

## 1. Einleitung

Die Analyse akustischer Aufnahmesituationen mit mehr als einer Schallquelle bzw. Reflexionen stellt auch heute noch ein sowohl akustisches als auch nachrichtentechnisches Problem dar.

So ist es oftmals erwünscht, von mehreren vorhandenen Schallquellen nur die Informationen einer Quelle zu erhalten und andere Quellen als Störquellen zu unterdrücken.

Bei Sprachverarbeitungssystemen, wie Spracherkennern oder Freisprechtelefonen z.B. in einem Großraumbüro, besteht das Problem, nur den Sprecher, der mit dem System kommuniziert, zu extrahieren und andere Sprecher sowie Reflexionen und Nachhall aus der Übertragung auszuschließen.

In der Lärmmeßtechnik ist es erwünscht, bei einer Vielzahl von Lärmquellen einzelne richtungsmäßig zu lokalisieren und den Geräuschanteil dieser Quellen zu extrahieren, um so gezielt gegen bestimmte Lärmanteile vorgehen zu können (Genuit [21]).

Hörbehinderte (z.B. Menschen mit nur einem funktionsfähigen Ohr) haben gegenüber Menschen mit gesundem Gehör die Schwierigkeit, sich in Gesprächen ("Cocktail-Party") auf den Gesprächspartner zu konzentrieren, da sie wesentlich stärker als gesunde Menschen durch andere Sprecher und Schallquellen gestört werden. Hörgeräte, die diesen Mangel ausgleichen könnten, wären hier eine große Hilfe.

In den oben geschilderten Situationen treten oft mehrere Störungsursachen gemeinsam auf: Störungen durch andere Schallquellen sowie Störungen durch Reflexionen und Nachhall. Eine Lösung dieser Probleme könnte darin bestehen, Schallquellen einer bestimmten Einfallrichtung selektiv herauszufiltern.

Eine Reihe akustisch-nachrichtentechnischer Verfahren zur Lösung des Problems sind bereits entwickelt worden, für einige der Probleme ist aber bisher noch keine befriedigende Lösung gefunden worden:

Bei der Array-Technik werden Mikrofone auf einer Strecke oder räumlich kombiniert. Durch entsprechende Verarbeitung der Mikrofonensignale kann die Empfangscharakteristik des Arrays verändert werden, so daß idealerweise nur Schallquellen einer bestimmten Einfallrichtung herausgefiltert werden. Um eine hohe Richtungsselektivität zu erreichen, müssen die Abmessungen der Arrays größer als die größte ausgewertete Wellenlänge sein. Um Richtungen eindeutig extrahieren zu können, muß der Mikrofonabstand kleiner als die kleinste ausgewertete Wellenlänge werden. Die richtungsselektive Übertragung des gesamten Hörbereichs des Menschen von 20..20000 Hz mit Wellenlängen zwischen 17 mm und 17 m bereitet mit dieser Technik allerdings Probleme. Lineare Arrays eignen sich eher für die richtungsselektive Übertragung bandbegrenzter Signale. Doch selbst für die richtungsselektive Übertragung des Telefonbereichs mit Frequenzen von 300...3400 Hz müssen Arrays noch über einen Meter groß werden und Mikrofonabstände von unter 10 cm aufweisen.

Richtmikrofone liegen ähnliche Mechanismen zugrunde wie der Array-Technik (Richtungsfilterung durch Phasenauslöschung für unerwünschte Richtungen). Somit liegen auch die Abmessungen in ähnlicher Größenordnung. Handhabbare Mikrofone erreichen erst ab einigen 100 Hz eine akzeptable Richtwirkung, die zu hohen Frequenzen weiter zunimmt. Die selektierte Einfallrichtung kann aber nur durch Schwenken des Mikrofons geändert werden. Simultanes bzw.

nachträgliches Herausfiltern weiterer Einfallrichtungen (wie u.a. in der Meßtechnik erwünscht) ist hiermit nicht möglich.

Adaptive Filter-Verfahren trennen Nutzsignal und Störungen aufgrund unterschiedlicher spektraler Eigenschaften. Voraussetzung für diese Verfahren ist, daß die Signaleigenschaften der Störungen oder der Nutzsignale bekannt sind (z.B. durch ein Mikrofon nahe an der Störquelle) oder aus den Signalen geschätzt werden können. Durch Filter, deren Übertragungsfunktion an das Signal-/Stör-Verhältnis angepaßt wird, lassen sich spektrale Störanteile aus dem Gesamtsignal eliminieren. Sind allerdings die Signale und Störungen unbekannt oder laufenden Änderungen unterworfen oder unterscheiden sich Nutz- und Störspektrum nicht signifikant voneinander (z.B. zwei Sprecher), treten bei der adaptiven Filterung Probleme auf.

Durch Kombinationen unterschiedlicher Verfahren, z.B. Array-Technik für hohe, adaptive Filter für niedrige Frequenzen lassen sich die Eigenschaften der einzelnen Verfahren kombinieren und die Ergebnisse verbessern.

Das menschliche Gehör ist mit nur 2 Ohren in der Lage, nahezu im gesamten Bandbereich Nutzsingale gegenüber Störungen aus anderen Einfallrichtungen hervorzuheben und störenden Nachhall zu entfernen, und das bei sehr hoher Übertragungsqualität. Bisher entwickelte technische Systeme zur Störsignal-Unterdrückung erreichen weder den Bandbereich noch die Übertragungsqualität des menschlichen Gehörs.

So ist die Erforschung des menschlichen Gehörs auch vom technischen Gesichtspunkt von Interesse, um Informationen über die "Algorithmen" des Gehörs zu erhalten, die eine Signalverarbeitung in solch hoher Qualität ermöglichen.

Das Entschlüsseln der Eigenschaften des Gehörs vollzieht sich auf mehreren Ebenen:

Ziel der physiologischen Akustik ist es, die Funktionsweise der einzelnen Bestandteile des Gehörsinns zu erfassen. Das bedeutet z.B.: Untersuchung der mechanischen, elektrischen, physikalischen und biologischen Eigenschaften von Außen-, Mittel- und Innenohr, Analyse von Nervenpotentialen am Hörnerv und den beteiligten Gehirnbereichen. Untersuchungsgegenstände sind hier z.B. die Umsetzung von Schallreizen in Nerven-Erregungen, Probleme der Informationsübertragung, die neuronale Verarbeitung des Gehörsinns. Man erhofft sich Auskünfte über die Auswirkung unterschiedlicher Verarbeitungsstufen (z.B. im Innenohr) auf die Schall-Informationen und über den übertragenen Informationsinhalt. Das Ergebnis sind physiologische Modelle, die möglichst exakt die Vorgänge im Gehör und damit die dort stattfindende Signalverarbeitung beschreiben sollen.

Die Psychoakustik betrachtet die Wirkung des Schalls auf die Wahrnehmung des Menschen. In Hörversuchen werden die Schallinformationen abgefragt, die das Bewußtsein des Menschen erreichen. Die neuronale Verarbeitung wird oftmals als "black box" angesehen. Das Interesse gilt den Ergebnissen der Verarbeitung. Es wird oft mit einfachen, gut zu reproduzierenden und in ihren Signaleigenschaften leicht beschreibbaren Signalen gearbeitet (Sinussignale, weißes Rauschen, periodische Signale, Clicks). Hierdurch sollen die Reaktionen des Gehörs auf einzelne Einflußparameter gezielt abgefragt werden, um durch Kombination vieler solcher Einzelinformationen schließlich auch komplexe Situationen erklären zu können. Hieraus abgeleitete psychoakustische Modelle sollen die Reaktionen auf Schallreize und die Wahrnehmungen des Menschen möglichst exakt beschreiben, quasi als eine Art "Übertragungsfunktion" Schall-Wahrnehmung. In Verbindung

mit physiologischen Erkenntnissen über die frühe Verarbeitung von Schallsignalen sind hieraus Rückschlüsse auf die Verarbeitung in höheren Gehörzentren möglich.

Die Signalverarbeitung versucht, psychoakustisch vorgefundene Effekte für technische Anwendungsfälle nutzbar zu machen (Erkennung von Einfallrichtungen, Enthüllung, richtungsselektive Filter). Ziel ist es, zu ergründen, welche Art von Signalverarbeitung den psychoakustischen Wahrnehmungen zugrunde liegen könnte, um hieraus Algorithmen zur Verarbeitung technischer Signale zu entwickeln. Das Ergebnis der Entwicklung sind Signalverarbeitungsmodelle, die Signale ähnlich gut (oder vielleicht sogar besser) verarbeiten sollen wie das menschliche Gehör, aber auf technische Systeme zugeschnitten sind. Es ist nicht das Ziel, die vom Gehör verwendeten Mittel der Verarbeitung im Detail nachzuvollziehen, sondern Verfahren zu entwickeln, die zu ähnlich guten Ergebnissen kommen.

Allen Forschungsansätzen gemeinsam ist das methodische Vorgehen: Nach einer experimentellen Bestandsaufnahme der Eigenschaften und Fähigkeiten des Gehörs wird versucht, aus Versuchsergebnissen Regelmäßigkeiten und Trends zu erkennen und verallgemeinernde Aussagen über die Fähigkeiten des Gehörs abzuleiten. Hieraus können Vorstellungen über mögliche Verarbeitungsverfahren, die solche Ergebnisse verursachen, entwickelt werden und zu einem Modell des Gehörs zusammengefügt werden. Solch ein Modell sollte die Verarbeitung der Schallsignale nachvollziehen und bei gleichen Eingangssignalen die Ergebnisse der Experimente nachvollziehen. In der Signalverarbeitung bewegt sich der letzte Schritt dann fort von der Beschreibung des Gehörs und hin zur Anwendung der Modellvorstellungen auf andere, gehörunabhängige Fälle.

Ähnliche Prinzipien liegen auch dem Aufbau der vorliegenden Arbeit zugrunde:

Nach einer kurzen Beschreibung psychoakustischer Erkenntnisse, die den Rahmen der folgenden Modellbildung abstecken sollen (Kapitel 2), werden Hörversuche vorgestellt, die Wahrnehmungen von Versuchspersonen bei der Anwesenheit mehrerer Schallquellen untersuchen. Die Ergebnisse der Versuche lassen Rückschlüsse auf mögliche Signalverarbeitungsschritte des Gehörs zu (Kapitel 3).

In Kapitel 4 werden die Grundlagen der folgenden Modellbildung vorgestellt: Die Beschreibung der Übertragungstrecke Schallquelle-Gehör, Nomenklatur und mathematische Beschreibung der zu verarbeitenden Signale sowie das Verhalten der Ohrsignale und der interauralen Differenzen bei Beschallung mit mehreren Schallquellen. Aus einem Vergleich der Hörversuche mit den Ergebnissen vorhandener binauraler Modelle ergibt sich die Notwendigkeit, Verfahren anzuwenden, die über die reine Auswertung interauraler Differenzen hinausgehen. Für Kreuzkorrelationsmodelle wird ein Verfahren im Frequenzbereich vorgestellt, mit dem die nach den Hörversuchen geforderte richtungsselektive Trennung zweier Schallquellen mit ähnlichen Spektren möglich ist. Durch den Übergang von der interauralen Kreuzkorrelationsfunktion im Frequenzbereich zum interauralen Kreuzprodukt im Zeitbereich kann dieses Verfahren beschleunigt und können Totzeiten verringert werden.

Das interaurale Kreuzprodukt stellt den Kern der folgenden Cocktail-Party-Prozessor-Algorithmen dar, die in den Kapiteln 5 und 6 beschrieben werden.

Der "Phasendifferenz-Cocktail-Party-Prozessor" (Kapitel 5) bestimmt mit Hilfe des interauralen Kreuzprodukts die Einfallrichtungen und Signalamplituden von zwei beteiligten Schallquellen. Die Eigenschaften dieses Prozessors werden vorgestellt, ebenso das Verhalten in komplexeren Schall-

feldern (mehr als 2 simultane Quellen, diffuses Schallfeld) sowie Algorithmen zur Schätzung der Schallfeld-Leistungen einer vorgegebenen Einfallsrichtung.

Der "Pegeldifferenz-Cocktail-Party-Prozessor" (Kapitel 6) bestimmt nach einer ähnlichen Methode wie der Phasendifferenz-Prozessor interaurale Pegeldifferenzen und Leistungen zweier Schallquellen aus den Beträgen der Ohrsignale. Die Eigenschaften dieses Prozessors auch in komplexeren Schallfeldern werden diskutiert und Algorithmen zur Bestimmung der Schallfeld-Leistung einer bestimmten Richtung vorgestellt. Pegel- und Phasendifferenz-Prozessor können kombiniert werden zu einer gemeinsamen Analyseeinheit, die Ungenauigkeiten und Mehrdeutigkeiten der einzelnen Prozessoren durch Kombination der Ergebnisse beseitigen kann.

In Kapitel 7 wird der Signalverarbeitungs-Rahmen, in den die Cocktail-Party-Prozessoren eingebettet sind, vorgestellt, sowie Signalverarbeitungs-Ergebnisse präsentiert. Zum Signalverarbeitungs-Rahmen wird hier die Filterung und Aufbereitung der Eingangssignale für die binaurale Analyse sowie die Nachverarbeitung der Analyse-Ergebnisse zu richtungsgefilterten Breitband-Signalen gezählt. In diesem Zusammenhang werden auch Möglichkeiten zur Datenreduktion und Rechenzeiterparnis diskutiert. Diese binauralen Vor- und Nachbearbeitungs-Stufen stellen eine auch für andere Signalverarbeitungs-Prozesse nutzbare "Peripherie-Schnittstelle" dar.

Die Steuerung der Cocktail-Party-Prozessoren kann von einem sogenannten "Präzedenz-Effekt-Prozessor" übernommen werden (Kapitel 8), der, ähnlich wie der Präzedenz-Effekt beim menschlichen Gehör, auch in komplexen Schallfeldern Einfallsrichtungen bestimmen und eine Analyse-richtung für die Prozessoren vorgeben soll. Mit Hilfe dieses Prozessors wären Cocktail-Party-Prozessoren in der Lage, ohne externe Vorgabe einer "Hörrichtung" sich selbständig im Raum zu orientieren und Quellen-Einfallsrichtungen zu suchen. Für einen weiteren Modellausbau bietet es sich an, diese Steuerungseinheit als Schnittstelle zum Einbeziehen weiterer Informationsquellen (z.B. optische Informationen) auszubauen.

## 2. Psychoakustische Grundlagen

Ziel psychoakustischer Forschungen ist es unter anderem, Informationen über die Signalverarbeitung des menschlichen Gehörs zu sammeln. Für technische Anwendungen interessant sind besonders die Eigenschaften des binauralen Systems zur Richtungserkennung und zur Verarbeitung von Signalen unterschiedlicher Einfallrichtungen.

### 2.1. Erkennen von Richtungen

#### Richtungscodierung durch Kopf und Ohr

Schall von einer seitlichen Einfallrichtung erreicht das zugewandte Ohr früher als das abgewandte. Bei natürlichen Ohrabständen von ca. 20 cm erreichen diese Laufzeit-Unterschiede Werte bis zu 600...700  $\mu$ s. Solange die halbe Schallwellenlänge größer ist als der Ohrabstand, also für Frequenzen unterhalb von 800 Hz, reichen interaurale Phasenunterschiede aus, die Einfallrichtung eindeutig zu beschreiben. Für höhere Frequenzen muß die interaurale Gruppenlaufzeit für eine eindeutige Beschreibung benutzt werden.

Schallwellen werden um den Kopf herum gebeugt bzw. durch den Kopf abgeschattet. Der Schallpegel im abgewandten Ohr wird niedriger als im zugewandten. Da die Beugungseffekte vom Verhältnis der Kopfabmessungen zur Wellenlänge abhängen, sind die hierdurch entstehenden interauralen Pegeldifferenzen frequenzabhängig. Während bei niedrigen Frequenzen die interauralen Pegeldifferenzen nur wenige dB betragen ( $\leq 5$  dB bei 500 Hz), können sie für hohe Frequenzen auf über 40 dB anwachsen.

Durch Reflexionen an Schultern und Oberkörper und die Struktur der Ohrmuschel werden die Übertragungsfunktionen Schallfeld-Trommelfell zusätzlich beeinflusst. Ohrmuschel und Gehörgangseingang wirken als akustisches Resonatorsystem, dessen Eigenschaften von der Einfallrichtung des Schalls abhängig sind; den Spektren der Ohrsignale werden so richtungsabhängige Maxima und Minima aufgeprägt. Diese Effekte treten allerdings erst dann in Erscheinung, wenn die Schallwellenlänge die Abmessungen der Ohrmuschel erreicht oder unterschreitet (d.h. ab Frequenzen von ca. 3 kHz) (Shaw/Teranishi [36]).

Die Einflüsse von Kopf und Ohrmuschel auf die Schallsignale am Trommelfell werden durch die Freifeld-Außenohr-Übertragungsfunktionen beschrieben. Sie sind definiert als Quotient zwischen dem Spektrum des Schallsignals am Trommelfell zum Spektrum des Schallsignals, das am gleichen Ort ohne den Kopfeinfluß aufträte.

#### Wahrnehmung von Richtungen

Beim natürlichen Hören im Freifeld sind sowohl interaurale Zeit- und Pegeldifferenzen als auch spektrale Eigenschaften der Ohrsignale an der Richtungserkennung und Signalverarbeitung beteiligt. Ziel psychoakustischer Untersuchungen ist es, die Bedeutung dieser Parameter abzuschätzen. Für technische Anwendungen (Cocktail-Party-Prozessor u.ä.) ist es interessant, welche Parameter sozusagen "aus der Sicht des Gehörs" es wert sind, ausgewertet zu werden. Eine Übersicht über entsprechende Ergebnisse psychoakustischer Forschungen findet man z.B. bei Blauert [5]:

Die Wahrnehmung des Azimutalwinkels von Schallquellen wird sehr stark von interauralen Zeit- und Pegeldifferenzen beeinflusst, die Wahrnehmung der Elevation fußt dagegen zu wesentlichen

Teilen auf der Auswertung der Spektren der Ohrsignale. Innerhalb von "richtungsbestimmenden Bändern" (Blauert [3]) rufen spektrale Maxima und Minima der Schalleistung definierte Elevations-Wahrnehmungen hervor.

Je nach Frequenzbereich ergeben sich unterschiedliche physikalische Einflüsse auf die Ohrsignal-Parameter. Dies spiegelt sich auch in der Wahrnehmung von Schallsignalen wider.

Für Frequenzen unterhalb von 800 Hz, bei denen ein eindeutiger Zusammenhang zwischen Einfallswinkel und interauraler Phase besteht, werden vor allem interaurale Phasendifferenzen zur Bestimmung interauraler Laufzeiten herangezogen. Interaurale Pegeldifferenzen sind in diesem Frequenzbereich noch gering und nur mit großer Ungenauigkeit auszuwerten. Z.B. beträgt bei 500 Hz die interaurale Pegeldifferenz maximal 5 dB bei einer Lateralisationsunschärfe von ca. 1 dB (Blauert [5], S.161). Das Spektrum der Freifeld-Außenohr-Übertragungsfunktionen ist relativ glatt, zeigt aber bereits leichte Maxima und Minima. Hier liegt ein richtungsbestimmendes Band für die Vorne-Richtung (Blauert [3]).

Im Bereich von ca. 800 Hz bis 1.6 kHz erfolgt bei den interauralen Zeitdifferenzen der Übergang von der Auswertung der Phasendifferenzen auf Laufzeit-Differenzen der Signaleinhüllenden. Da die Signalperiode die Größe der interauralen Laufzeit erreicht, ist eine eindeutige Bestimmung der Einfallrichtung aus der interauralen Phase nicht mehr möglich. Die Bedeutung interauraler Pegeldifferenzen nimmt in diesem Bereich zu ( $\Delta L \leq 15$  dB). In den Spektren der Freifeld-Außenohr-Übertragungsfunktion liegt hier ein richtungsbestimmendes Band für die Hinten-Richtung begründet.

Oberhalb von 1.6 kHz wird als Laufzeitparameter die interaurale Gruppenlaufzeit ausgewertet. Ebenso gewinnt die interaurale Pegeldifferenz an Bedeutung. Die Freifeld-Außenohr-Übertragungsfunktionen werden zu hohen Frequenzen zerklüfteter und richtungsabhängiger und enthalten wesentliche Informationen z.B. über die Elevation der Schallquelle. So ist z.B. ein sehr scharfes Maximum der Übertragungsfunktion bei 8 kHz ein Kennzeichen für die Oben-Richtung.

Stimmen die Informationen der richtungsbestimmenden Parameter überein, wird ein Signal aus der entsprechenden Richtung wahrgenommen. Treten Widersprüche auf oder sind die Informationen unvollständig, kann es zu Fehl-Ortungen kommen. Stimmen die interauralen Zeit- und Pegeldifferenzen nicht mit den natürlichen Kombinationen überein, werden zwei Hörereignisse oder ein diffuses Hörereignis wahrgenommen (Gaik [20]). Stimmen die Freifeld-Außenohr-Übertragungsfunktionen der Beschallung nicht mit den persönlichen überein, können Im-Kopf-Lokalisationen und falsche Richtungs-Wahrnehmungen (Vorne-Hinten-Vertauschungen) auftreten (Fornefeld [15]).

### **Dynamische Eigenschaften der Richtungserkennung**

In Räumen, wo Reflexionen und Nachhall auftreten, ist das menschliche Gehör in der Lage, die zuerst eintreffende Richtungsinformation des Direktschalls auszuwerten und die Richtung der zugehörigen Schallquelle zu bestimmen, aber später eintreffende Richtungsinformation aus Reflexionen und Nachhall nicht mehr zur Wahrnehmung kommen zu lassen. Psychoakustisch werden solche dynamischen Richtungserkennungs-Effekte unter dem Begriff "Präzedenz-Effekt" zusammengefaßt.

Bei zeitlich aufeinander folgenden unterschiedlichen Richtungsinformationen reagiert das Gehör folgendermaßen (vergl. Blauert [5]): Liegt der Zeitraum zwischen zwei Richtungsinformationen im

Bereich der interauralen Laufzeiten, kann es bei kohärenten Signalen zur Summenlokalisation kommen, bei der dann eine "mittlere Richtung" wahrgenommen wird. Bei größeren Zeitunterschieden bis zu 5..70 ms wird unter bestimmten Umständen nur die Richtung der Quelle erkannt, deren Schall zuerst am Ohr eintrifft (früher "Gesetz der ersten Wellenfront"). Dieses kann auch für nichtkohärente Signale in unterschiedlichen Frequenzbereichen gelten (Blauert/Divenyi [6]). Dieser Effekt kann bei bestimmten Signalkonfigurationen aber auch aussetzen (Clifton [11], Wolf [47], Blauert/Col [8]). Für Zeitunterschiede oberhalb der Echschwelle werden beide Signalrichtungen wahrgenommen.

Bei Anwesenheit von Reflexionen und Nachhall ist gemäß Franssen [16] das Gehör unter bestimmten Umständen nur bei Signaländerungen (z.B. während ansteigender Flanken) in der Lage, die Richtungen von Schallquellen zu bestimmen. Bei stationären Signalen ist unter Umständen keine Richtungserkennung möglich.

### **Spektrale Eigenschaften der Richtungserkennung**

Nach Scharf u.a.[35] werden Richtungsinformationen innerhalb von Frequenzgruppen verarbeitet. Die Bandbreiten dieser binauralen Frequenzgruppen entsprechen den monauralen Frequenzgruppen, wie sie von Zwicker u.a.[52] zur Beschreibung der Lautheitswahrnehmung bestimmt wurden (siehe Anhang B).

## **2.2. Verarbeiten von Signalen bestimmter Richtungen**

Untersuchungen der *BMLD* (Binaural Masking Level Difference) geben Auskunft über Verbesserungen des Signal-/Stör-Abstands durch die richtungsselektive Verarbeitung des Gehörs. Untersuchungsgegenstand ist die Erkennung von Signalen bei Anwesenheit eines Maskierers, wobei einmal Signal und Maskierer gleiche, einmal unterschiedliche interaurale Parameter besitzen. Als interaurale Parameter werden oft zwei unterschiedliche interaurale Phasen benutzt ( $0$  und  $\pi$ ).

Bei unterschiedlichen Einfallrichtungen bzw. unterschiedlichen interauralen Parametern kann das Gehör bis zu 10..18 dB schwächere Testsignale im Maskierer erkennen als ohne diese Möglichkeit (Blauert [5]). Die größte *BMLD* von 18 dB tritt im Bereich um 300 Hz auf, zu hohen Frequenzen sinkt sie auf ca. 0..3 dB oberhalb 5 kHz ab. Die Detektion der Testsignale erscheint hier vor allem in relativen Minima der Maskierer-Hüllkurve zu erfolgen (Langhans/Kohrausch [24]).

Mithörschwellen-Untersuchungen wie *BMLD*-Versuche machen Aussagen darüber, ob durch Hinzufügen eines Testsignals Änderungen wahrnehmbar werden, unabhängig davon, ob der Informationsgehalt des Testsignals erfaßt werden kann. Untersuchungen der *BILD* (Binaural Intelligibility Level Difference) kombinieren die Methodik der *BMLD*-Untersuchungen mit Sprachverständlichkeitstests.

Unterscheiden sich die interauralen Parameter von Testsignal und Maskierer, so daß die Möglichkeit zur richtungsselektiven Verarbeitung gegeben ist, können bei gleicher Verständlichkeit Sprachsignale in gestörter Umgebung um bis zu 9 dB schwächer dargeboten werden als bei Präsentation mit gleichen interauralen Parametern.

Beim Hören im Freifeld beruht die *BILD* zu wesentlichen Teilen auf der Abschattung von Störquellen durch den Kopf. Die *MILD* (Monaural Intelligibility Level Difference), die aussagt, wieviel

Verständlichkeits-Gewinn das Gehör z.B. aus der Kopf-Abschattung einer Störquelle erzielt, kann Werte bis zu 6 dB erreichen (vom Hövel/Platte [45]).

Der Begriff *Cocktail-Party-Effekt* benennt die Fähigkeit des menschlichen Gehörs, richtungsselektiv Informationen zu verarbeiten und so aus einem Sprechergemisch die Sprachinformationen eines Sprechers zu extrahieren. Der Cocktail-Party-Effekt läßt sich durch die BILD quantitativ beschreiben.

Das binaurale System verfügt über die Möglichkeit zur Reflexions- und *Nachhallunterdrückung*. Bei binauraler Verarbeitung können so modulierte Signale mit wesentlich kleinerem Modulationsgrad wahrgenommen werden als bei monotonischer Wahrnehmung (Danilenko [12]).

Die Fähigkeit zur richtungsselektiven Verarbeitung von Informationen nimmt mit zunehmendem Reflexions- und Nachhall-Anteil ab. So nimmt die BILD stark ab, wenn sich der Hörer außerhalb des Hallradius der Quelle befindet, wo die Leistung des Direktschalls kleiner wird als die Leistung der Reflexionen (Plomp [32]). Franssen [16] zeigte, daß bei starkem Reflexionsfeld ein sanftes Überblenden von einem Lautsprecher zu einem zweiten weder richtungs- noch informationsmäßig wahrgenommen wird.

### **2.3. Merkmale und Leistungen der binauralen Signalverarbeitung**

Ein Signalverarbeitungsmodell, das an die Verarbeitung des Gehörs angepaßt ist, sollte also folgende Eigenschaften haben:

- Verarbeiten der Signale innerhalb von Frequenzgruppen.
- Verwenden von interauralen Zeit- und Pegeldifferenzen zur Bestimmung der Einfallsrichtung und als Grundlage der richtungsselektiven Verarbeitung; bei Bearbeitung von Elevationsinformationen Analyse der spektralen Struktur der Ohrsignale.
- Vergleich der interauralen Zeit- und Pegeldifferenzen mit natürlichen Kombinationen zur Plausibilitätsuntersuchung der Hörereignisse (Mehrfach-Hörereignisse).
- Möglichkeit zur Extraktion von Richtungen und Signalparametern aus kurzen Signalabschnitten, um in reflexionsbehafteter Umgebung die Parameter des Direktschalls erfassen zu können.
- Steuerungseinheit auf "höherer Ebene" zur Bestimmung der gewünschten Verarbeitungsrichtung (Anpassung der Steuerung an Ergebnisse zum Präzedenz-Effekt).

Die Fähigkeiten des Gehörs zeigen, welche Signalverarbeitungs-Möglichkeiten im Prinzip bestehen. Ein technisches Signalverarbeitungs-Modell muß sich an diesen Vorgaben messen lassen:

- Verbesserung der Signaldetektion um bis zu 10-18 dB,
- Verbesserung der Sprachverständlichkeit um 6-9 dB,
- Reduktion der Halligkeit.

Bei starkem Nachhall nehmen allerdings die Fähigkeiten des Gehörs zur richtungsselektiven Verarbeitung ab.



## **3. Hörversuche zur Lokalisation mehrerer Schallquellen**

### **3.1. Aufbau der Hörversuche**

#### **Hintergrund**

Wirken mehrere Schallquellen aus unterschiedlichen Richtungen gleichzeitig auf das menschliche Gehör ein, ist dieses in der Lage, hieraus die Signale einer Einfallrichtung herauszufiltern. Hierbei sind Wiedergabequalität, Richtungsaflöse- und Trennungsvmögen des "menschlichen Cocktail-Party-Prozessors" bisher vergleichbaren technischen Systemen mit zwei Empfängern überlegen. Das menschliche Gehör demonstriert sehr eindrucksvoll, wie leistungsfähig solche Systeme sein könnten. Will man ein ähnlich hochwertiges technisches System konstruieren, besteht ein Ansatz darin, mit Hilfe von psychoakustischen Untersuchungen die Eigenschaften des Gehörs zu studieren und hierbei gefundene Signalverarbeitungs-Prozesse in technische Systeme zu übertragen.

Werden die psychoakustischen Untersuchungen in Situationen durchgeführt, in denen die Signalverarbeitung des Gehörs nicht "perfekt" ist und die Hörereignisse nicht einer genauen Analyse der Schallsignale entsprechen, werden Rückschlüsse auf die Signalverarbeitungs-Methoden des Gehörs möglich, z.B. durch Vergleich des Fehlerverhaltens mit technisch realisierbaren Prozessen. Auf der anderen Seite kann aus den Begrenzungen der Fähigkeiten des Gehörs die notwendige Güte der Modellbildung für technische Anwendungen abgeschätzt werden. Thesen und Modellvorstellungen über das Hören lassen sich anhand psychoakustischer Ergebnisse überprüfen und schließlich zu technischen Prozessen mit gehörverwandten Eigenschaften weiterentwickeln.

#### **Untersuchungsziel, verwendete Signale**

Untersucht werden soll die Fähigkeit des menschlichen Gehörs, einzelne Quellen eines Quellengemisches zu lokalisieren und deren Signale zu extrahieren. Fragestellungen sind hierbei unter anderem:

- Wie groß ist die wahrgenommene Anzahl der Hörereignisse, wenn mehrere Schallquellen gleichzeitig Schall abstrahlen? Wann ist das Gehör in der Lage, die korrekte Anzahl der Schallquellen zu bestimmen?
- Welche Richtungen werden wahrgenommen? Wann ist das Gehör in der Lage, die Schallquellen korrekt zu lokalisieren?
- Welche Hörereignismerkmale werden wahrgenommen? Wann ist das Gehör in der Lage, die Signaleigenschaften der Schallquellen korrekt zu bestimmen?

Von Interesse ist hierbei, wie ähnlich Signale unterschiedlicher Richtung sein dürfen, damit das Gehör diese noch als unterschiedliche Quellen wahrnimmt. Weiterhin wäre die Frage zu beantworten, ob die Möglichkeiten des Gehörs zur Lokalisation und zur Signalverarbeitung auf voneinander abhängigen Prozessen beruhen, d.h. ob bei lokalisierten Schallquellen auch alle Eigenschaften der Schallquellen korrekt bestimmt werden können.

Es wird davon ausgegangen, daß das Gehör für alle Analysezwecke die Signale innerhalb von Frequenzgruppen auswertet (Zwicker u.a. [52]; Scharf u.a. [35]; Kohlrausch [23]) und daß Schallquellen, deren Signale in unterschiedlichen Frequenzgruppen liegen, leicht zu trennen sind (Blauert [4]). Hier soll nun untersucht werden, wie das menschliche Gehör reagiert, wenn das

Gesamtspektrum aller Testsignale die Bandbreite einer Frequenzgruppe nicht überschreitet. Aus diesem Grund werden schmalbandige Testsignale ausgesucht (Sinussignale, Schmalbandrauschen mit 7% relativer Bandbreite).

Weiterhin ist zu beachten, daß die Auswertung interauraler Parameter in unterschiedlichen Frequenzbereichen auf unterschiedlichen Prinzipien beruht: Für Frequenzen unter 800 Hz beruht die Auswertung interauraler Zeitunterschiede vor allem auf der interauralen Phase, für Frequenzen über 1600 Hz auf Zeitunterschieden der Einhüllenden. Somit werden Untersuchungen in zwei Frequenzbereichen durchgeführt, im Bereich um 500 Hz und im Bereich oberhalb 2 kHz. Die verwendeten Signale sind den unterschiedlichen Auswertemethoden angepaßt: Signale ohne Hüllkurveninformation (Sinussignale) und Signale mit strukturierter Hüllkurve (Schmalbandrauschen).

Es werden Versuche durchgeführt, bei denen jeweils 2 Lautsprecher simultan unterschiedliche Signale abstrahlen. Und zwar:

- 1) Signal 1: Sinus 500 Hz; Signal 2: Sinus 500 Hz+ $\Delta f$ ;  $\Delta f = 10..160$  Hz; 8 Versuchspersonen.
- 2) Signal 1: Schmalbandrauschen 500 Hz; Signal 2: Schmalbandrauschen 500 Hz+ $\Delta f$ ;  $\Delta f = 0..160$  Hz; 8 Versuchspersonen.
- 3) Signal 1: Sinus 2 kHz; Signal 2: Sinus 2 kHz+ $\Delta f$ ;  $\Delta f = 10..1200$  Hz; 10 Versuchspersonen.
- 4) Signal 1: Schmalbandrauschen 2 kHz; Signal 2: Schmalbandrauschen 2 kHz+ $\Delta f$ ;  $\Delta f = 0..1200$  Hz; 10 Versuchspersonen.

Die Versuche wurden im reflexionsarmen Raum bei einem Schalldruckpegel von 50 dB durchgeführt.

## Versuchsaufbau

Der Versuchsaufbau sollte so gestaltet werden, daß nur die zu untersuchenden Effekte das Ergebnis beeinflussen können. Folgende Forderungen sind an die Hörversuche zu stellen:

1. Bei Beschallung mit einer Schallquelle sollte nur ein Hörereignis auftreten. Nach Gaik [20] können Mehrfach-Hörereignisse oder diffuse Hörereignisse auftreten, wenn interaurale Zeit- und Pegeldifferenzen in unnatürlichen Kombinationen auftreten. Um Effekte dieser Art auszuschließen, wurden die Signale über Lautsprecher im reflexionsarmen Raum dargeboten.
2. Es sollten nur akustische Parameter von den Versuchspersonen ausgewertet werden können. Um optische Zusatzinformationen über die Richtungen der Schallquellen auszuschließen, wurden die Lautsprecher hinter einem Gaze-Vorhang verborgen (Bild 3.1).
3. Die Ohrsignale sollten bei unterschiedlichen Versuchspersonen reproduzierbar sein, die Ohrparameter sollten sich während einer Versuchsreihe nicht ändern. Am Stuhl der Versuchspersonen war deshalb eine Kopfstütze montiert, an die die Versuchspersonen den Kopf anzulehnen hatten. Hierdurch konnten Kopfbewegungen reduziert werden, kleinere Bewegungen (z.B. im Millimeterbereich) konnten aber nicht ausgeschlossen werden.
4. Es sollten nur Signale im stationären Zustand betrachtet werden. Einschwingvorgänge, die bei Vorversuchen zu einer besseren Lokalisation der Schallquellen führten, sollten von den Versuchspersonen nicht ausgewertet werden können. Die Signale wurden deshalb sanft ein- und ausgeblendet und hierbei von einem Störsignal überlagert (rosa Rauschen + die Summe aller Testsignale, abgestrahlt von allen aufgebauten Lautsprechern, Störsignal-Pegel= 60 dB, 10 dB über dem Signalpegel im eingeschwungenen Zustand, ergibt dominante Summenlokalisierung aus der nicht vorkommenden Richtung 0°).

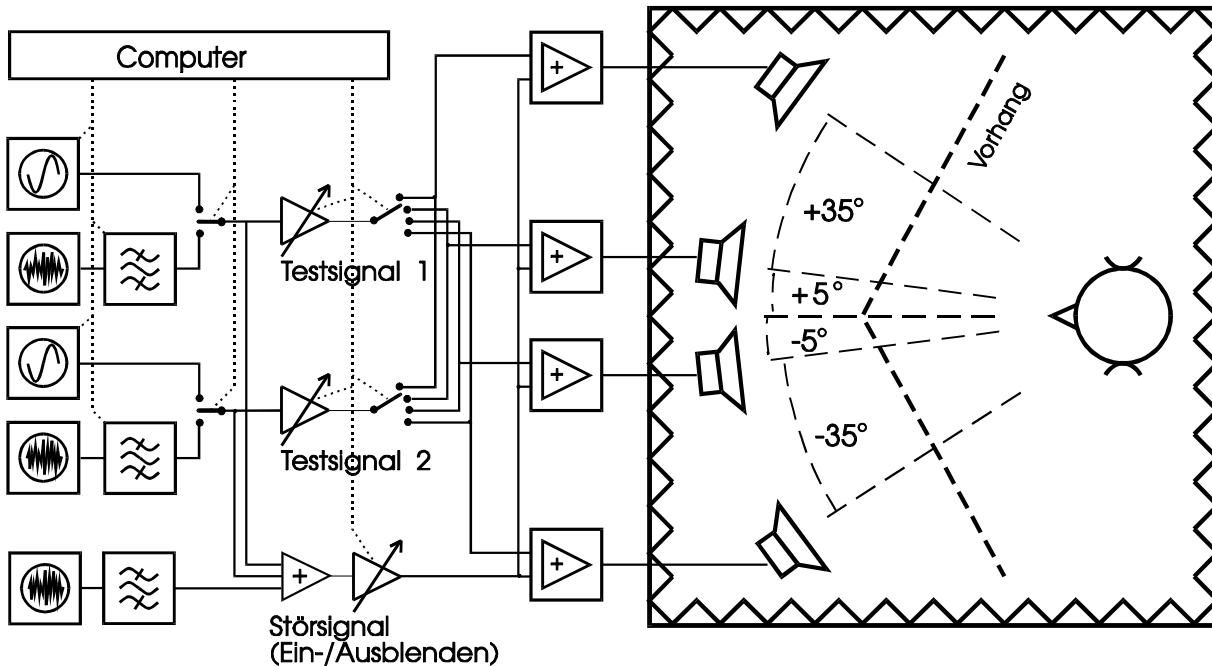


Bild 3.1: Versuchsaufbau, Geräte

5. Vorwissen über die Versuchsbedingungen sollte die Versuchspersonen möglichst nicht beeinflussen. Die verwendeten Signale, die Anzahl und Verteilung der Schallquellen blieb unbekannt. Die verwendeten Lautsprecher wurden von Versuch zu Versuch gewechselt.
6. Zufallslokalisationen sollten nach Möglichkeit ausgeschlossen werden. Bei Testsignalen mit nur geringem (Mitten-)Frequenzunterschied traten in den Bereichen  $50^\circ..90^\circ$  und  $-50^\circ..-90^\circ$  bei vielen Versuchspersonen immer Hörereignisse auf, unabhängig davon, ob an diesen Positionen Schallquellen auftraten oder nicht (Slatky [39]). Um Verfälschungen der Ergebnisse hierdurch auszuschließen, wurden die Lautsprecher nur im Einfallswinkel-Bereich  $-35^\circ..35^\circ$  aufgestellt.
7. Da die dargebotenen Signale und die Hörsituation für die Versuchspersonen sehr ungewohnt waren, wurde für jede Versuchsperson zu Beginn eine Trainingsphase durchgeführt, in der sie ihre Hörereignisse schilderte und dazu angeleitet wurde, eine komplexe Hörsituation bei Überlagerung mehrerer Schallquellen zu analysieren.
8. Die Versuchspersonen sollten die Hörereignis-Richtungen möglichst genau angeben können. So befanden sich Zahlen von -10 bis +10 auf dem Vorhang, die die seitliche Auslenkung der entsprechenden Richtung charakterisierten.
9. Die Auswertung sollte auf psychoakustisch relevanten Parametern fußen. Die Einfallswinkel der Schallquellen  $\Theta$  und die Antworten der Versuchspersonen wurden deshalb auf eine Lateralisations-Skala  $L_t$  (Wertebereich: -10(links)..+10(rechts)) oder auf die *normierte interaurale Zeitdifferenz*  $\tau_L$  bezogen :

$$L_t = 10 \sin(\Theta) \qquad \tau_L = 625 \mu s \sin(\Theta) \qquad (3.1/1)$$

10. Einzeln dargebotene Signale sollten lokalisierbar sein. Bei Vorversuchen konnten fast alle Versuchspersonen die Richtung eines 500 Hz-Sinus-Signals auf  $10^\circ$  genau angeben. Bei 2500 Hz-Signalen sank die Trefferquote auf 60% ( $20^\circ$  von rechts, erlaubte Ungenauigkeit:  $\pm 15^\circ$ ).

## **Fragestellungen an die Versuchspersonen**

Die Versuchspersonen sollten ihre Hörereignisse möglichst umfassend beschreiben. Auf einem Antwortbogen sollten zu jedem richtungsmäßig separierbaren Hörereignis Angaben über Richtung, Ausdehnung, relative Tonhöhe, relative Lautheit (verglichen mit den anderen Hörereignissen) sowie bei Sinussignalen zum Klang (Einzelton, Tongemisch, "rauh") gemacht werden. Aus der Auswertung dieser Parameter werden Aufschlüsse über die Signalverarbeitung des menschlichen Gehörs erwartet. Von folgenden Thesen über Signalverarbeitungs-Mechanismen des Gehörs wurde ausgegangen:

- Als erster Schritt der Richtungsanalyse wird das Erkennen der Schallquellen-Anzahl angesehen. Hierüber kann die Anzahl der Hörereignisse Aufschluß geben.
- Der nächste Schritt wäre das Erkennen der Einfallsrichtungen der Schallsignale; die korrekte Schätzung der Schalleinfallsrichtungen stellt die Voraussetzung für die richtungsselektive Verarbeitung der Schallsignale dar. Informationen hierüber können aus Richtung und Ausdehnung der Hörereignisse gewonnen werden.
- Als erster Schritt der richtungsselektiven Verarbeitung wird die korrekte Identifikation einer Schallquelle angesehen. Aussagen über die korrekte Zuordnung von Schallsignalen zu Richtungen kann die Auswertung der relativen Tonhöhe der Hörereignisse liefern.
- Weitere Schritte der Signalverarbeitung beinhalten die Zuordnung von Schallquellenparametern zu Einfallsrichtungen. Hierüber können die relativen Lautheiten der Hörereignisse Aufschluß geben.
- Als Ergebnis der Signalverarbeitung gilt die richtungsselektive Trennung der Schallsignale. Der Klang der Hörereignisse kann darüber Auskunft geben, wie weit die Charakteristika der Schallsignale vom Gehör richtungsselektiv auswertbar sind.

Hierbei interessiert nicht nur die Fähigkeit des Gehörs, einzelne Richtungs- und Signalparameter zu erfassen, sondern auch das Zusammenspiel und die Reihenfolge der Erkennung. Hieraus können für eine Modellierung wichtige Informationen über mögliche Signalverarbeitungsstufen des Gehörs zusammengetragen werden.

## **3.2. Ergebnisse der Hörversuche**

### **Beschreibungen der Versuchspersonen**

Die Hörversuche wurden von allen Versuchspersonen, auch von geübten, als sehr schwierig eingestuft. Einige Versuchspersonen benötigten mehrere Minuten, um eine dargebotene Schallsituation zu analysieren. Einige Versuchspersonen berichteten, daß sich bei längerem Zuhören Position und Charakter der Hörereignisse änderten. Relativ wenig Schwierigkeiten hatten nur die Personen, die den Versuch leiteten. Da nicht geklärt werden konnte, ob dies auf größere Hörerfahrung oder das Vorwissen um den Aufbau der Experimente (Schallquellen-Anzahl und mögliche Positionen) zurückzuführen war, wurden deren Ergebnisse nicht gewertet.

Alle Versuchspersonen nahmen bei Sinussignalen mit sehr geringen Frequenzunterschieden ( $f_1=500$  Hz;  $f_2=501$  Hz) durch den gesamten Raum wandernde Hörereignisse wahr, ähnlich "binaural beats" (binauralen Schwebungen) bei Kopfhörerversuchen (Blauert [5]). Bewegungsgeschwindigkeit und Lautheit waren hierbei abhängig von der momentan wahrgenommenen Position im Raum. Der Charakter dieser Hörereignisse war nahezu unabhängig

von den tatsächlichen Schalleinfallrichtungen. Bis zu Frequenzdifferenzen von 5-10 Hz konnte - wenn auch unter Schwierigkeiten - der Weg des Hörereignisses verfolgt werden.

Bei kleinen Frequenzdifferenzen ( $\Delta f=5-10$  Hz bei Sinussignalen um 500 Hz) wurde das Hörereignis oft als diffus und "den gesamten Raum erfüllend" beschrieben, wobei aber auch räumliche Lautheitsmaxima und -minima ausgemacht werden konnten. Lautheitsmaxima wurden besonders an den Rändern des Beurteilungsbereiches ( $\pm 90^\circ$ ) wahrgenommen. Hieran änderte sich bei Wechsel der angesteuerten Lautsprecher nur wenig.

Bei mittleren Frequenzdifferenzen ( $\Delta f \leq 30$  Hz bei Sinussignalen um 500 Hz) wurden die Hörereignisse weniger oft als diffus beschrieben. Enger umreißbare Hörereignis-Positionen bildeten sich z.T. nach längeren Hinhören heraus. Hierbei änderte sich oft auch der Charakter der Hörereignisse (Klang, Lautheit). Anschließend blieb der Hörereignis-Charakter meist konstant.

Bei großen Frequenzdifferenzen, d.h. oberhalb einer Frequenzgruppen-Breite ( $f_1=500$  Hz;  $f_2=660$  Hz), konnte den Hörereignissen meist eine eng umgrenzte Position zugeordnet werden.

Die Beschreibungen der Hörereignisse waren subjektiv unterschiedlich. Aus der Beantwortung der Fragebögen lassen sich zumindest 4 unterschiedliche Arten der Wahrnehmung bzw. Beschreibung ableiten:

- Angabe großer Hörereignis-Ausdehnungen. Die Beschreibung der Hörereignisse umfaßt relativ große Winkelbereiche (z.B.  $-20^\circ..+90^\circ$ ). Zu Schwerpunkten innerhalb dieses Bereiches werden keine Angaben gemacht.
- Angabe einzelner Hörereignis-Schwerpunkte. Die Beschreibung der Hörereignisse umfaßt oft nur einzelne Punkte auf der Lateralisations-Skala. Zur Ausdehnung der Hörereignisse werden keine Angaben gemacht.
- Angabe von Hörereignis-Grenzen. Die Beschreibung der Hörereignisse umfaßt einige markante Punkte (Maxima oder Randpunkte). Über den Bereich dazwischen werden keine Aussagen gemacht.
- Eindeutige Angaben. Jeweils nur ein Hörereignis wird genannt. Zu Mehrfach-Hörereignissen werden keine Angaben gemacht.

Hierbei kann grundsätzlich nicht entschieden werden, ob unterschiedliche Angaben bei gleicher Schallsituation auf unterschiedlichen Wahrnehmungen beruhen oder auf unterschiedlichen Arten, ähnliche Wahrnehmungen zu beschreiben.

### **Auswerteverfahren**

Da die Richtungsangaben quantisiert erfolgten (z.B. in Einer-Schritten auf der Lateralisations-Skala von -10 bis +10), haftet den Angaben eine Ungenauigkeit der halben Schrittweite an, in Bild 3.2 als *Lokalisationsunschärfe* bezeichnet. Mit der von den Versuchspersonen angegebenen Hörereignis-Ausdehnung wird diese Ungenauigkeit zu einer sogenannten *effektiven Hörereignis-Ausdehnung* zusammengefaßt (Bild 3.2).

Da die Lokalisations-Angaben mit Unsicherheiten behaftet sind und die Wahrnehmungen bzw. Beschreibungen der Versuchspersonen sehr unterschiedlich sind, wird um die Schallereignis-Richtungen ein sogenannter *Fangbereich* definiert und alle Hörereignisanteile, die in diesen Fangbereich fallen, als korrekt lokalisiert bewertet. Dieser Fangbereich muß für alle Versuche gleich bleiben. Die Fangbereiche zweier aktiver Quellen dürfen sich nicht überlappen (Bild 3.2).

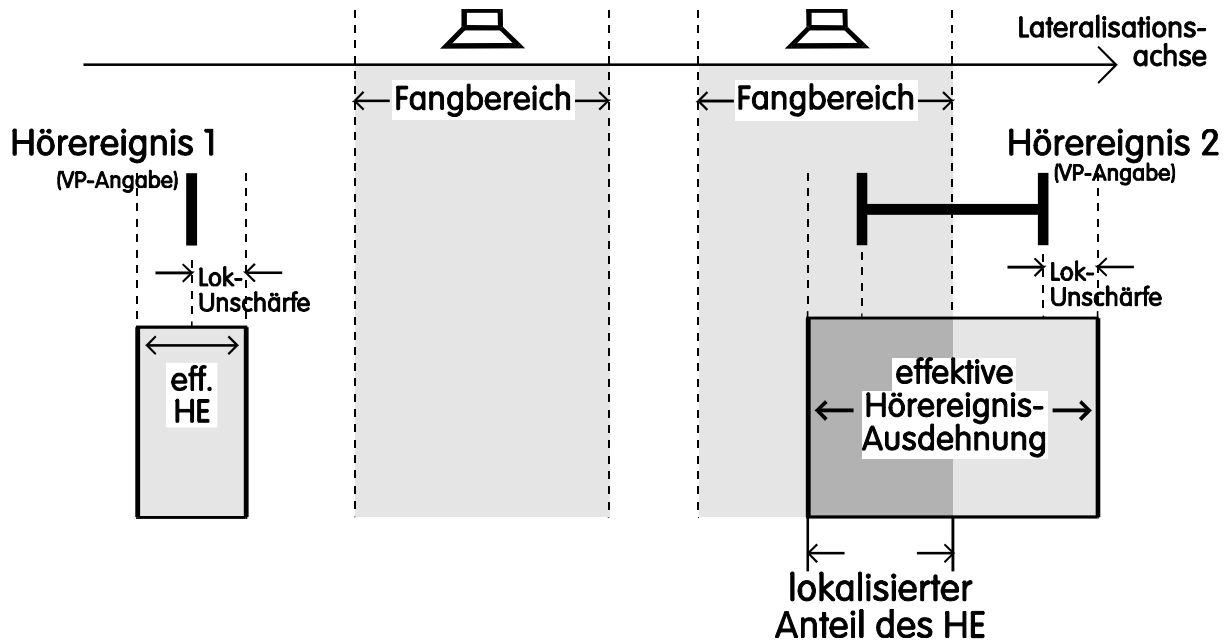


Bild 3.2: Fangbereich, effektive Hörereignis-Ausdehnung (HE=Hörereignis).

Der Begriff *korrekte Lokalisation* wird im folgenden dazu benutzt, Übereinstimmung zwischen Hörereignis- und Schallereignis-Richtung (incl. Fangbereich) zu bezeichnen.

Als *Lokalisationsgrad LG* wird im folgenden der Anteil der effektiven Hörereignis-Ausdehnung bezeichnet, der mit dem Fangbereich einer Schallquelle übereinstimmt. Der Lokalisationsgrad kann als *Wahrscheinlichkeit* interpretiert werden, mit der ein Hörereignis mit einem bestimmten Schallereignis in Beziehung steht. Wenn alle Quellen korrekt lokalisiert werden, entspricht die Summe der Lokalisationsgrade über alle Hörereignisse der Anzahl der vorhandenen Schallquellen. Übersteigt die Anzahl der Hörereignisse  $N_{HE}$  die Anzahl der Schallquellen  $N_{SQ}$ , muß der Lokalisationsgrad auf die Anzahl der Schallquellen normiert werden.

$$LG = \frac{\text{Fangbereich} \cap (\text{eff. HE-Ausdehnung})}{(\text{eff. HE-Ausdehnung})} \cdot \frac{N_{SQ}}{\text{Max}(N_{SQ}, N_{HE})} \quad (3.2/1)$$

Eine Schallquelle gilt dann als korrekt lokalisiert, wenn sie im Mittel besser lokalisiert wird als bei zufälligen Angaben. Der Lokalisationsgrad bei zufälligen Angaben der Einfallsrichtung wird als *Ratewahrscheinlichkeit*  $w_r$  bezeichnet. Ist bei einem möglichen Antwortbereich  $\text{AntB}$  (z.B.  $\pm 10$ ) ein Fangbereich  $F$  um jede Schallquelle definiert (z.B.  $\pm 1.5$ ) und werden Angaben zu Mehrfach-Hörereignissen bis zur Anzahl der Schallquellen  $N_{SQ}$  ohne Abwertung zugelassen, erhält man für die Ratewahrscheinlichkeit  $w_r$ :

$$w_r = \frac{N_{SQ} \cdot F}{\text{AntB}} \quad (3.2/2)$$

Mit den obigen Werten erhält man bei 2 Schallquellen eine Ratewahrscheinlichkeit von 30%.

Zur Auswertung werden die Hörereignisse aller Versuchspersonen zahlenmäßig bewertet (z.B. über den Lokalisationsgrad) und über alle betreffenden Versuche und Versuchspersonen gemittelt. Hierbei ist es auch möglich, die Ergebnisse auf eine Referenz-Situation zu beziehen. In Anhang A ist das Auswertungsverfahren beschrieben.

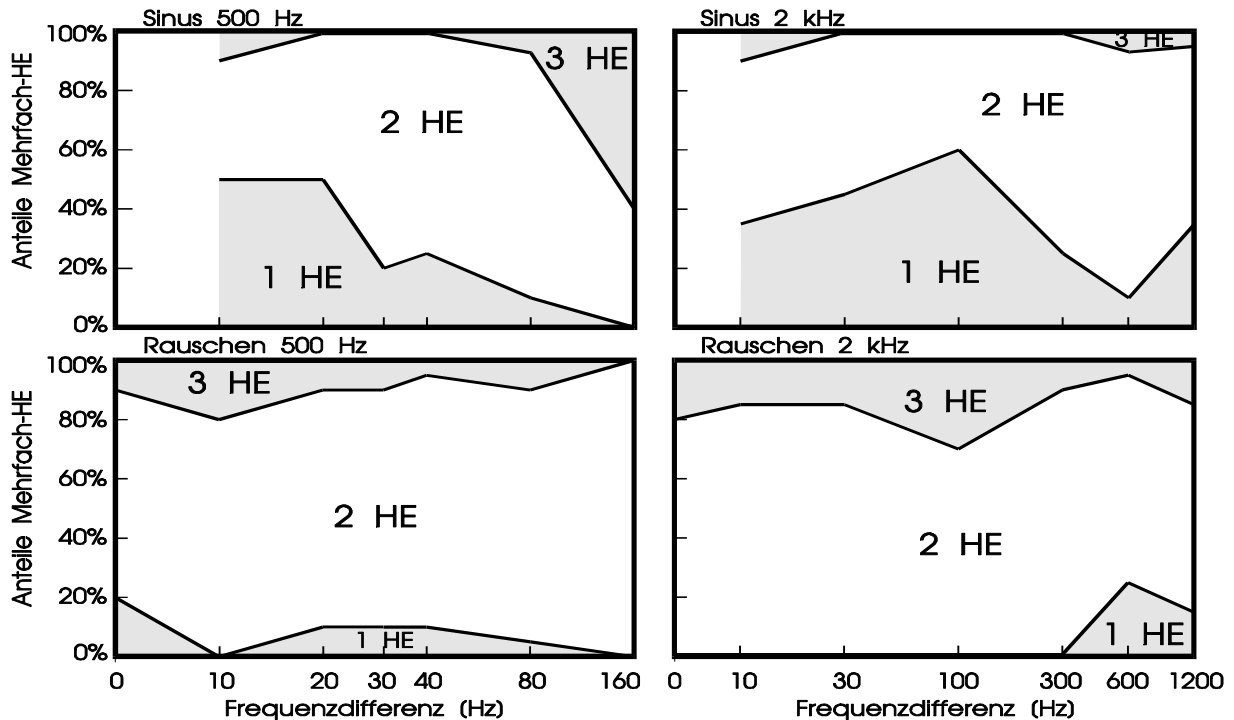


Bild 3.3: Anzahl der Hörereignisse (HE) bei den Hörversuchen mit 2 Schallquellen, aufgetragen über die (Mitten-)Frequenzdifferenz zwischen den Testsignalen

Ausgewertet wurden: Die Anzahl der Hörereignisse, die Lokalisations-Angaben, die Tonhöhe der Hörereignisse, der Klang und die Lautheit. Dargestellt werden diese Größen üblicherweise als Funktion der (Mitten-)Frequenzdifferenz.

### Anzahl der Hörereignisse

Die korrekte Schätzung der Schallquellen-Anzahl wird als erster Schritt zu einer korrekten Analyse der Schallsituation angesehen. Die Anzahl der wahrgenommenen Hörereignisse ist in Bild 3.3 dargestellt.

Bei den Versuchen mit Rauschsignalen stimmten Schallquellen- und Hörereignis-Anzahl meistens überein. Bei den Sinusversuchen trat bei geringen Frequenzdifferenzen häufig nur ein einzelnes diffuses Hörereignisse auf (für  $\Delta f < 30$  Hz bei  $f \approx 500$  Hz,  $\Delta f < 300$  Hz bei  $f \approx 2$  kHz). Bei Sinussignalen von 500 Hz und 660 Hz wurden oft 3 Hörereignisse wahrgenommen, wobei eines als relativ tieffrequent beschrieben wurde, mit Tonhöhen im Bereich der Differenzfrequenz der Signale.

Das Gehör ist somit in der Lage, bei Anwesenheit von zwei Schallquellen, deren Frequenzabstand eine Frequenzgruppen-Breite ( $\approx 110$  Hz um 500 Hz,  $\approx 340$  Hz um 2 kHz) unterschreitet, zwei separate Hörereignisse wahrzunehmen und die Anzahl der Schallquellen korrekt zu schätzen. Die Voraussetzungen für eine weitergehende Verarbeitung wären somit erfüllt.

Oberhalb einer Frequenzgruppen-Breite tritt allerdings bei den 500 Hz-Sinus-Versuchen ein drittes Hörereignis aus Richtungen  $> \pm 70^\circ$  auf. Der Klang entsprach einem Sinuston, die Tonhöhe etwa der Differenzfrequenz zwischen den beiden Testsignalen. Die Eigenschaften dieses Hörereignisses entsprechen den Parametern der Ohrsignal-Hüllkurven. Dies könnte auf nichtlineare Effekte (Bildung eines Differenztons) oder die Auswertung von Hüllkurven-Informationen

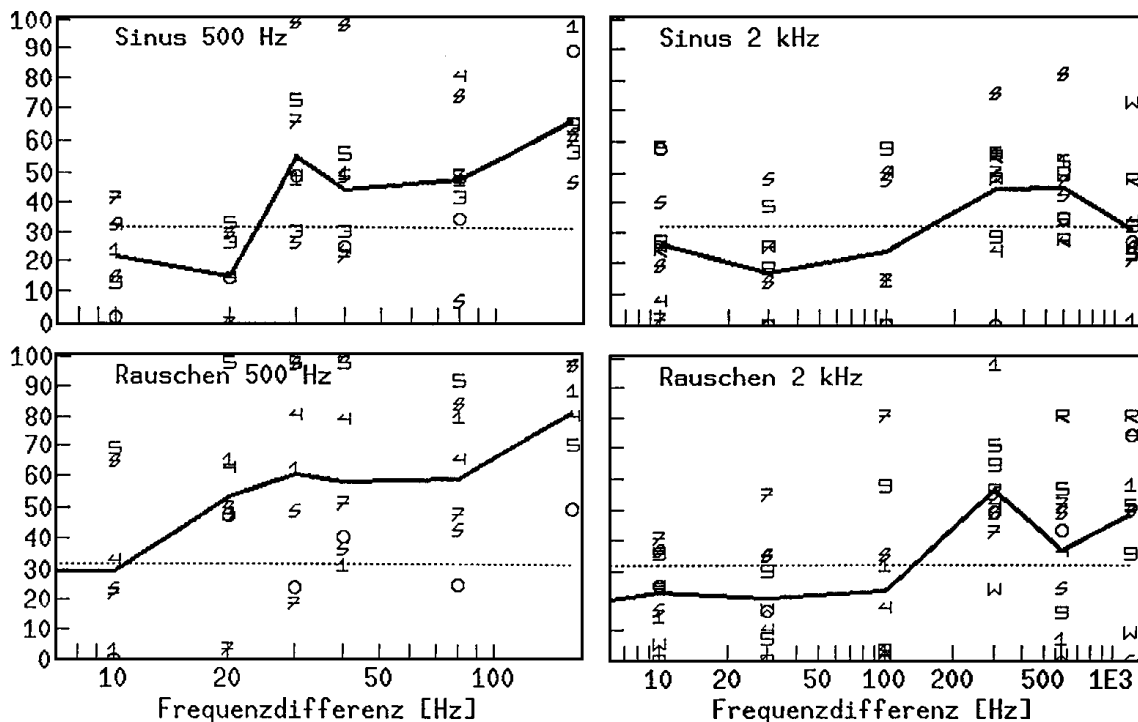


Bild 3.4: Übereinstimmung von Schall- und Hörereignis-Richtung.  
 1345789RSWO individuelle Ergebnisse der Versuchspersonen  
 ————— mittlerer Lokalisationsgrad  
 ..... Ratewahrscheinlichkeit

hinweisen. Eine genauere Eingrenzung dieses Phänomens wurde in diesem Rahmen aber nicht durchgeführt.

### Lokalisation der Schallquellen

Hauptfragestellung der Versuche war, wann das menschliche Gehör in der Lage ist, die Einfallsrichtungen mehrerer Schallquellen korrekt zu bestimmen und ob zur Beschreibung dieser Fähigkeiten eine Auswertung in Frequenzgruppen ausreicht oder ob eine feinere Frequenzauflösung oder zusätzliche Signalverarbeitungs-Verfahren erforderlich sind.

Bei der Auswertung der Lokalisations-Angaben wurde von einem Fangbereich um die Quellenrichtung von  $\pm 100 \mu s$  ausgegangen ( $\pm 10^\circ$  für die Vorne-Richtung) und einer Ungenauigkeit der Lokalisations-Angaben von  $\pm 50 \mu s$  ( $\pm 5^\circ$  für die Vorne-Richtung). Die Angaben in  $\mu s$  entsprechen der normierten interauralen Zeitdifferenz, bezogen auf ein Maximum von  $625 \mu s$  (Formel 3.1/1). Unter diesen Voraussetzungen ergibt sich eine Ratewahrscheinlichkeit von 0.32 (gepunktete Linie). In Bild 3.4 ist der mittlere Lokalisationsgrad als Funktion der Frequenzdifferenz zwischen den Schallquellen dargestellt. Zusätzlich sind auch die individuellen Ergebnisse der Versuchspersonen dargestellt.

Bei den Versuchen im Frequenzbereich oberhalb 2 kHz sinkt allerdings der mittlere Lokalisationsgrad für höhere (Mitten-)Frequenzdifferenzen teilweise wieder ab (Sinus 3200/2000 Hz, Rauschen 2600/2000 Hz).

Bei den Versuchen mit Sinussignalen ab 500 Hz wurde die höherfrequente Quelle (Kurve H) wesentlich besser lokalisiert als die tieferfrequente (Kurve T), während bei den 500 Hz-Rauschsignalen eher eine Tendenz zugunsten der tieferfrequenten Quelle bestand. Bei den Versuchen um 2 kHz konnte keine eindeutige Bevorzugung eines Schallquellen-Typs festgestellt werden.



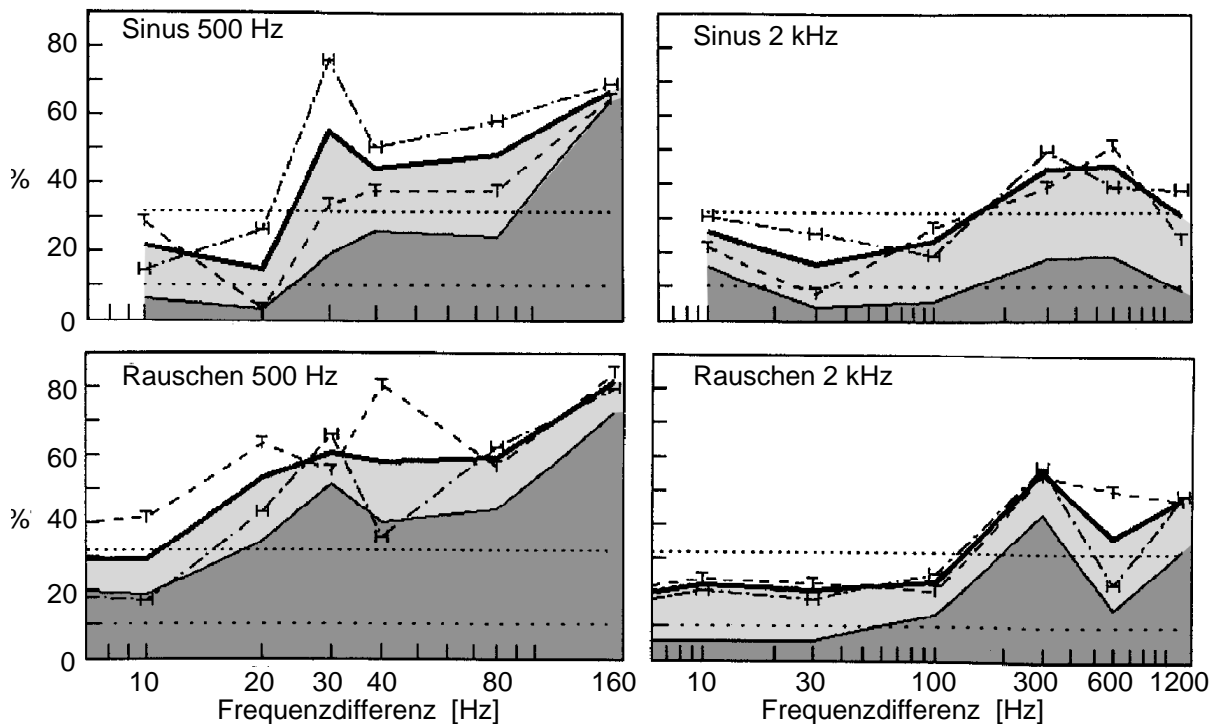


Bild 3.5: Lokalisationsgrad, bezogen auf einzelne Lautsprecher(LS) oder Lautsprecher-Kombinationen

- (hellgrau unterlegt) mittlerer Lokalisationsgrad
- · - H - · - Lokalisationsgrad höherfrequenter Lautsprecher
- - T - - Lokalisationsgrad tieferfrequenter Lautsprecher
- ..... Ratewahrscheinlichkeit
- (dunkelgrau) Lokalisationsgrad beide Lautsprecher gemeinsam lokalisiert
- · · · · Ratewahrscheinlichkeit beide LS lokalisiert

Wesentlich (d.h. mindestens 50%) oberhalb der Ratewahrscheinlichkeit liegt der mittlere Lokalisationsgrad ab folgenden Frequenzdifferenzen  $\Delta f_L$ :

Sinus	500 Hz	$\Delta f_L =$	30 Hz
Rauschen	500 Hz	$\Delta f_L =$	20 Hz
Sinus	2000 Hz	$\Delta f_L =$	300 Hz
Rauschen	2000 Hz	$\Delta f_L =$	300 Hz

In Bild 3.5 sind die Ergebnisse zusätzlich nach Lautsprechern aufgeschlüsselt.

Bestimmt man den Lokalisationsgrad für den Fall, daß beide Schallquellen gleichzeitig korrekt lokalisiert werden (dünne durchgezogene Linie), wird die zugehörige geringere Ratewahrscheinlichkeit von 0.10 ab folgenden Frequenzdifferenzen wesentlich, d.h. um mindestens 50% überschritten:

Sinus	500 Hz	$\Delta f_{L2} =$	30 Hz
Rauschen	500 Hz	$\Delta f_{L2} =$	0 Hz
Sinus	2000 Hz	$\Delta f_{L2} =$	300 Hz
Rauschen	2000 Hz	$\Delta f_{L2} =$	300 Hz (100 Hz)

Im Frequenzbereich um 500 Hz konnten ab Frequenzabständen von 30 Hz zwei Sinusschallquellen gleichzeitig korrekt lokalisiert werden, bei 7% Schmalbandrauschen (Bandbreite 35 Hz) schon bei unabhängigen Quellen gleicher Mittenfrequenz. Der spektrale Abstand der Quellen liegt also bedeutend unter einer Frequenzgruppen-Breite (ca. 110 Hz in diesem Frequenzbereich).

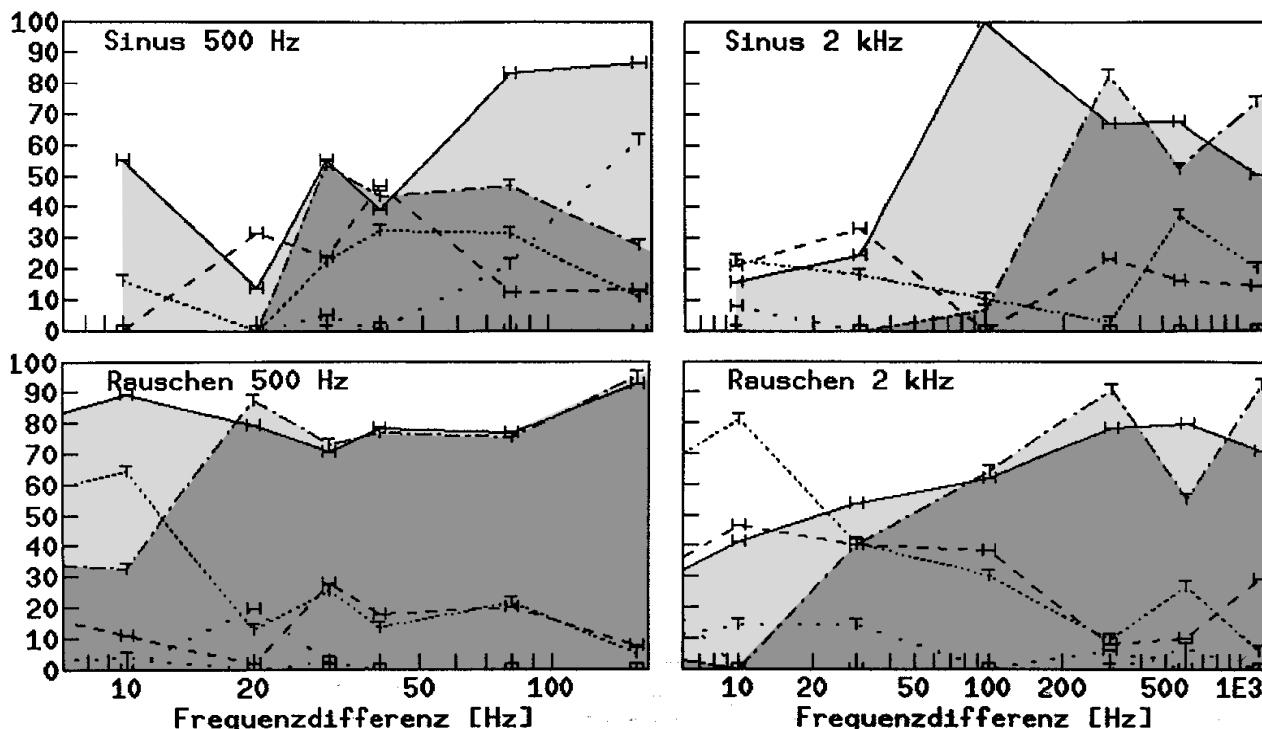


Bild 3.6: Vergleich der relativen Tonhöhen der Hörereignisse (HE) und Schallquellen (LS) für den Fall, daß beide Quellen korrekt lokalisiert werden

rel. Tonh. gleich: ---H--- Tonhöhe LS hoch, Tonh. HE hoch    ···T··· Tonh. LS tief, HE tief

" " verschieden: -H- - Tonhöhe LS hoch, Tonh. HE tief    ···T··· Tonh. LS tief, HE hoch

· ·T· · Tonh. LS tief, HE mittel

hellgrau unterlegt: Übereinstimmung der relativen Tonhöhen für 1 Hörereignis

dunkelgrau unterlegt: Übereinstimmung der relativen Tonhöhen für 2 Hörereignisse

Bei Signalen oberhalb von 2 kHz ist eine korrekte Lokalisation erst bei Frequenzabständen von 300 Hz möglich, dies entspricht nahezu einer Frequenzgruppen-Breite. Auch hier ist ein Abfall des Lokalisationsgrades für einige Signalkonfigurationen zu beobachten.

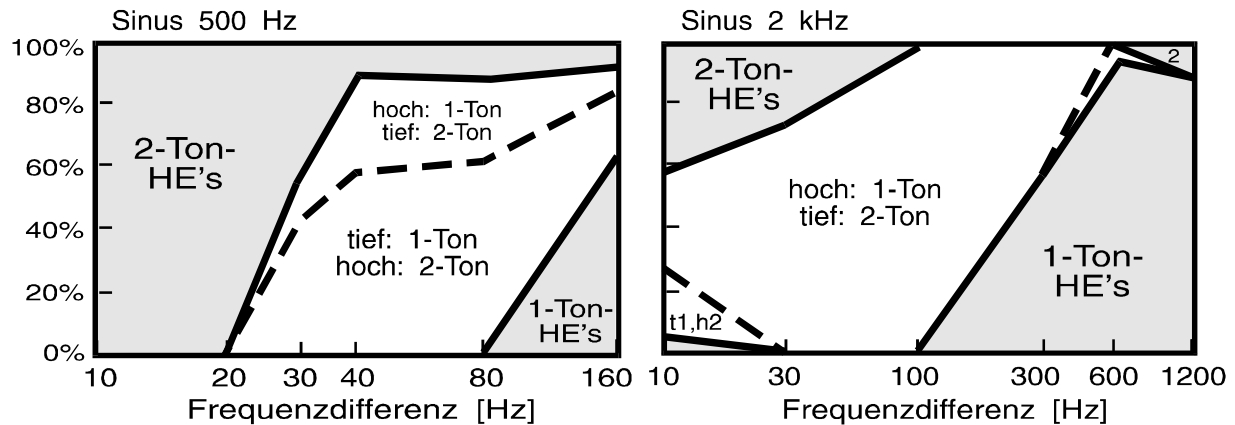
Eine Forderung an binaurale Modelle, die diese Effekte nachvollziehen sollen, ist also, daß sie (zumindest für den Bereich niedriger Frequenzen) innerhalb einer Frequenzgruppe zwei Einfallsrichtungen parallel detektieren und verarbeiten können müssen. (Siehe auch Kapitel 3.3 und 3.4)

### Relative Tonhöhen-Wahrnehmung

Die wahrgenommene Tonhöhe korrekt lokalisierter Schallquellen gibt Auskunft darüber, ob die zu einer Richtung gehörenden Signale identifiziert werden können. Eine korrekte Zuordnung von Tonhöhen zu Richtungen stellt die erste Stufe einer richtungsselektiven Signalverarbeitung dar.

In Bild 3.6 ist dargestellt, wie häufig bei Hörereignissen, bei denen die Richtungen des Hörereignisses der Schallquellen-Richtung entspricht, auch die relativen Tonhöhen (d.h. höher/tiefer als andere Schall- bzw. Hörereignisse) übereinstimmen. Eine solche Übereinstimmung ist ab folgenden Frequenzdifferenzen zu beobachten:

- 30/80 Hz bei den Versuchen zu Sinus 500 Hz
- 20 Hz bei den Versuchen zu Rauschen 500 Hz
- 300 Hz bei den Versuchen zu Sinus 2000 Hz
- 100 Hz bei den Versuchen zu Rauschen 2000 Hz



**Bild 3.7:** Klang der Hörereignisse (HE) bei Sinusversuchen für den Fall, daß beide Quellen korrekt lokalisiert werden

1-Ton-HE's: beide Hörereignisse mit tonalem (sinusartigem) Klang

2-Ton-HE's: beide Hörereignisse mit dem Klang eines Tongemisches (modulierter, rauher bzw. akkordartiger Klang)

1-Ton/2-Ton: jeweils ein HE mit tonalem, bzw. Gemisch-Klang

tief /hoch : Klang des HE aus der Richtung der tiefen / hohen Quelle

Bei den Sinusversuchen um 500 Hz tritt bei hohen Frequenzdifferenzen oft ein drittes, tieffrequentes Hörereignis auf (Differenzton). Die Tonhöhe des Lautsprechers mit der tieferen Signalfrequenz wird dann als mittlere Tonhöhe (zwischen Differenzton und hoher Signalfrequenz) wahrgenommen.

Besonders bei niedrigen Frequenzdifferenzen wurde häufiger die Tonhöhe des höheren Lautsprechers korrekt wahrgenommen als die des tieferen.

Erst ab Frequenzdifferenzen, bei denen korrekte Lokalisation möglich war, wurden die Schallquellen tonhöhenmäßig korrekt identifiziert.

Dies deutet darauf hin, daß die Fähigkeiten zur Lokalisation und zur Identifikation von Tonhöhen miteinander gekoppelt sind. Bei den wenigen Fälle korrekter Lokalisation bei geringen Frequenzdifferenzen scheint es sich eher um Zufalls-Lokalisationen zu handeln; die Tonhöhe der entsprechenden Quelle konnte hier zumindest nicht korrekt bestimmt werden.

### Klang der lokalisierten Schallquellen

Bei Sinussignalen wurden die Hörereignisse daraufhin ausgewertet, ob ihr klanglicher Charakter eher einem einzelnen Sinuston oder einem Gemisch unterschiedlicher Sinustöne (moduliertes Signal, rauher Klang, Tongemisch, Akkord) entsprach. Diese, im folgenden als "Klang" bezeichneten Hörereignis-Eigenschaften, können Hinweise darauf liefern, ob unterschiedliche Signale bei der Richtungsverarbeitung spektral zusammengefaßt oder getrennt verarbeitet werden. Ein Misch-Klang (moduliertes Signal, rauher Klang, Tongemisch, Akkord) deutet somit auf eine gemeinsame spektrale Verarbeitung unterschiedlicher Quellensignale hin, eine Wahrnehmung als Einzelton eher auf getrennte spektrale Verarbeitung. Während bei Sinussignalen Unterschiede dieser Art leicht feststellbar sind, führen bei Rauschsignalen unterschiedliche Verarbeitungs-Strategien lediglich zu einer Veränderung der Bandbreite, die klanglich nur schwer zu erkennen ist.

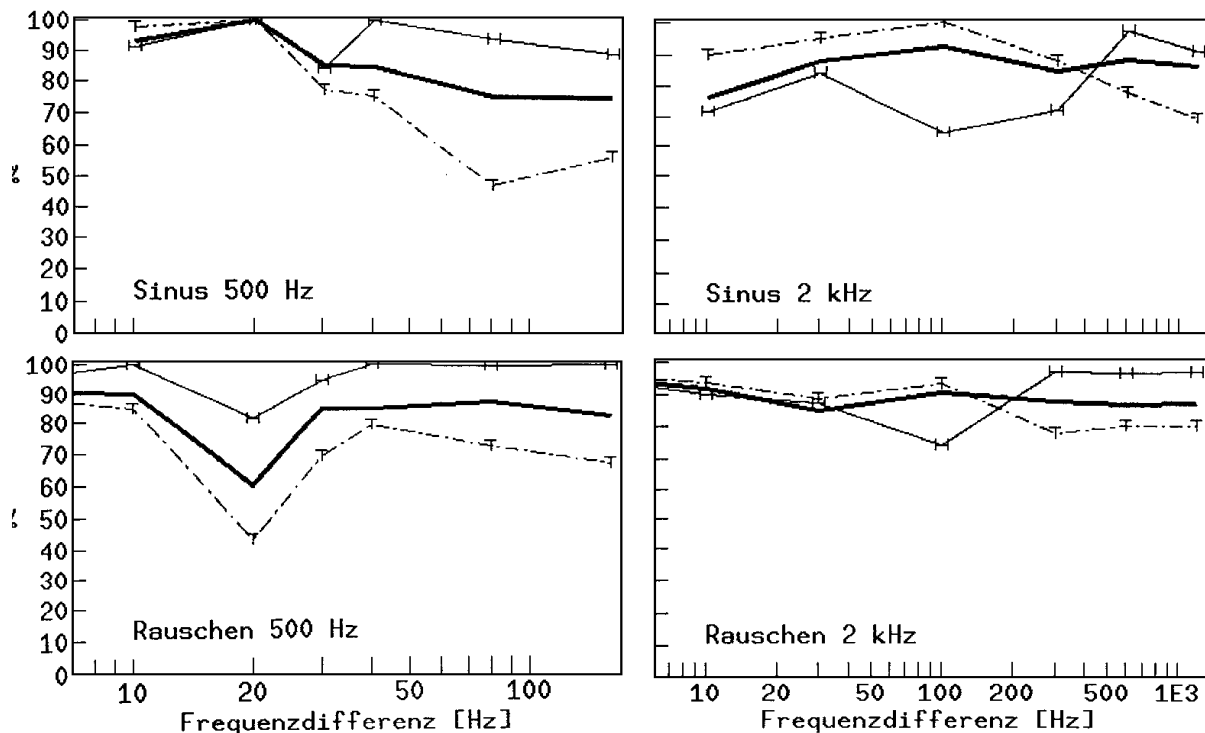


Bild 3.8: relative Lautheit der Hörereignisse (HE) für den Fall, daß alle Quellen korrekt lokalisiert wurden.

- H----- relative Lautheit des HE aus Richtung der höheren Quelle
- . . . - T . . . relative Lautheit des HE aus Richtung der tiefen Quelle
- mittlere relative Lautheit der lokalisierten Quellen

Bild 3.7 zeigt den Klang der Hörereignisse bei korrekter Lokalisation beider Schallquellen. Dargestellt ist, welche Kombination von Hörereignis-Klängen (Einzel-/Mischklang) auftrat. Zusätzlich werden die Hörereignisse noch zur Tonhöhe der Schallquellen in Beziehung gesetzt.

Übersteigt die Frequenzdifferenz zwischen beiden Quellsignalen eine Frequenzgruppen-Breite, entsprechen bei korrekter Lokalisation beider Schallquellen die Klänge der Hörereignisse den Schallsignalen (sinusförmige Einzeltöne). Für niedrigere Frequenzdifferenzen entspricht nur der Klang eines Hörereignisses dem Quellsignal, bei den 500 Hz-Versuchen ist dies für Frequenzdifferenzen zwischen 30 Hz und 80 Hz meist das Hörereignis aus der Richtung der tieffrequenten Quelle, bei den 2 kHz-Versuchen das Hörereignis aus der Richtung der hochfrequenten Quelle. Der Klang des anderen Hörereignisses entspricht dem Gemisch beider Schallsignale. Bei Frequenzdifferenzen unterhalb von 30 Hz entspricht in den wenigen Fällen, in denen beide Schallquellen korrekt lokalisiert werden, der Klang jedes Hörereignisses einem Gemisch beider Schallquellen-Signale.

Eine richtungsselektive Verarbeitung beider Schallsignale ist also nur möglich, wenn die Signale in unterschiedlichen Frequenzgruppen liegen. Überlappen sich spektrale Anteile der Schallsignale innerhalb einer Frequenzgruppe, können zwar die Richtungen der Schallquellen korrekt bestimmt werden, die zu einer Richtung gehörenden Schallsignale können aber nicht mehr vollständig extrahiert werden. Mit der Fähigkeit zur Lokalisation ist in diesem Fall also nicht die Fähigkeit zur richtungsselektiven Trennung der Signalspektren verbunden.

## **Lautheiten der lokalisierten Schallquellen**

In Bild 3.8 ist die wahrgenommene relative Lautheit der Hörereignisse dargestellt (lautestes Hörereignis = 100%), wenn beide Schallquellen gleichzeitig korrekt lokalisiert werden.

Bei Frequenzdifferenzen, bei denen die meisten Versuchspersonen beide Schallquellen korrekt lokalisieren können, wird die höherfrequente Schallquelle meist lauter als die tieferfrequente wahrgenommen. Für geringere Frequenzdifferenzen wird entweder kein Lautheits-Unterschied wahrgenommen (Sinus 500 Hz, Rauschen 2 kHz) oder die Lautheit der tieferfrequenten Quelle überwiegt (Sinus 2 kHz).

## **Ergebnisse der Versuche im Bereich um 500 Hz**

Werden im Bereich um 500 Hz zwei Signale aus unterschiedlichen Richtungen dargeboten, so sind unterhalb einer Frequenzdifferenz von 30 Hz Sinussignale nicht korrekt lokalisierbar, der Lokalisationsgrad liegt unterhalb der Ratewahrscheinlichkeit. Hier treten meist relativ diffuse Hörereignisse auf. Stimmen Hörereignis- und Schalleinfallrichtung einmal überein, können die Quellen tonhöhenmäßig nicht korrekt identifiziert werden, was die Vermutung nahelegt, daß die Lokalisation nur auf zufälligen Übereinstimmungen beruht. Der Klang entspricht in allen Fällen dem Gemisch aus beiden Quellsignalen. Entsprechend bleiben die Lautheiten beider Hörereignisse gleich.

Ab einer Frequenzdifferenz von 30 Hz bei Sinussignalen bzw. bei unabhängigen Rauschsignalen gleicher Mittenfrequenz können die Versuchspersonen die Anzahl der Schallquellen korrekt angeben (Ausnahme: Sinussignale oberhalb einer Frequenzgruppen-Breite) und die Quellen korrekt lokalisieren. Der zur Lokalisation erforderliche Frequenzabstand liegt somit bedeutend unter einer Frequenzgruppen-Breite von ca. 110 Hz. Die korrekte Zuordnung der Tonhöhe zu den Hörereignis-Richtungen ist hier noch nicht in allen Fällen möglich. Bei korrekter Lokalisation kann meist der Klang einer Quelle aus dem Gemisch extrahiert werden und als Einzelton wahrgenommen werden, während ansonsten der wahrgenommene Klang der Summe der Signale entspricht.

Ab Frequenzdifferenzen von ca. 80 Hz (d.h. immer noch unterhalb der Frequenzgruppen-Breite von ca. 110 Hz) können die Richtungen und Tonhöhen beider Signale korrekt bestimmt werden. Bei Sinusversuchen kann ein Signal im Originalklang wahrgenommen werden, der Klang des anderen Hörereignisses entspricht dem Gemisch beider Signale. Das Signal aus der Richtung der höheren Quelle wird meist als das lauteste empfunden, was auch dem Klang dieses Hörereignisses, der Summe beider Signale, entspricht.

Überschreitet die (Mitten-)Frequenzdifferenz eine Frequenzgruppen-Breite, kann die Mehrzahl der Versuchspersonen die Richtungen beider Quellen korrekt angeben. Allerdings treten bei den Sinus-Versuchen in 60% aller Fälle 3 Hörereignisse auf, deren Tonhöhe etwa der Differenzfrequenz der Signale, also der Periode der Hüllkurve entspricht. Unter Einbeziehung dieses Effekts können die Quellen tonhöhenmäßig korrekt identifiziert werden und beide Schallsignale klanglich getrennt werden. Die Lautheit der höheren Quelle bleibt wesentlich größer als die der tieferen Quelle, was allerdings nur in der Tendenz den Kurven gleicher Lautheit entspricht, der empfundene Lautheits-Unterschied geht z.T. wesentlich darüber hinaus.

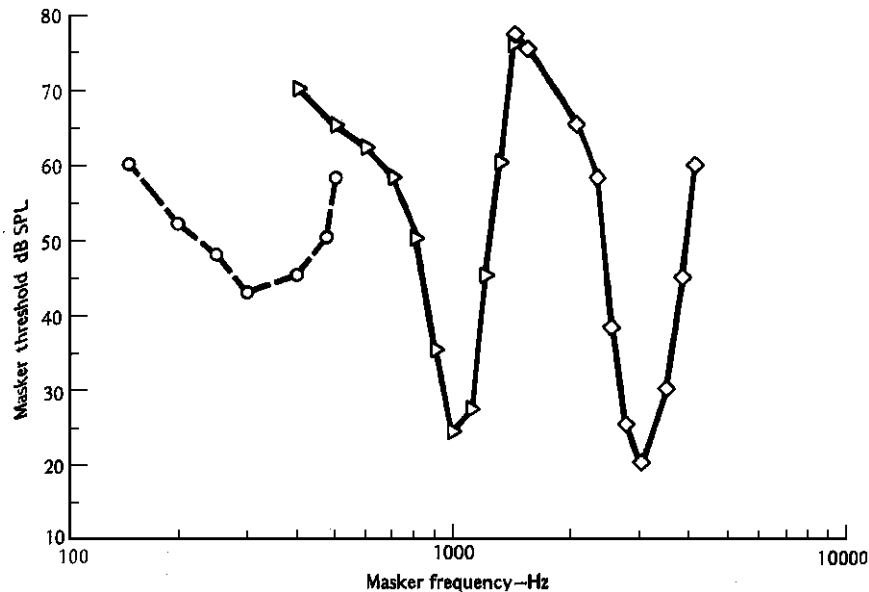


Bild 3.9:  
Psychoakustisch  
gemessene Tuning-  
Curves:  
Verdeckungsschwelle  
von Testtönen von  
300 Hz, 1000 Hz  
und 3000 Hz  
in Abhängigkeit von  
der Frequenz eines  
Sinus-Maskierers  
(Wightman/McGee/  
Kramer, [46])

### Ergebnisse der Versuche im Bereich um 2 kHz

Bei den 2000-Hz-Versuchen mit Sinussignalen können die Versuchspersonen bei erst ab Frequenzdifferenzen von 300 Hz, etwa einer Frequenzgruppen-Breite, die Schallsituation korrekt erfassen und 2 Hörereignisse wahrzunehmen. Bei Rauschsignalen bestehen in dieser Beziehung keine Probleme.

Ab spektralen Differenzen von ca. 300 Hz können die Quellen korrekt lokalisiert werden (bei Sinussignalen ab  $\Delta f=300$  Hz, bei Rauschsignalen ab  $\Delta f=100$  Hz, also ab einer Gesamtbandbreite von etwa 240 Hz). Hier ist auch die Identifikation der Quellen-Tonhöhe möglich.

Bei Sinussignalen wird unterhalb dieser 300 Hz-Schwelle das höher klingende Hörereignis meist als Einzelton beschrieben. Der Klang des tieferfrequenten Hörereignisses entspricht dagegen dem Gemisch beider Signale (2 Töne, rauh, moduliert). Oberhalb einer Frequenzgruppen-Breite werden die Schallsignale der lokalisierten Quellen korrekt als Einzeltöne wahrgenommen.

Lautheit und Klang der Hörereignisse entsprechen sich zumindest teilweise. Hörereignisse, deren Klang der Summe beider Signale entspricht, werden auch als lauter empfunden.

Bei Überschreiten von weiteren Frequenzgruppen-Grenzen (bei ca. 700 Hz und 1100 Hz Frequenzdifferenz) ändern sich die Ergebnisse kaum. Bei Sinussignalen nimmt bei sehr großen Frequenzdifferenzen der Lokalisationsgrad ab und entspricht für  $\Delta f=1200$  Hz wieder der Ratewahrscheinlichkeit.

### 3.3. Thesen zur Signalverarbeitung des Gehörs

Es wird von folgenden Thesen über die Signalverarbeitung des Gehörs ausgegangen:

- Die binaurale Analyse der Ohrsignale erfolgt innerhalb von frequenzgruppen-breiten Frequenzbereichen.

Lage der Signale in der Frequenzgruppe	
obere Frequenzgruppe	
untere Frequenzgruppe	
Lokalisierbarkeit	(leicht)
höheres Signal	schwierig
tiefere Signal	schwierig
Klang	hoch+tief   hoch (+tief)
Lautheit	hoch+tief   abgeschwächt

*Bild 3.10: Erwartete Ergebnisse der Hörversuche bei separater Auswertung der Signale in Frequenzgruppen.*

- Frequenzgruppen werden gebildet durch Zusammenfassen entsprechender Haarzellen-Bereiche des Innenohres. Die Flankensteilheit der Frequenzgruppen-Filter entspricht den sogenannten Tuning-Curves: 30..100 dB/Okt für die niederfrequente, bis zu 300 dB/Okt für die hochfrequente Flanke. Bild 3.9 stellt psychoakustisch gemessene Tuning-Curves in 3 Frequenzbereichen dar. Die Tuning-Curves geben an, wie sich die Wahrnehmungsschwelle eines sinusförmigen Testsignals bei Anwesenheit eines Signals anderer Frequenz verändert. Aus den Tuning-Curves kann entnommen werden, ob und in welchem Umfang Signale unterschiedlicher Frequenz gemeinsam verarbeitet werden. (Sie beschreiben sozusagen die Trennschärfe des Gehörs.)
- Frequenzgruppen werden in gleicher Weise für beide Ohren und alle Analyseziele des Gehörs (Richtung, Lautheit, Klänge, Zuordnung von Lautheiten, Tonhöhen, Klängen zu Richtungen) gebildet. Die Frequenzgruppen-Breiten von Zwicker u.a.[52] (Anhang B) müßten demnach auf Modelle der binauralen Verarbeitung übertragbar sein (ca. 100 Hz für Frequenzen unterhalb 500 Hz, ca. eine Terz darüber).

Somit wird bei Darbietung eines Einzeltons von 50 dB ein Bereich von mindestens einer Oktave (3 Frequenzgruppen) überschwellig erregt. Gleiches gilt für zwei Signale mit geringem spektralen Unterschied.

Es wird davon ausgegangen, daß zur Bildung von Frequenzgruppen der Haarzellen-Bereich mit der größten Erregung zur Bildung der Haupt-Frequenzgruppe zusammengefaßt wird. Hieran anschließende obere und untere Neben-Frequenzgruppen beinhalten den Teil der Erregung, der von der Haupt-Frequenzgruppe nicht erfaßt wird. Diese These wird unterstützt von Untersuchungen zur Lautheitsbildung (z.B. Zwicker u.a.[52]), wonach Frequenzlinien solange gemeinsam ausgewertet werden, bis die Bandbreite des Signals die Breite einer Frequenzgruppe erreicht. Ähnliche Ergebnisse ergaben auch Untersuchungen zu binauralen Frequenzgruppen (Scharf u.a.[35]; Kohlrausch [23]), sowie die Auswertung binauraler Hörversuche (Slatky [39]).

Einen Überblick über weitere Modelle zur Lage von Frequenzgruppen gibt Anhang B.

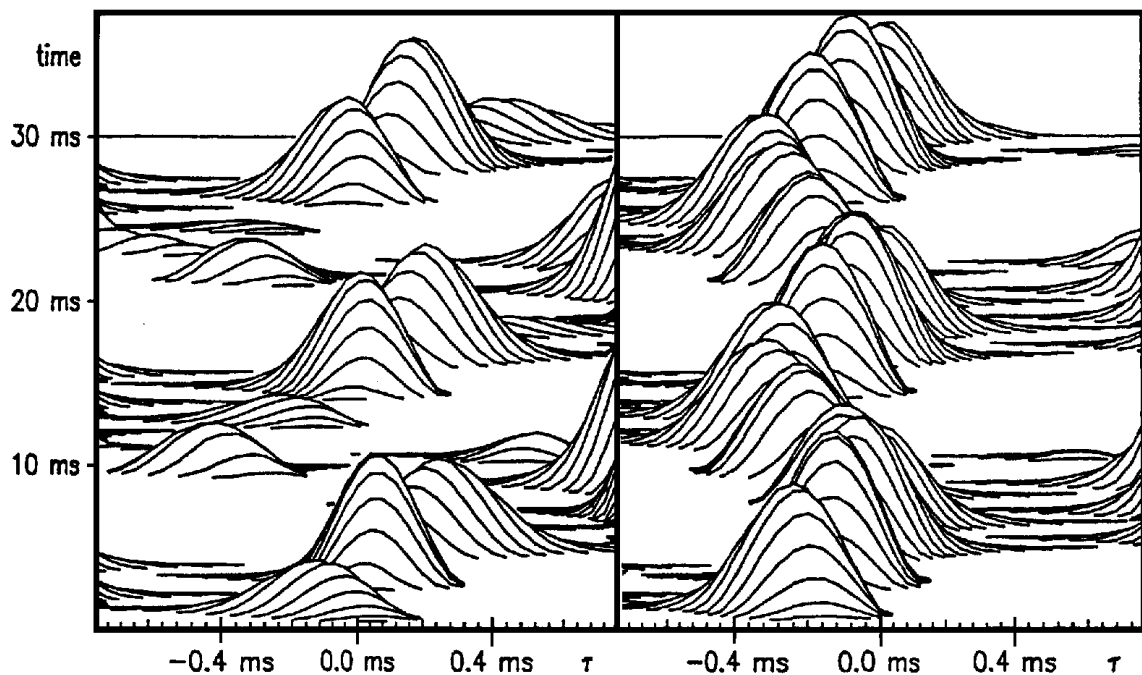


Bild 3.11: Normierte Korrelationsmuster des binauralen Modells nach Lindemann [25] bei 2 Schallquellen gleicher Amplitude und unterschiedlicher interauraler Laufzeit  $\tau$ . Signale: Sinus 580 Hz,  $\tau_a = -200 \mu\text{s}$  + Sinus 500 Hz,  $\tau_b = 600 \mu\text{s}$   
links: Haupt-Frequenzgruppe; rechts: obere Neben-Frequenzgruppe

### 3.4. Vergleich Hörversuche - binaurale Modelle

Zur Diskussion steht die Frage, ob die Ergebnisse der Hörversuche durch separate Analyse der interauralen Parameter in Frequenzgruppen ohne zusätzliche Verarbeitungsschritte erklärt werden können bzw. welche zusätzlichen Verarbeitungsschritte zu fordern wären. In [Bild 3.10](#) sind die Signalanteile in unterschiedlichen Frequenzgruppen dargestellt, sowie erwartete Ergebnisse bei direkter Bestimmung der Signalparameter aus den Frequenzgruppen-Signalen.

In [Bild 3.11](#) sind zum Vergleich die Ergebnisse eines binauralen Kreuzkorrelations-Modells nach Lindemann [25] bei Beschallung mit Signalen der Hörversuche und Einsatz der obigen Frequenzgruppen-Filter dargestellt. Die Korrelationsmuster stellen hierbei ein Maß für die Signalleistung in Abhängigkeit von der interauralen Zeitdifferenz und der Zeit dar.

Bei den angenommenen Frequenzgruppen-Filtern wären bei den Signalen der Hörversuche mindestens 2 Frequenzgruppen erregt, die Haupt-Frequenzgruppe, in der die Signale beider Quellen mit gleicher Amplitude auftreten, und die obere Neben-Frequenzgruppe, in der das tiefere Signal über die Filterflanken stärker abgeschwächt wird als das höhere. Ebenso wären in der unteren Neben-Frequenzgruppe noch (relativ schwache) Anteile besonders des tieferen Signals zu erwarten.

Durch die Überlagerung der Signale unterschiedlicher Frequenz ergeben sich zeitveränderliche interaurale Phasen- und Pegeldifferenzen. In der Haupt-Frequenzgruppe, in der beide Signale mit gleicher Amplitude auftreten, schwanken hierdurch die Positionen des Maximums der Kreuzkorrelationsfunktion ([Bild 3.11](#)) stark mit der Zeit. Der Schwerpunkt des Musters entspricht eher den mittleren interauralen Differenzen als einer vorhandenen Einfallsrichtung. Ein direktes Ablesen der Einfallsrichtung oder Amplitude einer Schallquelle erscheint aus diesem Muster nicht



möglich. (Eine genauere Betrachtung interauraler Parameter bei der Überlagerung zweier Schallquellen befindet sich in Kapitel 4.2.)

In den Neben-Frequenzgruppen werden bei geringen Frequenzdifferenzen beide Signale nahezu gleich stark über die Flanke der Frequenzgruppen-Filter abgeschwächt: Das Korrelationsmuster ähnelt dem der Haupt-Frequenzgruppe. Steigt der Frequenzunterschied, werden die Signale unterschiedlich stark abgeschwächt. Richtung und Parameter eines Signals werden direkt bestimmbar.

Bei Signalen von 500 Hz und 530 Hz, bei denen in den Hörversuchen beide Quellen lokalisierbar waren, würden in der oberen Neben-Frequenzgruppe (Filterflanke 30-100 dB/Okt) sich die Pegel der Signale um 3-8 dB voneinander unterscheiden, zu wenig zu einer stabilen Richtungsschätzung, und in der unteren Neben-Frequenzgruppe (Filterflanke 100-300 dB/Okt) um 8-25 dB.

Hier ist die korrekte *Lokalisation* beider Signale nicht durch unabhängige Auswertung der Frequenzgruppen-Ergebnisse zu erklären. Die interauralen Parameter eines Signals können zwar aus einer Neben-Frequenzgruppe bestimmt werden. Informationen über die Einfallsrichtung der anderen Schallquelle wären dann nur aus der Analyse des varianten Korrelationsmusters der Haupt-Frequenzgruppe gewinnbar. Zum Nachvollziehen der Hörversuche sind also Signalverarbeitungs-Algorithmen erforderlich, die aus varianten Kreuzkorrelationsfunktionen Schallquellen-Richtungen bestimmen können (Tabelle 3.1).

Zur Bestimmung der relativen *Tonhöhe* beider Signale reicht eine separate Auswertung in Frequenzgruppen aus. Ist aus einer Neben-Frequenzgruppe die Richtung einer Quelle bestimmbar, können, allein über die Lage der Frequenzgruppe, auch Aussagen über die Tonhöhe dieses Signals gemacht werden. Die Tonhöhe des Signalgemisches der Haupt-Frequenzgruppe ergibt sich aus der relativen Lage zur dieser Neben-Frequenzgruppe. Hierbei wäre eine Tonhöhenbestimmung erst möglich, wenn zumindest ein Signal lokalisiert werden kann.

Tabelle 3.1:  
Gegenüberstellung Hörversuche -  
Ergebnisse des Kreuzkorrelations-Modells (KKF-Modell).  
FG= Frequenzgruppe

Ergebnisse der Hörversuche	Ergebnisse des KKF-Modells		Folgerung
	Haupt-FG	o. Neben-FG	
<p><b>Klang</b></p> <p>Bei Signalfrequenzen unter einer Frequenzgruppen-Breite nur 1 Signal im Originalklang</p> <p>-----</p> <p>über FG-Breite beide Signale klanglich identifizierbar</p>	<p>Gemisch beider Signale</p> <p>-----</p> <p>tiefere Signal Klang bestimmbar</p>	<p>höheres Signal überwiegt Klang bestimmbar</p> <p>-----</p> <p>höheres Signal Klang bestimmbar</p>	<p>Auswertung erfolgt in Frequenzgruppen</p>
<p><b>Lokalisation</b></p> <p>Versuche um 500 Hz: ab 20...30 Hz Frequenzdifferenz beide Quellen lokalisierbar</p> <p>-----</p> <p>Versuche um 2 kHz: ab 300 Hz Frequenzdifferenz beide Quellen lokalisierbar</p>	<p>variante KKF: Quellenrichtung nicht direkt bestimmbar</p> <p>-----</p> <p>invariante KKF</p>	<p>invariante KKF ab <math>\Delta f = 20 \dots 30</math> Hz: Richtung des höheren Signals bestimmbar</p> <p>-----</p> <p>invariante KKF</p>	<p>"Rückrechen-Mechanismus": Bestimmung der Richtung des tieferen Signals aus varianter KKF</p> <p>-----</p> <p>Quellenrichtungen direkt aus KKF's bestimmbar (<math>\Delta f &gt; \text{FG-Breite}</math>)</p>
<p><b>Lautheit</b></p> <p>bei korrekter Lokalisation höhere Quelle lauter als tiefere (Lautheit tieferes Signal 70% des höheren)</p>	<p>Amplitude entspricht Summe der Signale</p>	<p>Amplitude gering (über Flanke des FG-Filters abgeschwächt)</p>	<p>"Rückrechen-Mechanismus": Zuordnung von Lautheiten zu Quellenrichtungen</p>

Auch der *Klang* der Hörereignisse entspricht einer separaten Auswertung der Frequenzgruppen. Für Frequenzdifferenzen unterhalb einer Frequenzgruppen-Breite kann nur der Klang einer lokalisierten Schallquelle bestimmt werden, während aus der Richtung der anderen Schallquelle als Klang ein Gemisch der Quellensignale wahrgenommen wird. Dies entspricht dem Signalinhalt der Frequenzgruppen: Einzelton in der Neben-Frequenzgruppe, Signalgemisch in der Haupt-Frequenzgruppe. Auch hierbei ist die Klangbestimmung einer Quelle erst möglich, wenn diese Quelle lokalisierbar wird.

Die *Lautheit* der Hörereignisse stimmt nicht mit den erwarteten Ergebnissen bei separater Auswertung von Frequenzgruppen überein. Bei korrekter Lokalisation wurde z.B. bei den 500-Hz-Versuchsreihen eine höherfrequente Quelle als etwas lauter wahrgenommen als eine tieferfrequente Quelle. Im binauralen Modell erfaßt die obere Neben-Frequenzgruppe, aus der die Richtung der höherfrequenten Schallquelle direkt bestimmbar ist, aber einen wesentlich geringeren Teil der Gesamtleistung als die Haupt-Frequenzgruppe, was den gegenteiligen Lautheitseindruck nahelegen würde. Die direkte Zuordnung von Frequenzgruppen-Leistungen zu Lautheiten führt somit zu

falschen Ergebnissen. Es ist also ein zusätzlicher Mechanismus erforderlich, der im Modell den berechneten Einfallrichtungen Lautheiten zuweist.

Zum Nachvollziehen der Hörversuche ist also ein Mechanismus erforderlich, der Richtungsinformationen aus Frequenzgruppen mit varianten interauralen Parametern bzw. varianten Kreuzkorrelationsfunktion bestimmen kann und diesen Richtungen Lautheiten zuweist. Die Integration der richtungsselektiven Verarbeitung von Signalphasen oder -frequenzen in diesen Mechanismus ist nach den Hörversuchen nicht zwingend erforderlich. Dieser Mechanismus sollte zumindest die Möglichkeit offenhalten, auch zusätzliche Richtungsinformationen zu berücksichtigen, z.B. aus anderen Frequenzgruppen, aber auch optische Informationen (Sehen der Schallquelle), Erinnerung usw.

In den folgenden Kapiteln sollen mögliche Algorithmen für diese Zwecke untersucht werden. Schwerpunkt der Untersuchung sollen hierbei nicht nur Möglichkeiten zur modellmäßigen Beschreibung der Hörversuche sein, sondern vor allem Anwendungsmöglichkeiten zu Signalverarbeitungszwecken.