

4. Grundlagen der binauralen Signalverarbeitung

Dieses Kapitel soll die Rahmenbedingungen für die weiteren modellmäßigen Beschreibungen festlegen, und enthält z.B. die Beschreibung der Ohrsignale, Formelzeichen und Nomenklatur.

Mit diesen Voraussetzungen sollen binaurale Modelle beschrieben werden und Bedingungen abgeleitet werden, unter denen Schallquellen-Richtungen und -Leistungen bei der Präsentation mehrerer Quellen bestimmt werden können. Anhand eines aus der Kreuzkorrelationsfunktion abgeleiteten Frequenzbereichs-Modells werden Möglichkeiten zur Konstruktion eines Cocktail-Party-Prozessors aufgezeigt.

Im folgenden werden Übertragungsfunktionen und Ohrsignale innerhalb von Frequenzgruppen betrachtet. Hierbei wird meist von "kopfbezogenen", "binauralen" Übertragungsfunktionen ausgegangen, wobei die Betrachtung von natürlichen Köpfen oder von Kunstköpfen eher als (relativ kompliziertes) Beispiel zu verstehen ist. Andere Aufnahmesysteme wie Stereomikrofon-Anordnungen oder mikrofonbesetzte Gegenstände (z.B. Telefongehäuse mit Mikrofonen) lassen sich auf dieselbe Art und Weise beschreiben.

Als grundlegende richtungsbestimmte Größe wird im folgenden die interaurale Laufzeit τ betrachtet. Andere binaurale Größen werden auf τ bezogen.

Es wird davon ausgegangen, daß sich der Empfänger (Ohr/Mikrofon) immer im Fernfeld der Schallquellen befindet und daß der Empfängerabstand klein gegen die Quellenentfernung ist, so daß die Wellenfronten einer Punktquelle am Empfänger eben sind und senkrecht zur Strecke Quelle-Empfänger stehen.

4.1. Die Übertragungstrecke Schallquelle-Ohr

Kopfbezogene Signale und Außenohr-Übertragungsfunktionen

Bei Beschallung eines Kopfes durch eine Schallquelle im Freifeld kann die Übertragungstrecke von der Quelle zum linken bzw. rechten Ohr durch die Außenohr-Übertragungsfunktionen $\underline{H}_{ql}(f, \tau)$, $\underline{H}_{qr}(f, \tau)$ beschrieben werden. Diese setzen sich jeweils zusammen aus der Übertragungsfunktion Schallquelle-Kopfposition $\underline{H}_{qk}(f)$ und den Freifeld-Außenohr-Übertragungsfunktionen $\underline{H}_{lf}(f, \tau)$, $\underline{H}_{rf}(f, \tau)$. Sind lediglich die Ohrsignale und Freifeld-Außenohr-Übertragungsfunktionen bekannt und keine Informationen über den Übertragungsweg Schallquelle-Kopfposition gegeben, kann statt des Signals am Ort der Schallquelle $\underline{A}_q(f)$ nur das Freifeld-Schallsignal am Ort des Kopfes $\underline{A}(f)$ bestimmt werden.

In den folgenden Betrachtungen werden die Freifeld-Außenohr-Übertragungsfunktionen (Übertragungstrecke Kopfposition-Ohr) $\underline{H}_{lf}(f, \tau)$, $\underline{H}_{rf}(f, \tau)$ aufgespalten in Übertragungstrecken zu einem Bezugspunkt "Kopfmittle" mit den Übertragungsfunktionen $\underline{H}_m(f, \tau)$, $\underline{H}_l(f, \tau)$ und $\underline{H}_r(f, \tau)$ (siehe Bild 4.1). "Kopfmittle" bezeichnet hierbei einen konstruierten Punkt, an dem sich die interaurale Übertragungsfunktion $\underline{H}_{rl}(f, \tau)$ so aufspalten läßt, daß zwei zueinander reziproke Übertragungsfunktionen $\underline{H}_l(f, \tau)$ und $\underline{H}_r(f, \tau)$ entstehen, die die Übertragung von der "Kopfmittle" zu den Ohren beschreiben:

$$\underline{H}_r(f, \tau) = 1/\underline{H}_l(f, \tau) = \sqrt{\underline{H}_{rl}(f, \tau)}$$

$\underline{H}_m(f, \tau)$ beschreibt die Übertragungstrecke Freifeld-"Kopfmittle". Durch die Wahl auf "Kopfmittle" bezogener Übertragungsfunktionen ergeben sich erhebliche algorithmische Vereinfachungen.

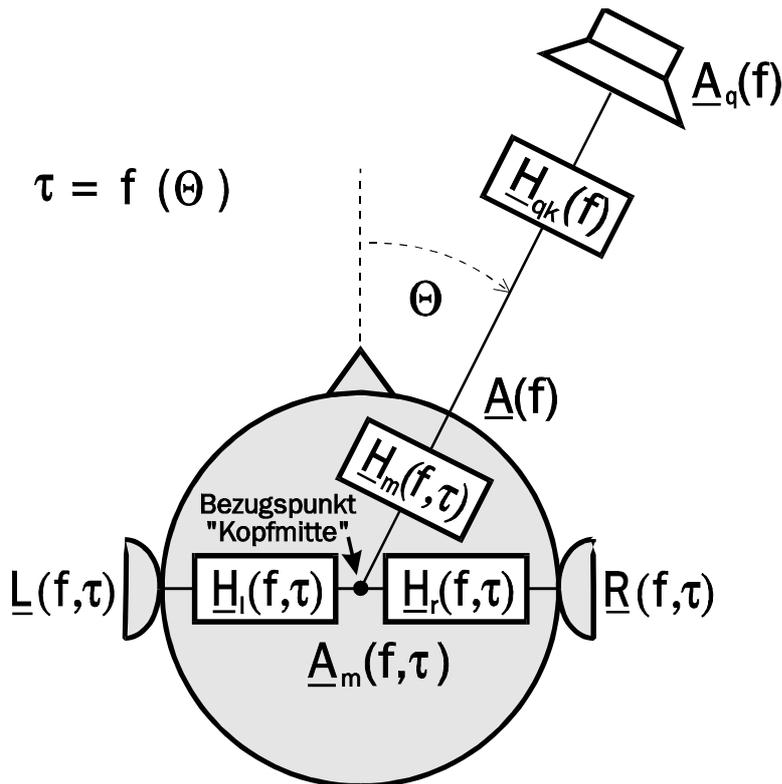


Bild 4.1:
Beschreibung der
Übertragungsstrecke
Schallquelle-Kopf

Alle Übertragungsfunktionen sind sowohl von der Frequenz als auch vom Einfallswinkel abhängig. Im folgenden werden Abhängigkeiten vom Einfallswinkel durch Abhängigkeiten von der interauralen Zeitdifferenz τ dargestellt. Die Übertragungsstrecken ergeben sich somit wie folgt:

Kopfposition-"Kopfmittle":
$$\underline{H}_m(f, \tau) = \sqrt{\underline{H}_{lf}(f, \tau) \underline{H}_{rf}(f, \tau)}$$

"Kopfmittle"-Ohren:
$$\underline{H}_l(f, \tau) = \sqrt{\underline{H}_{lf}(f, \tau) / \underline{H}_{rf}(f, \tau)}$$

$$\underline{H}_r(f, \tau) = \sqrt{\underline{H}_{rf}(f, \tau) / \underline{H}_{lf}(f, \tau)}$$
 (4.1/1)

Ohr-Ohr (interaurale Übertragungsfunktion)

$$\underline{H}_{rl}(f, \tau) = \underline{H}_{rf}(f, \tau) / \underline{H}_{lf}(f, \tau)$$

Kopfposition-Ohren (Freifeld-Außenohr-Übertragungsfunktionen)

$$\underline{H}_{lf}(f, \tau) \text{ bzw. } \underline{H}_{rf}(f, \tau)$$

Schallquelle-Kopfposition (Freifeld-Übertragungsfunktion)

$$\underline{H}_{qk}(f, \tau)$$

In den folgenden Darstellungen werden, der einfacheren Darstellung halber, auf Kopfmittle transformierte Signale (z.B. $\underline{A}_m(f, \tau)$) betrachtet. Hieraus läßt sich mit Hilfe der Freifeld-Außenohr-Übertragungsfunktionen leicht das Freifeld-Signal $\underline{A}(f)$ bestimmen.

$$\underline{A}(f) = \underline{H}_m(f, \tau)^{-1} \underline{A}_m(f, \tau)$$

Die Übertragungsfunktion Kopf-Kopfmittle läßt sich in Betrag und Phase aufspalten, wobei der Betrag die mittlere Dämpfung α_0 und die Phase die mittlere Laufzeit τ_0 der Freifeld-Außenohr-Übertragungsfunktionen widerspiegelt.

$$\underline{H}_m(f, \tau) = e^{\alpha_o(f, \tau) + j2\pi f \tau_o(f, \tau)}$$

Wird statt eines (Kunst-)Kopfes ein *Mikrofonpaar* (punktförmige Mikrofone) im Freifeld als Empfänger verwendet und ist der Abstand Quelle-Empfänger groß gegen den Empfängerabstand (ebene Wellen), so werden in diesem Spezialfall die Beträge der Übertragungsfunktionen $\underline{H}_m(f, \tau)$, $\underline{H}_l(f, \tau)$, $\underline{H}_r(f, \tau)$ zu Eins. Die Phasen sind dann proportional zum jeweiligen Wegunterschied der einfallenden Welle:

$$\begin{aligned} \underline{H}_{lo}(f, \tau) &= e^{-j\pi f \tau} & \underline{H}_{ro}(f, \tau) &= e^{+j\pi f \tau} & \underline{H}_{mo}(f, \tau) &= 1 \\ \tau &= d/c_{\text{schall}} \sin \theta & & & d &= \text{Mikrofonabstand, } \theta = \text{Einfallswinkel} \\ & & & & c_{\text{schall}} &= \text{Schallgeschwindigkeit} \end{aligned}$$

Interaurale Phase und interaurale Dämpfung

Die Beträge der Funktionen $\underline{H}_l(f, \tau)$ und $\underline{H}_r(f, \tau)$ sind abhängig von der interauralen Pegeldifferenz ΔL , die Phase von der interauralen Zeitdifferenz τ :

$$\begin{aligned} \underline{H}_r(f, \tau) &= 10^{+\Delta L(f, \tau)/40\text{dB}} e^{+j\pi f \tau(f)} \\ \underline{H}_l(f, \tau) &= 10^{-\Delta L(f, \tau)/40\text{dB}} e^{-j\pi f \tau(f)} \end{aligned}$$

Hiermit lassen sich eine *interaurale Dämpfung* $\alpha(f, \tau)$ und eine *interaurale Phase* $\beta(f, \tau)$ definieren, mit denen diese beiden Übertragungsfunktionen wesentlich einfacher beschrieben werden können:

$$\begin{aligned} \alpha(f, \tau) &= \Delta L(f, \tau) \ln(10) / 20\text{dB} \\ \beta(f, \tau) &= 2\pi f \tau(f) \end{aligned} \quad (4.1/2)$$

$$\begin{aligned} \underline{H}_r(f, \tau) &= e^{+1/2 \alpha(f, \tau) + j1/2 \beta(f, \tau)} \\ \underline{H}_l(f, \tau) &= e^{-1/2 \alpha(f, \tau) - j1/2 \beta(f, \tau)} \end{aligned} \quad (4.1/3)$$

Ohrsignale

Ist das Signal einer Schallquelle an der Position des Kopfes unter Freifeld-Bedingungen $\underline{A}(f)$ gegeben, ergeben sich die Ohrsignale im *Frequenzbereich* mit den interauralen Parametern der Signale der Quelle $\alpha_a(f) = \alpha(f, \tau_a)$ und $\beta_a(f) = \beta(f, \tau_a)$ zu:

$$\begin{aligned} \underline{R}(f, t) &= \underline{A}_m(f, \tau) e^{+1/2 \alpha_a(f) + j1/2 \beta_a(f)} & ; \underline{A}_m(f, t) &= \underline{A}(f) \underline{H}_m(f, \tau_a) \\ \underline{L}(f, t) &= \underline{A}_m(f, \tau) e^{-1/2 \alpha_a(f) - j1/2 \beta_a(f)} \end{aligned} \quad (4.1/4)$$

Im *Zeitbereich* ergeben sich die Ohrsignale $r(t)$ und $l(t)$ durch Faltung des Freifeld-Schallsignals $a(t)$ mit den Impulsantworten der entsprechenden Übertragungsfunktionen:

$$\begin{aligned} r(t) &= a_m(t) * h_r(t) \\ l(t) &= a_m(t) * h_l(t) & ; a_m(t) &= a(t) * h_m(t) \end{aligned} \quad (4.1/5)$$

Sind für die betrachtete Einfallrichtung die Freifeld-Außenohr-Übertragungsfunktionen (d.h. interaurale und mittlere Laufzeit τ , τ_o , interaurale und mittlere Dämpfung α , α_o) innerhalb einer Frequenzgruppe frequenzunabhängig, so kann die Faltung mit den Ohr-Impulsantworten durch eine Multiplikation mit einem konstanten Faktor und eine Zeitverschiebung der Signale dargestellt werden:

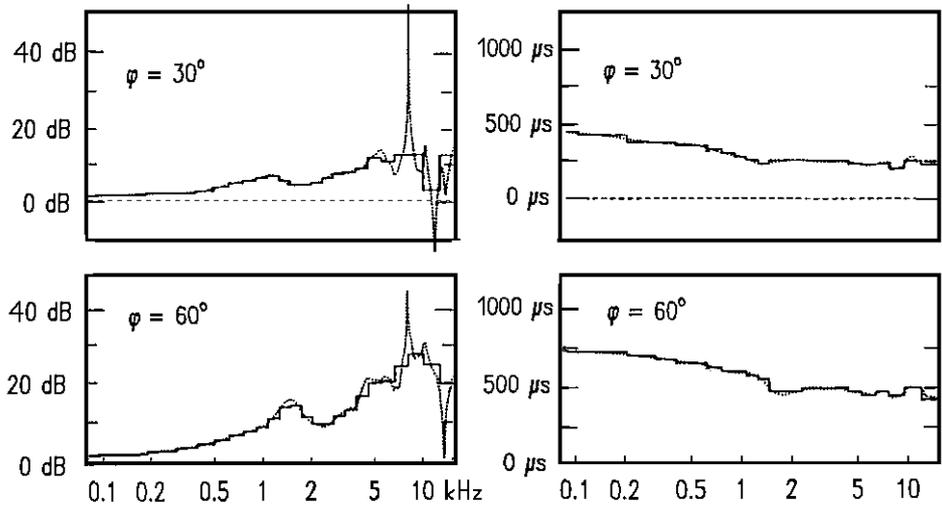


Bild 4.2: Vergleich gemessener interauraler Pegeldifferenzen und Phasenlaufzeiten mit über jeweils eine Frequenzgruppe gemittelten interauralen Parametern für die Azimutalwinkel 30° und 60° . Bis zu Frequenzen von ca. 7 kHz lassen sich die interauralen Parameter in Frequenzgruppen hinreichend genau durch konstante Werte beschreiben (nach GAIK [20]).

..... gemessene Werte
 ----- frequenzgruppen-gemittelte Werte

$$r(t) = a(t - \tau_o(\tau) + \frac{1}{2}\tau_a) e^{\alpha_o(\tau) + \frac{1}{2}\alpha_a} \quad \text{für } \alpha, \alpha_o, \tau, \tau_o \neq f(f)$$

$$l(t) = a(t - \tau_o(\tau) - \frac{1}{2}\tau_a) e^{\alpha_o(\tau) - \frac{1}{2}\alpha_a} \quad (4.1/6)$$

Bei niedrigen Frequenzen ist diese Voraussetzung in erster Näherung erfüllt (siehe Bild 4.2). Gaik [20] hat für einen großen Frequenzbereich konstante Beziehungen zwischen interauralen Parametern aufgestellt: Für jede Frequenzgruppe ließ sich, gemittelt über unterschiedliche Elevationswinkel, die interaurale Pegeldifferenz ΔL als frequenzunabhängige Funktion der interauralen Zeitdifferenz beschreiben, wobei die Abweichungen dieser mittleren Funktion gegenüber Meßwerten relativ gering blieben.

Das analytische Zeitsignal

Die im folgenden meist verwendete Darstellungsart ist die Beschreibung durch *analytische Zeitsignale*. Für positive Frequenzen stimmen die Fourier-Transformierten von analytischem Zeitsignal $\underline{A}_a(f)$ und reellem Zeitsignal $\underline{A}(f)$ überein. Für negative Frequenzen ist die Fourier-Transformierte des analytischen Zeitsignals Null:

$$\underline{A}_a(f) = \underline{A}(f)$$

$$\underline{A}_a(-f) = 0 \quad ; f > 0$$

Wegen der Symmetrieeigenschaften der Fourier-Transformation für reelle und rein imaginäre Funktionen

$$\underline{X}(f) = \underline{X}(-f) \quad \text{für } \underline{X}(f) = \mathcal{F}\{x(t)\}; x(t) \text{ reell}$$

$$\underline{X}'(f) = -\underline{X}'(-f) \quad \text{für } \underline{X}'(f) = \mathcal{F}\{j x(t)\}; x(t) \text{ reell}$$

ergibt sich die Fourier-Transformierte des analytischen Zeitsignals $\underline{A}_a(f)$:

$$\underline{A}_a(f) = \mathcal{F}\{\text{Re}\{\underline{a}(t)\}\}_{(f>0)} + \mathcal{F}\{\text{Im}\{\underline{a}(t)\}\}_{(f>0)}$$

$$\underline{A}_a(-f) = \mathcal{F}\{\text{Re}\{\underline{a}(t)\}\}_{(f>0)} - \mathcal{F}\{\text{Im}\{\underline{a}(t)\}\}_{(f>0)} \quad ; f > 0$$

Somit entspricht der Realteil des analytischen Zeitsignals $\underline{a}(t)$ dem halben reellen Zeitsignal $\frac{1}{2}a(t)$, und der Imaginärteil ergibt sich aus der Fourier-Transformierten des reellen Zeitsignals $\mathcal{F}\{a(t)\}$ wie folgt:

$$\operatorname{Im}\{\underline{a}(t)\} = \mathcal{F}^{-1}\{\mathcal{F}\{a(t)\} \underline{H}_a(f)\} \quad \text{mit} \quad \begin{array}{ll} \underline{H}_a(f) = j\frac{1}{2} & \text{für } f > 0 \\ \underline{H}_a(f) = 0 & \text{für } f = 0 \\ \underline{H}_a(f) = -j\frac{1}{2} & \text{für } f < 0 \end{array}$$

Für die Darstellung der Übertragungseigenschaften des Kopfes ergeben sich bei Verwendung von analytischen Zeitsignalen der Ohrsignale $\underline{r}(t), \underline{l}(t)$ ähnliche formelmäßige Zusammenhänge wie für reelle Zeitsignale. Bezeichnet man mit $\underline{h}(t)$ das zur Impulsantwort $h(t)$ einer Übertragungsfunktion $\underline{H}(f)$ zugehörige analytische Zeitsignal, ergibt sich:

$$\begin{aligned} \underline{r}(t) &= \underline{a}_m(t) * \underline{h}_r(t) \\ \underline{l}(t) &= \underline{a}_m(t) * \underline{h}_l(t) \quad ; \quad \underline{a}_m(t) = \underline{a}(t) * \underline{h}_m(t) \end{aligned} \quad (4.1/7)$$

Analog zu Formel 4.1/6 lassen sich bei innerhalb einer Frequenzgruppe konstanten Übertragungsfunktionen $\underline{H}_r(f), \underline{H}_l(f), \underline{H}_m(f)$ die Ausdrücke für die analytischen Zeitsignale der Ohrsignale vereinfachen:

$$\begin{aligned} \underline{r}(t) &= \underline{a}_m(t + \frac{1}{2}\tau) e^{+\frac{1}{2}\alpha_a} \\ \underline{l}(t) &= \underline{a}_m(t - \frac{1}{2}\tau) e^{-\frac{1}{2}\alpha_a} \quad \text{für } \alpha_a, \tau \neq \text{Funktion von } f \end{aligned} \quad (4.1/8)$$

Die Phase des analytischen Zeitsignals läßt sich über eine zeitabhängige Momentankreisfrequenz Ω bzw. Momentanphase Φ beschreiben:

$$\arg\{\underline{a}_m(t)\} = \Omega_a(t) t + \Phi_a(t)$$

Mit $\beta_a = \Omega_a \tau$, sowie $|\underline{a}_m(t \pm \frac{1}{2}\tau)| \approx |\underline{a}_m(t)|$ vereinfacht sich Formel 4.1/8 zu:

$$\begin{aligned} \underline{r}(t) &= |\underline{a}_m(t)| e^{j\Omega_a(t)t + j\Phi_a(t)} e^{+\frac{1}{2}\alpha_a + j\frac{1}{2}\beta_a} \\ \underline{l}(t) &= |\underline{a}_m(t)| e^{j\Omega_a(t)t + j\Phi_a(t)} e^{-\frac{1}{2}\alpha_a - j\frac{1}{2}\beta_a} \end{aligned} \quad (4.1/9)$$

Die Näherungen für Formel 4.1/8 und 4.1/9 gelten nicht im Bereich hoher Frequenzen, wo Informationen über die Einfallswinkel zusätzlich über Minima und Maxima der Freifeld-Außenohr-Übertragungsfunktion kodiert werden und sich die interauralen Parameter innerhalb einer Frequenzgruppe ändern (vergl. Bild 4.2). Über die Frequenz gemittelte interaurale Parameter werden dann abhängig vom Signalspektrum.

4.2. Binaurale Informationen bei Anwesenheit mehrerer Schallquellen

Ohrsignale bei zwei Schallquellen

Sind mehrere Schallquellen anwesend, überlagern sich die Schallsignale der einzelnen Quellen. Bei zwei Quellen a und b gilt analog zu Formel 4.1/9 für die analytischen Zeitsignale der Ohrsignale am Bezugspunkt "Kopfmitte" bei innerhalb einer Frequenzgruppe frequenzunabhängigen Freifeld-Außenohr-Übertragungsfunktionen:

$$\begin{aligned} \underline{r}(t) &= a_m(t) e^{+\frac{1}{2}\alpha_a + j\frac{1}{2}\beta_a} e^{j\Omega_a t + j\Phi_a} + b_m(t) e^{+\frac{1}{2}\alpha_b + j\frac{1}{2}\beta_b} e^{j\Omega_b t + j\Phi_b} \\ \underline{l}(t) &= a_m(t) e^{-\frac{1}{2}\alpha_a - j\frac{1}{2}\beta_a} e^{j\Omega_a t + j\Phi_a} + b_m(t) e^{-\frac{1}{2}\alpha_b - j\frac{1}{2}\beta_b} e^{j\Omega_b t + j\Phi_b} \end{aligned} \quad (4.2/1)$$

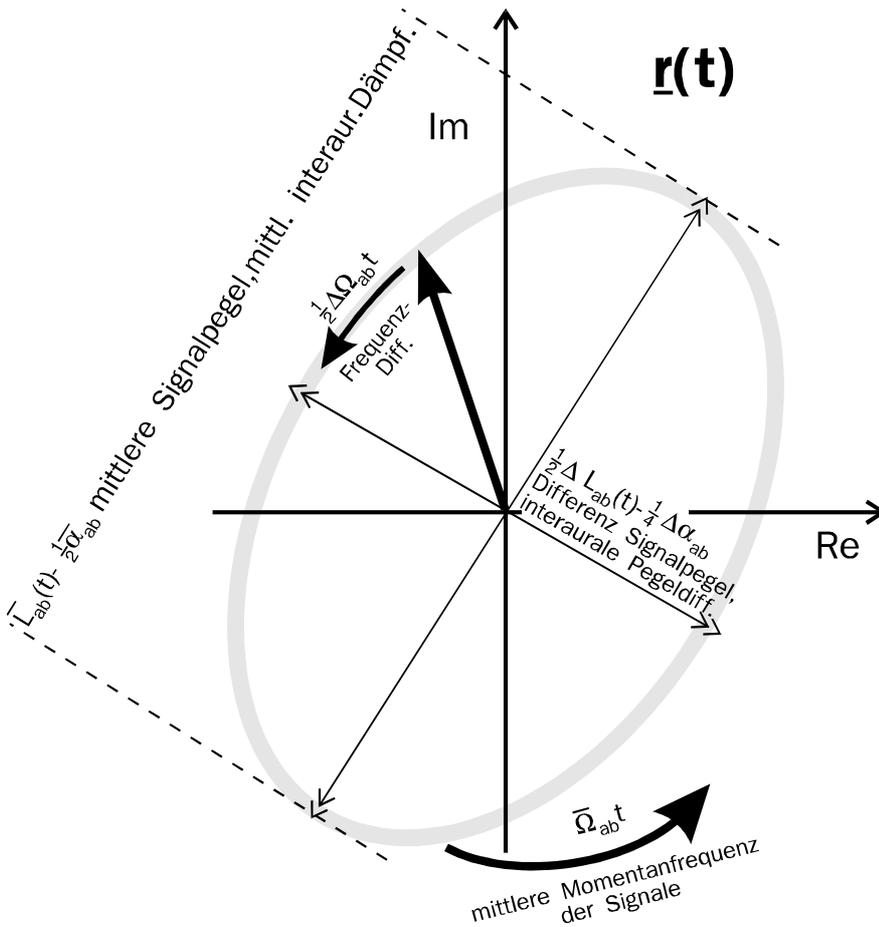


Bild 4.3:
Zeit-Ortskurve der
analytischen Zeitsignale
der Ohrsignale.
Hier: rechtes Ohrsignal $\underline{r}(t)$

Hieraus folgt:

$$\underline{r}(t) = 2 \sqrt{a_m(t) b_m(t)} e^{+1/4(\alpha_a + \alpha_b)} e^{j1/2[(\Omega_a + \Omega_b)t + \Phi_a + \Phi_b + 1/2(\beta_a + \beta_b)]} \cosh\left(\ln \sqrt{a_m(t)/b_m(t)} + 1/4(\alpha_a - \alpha_b) + j1/2[(\Omega_a - \Omega_b)t + \Phi_a - \Phi_b + 1/2(\beta_a - \beta_b)]\right)$$

$$\underline{l}(t) = 2 \sqrt{a_m(t) b_m(t)} e^{-1/4(\alpha_a - \alpha_b)} e^{j1/2[(\Omega_a + \Omega_b)t + \Phi_a + \Phi_b - 1/2(\beta_a - \beta_b)]} \cosh\left(\ln \sqrt{a_m(t)/b_m(t)} - 1/4(\alpha_a - \alpha_b) + j1/2[(\Omega_a - \Omega_b)t + \Phi_a - \Phi_b - 1/2(\beta_a - \beta_b)]\right) \quad (4.2/2)$$

Die Beschreibung der Ohrsignale besteht jeweils aus einem Term, der von den mittleren Quellsignalen und den mittleren interauralen Parametern gebildet wird, und einem "cosh"-Ausdruck, der von den Unterschieden der Signale sowie den Unterschieden der interauralen Parameter beeinflusst wird.

Die cosh-Funktion eines komplexen Arguments mit zeitveränderlichem Imaginärteil ergibt als Zeit-Ortskurve in der komplexen Ebene eine Ellipse; der Realteil des Arguments bestimmt hierbei das Hauptachsenverhältnis, der Imaginärteil den Winkel, unter dem ein Punkt der Ellipse vom Ellipsen-Schwerpunkt aus gesehen wird. Da der cosh-Term mit einer Funktion der Form $e^{j\Omega t + \Phi}$ multipliziert wird, also einer Funktion, die eine zeitabhängige Rotation in der komplexen Ebene durchführt, ergeben sich die Ohrsignale, in der komplexen Ebene als Funktion der Zeit dargestellt, als um den Ursprung rotierende Ellipsen: (Bild 4.3)

Die Größe und Position der Ellipse ist abhängig von den mittleren Signalpegeln und mittleren interauralen Dämpfungen. (konstanter Faktor $2(a_m(t)b_m(t)e^{\pm 1/2\alpha_a \pm 1/2\alpha_b})^{1/2}$) Die Neigung zum Zeitpunkt $t=0$ ist abhängig von der mittleren Signalphase und der mittleren interauralen Laufzeitdifferenz. (Faktor $e^{j(\Phi_a + \Phi_b \pm 1/2\beta_a \pm 1/2\beta_b)}$)

Das Hauptachsenverhältnis hängt von den Unterschieden der Signalpegel und der interauralen Pegeldifferenzen ab. (Realteil des cosh-Terms)

Der (Null-)Phasenwinkel ist wiederum von der Differenz der interauralen Phasen und Signalphasen abhängig. (zeitunabhängiger Teil des Imaginärteils des cosh-Terms)

Die Rotationsfrequenz der Ellipse entspricht der mittleren Momentanfrequenz der Signale. (Faktor $e^{j/2(\Omega_a+\Omega_b)t}$)

Die Kreisfrequenz, mit der die Ellipse durchlaufen wird, hängt von der Differenz der Momentanfrequenzen der Signale ab. (zeitabhängiger Teil des Imaginärteils des cosh-Terms)

Die Ellipse rotiert um den Ursprung mit der mittleren Signalfrequenz und wird dabei gleichzeitig mit der Differenz der Signalfrequenzen durchlaufen. Da hier die Rotationsfrequenz größer ist als die Umlauffrequenz der Ellipse, wird man als Zeit-Ortskurve eher eine Spirale erhalten.

Der interaurale Quotient

Die folgenden Überlegungen sollen Aufschluß darüber geben, ob bei Anwesenheit zweier Schallquellen die Ergebnisse der Hörversuche durch eine einfache Analyse interauraler Parameter nachvollzogen werden können oder ob zusätzliche Signalverarbeitungsschritte erforderlich werden.

Folgende interaurale Parameter werden, jeweils in Frequenzgruppen, untersucht:

- *Interaurale Gruppenlaufzeiten* entsprechen den Laufzeiten zwischen korrespondierenden Zeitpunkten der Signalhüllkurve. Der Verlauf der Signalhüllkurve entspricht dem Verlauf des Betrags des analytischen Zeitsignals.
- *Interaurale Phasenlaufzeiten* entsprechen den Laufzeiten zwischen korrespondierenden Zeitpunkten gleicher Phase. Dies entspricht der Phase des analytischen Zeitsignals.
- *Interaurale Pegeldifferenzen* entsprechen den Pegelunterschieden zwischen den Ohrsignalen. Dies entspricht dem Quotienten der Beträge des analytischen Zeitsignals.

Die interauralen Pegeldifferenzen ΔL und die Phasen-Laufzeitdifferenzen τ_{Φ} lassen sich über den Quotienten der analytischen Zeitsignale der Ohrsignale beschreiben. Dieser Quotient wird im folgenden *interauraler Quotient* $\underline{d}(t)$ genannt:

$$\underline{d}(t) = \frac{\underline{r}(t)}{\underline{l}(t)} = \frac{|r(t)|}{|l(t)|} e^{j \arg\{r(t)\} - j \arg\{l(t)\}}$$

$$\Delta L(t) = 20 \text{ dB } \lg |\underline{d}(t)| \qquad \tau_{\Phi}(t) = \arg\{ \underline{d}(t) \} / \Omega \qquad (4.2/3)$$

Bei einer Schallquelle mit der Einfallrichtung θ lassen sich die Ohrsignale $\underline{r}(t)$ und $\underline{l}(t)$ innerhalb einer Frequenzgruppe gemäß Formel (4.1/9) beschreiben. Der interaurale Quotient $\underline{d}(t)$ entspricht dann der interauralen Übertragungsfunktion dieser Einfallrichtung.

Interaurale Gruppenlaufzeiten bei zwei Schallquellen

Die interauralen Gruppenlaufzeiten entsprechen den Laufzeiten zwischen zwei entsprechenden Maxima der Hüllkurve, das heißt zwischen entsprechenden Betrags-Maxima der analytischen Zeitsignale der Ohrsignale. Solche Betrags-Maxima treten zu den Zeitpunkten t_r (rechts) und t_l (links) auf, in denen der Phasenwinkel der obigen Ellipse zu einem geraden bzw. ungeraden Vielfachen von $\pi/2$ wird, also der Imaginär- bzw. Realteil des "cosh" zu Null wird. "Gerade" oder "ungerade" hängt davon ab, ob das Hauptachsenverhältnis größer oder kleiner als Eins ist. Aus der Bedingung $\text{Im}\{ \text{Arg}\{\cosh(\dots)\} \} = j\pi/2$ erhält man aus Formel 4.2/2:

$$t_r = \frac{i_r \pi/2 - (\Phi_a - \Phi_b) - 1/2(\beta_a - \beta_b)}{\Omega_a - \Omega_b} \quad i_r \in \mathbf{G} \text{ bzw. } i_r \in \mathbf{U}$$

$$t_l = \frac{i_l \pi/2 - (\Phi_a - \Phi_b) + 1/2(\beta_a - \beta_b)}{\Omega_a - \Omega_b} \quad i_l \in \mathbf{G} \text{ bzw. } i_l \in \mathbf{U}$$

Mit $\beta = \Omega\tau$ erhält man für die interaurale Zeitdifferenz zwischen den Hüllkurven-Maxima:

$$t_r - t_l = \frac{(i_r - i_l)\pi/2 - \Omega_a \tau_a + \Omega_b \tau_b}{\Omega_a - \Omega_b}$$

Sind die Hauptachsenverhältnisse so, daß $i_r = i_l$ ist (Kombination mit minimaler Phasendifferenz), entspricht diese interaurale Gruppenlaufzeit den Unterschieden zwischen den interauralen Phasen der Signale, normiert auf die Differenz der Momentanfrequenzen. Werden Bandpaßfilter mit Bandbreite Δf_{FG} und Mittenfrequenz f_m verwendet (z.B. Frequenzgruppen-Filter), lassen sich die minimalen Gruppenlaufzeiten abschätzen:

$$t_r - t_l = -1/2 \frac{\Omega_a \tau_a + \Omega_b \tau_a - \Omega_a \tau_b - \Omega_b \tau_b + \Omega_a \tau_a - \Omega_b \tau_a + \Omega_a \tau_b - \Omega_b \tau_b}{\Omega_a - \Omega_b}$$

$$t_l - t_r = 1/2 \frac{\Omega_a + \Omega_b}{\Omega_a - \Omega_b} (\tau_a - \tau_b) + 1/2 (\tau_a + \tau_b)$$

$$t_l - t_r \geq f_m / \Delta f_{FG} (\tau_a - \tau_b) + 1/2 (\tau_a + \tau_b) \quad (4.2/4)$$

Die Gruppenlaufzeit-Differenz unterscheidet sich gegenüber der mittleren interauralen Laufzeitdifferenz $1/2(\tau_a + \tau_b)$ um den Term $f_m / \Delta f_{FG} (\tau_a - \tau_b)$. Bei schmalbandigen Signalen wird $f_m / \Delta f_{FG}$ größer als Eins. Somit können schon bei kleinen Einfallswinkel-Unterschieden zwischen den Schallquellen interaurale Gruppenlaufzeiten außerhalb des natürlichen interauralen Laufzeitbereichs entstehen.

Bei Terzfiltern mit ca. 25% relativer Bandbreite würde $f_m / \Delta f_{FG} = 4$. Für zwei Quellen mit den Einfallswinkeln $+20^\circ$ und -20° und normierten interauralen Laufzeiten (Formel 3.1/1) von $\pm 210 \mu s$ entstünden gemäß Formel 4.2/4 interaurale Gruppenlaufzeit-Differenzen von $\pm 840 \mu s$, also außerhalb des natürlich vorkommenden Bereichs (von $\pm 625 \mu s$).

Interaurale Phasen und Pegeldifferenzen bei zwei Schallquellen

Der interaurale Quotient als Maß für die interauralen Parameter ergibt sich bei der Beschallung mit 2 Quellen entsprechend Formel 4.2/1 zu:

$$\underline{d}(t) = e^{1/2(\alpha_a + \alpha_b)} e^{j1/2(\beta_a + \beta_b)} \frac{\cosh\{1/2 \ln(a_m(t)/b_m(t)) + j1/2(\Omega_a t - \Omega_b t + \Phi_a - \Phi_b) + 1/4(\alpha_a - \alpha_b) + j1/4(\beta_a - \beta_b)\}}{\cosh\{1/2 \ln(a_m(t)/b_m(t)) + j1/2(\Omega_a t - \Omega_b t + \Phi_a - \Phi_b) - 1/4(\alpha_a - \alpha_b) - j1/4(\beta_a - \beta_b)\}} \quad (4.2/5)$$

Im zeitlichen Mittel entspricht der interaurale Quotient den mittleren interauralen Phasen und Pegeldifferenzen der beteiligten Quellen.

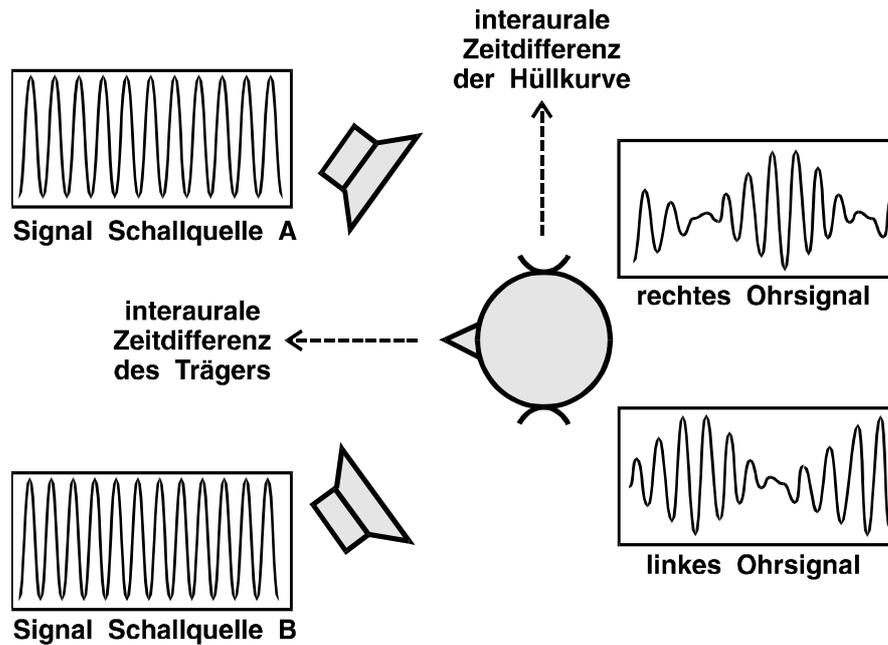


Bild 4.4: Ohrsignale, interaurale Parameter und erwartete Hörereignisse bei 2 sinusförmigen Schallsignalen unterschiedlicher Richtung und Frequenz.

Interaurale Parameter und Ergebnisse der Hörversuche

Geht man von der These aus, daß die wahrgenommene Richtung von interauralen Phasen-, Gruppenlaufzeit- und Pegeldifferenzen bestimmt wird, wären aus der direkten Analyse der interauralen Parameter 2 Hörereignisse zu erwarten: ein auf interauralen Phasen und Pegeldifferenzen basierendes Hörereignis aus der Mitte zwischen beiden Schallquellen und ein auf der interauralen Gruppenlaufzeit basierendes Hörereignis mit sehr großen interauralen Zeitdifferenzen, also von der Seite oder als monaurales Hörereignis (Bild 4.4). Ähnliche Schlüsse ergeben sich bei binauralen Modelle, die auf der Analyse der interauralen Differenzen beruhen:

Bild 3.11 zeigt die Ergebnisse einer Modellsimulation mit dem binauralen Modell nach Lindemann [25]. Dieses Modell wertet die interauralen Differenzen (Phasen-, Gruppenlaufzeit- und Pegeldifferenzen) aus und zeigt durchaus das erwartete Verhalten: Maxima der Kreuzkorrelationsfunktion genau zwischen den Schallquellen-Einfallswinkeln und am Rand des analysierten Bereichs.

Für ungeübte Versuchspersonen oder nur sehr kleine spektrale Differenzen zwischen den Testsignalen entspricht die Analyse der interauralen Parameter durchaus den wahrgenommenen Hörereignissen: Es traten Hörereignisse ganz rechts oder links außen auf (wahrgenommener Einfallswinkel $>70^\circ$ bei Lautsprecher-Auslenkungen $\leq 35^\circ$). Diese Hörereignisse waren unabhängig von der Aufstellung der Lautsprecher (Slatky [39]), die Tonhöhe dieser Hörereignisse wurde oft als sehr tief beschrieben, entsprechend der Periode der Hüllkurve. Ebenso traten auch Hörereignisse aus der Mitte zwischen den Lautsprechern auf. Geübte Versuchspersonen waren dagegen in der Lage, schon bei geringen Frequenzdifferenzen beide Quellen zu lokalisieren.

Die reine Betrachtung der interauralen Differenzen reicht also nicht aus, die Ergebnisse der Hörversuche zu erklären, da anhand der interauralen Parameter nur Hörereignisse aus der Richtung zwischen beiden Quellen oder rechts oder links außen zu erwarten wären, nicht aber, wie in den Hörversuchen, aus der Richtung der Schallquellen.

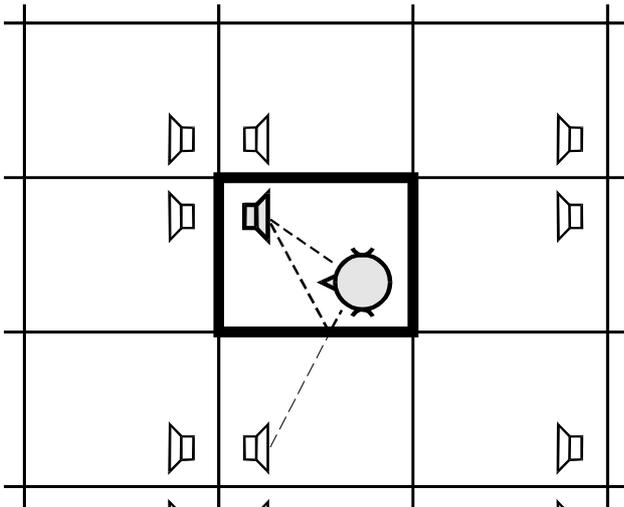


Bild 4.5:
Spiegelschallquellen-Modell:
Beschreibung reflektierter Schallanteile
durch Spiegelschallquellen
- - - reflektierter Schallstrahl
- - - Schallstrahl der entsprechenden
Spiegelschallquelle

Summenlokalisation

Werden zwei identische Signale $\underline{A}(f)=\underline{B}(f)$ aus unterschiedlichen Richtungen präsentiert, werden die Argumente der cosh-Terme in Formel 4.2/1 für beide Ohrsignale gleich. Die interauralen Parameter der Ohrsignale werden dann nur noch durch die mittleren interauralen Differenzen der Schallquellen bestimmt.

$$\underline{d}(t) = e^{\frac{1}{2}(\alpha_a+\alpha_b)} e^{j\frac{1}{2}(\beta_a+\beta_b)} \quad (4.2/6)$$

Entsprechend diesen interauralen Parametern wäre nur ein Hörereignis aus der Mitte zwischen beiden Signalrichtungen zu erwarten, was auch den psychoakustischen Beobachtungen entspricht.

Intensitäts-Stereophonie

Werden bei ansonsten identischen Quellsignalen und einer symmetrischen Aufstellung ($\tau_a=-\tau_b=\tau$; $\alpha_a=-\alpha_b=\alpha$; $\beta_a=-\beta_b=\beta$) die Pegel der Quellen antisymmetrisch geändert, gilt $\ln|A(f)|=-\ln|B(f)|$. Führt man eine aus der Pegeldifferenz zwischen beiden Schallquellen gebildete Größe L_s ein, mit $L_s=\ln\{|\underline{A}(f)/\underline{B}(f)|\}$, so ergibt sich ein interauraler Quotient zu:

$$\underline{d}(t) = \frac{\cosh(\frac{1}{2}L_s + \frac{1}{2}\alpha_a + j\frac{1}{2}\beta)}{\cosh(-\frac{1}{2}L_s + \frac{1}{2}\alpha_a + j\frac{1}{2}\beta)} \quad (4.2/7)$$

Durch den Pegelunterschied der Schallquellen ändern sich die Realteile der jeweiligen cosh-Terme. Hiermit werden die Hauptachsen der hierdurch repräsentierten Ellipse in beiden Ohrsignalen unterschiedlich geändert. Bei Änderungen des Pegelunterschieds zwischen den Schallquellen (d.h. Änderungen von L_s) entstehen geänderte interaurale Parameter. Entspricht das Verhältnis von interauralen Phasen zu Pegeln natürlichen Kombinationen, entsteht ein fusioniertes Hörereignis mit von L_s abhängigem Einfallswinkel.

Reflexionen und Nachhall

Bei Anwesenheit reflektierender Flächen ergeben sich die Ohrsignale durch die Überlagerung von Direktschall und reflektierten Anteilen. Während frühe Reflexionen eine ähnliche Schallfeld-Charakteristik besitzen wie eine begrenzte Anzahl zusätzlicher Einzelschallquellen, entspricht der späte Nachhall eher einem diffusen Schallfeld.

Die Ohrsignale ergeben sich aus der Überlagerung der Spiegelschallquellen (Bild 4.5):

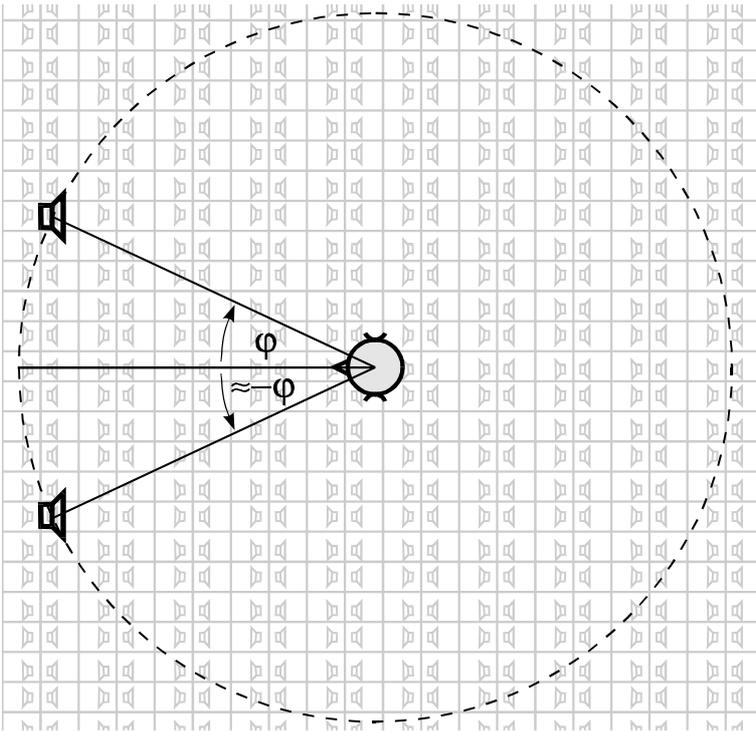


Bild 4.6:
Spiegelschallquellen-Modell:
 Im späten Nachhall (d.h. bei großen Abständen Spiegelschallquelle-Kopf) lassen sich die Spiegelschallquellen zu Paaren kombinieren mit gleicher Laufzeit zum Empfänger aber reziproken interauralen Parametern.

$$\underline{L}(f) = \sum_i A_q(f) H_{qki}(f) H_{lf}(f, \tau_i)$$

$$\underline{R}(f) = \sum_i A_q(f) H_{qki}(f) H_{rf}(f, \tau_i)$$

Ist der Kopf symmetrisch, d.h. $H_{rf}(f, \tau) = H_{lf}(f, -\tau)$, lassen sich jeweils Spiegelschallquellen mit entgegengesetzter interauraler Laufzeit i_1, i_2 zusammen betrachten:

$$\underline{L}(f) = A_q(f) \sum_i H_{qki1}(f) H_{lf}(f, \tau_i) + H_{qki2}(f) H_{lf}(f, \tau_i)$$

$$\underline{R}(f) = A_q(f) \sum_i H_{qki1}(f) H_{rf}(f, \tau_i) + H_{qki2}(f) H_{rf}(f, \tau_i)$$

Entspricht die Kopfachse einer Symmetrieachse des Raumes, so daß gilt: $H_{qki1}(f) = H_{qki2}(f)$, werden die Ohrsignale identisch, eine Lokalisation ist somit nur in der Medianebene möglich.

Im idealen diffusen Schallfeld ist die Energie im Mittel über alle Einfallsrichtungen und alle Zeitpunkte gleich verteilt, so daß jede Achse eine Symmetrieachse des Schallfeldes darstellt. Gemäß den obigen Betrachtungen wären hieraus im Mittel hier keine interauralen Differenzen zu erwarten, eine hieraus resultierende Lokalisation könnte nur in der Medianebene stattfinden.

In realen Räumen ist solch ein Feld angenähert im späten Nachhall zu erwarten, wenn durch die hohe Reflexionsdichte alle Spiegelschallquellen mit gleicher Laufzeit Spiegelschallquelle-Kopf zu Paaren mit entgegengesetzten interauralen Differenzen zusammengefaßt werden können (Bild 4.6).

Auf andere Fälle übertragen bedeutet dieses Ergebnis:

- Läßt sich der Nachhall als ideales diffuses Schallfeld beschreiben, unterstützt dieser Richtungsinformationen der Medianebene.
- Von der Medianebene abweichende Richtungsinformationen müssen in geschlossenen Räumen aus Direktschall und frühen Reflexionen stammen.
- Bei binauralen Signalverarbeitungsmodellen könnte dies zum Vortäuschen zusätzlicher Signale der Vorne-Richtung führen und damit zu Fehlern bei der Verarbeitung der Vorne-Richtung.

4.3. Algorithmen zur binauralen Verarbeitung der Ohrsignale

Im folgenden soll versucht werden, die Ergebnisse der Hörversuche und der Vorüberlegungen zu auftretenden Ohrsignalen mit existierenden binauralen Modellen in Beziehung zu bringen und hieraus Möglichkeiten zur modellmäßigen Beschreibung von Cocktail-Party-Prozessor-Phänomenen abzuleiten.

Die wichtigste Kategorie binauraler Modelle sind Kreuzkorrelations-Modelle oder Modelle verwandter Funktionen. Aufgabe dieser Modelle ist es, die Eingangssignale so zu transformieren, daß binaural relevante Informationen direkt ablesbar werden, wie Einfallrichtungen, interaurale Parameter, die Leistungen der Schallquellen. Signalphasen sind hierbei weniger von Interesse.

Die Kreuzkorrelationsfunktion $S_{rl}(\tau)$ zwischen den reellen Zeitfunktionen des rechten und linken Ohrsignals $r(t)$ und $l(t)$ ist definiert als:

$$S_{rl}(\tau) = \lim_{T \rightarrow \infty} 1/2T \int_{-T}^T r(t) l(t+\tau) dt$$

Sind $r(t)$ und $l(t)$ zeitverzögerte Repräsentationen desselben Signals $s(t)$, erhält man:

$$\begin{aligned} r(t) &= s(t-\tau_r) & l(t) &= s(t-\tau_l) \\ S_{rl}(\tau_r-\tau_l) &= \lim_{T \rightarrow \infty} 1/2T \int_{-T}^T s^2(t) dt & & \text{(mittlere Leistung des Signals)} \quad (4.3/1) \\ S_{rl}(\tau_r-\tau_l) &\geq S_{rl}(\tau) & & \text{für alle } \tau \end{aligned}$$

Wo der Verschiebungsparameter τ der Zeitverzögerung zwischen den Signalen entspricht, entsteht ein Maximum der Kreuzkorrelationsfunktion. Der Betrag an dieser Stelle entspricht der mittleren Leistung des Signals. Aus Verschiebung und Amplitude des Maximums der Kreuzkorrelationsfunktion lassen sich somit die interaurale Zeitdifferenz und die mittlere Leistung des Signals ablesen.

Gleiches gilt für die gleitende Kreuzkorrelationsfunktion, wenn auch mit gewissen Einschränkungen durch den Gebrauch einer Fensterfunktion $w(t)$ ($w(t)=0$ für $|t|>T$):

$$S_{rl}(\tau) = 1/2T \int_{t-T}^{t+T} r(t') l(t'+\tau) w(t-t') dt' \quad (4.3/2)$$

Durch Modifikation des Algorithmus kann die Kreuzkorrelationsfunktion dem menschlichen Gehör besser angepaßt werden. Durch Einführung inhibitorischer Elemente (Lindemann [25] bzw. Gaik [20]) bzw. Nachbearbeitung der Korrelationsmuster (Stern/Colburn [43]) kann es erreicht werden, daß Kreuzkorrelationsfunktionen, ähnlich wie das menschliche Gehör, auf interaurale Pegeldifferenzen reagieren. Funktionen dieser Art eignen sich somit gut zur Modellierung von Lokalisationsphänomenen. Der Vorteil einer gehörgerechteren Darstellung wird hierbei allerdings durch Einbußen in der Signalverarbeitung erkaufte (größere Rechenzeiten, keine leistungs-proportionalen Ausgangsdaten).

Daneben existieren noch eine Reihe weiterer Modelle zur Untersuchung binauraler Phänomene, die auf anderen Transformationen der Ohrsignale aufbauen, wie das "central-spectrum-model" von Raatgever/Bilson [33] zur Modellierung binauraler Pitch-Phänomene, wie das EC-Modell von Durlach [14] zur Modellierung binauraler Maskierungs-Experimente. Eine Sonderstellung nehmen

hier physiologisch motivierte Modelle ein, die binaurale Interaktionen durch neuronale Spike-Folgen nachzubilden versuchen. Wolf [49] hat gezeigt, daß die Ergebnisse psychoakustisch motivierter Modelle wie das von Lindemann[25]/Gaik[20] auch auf der Ebene neuronaler Interaktionen nachvollziehbar und somit physiologisch möglich sind.

Gaik [17] und Bodden [10] haben Anwendungen von Kreuzkorrelations- bzw. binauralen Modellen zur Signalverarbeitung vorgestellt, bei denen bei Anwesenheit mehrerer Quellen die Leistung einer Schallquelle aus den Kreuzkorrelationsmustern geschätzt wird.

Im folgenden soll untersucht werden, ob es mit Hilfe von Kreuzkorrelations- oder verwandten Funktionen möglich ist, die Ergebnisse der Hörversuche nachzuvollziehen, also aus der Analyse der Ohrsignale einer Frequenzgruppe Informationen über Einfallsrichtungen und Signalleistungen zweier Schallquellen zu gewinnen.

Die Betrachtung beschränkt sich auf nicht-modifizierte Kreuzkorrelationsfunktionen, da diese mathematisch geschlossen darstellbar sind. Für Signalverarbeitungs-Anwendungen sind insbesondere Darstellungsmöglichkeiten im Frequenzbereich von Interesse, da hier für unmodifizierte Kreuzkorrelationsfunktionen schnelle Berechnungsmethoden existieren.

4.4. Kreuzkorrelationsmodelle bei einer Schallquelle

Bei Anwesenheit einer Schallquelle $a(t)$ mit interauraler Laufzeit τ ergibt sich entsprechend Formel 4.3/2 die Fourier-Transformierte der gleitenden Kreuzkorrelationsfunktion zu:

$$\begin{aligned} \underline{S}_r(f,t) &= \underline{R}(f,t) \underline{L}(f,t)^* * \underline{W}(f) \\ \underline{S}_r(f,t) &= \underline{A}_m(f,t) \underline{A}_m(f,t)^* e^{j\beta_a(f)} * \underline{W}(f) \\ \underline{S}_r(f,t) &= \underline{A}_m^+(f,t)^2 * \underline{W}(f) \end{aligned} \quad (4.4/1)$$

$\underline{A}_m^+(f,t)^2$ sei die interaurale Kreuzleistungsdichte des Signals $a_m(t)$. Die Kreuzkorrelationsfunktion entspricht hier der um die interaurale Zeitdifferenz verschobenen Auto-Korrelationsfunktion des Schallsignals am Bezugspunkt "Kopfmitte". Da bei reellen Signalen das Maximum der Auto-Korrelationsfunktion bei einer Verschiebung von Null auftritt, ist das Maximum der Kreuzkorrelationsfunktion um die interaurale Zeitdifferenz τ verschoben.

4.5. Die Kreuzkorrelationsfunktion bei mehreren Schallquellen

Die Kreuzkorrelationsfunktion von $-\infty$ bis ∞

Überlagern sich die Signale zweier Schallquellen $a(t), b(t)$ unterschiedlicher Einfallsrichtungen (interaurale Parameter Quelle a: $\alpha_a(f)=\alpha(f, \tau_a)$; $\beta_a(f)=\beta(f, \tau_a)$; Quelle b: $\alpha_b(f)=\alpha(f, \tau_b)$; $\beta_b(f)=\beta(f, \tau_b)$), ergeben sich die Spektren der Ohrsignale und die Kreuzkorrelationsfunktion zu:

$$\begin{aligned} \underline{R}(f) &= \underline{A}_m(f) e^{+1/2\alpha_a(f) + j1/2\beta_a(f)} + \underline{B}_m(f) e^{+1/2\alpha_b(f) + j1/2\beta_b(f)} \\ \underline{L}(f) &= \underline{A}_m(f) e^{-1/2\alpha_a(f) - j1/2\beta_a(f)} + \underline{B}_m(f) e^{-1/2\alpha_b(f) - j1/2\beta_b(f)} \end{aligned} \quad (4.5/1)$$

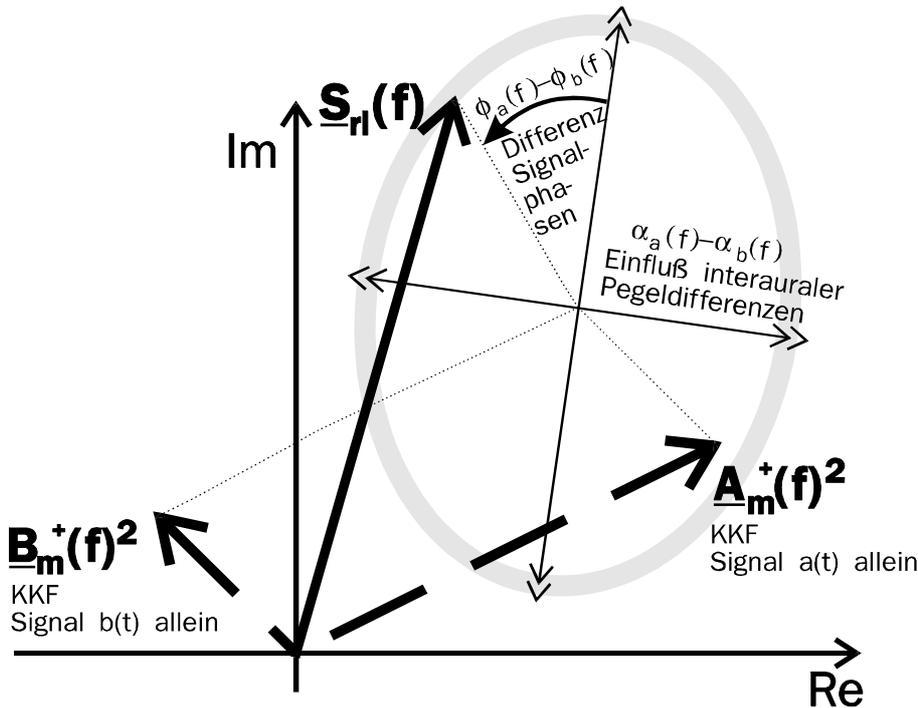


Bild 4.7:
Ortskurve einer
Frequenzlinie der
Kreuzkorrelations-
funktion $\underline{S}_{rl}(f)$ bei
Anwesenheit zweier
unterschiedlicher
Schallquellen.
 $\underline{A}_m^+(f)^2$, $\underline{B}_m^+(f)^2$ =
Kreuzleistungsdichten
der Einzelsignale
Laufparameter ist die
Phasendifferenz
 $\Phi_a(f) - \Phi_b(f)$

$$\begin{aligned} \underline{S}_{rl}(f) &= |A_m(f)|^2 e^{j\beta_a(f)} + |B_m(f)|^2 e^{j\beta_b(f)} \\ &+ \underline{A}_m(f) \underline{B}_m(f)^* e^{+1/2\alpha_a(f) - 1/2\alpha_b(f)} e^{j1/2\beta_a(f) + j1/2\beta_b(f)} \\ &+ \underline{A}_m(f)^* \underline{B}_m(f) e^{-1/2\alpha_a(f) + 1/2\alpha_b(f)} e^{j1/2\beta_a(f) + j1/2\beta_b(f)} \end{aligned}$$

Mit $\Phi_a(f) = \arg\{\underline{A}_m(f)\}$, $\Phi_b(f) = \arg\{\underline{B}_m(f)\}$ erhält man:

$$\underline{S}_{rl}(f) = \underline{A}_m^+(f)^2 + \underline{B}_m^+(f)^2 + 2 \underline{A}_m^+(f) \underline{B}_m^+(f) \cosh \frac{1}{2}\{\alpha_a(f) - \alpha_b(f) + j\Phi_a(f) - j\Phi_b(f)\} \quad (4.5/2)$$

$$\underline{S}_{rl}(f) = \underline{A}_m^+(f)^2 + \underline{B}_m^+(f)^2 + 2 \underline{A}_m^+(f) \underline{B}_m^+(f) \left(\cos \frac{1}{2}(\Phi_a(f) - \Phi_b(f)) \cosh \frac{1}{2}(\alpha_a(f) - \alpha_b(f)) + j \sin \frac{1}{2}(\Phi_a(f) - \Phi_b(f)) \sinh \frac{1}{2}(\alpha_a(f) - \alpha_b(f)) \right)$$

Die Kreuzkorrelationsfunktion ergibt sich als Summe aus den Korrelationsfunktionen der Einzelsignale und einem Misch-Term, der zusätzlich noch von der Differenz der Phasen und der Differenz der interauralen Pegeldifferenzen der Einzelquellen abhängt (siehe Bild 4.7). Die Ortskurve von $\underline{S}_{rl}(f)$ mit der Phasendifferenz zwischen beiden Quellsignalen $\Phi_a(f) - \Phi_b(f)$ als Laufparameter ergibt eine Ellipse, deren Mittelpunkt die Summe der Korrelationsfunktionen der Einzelsignale ist. Die Hauptachsen sind vom geometrischen Mittel der Einzel-Kreuzkorrelationsfunktionen und von der Differenz der interauralen Dämpfungen abhängig.

Einfallsrichtungen und Signalleistungen der Schallquellen sind nicht mehr direkt ablesbar. Betrag und Phase der Kreuzkorrelationsfunktion werden von insgesamt 6 Parametern beeinflusst (Beträge und interaurale Phasen beider Schallsignale, Differenzen zwischen den Signalphasen bzw. zwischen den interauralen Pegeldifferenzen beider Signale). Informationen über die beteiligten Schallquellen sind nur für Spezialfälle direkt aus der Kreuzkorrelationsfunktion bestimmbar:

- Bei Schallquellensignalen mit stark strukturierten Hüllkurven (Signale mit zahlreichen Pausen im betrachteten Frequenzbereich, z.B. Sprache) ist zu vielen Zeitpunkten jeweils ein Signal groß gegen das andere, die Kreuzkorrelationsfunktion entspricht dann der eines Einzelsignals: Die Leistung des starken Signals wird bei der zu diesem Signal gehörenden

Verschiebung aufintegriert. Es entsteht eine Kreuzkorrelationsfunktion mit mehreren relativen Maxima, deren Positionen und Amplitude als Schätzer für die interaurale Laufzeit und mittlere Signalleistung der beteiligten Quellen gelten können. Bodden [10] nutzt diese Eigenschaften aus, um die Positionen mehrerer Sprecher aus Korrelationsmustern zu bestimmen.

- Sind ein Signal und die interauralen Parameter beider Quellen bekannt, lassen sich Betrag und Phase der unbekanntenen Quelle aus der Kreuzkorrelationsfunktion bestimmen. (Gleichung 4.5/2 entspricht dann einer komplexen Gleichung mit einer komplexen Unbekannten) Dieses Problem könnte allerdings auch durch ein lineares Gleichungssystem über 2 Empfängersignale gelöst werden (gemäß Gleichung 4.5/1).

Sind die interauralen Parameter über einen Frequenzbereich, der mehrere Frequenzlinien umfaßt, konstant und die Signale benachbarter Frequenzlinien voneinander unabhängig, läßt sich ein Gleichungssystem über mehrere Frequenzlinien aufstellen, aus dem alle Signal- bzw. interauralen Parameter bestimmt werden können.

Beispiel: Glatte Signalspektren, glatte Freifeld-Außenohr-Übertragungsfunktionen, mindestens ein Signal mit gleichverteilter Phase: Durch Mittelung über einen größeren Frequenzbereich könnte der signalphasen-abhängige Misch-Term in Formel 4.5/1 eliminiert werden. Die interaurale Kreuzkorrelationsfunktion des Quellengemisches entspräche der Summe der interauralen Kreuzkorrelationsfunktionen der beiden Einzelquellen. Bei bekannten interauralen Laufzeiten könnten hieraus die Leistungen der Quellsignale bestimmt werden.

Eine Mittelung über einen breiteren Spektralbereich entspricht einer Verminderung der Frequenzauflösung. Dies ist äquivalent zum Einbringen eines kürzeren Zeitfensters, also zur Bildung einer gleitenden Kreuzkorrelationsfunktion.

Bei voneinander unabhängigen Signalen sind auch die Phasen unterschiedlicher Frequenzlinien voneinander unabhängig. Die Voraussetzungen dieser Analyseverfahren wären somit in vielen Fällen erfüllt. Diese skizzierte Methode soll im folgenden näher untersucht werden.

Analyse durch gleitende Kreuzkorrelationsfunktionen

Die gleitende Kreuzkorrelationsfunktion $S_{rl}(t, \tau)$ ist definiert als:

$$S_{rl}(t, \tau) = \frac{1}{2T} \int_{t-T}^{t+T} r(t_s) l(t_s + \tau) w(t_s - t) dt_s \quad ; \quad w(t) = \text{Fensterfunktion: } w(t) = 0 \text{ für } |t| > T$$

Im Frequenzbereich entspricht die Verwendung eines Zeitfensters $w(t)$ einer Faltung mit der Übertragungsfunktion dieses Fensters. Die hierbei erzielbare Frequenzauflösung entspricht der Bandbreite der Fourier-Transformierten des Fensters ($\sim 1/2T$). Gefensterter Signale lassen sich aus diesem Grund ohne Informationsverlust im Frequenzbereich durch ein Linienspektrum mit dem Linienabstand $1/2T$ beschreiben. Führt man $\underline{A}_m^+(f, t)^2$ und $\underline{B}_m^+(f, t)^2$ als Kreuzleistungsdichten gefensterter Signale bei Anwesenheit jeweils nur der entsprechenden Schallquelle ein, ergeben sich ähnliche Zusammenhänge wie nach Formel 4.5/1, allerdings mit zeitabhängigen Parametern:

$$\begin{aligned} \underline{S}_{rl}(f, t) = & \underline{A}_m^+(f, t)^2 + \underline{B}_m^+(f, t)^2 \\ & + 2 \underline{A}_m^+(f, t) \underline{B}_m^+(f, t) \cosh \frac{1}{2} \{ \alpha_a(f, t) - \alpha_b(f, t) + j\Phi_a(f, t) - j\Phi_b(f, t) \} \end{aligned} \quad (4.5/3)$$

Lösung im Spektralbereich

Im folgenden sollen Betrachtungen durchgeführt werden für interaurale Parameter und Signalparameter, die sich - gemessen an der Fensterlänge der Kreuzkorrelationsfunktion - nur relativ langsam ändern.

Im Spektralbereich repräsentiert eine Frequenzlinie einen Frequenzbereich, dessen Breite der Übertragungsfunktion des Fensters entspricht. Bleiben Signalparameter über mehrere Fensterlängen konstant, können die Signale mit größerer Frequenzauflösung beschrieben werden als die Kreuzkorrelationsfunktion. Stimmen die Frequenzlinien der Kreuzkorrelationsfunktion nicht mit den im Signal enthaltenen Frequenzen überein, entstehen zwischen aufeinanderfolgenden Korrelationsfenstern Phasenunterschiede, deren Größe von der Abweichung zwischen Signalspektrum und Analysefrequenzlinie abhängt. Ebenso können zwischen beiden Signalen Phasenunterschiede auftreten, wenn die Spektren im analysierten Bereich nicht übereinstimmen.

Die Zeitabhängigkeit der Signalphasen kann mit Hilfe von unterschiedlichen Momentan-Kreisfrequenzen $\Omega_a(t), \Omega_b(t)$ beschrieben werden mit $\Phi(t) = \Omega(t)t + \Phi_0$. Zwischen den Signalen entsteht eine zeitabhängige Differenzfrequenz $\Omega_a - \Omega_b$. Für den Mittelwert des Spektrums aufeinanderfolgender gleitender Kreuzkorrelationsfunktionen gilt:

$$\underline{\mu}(f, t) = 1/2T_\mu \int_{t-T_\mu}^{t+T_\mu} \underline{S}_{rl}(f, t_\mu) dt_\mu \quad (4.5/4)$$

$$\begin{aligned} \underline{\mu}(f, t) = & 1/2T_\mu \int_{t-T_\mu}^{t+T_\mu} \underline{A}_m^+(f, t_\mu)^2 dt_\mu + 1/2T_\mu \int_{t-T_\mu}^{t+T_\mu} \underline{B}_m^+(f, t_\mu)^2 dt_\mu \\ & + 1/T_\mu \int_{t-T_\mu}^{t+T_\mu} \underline{A}_m^+(f, t_\mu) \underline{B}_m^+(f, t_\mu) \cosh \frac{1}{2}(\alpha_a(f) - \alpha_b(f) + j\Phi_a(f, t_\mu) - j\Phi_b(f, t_\mu)) dt_\mu \end{aligned}$$

Sind \underline{A}_m^+ und \underline{B}_m^+ nahezu konstant und ist die Integrationszeit T_μ groß gegen $2\pi/[\Omega_a(f) - \Omega_b(f)]$, dann gilt:

$$\underline{\mu}(f, t) = \underline{A}_m^+(f, t)^2 + \underline{B}_m^+(f, t)^2 \quad (4.5/5)$$

Weitere Aussagen über die Schallquellen lassen sich durch Betrachtung der Standardabweichung aufeinanderfolgender Kreuzkorrelations-Fenster erhalten. Es läßt sich eine komplexe Standardabweichung $\underline{\sigma}$ definieren als:

$$\underline{\sigma}^2(f, t) = 1/2T_\mu \int_{t-T_\mu}^{t+T_\mu} (\underline{S}_{rl}(f, t_\mu) - \underline{\mu})^2 dt_\mu \quad (4.5/6)$$

$$\underline{\sigma}^2(f, t) = 2/T_\mu \int_{t-T_\mu}^{t+T_\mu} \underline{A}_m^+(f, t_\mu)^2 \underline{B}_m^+(f, t_\mu)^2 \cosh^2 \frac{1}{2}(\alpha_a(f) - \alpha_b(f) + j\Phi_a(f, t_\mu) - j\Phi_b(f, t_\mu)) dt_\mu$$

Unter gleichen Bedingungen wie oben ($\underline{A}_m^+, \underline{B}_m^+ = \text{const.}; T_\mu \gg 2\pi/[\Omega_a(f) - \Omega_b(f)]$) gilt:

$$\underline{\sigma}^2(f, t) = 2 \underline{A}_m^+(f, t)^2 \underline{B}_m^+(f, t)^2 \quad (4.5/7)$$

Aus Formel 4.5/5 und 4.5/7 lassen sich Beziehungen aufstellen, mit denen aus Mittelwert und Standardabweichung die Kreuzleistungsdichten der Einzelsignale bestimmt werden können. Man erhält Schätzer $\underline{A}_m'(f,t)$, $\underline{B}_m'(f,t)$ für die Frequenzlinien der Signale am Bezugspunkt "Kopfmittle":

$$\begin{aligned} 2 \underline{A}_m'(f,t) &= \sqrt{\underline{\mu} + \sqrt{2}\underline{\sigma}} + \sqrt{\underline{\mu} - \sqrt{2}\underline{\sigma}} \\ 2 \underline{B}_m'(f,t) &= \sqrt{\underline{\mu} + \sqrt{2}\underline{\sigma}} - \sqrt{\underline{\mu} - \sqrt{2}\underline{\sigma}} \end{aligned} \quad (4.5/8)$$

Die Phasen von $\underline{A}_m'(f,t)$, $\underline{B}_m'(f,t)$ entsprechen den interauralen Phasen der Signale, die Beträge den Beträgen der Schallsignale am Bezugspunkt "Kopfmittle". Unter Kenntnis der Freifeld-Außenohr-Übertragungsfunktionen sind hieraus die Einfallsrichtungen und Spektren der Quellensignale bestimmbar.

Im Kasten auf Seite 44 ist dieser Algorithmus zur Schätzung von interauralen Phasen und Beträgen (und damit der Einfallsrichtungen und Spektren) zweier sich überlagernder Schallquellen zusammenhängend dargestellt.

Durch die Analyse von Kreuzkorrelationsfunktionen lassen sich somit im Frequenzbereich Leistungen und Einfallsrichtungen zweier sich überlagernder Schallquellen bestimmen. Die Forderungen aus den Hörversuchen nach der Existenz von Cocktail-Party-Prozessor-Mechanismen sind somit erfüllbar.

Die Verwendung der Kreuzkorrelationsfunktion im Frequenzbereich zur richtungsselektiven Filterung wurde bereits von Gaik [17] vorgeschlagen. Gaik wertete den Momentanwert bzw. Mittelwert der Kreuzkorrelationsfunktion aus und sah ein Signal einer gewünschten Richtung als vorhanden an, wenn die Verschiebung der Werte der Kreuzkorrelationsfunktion mit einem Bewertungsfenster um die Sollrichtung übereinstimmte.

Dieser Algorithmus stellt in gewisser Weise eine wesentliche Erweiterung zu Gaik's Algorithmus dar. Durch die zusätzliche Analyse der Standardabweichung können nun gleichzeitig die Parameter zweier Quellen bestimmt werden. Außer zum parallelen Verarbeiten zweier Signale kann der zweite Schätzer zum Abfangen von Störsignalen benutzt werden. Mit diesem Verfahren können die Parameter einer Nutzquelle auch bei stark negativen Signal-Störabständen noch analysiert werden, wenn also die Verschiebung der Maxima der Kreuzkorrelationsfunktion nahezu der Verschiebung der Störquelle entspricht (also außerhalb eines Bewertungsfensters der Sollrichtung liegt) und der Einfluß der gewünschten Richtung sich lediglich in leichten Variationen der Störquellen-Parameter äußert.

Der Aufbau eines Cocktail-Party-Prozessors im Frequenzbereich mit Hilfe dieses Algorithmus ist an bestimmte Voraussetzungen gebunden:

- quasi-stationäre Signale, deren Parameter sich während der Integrationszeit $2T_\mu$ nicht ändern,
- spektrale Unterschiede zwischen den Quellen-Signalen in den betrachteten Frequenzbereichen, d.h. unterschiedliche Signale für alle Quellen,
- genügend lange Integrationszeit T_μ zur Bestimmung statistischer Parameter, d.h. T_μ muß größer sein als die Periode der Differenz der Momentanfrequenzen der Signale (Voraussetzung für Vereinfachungen bei der Herleitung des Algorithmus),
- gleitende Kreuzkorrelationsfunktion mit kurzem Zeitfenster T , um Änderungen der Signalparameter gut erfassen zu können,
- hieraus folgt: geringe Frequenzauflösung (Folgerung aus dem letzten Punkt gemäß Unschärfe-Relation).

Ein Cocktail-Party-Prozessor im Spektralbereich

1. Berechnung des interauralen Kurzzeit-Kreuzkorrelationsfunktion aus den Fourier-Transformierten der Ohrsignale

$$\underline{S}_{rl}(f,t) = \underline{R}(f,t) \underline{L}(f,t)^* * \underline{W}(f)$$

2. Berechnung von statistischen Parametern der Frequenzlinien der Kreuzkorrelationsfunktion

$$\underline{\mu}(f,t) = 1/2T_{\mu} \int_{t-T_{\mu}}^{t+T_{\mu}} \underline{S}_{rl}(f,t_{\mu}) dt_{\mu}$$

$$\underline{\sigma}^2(f,t) = 1/2T_{\mu} \int_{t-T_{\mu}}^{t+T_{\mu}} (\underline{S}_{rl}(f,t_{\mu}) - \underline{\mu})^2 dt_{\mu}$$

3. Schätzen von Betrag und interaurale Phase zweier Quellen aus den statistischen Parametern der Kreuzkorrelationsfunktion

$$\underline{A}_m'(f,t) = \frac{1}{2} \sqrt{\underline{\mu} + \sqrt{2\underline{\sigma}}} + \frac{1}{2} \sqrt{\underline{\mu} - \sqrt{2\underline{\sigma}}}$$

$$\underline{B}_m'(f,t) = \frac{1}{2} \sqrt{\underline{\mu} + \sqrt{2\underline{\sigma}}} - \frac{1}{2} \sqrt{\underline{\mu} - \sqrt{2\underline{\sigma}}}$$

4. Bestimmung von interauralen Zeitdifferenzen und Leistungen der geschätzten Quellsignale

$$\tau_a'(f,t) = \arg\{ \underline{A}_m'(f,t) \} / 2\pi f$$

$$\tau_b'(f,t) = \arg\{ \underline{B}_m'(f,t) \} / 2\pi f$$

$$A'(f,t)^2 = | \underline{A}_m'(f,t) / \underline{H}_m(f, \tau_a') |^2$$

$$B'(f,t)^2 = | \underline{B}_m'(f,t) / \underline{H}_m(f, \tau_b') |^2$$

4.6. Das komplexe Kreuzprodukt

Die zulässige Änderungsgeschwindigkeit der Signalparameter ist beim obigen Algorithmus begrenzt durch die Integrationszeit zur Bildung der statistischen Parameter. Diese sollte möglichst kurz sein. Die Integrationszeit kann eingeschränkt werden, wenn das Zeitfenster zur Bildung der Kreuzkorrelationsfunktion möglichst kurz wird. Setzt man als Zeitauflösung dieses Verfahrens die Länge des Zeitfensters der Fourier-Transformation an, so ist die Zeitauflösung maximal, wenn die Frequenzauflösung des Verfahrens minimal ist, also der Bandbreite des zu analysierenden Signals (z.B. eine Frequenzgruppen-Breite) entspricht. Dieser Zusammenhang besteht auch bei abgetasteten Signalen, wenn die Abtastfrequenz der Bandbreite des Signals entspricht, also für jeden Abtastwert interaurale Beziehungen betrachtet werden. In diesem Fall entspricht eine Frequenzlinie der Fourier-Transformation einem Abtastwert des analytischen Zeitsignals des entsprechenden bandbegrenzten Signals. Jede Frequenzlinie der Fourier-Transformierten der Kreuzkorrelationsfunktion entspricht dann einem Abtastwert des interauralen Kreuzprodukts der analytischen Zeitsignale der Ohrsignale in diesem Frequenzbereich. Hieraus ergibt sich eine weitere Analyseverfahren für einen Cocktail-Party-Prozessor:

- Frequenzgruppen-Filterung, Bildung der analytischen Zeitsignale der Ohrsignale,
- Bildung von Abtastwerten des analytischen Zeitsignals entsprechend der Bandbreite der betrachteten Signale (Frequenzgruppen-Breite), d.h. Taktreduktion.
- Bildung des interauralen Kreuzprodukts aus diesen Abtastwerten,
- Bestimmung von Mittelwert und Standardabweichung des interauralen Kreuzprodukts,

- Bestimmung von (auf "Kopfmittle" normierten) Signalbeträgen und interauralen Phasendifferenzen zweier Schallquellen aus einem aus Mittelwert und Standardabweichung gebildeten Gleichungssystem,
- Bestimmung der Originalleistungen der Signalquellen und Rekonstruktion der Zeitsignale der beiden Schallquellen,

Eine Analyse dieser Art unterliegt einer Reihe von Einschränkungen. Ein begrenzender Faktor ist die Zeitkonstante T_{μ} zur Bildung der statistischen Parameter des Kreuzprodukts:

- Die angegebene Lösung ist nur gültig, wenn der "cosh"-Mischterm in Gleichung 4.5/1 verschwindet. Hierzu muß die Zeitkonstante zur Bildung der statistischen Parameter größer sein als die Periode der Differenzfrequenz zwischen den Signalen: $2T_{\mu} > 2\pi/[\Omega_a - \Omega_b]$. Diese Differenzfrequenz bezeichnet die Differenz der Phasenänderungs-Geschwindigkeiten der Signale und kann als Maß für die Unähnlichkeit der Quellensignale angesehen werden.

Ähnliche Einschränkungen gelten auch für die Fähigkeit des Menschen, Signale unterschiedlicher Richtungen zu trennen (s. Hörversuche in Kap.3). Erst ab einer minimalen Differenzfrequenz zwischen den Signalen waren Versuchspersonen in der Lage, zwei dargebotene Schallquellen zu lokalisieren. Im Frequenzbereich um 500 Hz betrug die minimale Differenzfrequenz ca. 30 Hz, das entspricht einer Periode von etwas mehr als 30 ms.

- Spektrum und interaurale Parameter der betrachteten Signale dürfen sich während der Analysezeit $2T_{\mu}$ nicht ändern.
- Das Verfahren kann keine Aussagen über die Phasen der Signale machen. Eine Rekonstruktion der Quellsignale aus dem Signalgemisch ist nur in bezug auf die Leistungen möglich. Phasenfehler und Frequenz-Verschiebungen sind also möglich.

Dies entspricht den Ergebnissen der Hörversuche. Bei Differenzfrequenzen unterhalb einer Frequenzgruppen-Breite konnten die Versuchspersonen zwar die Schallquellen lokalisieren, aber nicht die Klänge beider Signale aus dem Gemisch zu extrahieren.

- Das Verfahren macht keine Aussagen über interaurale Pegeldifferenzen. Im Bereich hoher Frequenzen, wo die interaurale Phase nicht mehr eindeutig ist, können Mehrdeutigkeiten in der geschätzten Signalrichtung auftreten. Hierdurch können Fehler bei der Wahl der Freifeld-Außenohr-Übertragungsfunktion verursacht werden und die Signalleistung nicht mehr korrekt aus den normierten Schätzer-Beträgen rekonstruiert werden. (Näheres zur Behebung dieses Fehlers in Kapitel 6.)
- Schallsituationen mit 2 Schallquellen sind hiermit nach Quellen getrennt auflösbar. Bei drei und mehr Quellen entsprechen die geschätzten Leistungen und Richtungen nicht mehr der wirklichen Quellenverteilung. (Näheres zur Korrektur dieser Schätzfehler in Kapitel 5.)

Während die Begrenzungen durch die Integrationszeit $2T_{\mu}$ die absoluten Grenzen des Verfahrens darstellen, lassen sich andere Einschränkungen beseitigen:

- Interaurale Pegeldifferenzen können über ein paralleles Verfahren geschätzt werden (Anwendung des statistischen Verfahrens auf die Auto-Korrelationsfunktionen bzw. die Beträge der Ohrsignale).
- Bei drei oder mehr Quellen oder für diffuse Schallfelder können Korrekturverfahren für die Richtungsschätzer angewendet werden. Mehrquellen-Situationen sind an zeitvarianten Richtungsschätzern zu erkennen. Aus der Analyse dieses zeitabhängigen Verhaltens können Korrekturfaktoren für die Leistungsschätzer gewonnen werden.

Im folgenden soll ein binaurales Signalverarbeitungsmodell vorgestellt werden, das die hier abgeleiteten Signalverarbeitungsstufen enthält und zur Lösung des Cocktail-Party-Prozessor-Problems eingesetzt werden kann. Die Eigenschaften dieses Verfahrens, die Reaktion auf bestimmte Signale sowie Möglichkeiten zur Optimierung des Systems werden diskutiert.