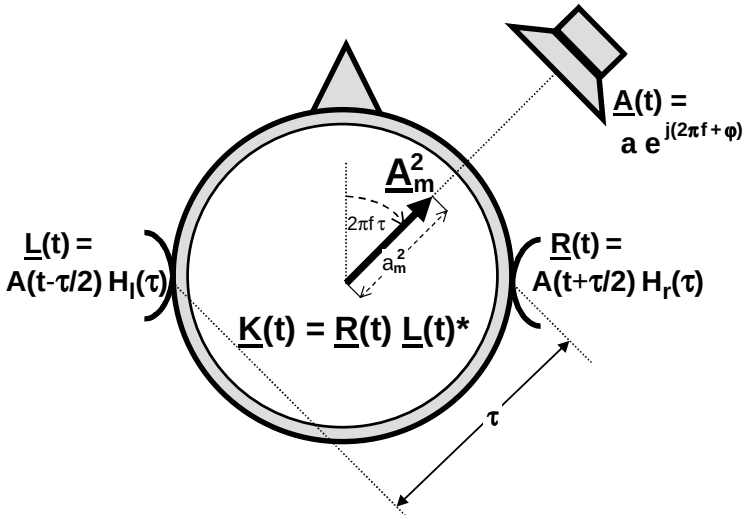


Abb.5:
Das interaurale Kreuzprodukt k(t) für eine Schallquelle mit konstanter Amplitude

Suche nach einer problemangepassten mathematischen Beschreibung

Eine weitere Methode, binaurale Interaktionen innerhalb von Frequenzgruppen zu beschreiben, ist das komplexe Kreuzprodukt aus den analytischen Zeitsignalen der Ohrsignale. Die Eigenschaften sind:

- Analytische Zeitsignale innerhalb von Frequenzgruppen können mit verminderter Datenrate verarbeitet werden, die Verarbeitung wird schneller.
- Die Abhängigkeit der binauralen Interaktionsmuster von den Ohrsignalen kann in mathematisch geschlossener Form beschrieben werden.
- Bei stationären Signalen aus 1 oder 2 Einfallsrichtungen ergeben sich binaurale Interaktionsmuster in einfachen geometrischen Formen (siehe unten).

Innerhalb von Frequenzgruppen kann man beliebige Signale als Amplituden- und Frequenz-modulierte Sinus-Signale beschreiben. Die zugehörige analytische Zeitfunktion $\underline{A}(t)$ ist: (f(t)=Frequenz, a(t)=Amplitude, φ(t)=Phase)

$$\underline{A}(t) = a(t)\, e^{+j2\pi f(t)t + j\varphi(t)}$$

Die zugehörigen Ohrsignale sind: (τ=interaurale Zeitdifferenz, $H_l(\tau)$, $H_r(\tau)$ Außenohr-Übertragungs-Funktionen)

$$\underline{L}(t) = \underline{A}(t-\tau/2)\, H_l(\tau) \qquad \underline{R}(t) = \underline{A}(t+\tau/2)\, H_r(\tau)$$

Das Kreuzprodukt $\underline{K}(t)$ aus linkem und rechtem Ohrsignal ist dann::

$$\underline{K}(t) = \underline{R}(t)\, \underline{L}(t)^* = a_m(t)^2\, e^{+j2\pi f(t)\tau} \qquad a_m(t)^2 = a(t)^2\, H_l(\tau)\, H_r(\tau)$$

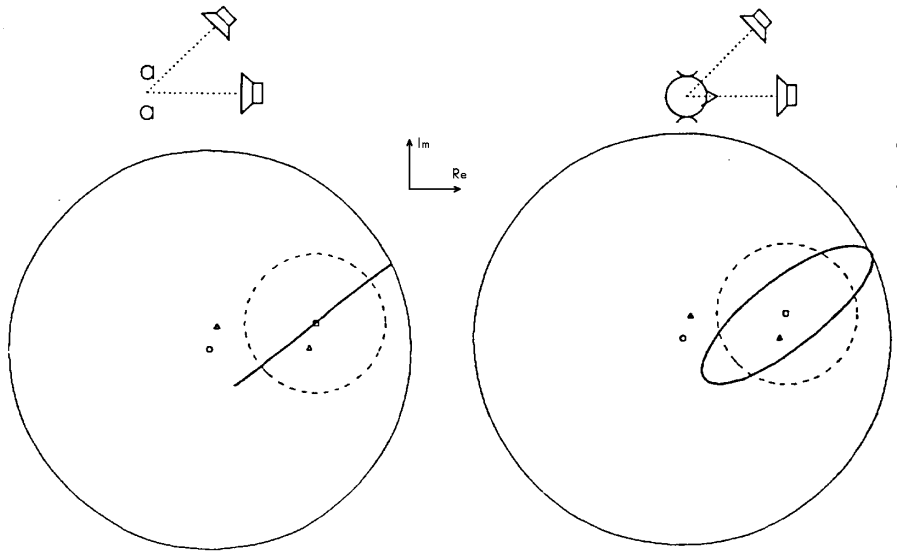


Abb. 6:
Ortskurve des Kreuzprodukts bei 2 Schallquellen:
a) Sinus 500Hz, a=1 , τ_a=0μs
b) Sinus 560Hz, b=0.5, τ_b=400μs
links: interaurale Pegeldifferenz 0dB
rechts: interaurale Pegeldifferenz 6dB
Δ Ortskurve für jede Schallquelle allein
☐ Komplexer Mittelwert
---Kreis um den Mittelwert, Radius =Standardabweichung

Für Sinussignale (a(t), f(t), φ(t)=const.) ergibt die Ortskurve des interauralen Kreuzprodukts $\underline{K}(t)$ einen Punkt in der komplexen Ebene. Der Betrag ist proportional zur mittleren Energie der Ohrsignale, die Phase entspricht der interauralen Phase. Dies entspricht den Ergebnissen von Kreuzkorrelationsmodellen bei Darstellung der Maxima in Polarkoordinaten.

Bei 2 Signalen $\underline{A}(t)$, $\underline{B}(t)$ aus unterschiedlichen Richtungen addieren sich die entsprechenden Ohrsignale, und es entstehen binaurale Schwebungen. Die Ortskurve des Kreuzprodukts wird zeitabhängig. Bei stationären Signalen hat die Ortskurve je nach interauraler Pegeldifferenz die Form einer Gerade oder einer Ellipse. Bildet man hiervon den komplexen Mittelwert $\underline{\mu}$ und die komplexe Standardabweichung $\underline{\sigma}$, erhält man ein komplexes Gleichungssystem, aus dem die interauralen Phasen $2\alpha=2\pi f_a(t)\tau_a$, $2\beta=2\pi f_b(t)\tau_b$, und die mittleren Amplituden $a_m(t)$, $b_m(t)$ der Schallquellen geschätzt werden können.

$$\underline{\mu}(t) = 1/2T \int_{t-T}^{t+T} \underline{K}(t')\, dt' \qquad \underline{\sigma}^2(t) = 1/2T \int_{t-T}^{t+T} [\, \underline{K}(t') - \underline{\mu}\,]^2\, dt'$$

$$\underline{\mu}(t) = a_m(t)^2\, e^{j2\alpha} + b_m(t)^2\, e^{j2\beta} \qquad \underline{\sigma}^2(t) = 2\, a_m(t)^2\, b_m(t)^2\, e^{j2(\alpha+\beta)}$$

Eigenschaften des Verfahrens

Die Genauigkeit des Verfahrens hängt von der Integrationszeit und der Änderungsgeschwindigkeit der Schallquellenparameter ab. Bei stationären Signalen (Sinus, harmonische Signale) und langer Integrationszeit ist die Schätzung selbst bei 100 dB Pegelunterschied zwischen den Schallsignalen noch hinreichend genau (Fehler < 1dB). Bei Signalen mit veränderlichen Amplituden (Rauschen, Sprache) müssen die Integrationszeiten kurz sein (10-20 ms). Hierbei lassen sich noch bis zu Pegeldifferenzen von 20 dB zwischen den Schallquellen genauere Schätzer von Amplituden und Richtungen der Schallsignale erzielen.

Verglichen mit anderen Methoden der Richtungsfilterung (Richtmikrofone, lineare Array-Technik) hat dieser Algorithmus den Vorteil, bei Empfängerabständen, die wesentlich kleiner als die Wellenlänge sind (Ohrabstand) noch scharfe Richtkeulen zu liefern. Im Bereich niedriger Frequenzen sind Richtkeulen-Breiten von +/-150 µs (+/-15° für die Vorne-Richtung) erreichbar.

Strahlen mehr als zwei Schallquellen innerhalb einer Frequenzgruppe Energie ab, können die Attribute der beiden stärksten Schallquellen durch das Verfahren abgeschätzt werden. Zusätzlich kann abgeschätzt werden, mit welcher Wahrscheinlichkeit die mit diesem Verfahren gewonnenen Quellenschätzer mit der Nutzrichtung übereinstimmen (Auswertung des Schätzfehlers). Hierüber kann die wahrscheinliche Amplitude eines Signals einer Nutzrichtung bestimmt werden.

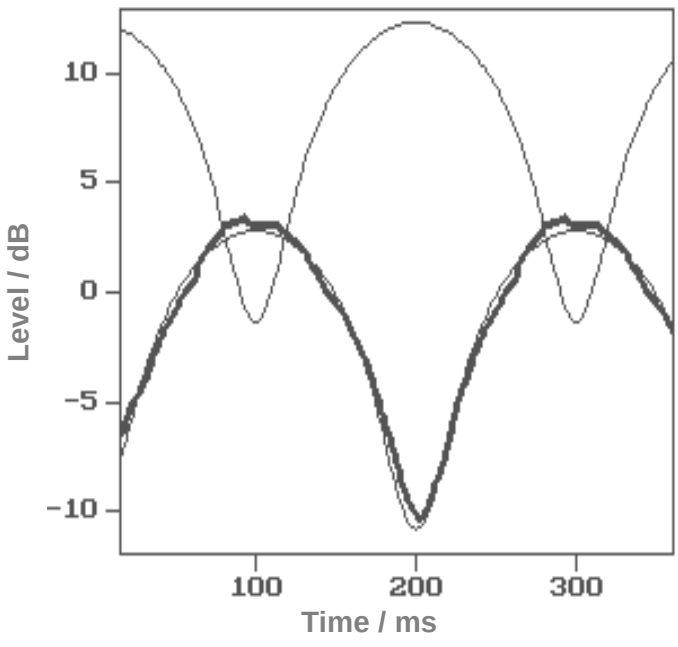


Abb. 7:
Richtungsfilterung von amplitudenmodulierten Signalen.
Nutzsignal: Pegel = 0dB
Sinus 560Hz, f_{mod}=5 Hz, τ=400µs

Störsignal Pegel=10dB
Sinus 500 Hz, f_{mod}=5 Hz, τ=0µs

—— *Hüllkurven von Nutz- und Störsignal*
━━ *Schätzer für die Hüllkurve des Nutzsignals*

x-Achse: Zeit in ms
y-Achse: Pegel in dB,
relativ zum mittleren Nutzsignalpegel

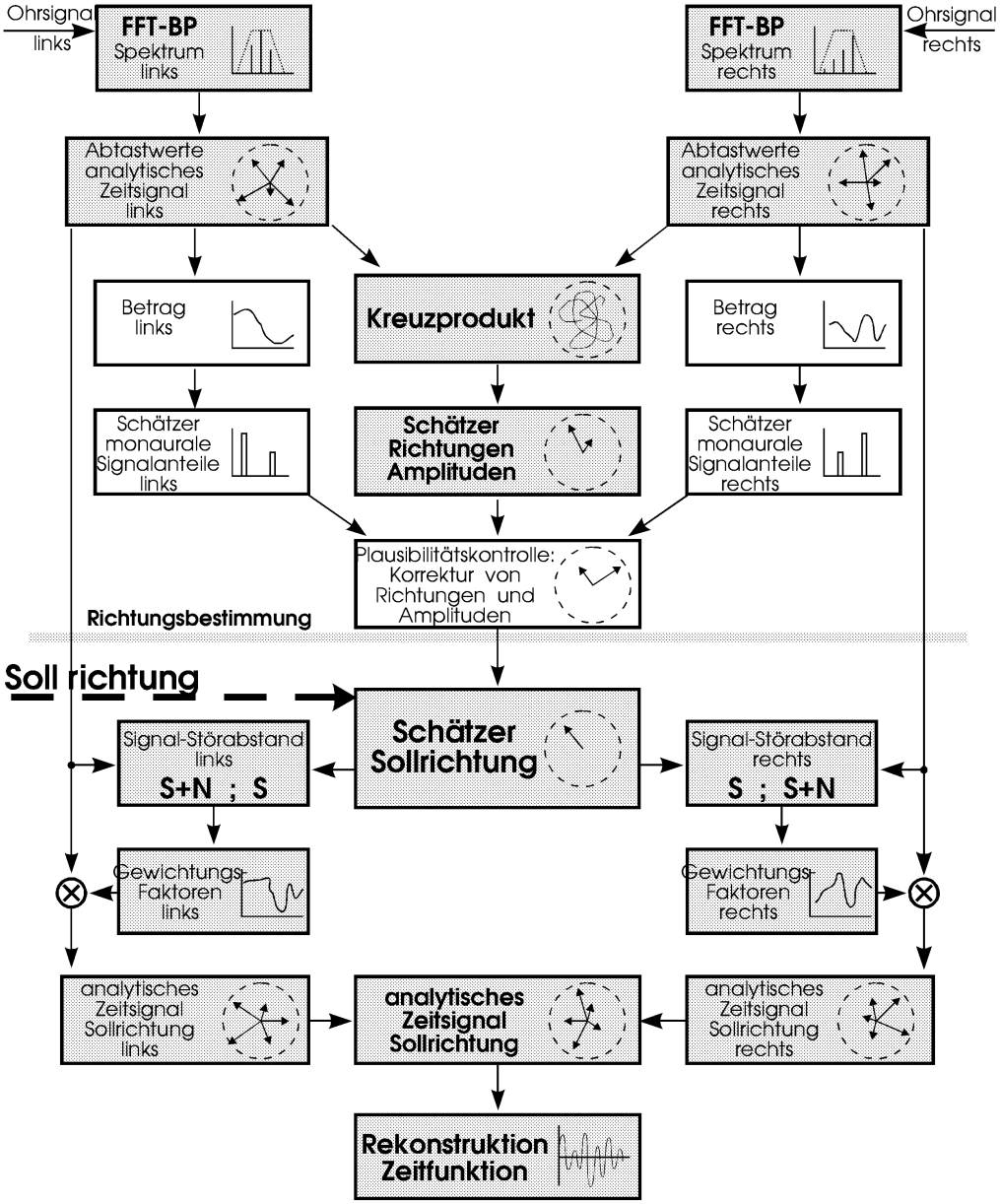


Abb.: 8: Aufbau des binauralen Signalver-arbeitungs-modells (für eine Frequenz-gruppe)

Aufbau eines binauralen Signalverarbeitungsmodells

Zum Aufbau eines binauralen Signalverarbeitungsmodells, das auf den oben beschriebenen Algorithmen basiert, sind folgende Bestandteile erforderlich:

- Signalvorverarbeitung: Frequenzgruppen-Filterung der Ohrsignale und Bildung der analytischen Zeitsignale.
- Bestimmung des Kreuzprodukts und der komplexen Mittelwerte und Standardabweichungen
- Schätzung der Schallquellen-Richtungen und -Amplituden aus den statistischen Parametern des Kreuzprodukts, Abschätzung des Fehlers und des Gültigkeitsbereichs der Schätzer.
- Entscheidung, welche Richtung als Nutzrichtung gewählt werden soll.
- Bestimmung der wahrscheinlichen Amplitude des Nutzsignals aus Schätzwert und Schätzfehler (Wahrscheinlichkeit, dass Schätzer zur Nutzrichtung gehört).
- Bestimmung des Signal-Störabstands für jedes Ohrsignal aus dem Schätzer der Nutzrichtung und den Amplituden der Ohrsignale => Gewichtungsfaktoren für die Ohrsignale.
- Bildung eines breitbandigen verarbeiteten Signals aus den gewichteten Signalen der einzelnen Frequenzgruppen.

Mit Hilfe dieses Prozesses lassen sich für 2 Sprecher im Freifeld bei Signal-Störabständen bis zu -30 dB Verbesserungen des Signal-Störabstands von bis zu 20 dB für den Nutzsprecher erzielen. Die Wahrnehmbarkeit und die Verständlichkeit des Nutzsprechers steigen erheblich.

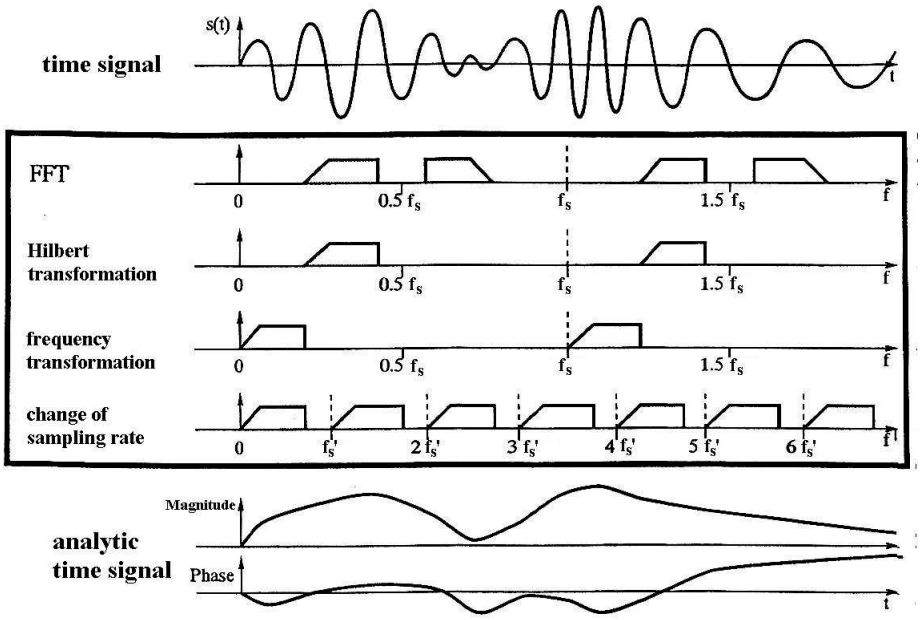


Abb. 9: Signalvorverarbeitung: Erzeugung des analytischen Zeitsignals bei gleichzeitiger Reduktion der Abtastrate

Durch die Verarbeitung analytischer Zeitsignale an Stelle von reellen Zeitfunktionen lässt sich die Datenrate und damit der Rechenaufwand stark reduzieren. Da alle Frequenzlinien im Bereich $f_S/2..f_S$ zu null werden (f_S=Abtastrate), können die Frequenzgruppen-gefilterten Signale zu niedrigen Frequenzen herunter transformiert werden und mit einer Abtastrate verarbeitet werden, die der Bandbreite der Frequenzgruppe entspricht. Gegenüber einer Standard-Digitalfilterbank lässt sich bei 24 Frequenzgruppen eine Reduktion der Datenrate auf 10-20% erreichen.

Ausblick

Dieser Algorithmus beruht auf der Auswertung interauraler Phasen. Bei höheren Frequenzen (f>800 Hz) ist die Beziehung zwischen Schalleinfallsrichtung und interauraler Phase nicht mehr eindeutig. Treffen dann in einzelnen Frequenzgruppen die interauralen Phasen von Nutz- und Störrichtung zusammen, ist eine Richtungsfilterung keinen Effekt mehr. Dieses Problem könnte dadurch gelöst werden, dass man einen Richtungsfiltermechanismus hinzunimmt, der auf der Auswertung interauraler Pegeldifferenzen basiert.

Dieses Modell könnte in der Psychoakustik benutzt werden, Mehr-Quellen-Phänomene zu interpretieren, aber auch den Präzedenz-Effekt. Eine zusätzliche "Richtungserkennungs-Stufe" würde hierbei entscheiden, welche Richtung als Grundlage der Verarbeitung dienen soll. Andere Richtungen (z.B. Echos) könnten als Störsignale markiert und ausgeblendet werden. Ausnahmen des Präzedenz-Effekts ließen sich als das Wählen einer neuen Nutzrichtung deuten.

Mehrfach-Hörereignisse, die entstehen, wenn interaurale Zeit- und Pegeldifferenzen nicht zueinander passen (GAIK[1]), wären durch das Modell interpretierbar als Folge inkompatibler Richtungsschätzer aus der Auswertung von Phasen- und Pegeldifferenzen.

Technische Anwendungen eines Richtungsfilters könnten sein: richtungsselektive Hörgeräte, richtungsselektive Front-Ends für Sprachverarbeitungs-Systeme (Spracherkenner, Freisprechtelefone) oder als niederfrequente Ergänzung für Richtmikrofone und Mikrofon-Arrays.

---

1 GAIK(1990); Untersuchungen zur binauralen Verarbeitung kopfbezogener Signale; Fortschritts-Berichte VDI, Reihe 17: Biotechnik, nr.63; VDI-Verlag, Düsseldorf

2 LINDEMANN(1986): Extensions of a binaural cross-correlation model by contralateral inhibition; JASA 80; p.1608

3 SLATKY (1990); Lokalisation simultan abstrahlender Schallquellen: Konsequenzen für den Aufbau binauraler Modelle; Fortschritte der Akustik DAGA'90, Wien; DPG-Verlag, Bad Honnef, Germany , p.751

4 Basiert auf: SLATKY(1991); Ein binaurales Modell zur Lokalisation und Signalverarbeitung bei Darbietung mehrerer Schallquellen; Fortschritte der Akustik DAGA'91, Bochum; DPG-Verlag, Bad Honnef, Germany