

Psychoakustische Messungen auf dem Gebiet der menschlichen Lokalisationsschärfe

Wersényi György

HU ISSN 1418-7108: HEJ Manuscript no.: EE-020110-A

Abstract

Virtual Acoustic Displays (VAD) haben ein breites Anwendungsgebiet. VAD bedeutet, mit virtuellen Schallquellen eine solche Umgebung zu erzeugen, daß der Hörer die Schallereignisse lokalisieren, identifizieren und diskriminieren kann. Die Schallereignisse, als akustische Darstellungen können auch als ergänzende Information neben den visuellen erscheinen (mit Bildschirm). VADs - man nennt sie auch Spatial Auditory Displays - können sowohl blinde Rechnerbenutzer als auch Flugsimulatorpiloten oder Computerspieler verwenden. In dieser Arbeit wird die Fähigkeit eines System für Kopfhörerwiedergabe präsentiert.

1. Einleitung

Bei der Realisierung eines VADs stellen wir zwei unabhängige Fragen: welche Schallereignisse (Töne, Klänge, Geräusche usw.) sind am besten geeignet um ein Ereignis des Bildschirms zu bezeichnen und identifizieren? Die zweite lautet: wie gut kann man während einer Wiedergabe diese Schallereignisse lokalisieren und diskriminieren. Die erste Frage ist psychologisch auszuwerten, die zweite ist eine reine (psycho)akustische.

Im Prinzip kann man virtuelle Schallereignisse entweder mit Hilfe von Lautsprechern (z.B. Standard-Stereophonie oder Mehrkanal-Lautsprecheranordnungen) oder mit einem Kopfhörer realisieren. Beide Möglichkeiten haben Vor- und Nachteile.

Es können natürlich mehrere, verschiedene 2D VADs vorhanden sein; wir konzentrieren uns auf eine rechteckförmige Darstellung. Prinzipiell kann man sich auch ein 3D-VAD vorstellen, bei dem das Merkmal „Entfernung“ oder „Tiefe“ durch die Entfernung des Schallereignisses interpretiert werden kann, z.B. ein Objekt, das auf den Hörer zukommt, oder bei überlappenden Objekten auf dem Bildschirm. Beim Entfernungshören sind weitere Aspekte zu berücksichtigen, deswegen bleiben wir bei unserer Untersuchung weiterhin zweidimensional (Abb.1.). Es sei auch angenommen, daß bei einer solchen Darbietung der Hörer immer in der Lage ist, den Lautstärkereglern zu betätigen, und die Lautheit („Entfernung“) der Quellen zu modifizieren.

2. Grundbegriffe und Definitionen

Das übliche Koordinatensystem für Hörversuche ist zum Kopf fixiert (Abb.2). Dieses s.g. kopfbezogene (Head-Related) System hat zwei ausgewählte Richtungen (Ebenen). Die Horizontalebene enthält die seitlichen Auslenkungen, bezeichnet mit dem Seitenwinkel φ . Die Position einer Schallquelle in der Medianebene - auch Vertikalebene genannt - wird mit dem Erhebungswinkel δ bezeichnet.

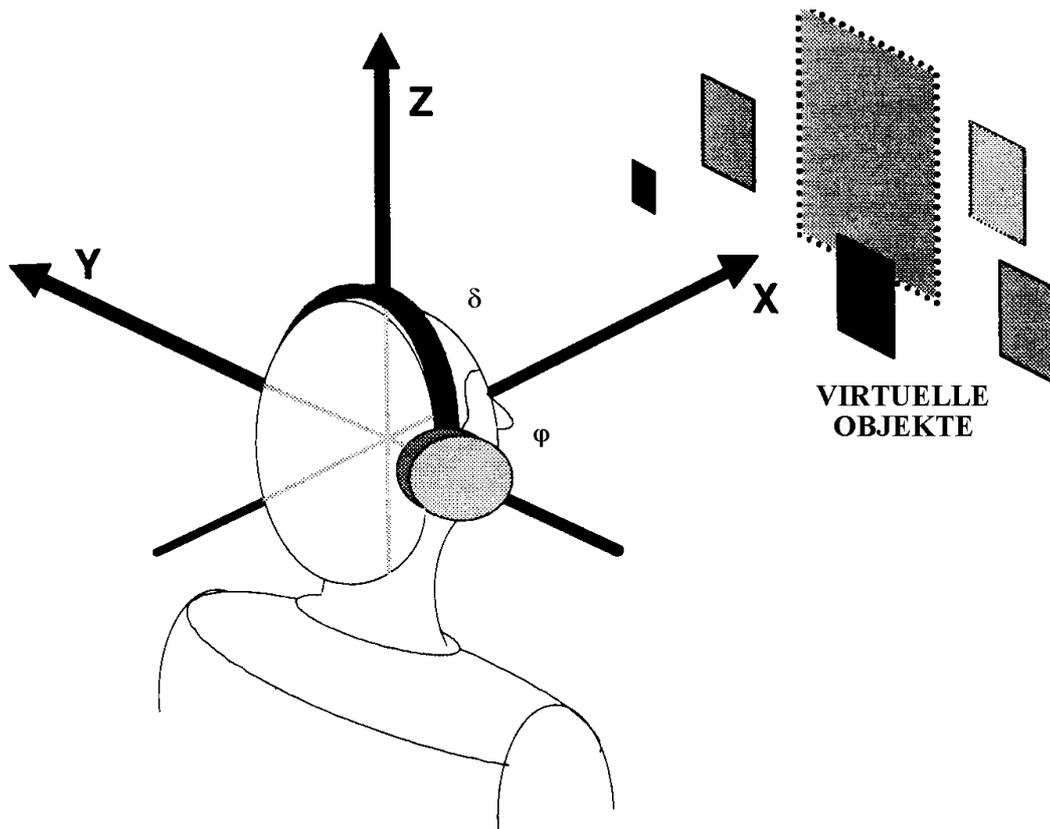


Abb.1. 2D akustische Darstellung. Die akustische Oberfläche ist parallel mit der Y-Z-Ebene. Ihr Mittelpunkt (Ursprung) ist in der Vorwärtsrichtung, $\varphi=\delta=0$. Virtuelle Schallquellen während der Untersuchung bewegen sich entweder parallel mit der Y-Achse, oder mit der Z-Achse.

In der Literatur sind bereits mehrere Ergebnisse über die menschliche Lokalisationsfähigkeit veröffentlicht worden. Am häufigsten setzen sie eine konstante Schallquellenentfernung vom Hörer voraus, d.h. die virtuellen Schallquellen sind rund um den Hörer platziert.

Die Definition der Lokalisationsunschärfe lautet: die kleinste Änderung eines oder mehrerer bestimmter Merkmale des Schallereignisses, die gerade zu einer wahrgenommenen Ortsänderung des Hörereignisses führt. Sie kennzeichnet die Tatsache, daß der Hörraum weniger differenziert ist als der Schallquellenraum. Die räumliche Auflösung des Gehörs ist geringer, als die mit physikalischen Messtechniken erzielbare. Als Lokalisationsunschärfe bezeichnen wir - wie in der Psychophysik üblich - diejenige Ortsänderung der Schallquelle, bei der gerade 50% der Versuchspersonen eine Änderung des Hörereignisortes bemerken. Andere Schwellendefinitionen sind auch möglich [14].

2.1. Untersuchungsmethoden

Die menschliche Lokalisation kann in unterschiedlicher Weise untersucht werden. Bei einer sog. „**absoluten**“ Messung hat die VP die Aufgabe, auf die Schallquelle zu zeigen, entweder mit einem Zeiger oder indem sie einfach den Kopf hinwendet („mit der Nase zeigen“). Das ist insbesondere für die Lautsprecherwiedergabe gedacht. Bei der Kopfhörerwiedergabe musste man Kopfbewegungssensoren anwenden. Die Frage ist: Wo ist die Schallquelle?

Viel einfacher ist die Frage nach einer Änderung der Schallquellenrichtung. Bei solchen Untersuchungen sind prinzipiell immer zwei Schallereignisse vorhanden: eine Referenz und eine bewegliche. Die VP muss sich entscheiden, ob sie eine Änderung wahrgenommen hat oder nicht. Bei diesem Fall wird die JND (Just Noticeable Difference) - auch als **MAA** (Minimum Audible Angle) bekannt – gemessen. In der Regel kann man sagen, daß dieses eine leichtere Aufgabe ist und bessere Resultate ergibt. Bei den sogenannten **Intervallurteil**-Methoden wird die Versuchsperson aufgefordert, auf Unterschiede zwischen dem Wahrgenommenen zu achten.

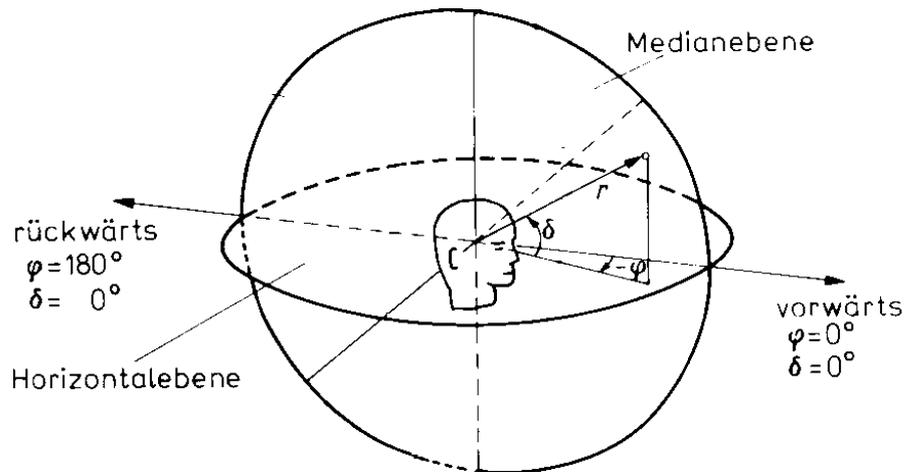


Abb.2. Kopfbezogenes Koordinatensystem (Head-Related Coordinate System) nach Blauert [14]. Erhebungswinkel $\delta=0^\circ$ bestimmt die Horizontalebene, Seitenwinkel $\varphi=0^\circ$ bestimmt die Medianebene.

2.2. Die HRTFs

Der größte Unterschied zwischen Lautsprecher- und Kopfhörerwiedergabe liegt in der Filterfunktion des äußeren Ohres. Unter normalen Hörbedingungen oder bei Lautsprecherwiedergabe kann das Außenohr seinen natürlichen Richtungs-Filtereffekt ausnutzen. Die Ohrmuschel, der Kopf, die Schultern und der Oberkörper beeinflussen die Übertragung. Die sog. Außenohr-Übertragungsfunktionen (HRTFs: Head-Related Transfer Functions) sind komplexe Funktionen, welche die Übertragung vom Freifeld bis zum Trommelfell richtungsabhängig beschreiben [14][31][32]. Es ist erwiesen, daß beim Eingang des Gehörs die HRTFs die selbe räumliche Information haben wie am Trommelfell [15].

Jeder Mensch hat sein eigenes, individuelles HRTF-Set. Bei einer Kopfhörerwiedergabe ist diese Funktion unterbunden und muss elektronisch nachgebildet werden. Um eine korrekte binaurale Wiedergabe zu erzeugen muß man alle unerwünschte Übertragungsfunktionen in der Übertragungskette eliminieren, so z.B. des Mikrofon und andere Elemente, die bei der HRTF-Aufnahme vorhanden waren, und vor allem muß der Kopfhörer gründlich angepaßt werden (Equalisation). Erst danach können die HRTFs implementiert werden. Das erfolgt am häufigsten durch die HRIRs (Head-Related Impulse Responses), die Zeitfunktions-Version der HRTFs. HRIRs sind die inversen Fourier-Transformierten der HRTFs und werden als FIR-Filters realisiert. Heutzutage kann man alle gespeicherten FIR-Filter und das Erregungssignal im Zeitbereich in Echtzeit konvolvieren [31][34][33]. Die Literatur beschäftigt sich weitgehend damit, wie man die benötigten HRTFs aufnehmen kann [14][31][32]. Es besteht die Möglichkeit, die eigene individuelle oder auch die von einer gut lokalisierenden Person, oder eben von einem Kunstkopf zu messen. Die Qualität der Lokalisation mit diesen HRTFs ist ein breites Forschungsgebiet [28][19][29]. Bis heute wissen wir, dass die beste Lokalisation durch die eigenen HRTFs erreicht werden kann. Da es eine langzeitige und aufwändige Arbeit ist, die individuellen HRTFs zu messen, suchen wir nach Möglichkeiten nicht-individuelle HRTFs wirksam anzuwenden [23][20][21][24][26]. Dazu zählen auch die Kunstkopf-HRTFs, welche die niedrigste binaurale Qualität haben, aber dennoch eine recht gute Lokalisation ermöglichen. Sie sollten den „durchschnittlichen“ Mensch modellieren - mit Erfolg, denn die Lokalisationsmöglichkeiten sind auch erst durchschnittlich. Auch andere Möglichkeiten existieren, um die HRTFs spektral „durchschnittlicher“ zu machen [35]. Die beste Möglichkeit ist vermutlich, die HRTFs von einer gut lokalisierenden Person zu verwenden.

2.3. Weitere Parameter

Zur Lokalisation nutzt das Gehör grundsätzlich zwei Merkmale. Die, welche mit nur einem Ohr wahrgenommen und ausgewertet werden können nennt man **monaurale** Merkmale, und treten am häufigsten in der Medianebene auf. Eine Schallquelle in der Medianebene bildet einen speziellen Fall, da von diesen

Richtungen im Idealfall beide Ohren gleichermaßen beschallt werden, und die Ohrsignale identisch sind. Das Gehör versucht den Schallquellenort in der Medianebene durch monaurale Merkmale zu finden. Dabei benutzt es die spektralen Eigenschaften der HRTFs. Bei einer seitlichen Auslenkung unterscheiden sich die Ohrsignale. Die **interauren** Merkmale sind die Pegeldifferenzen (IID oder ILD: Interaural Intensity oder Level Difference) und/oder die Zeitdifferenzen (ITD: Interaural Time Difference). Das Gehör lokalisiert in der Horizontalebene auf der Basis von ITDs und IIDs. Es ist also möglich, seitliche Auslenkungen virtuell so zu produzieren, daß zwischen den Ohrsignalen Pegelunterschiede und/oder Laufzeitunterschiede auftreten. In natürlichen Umgebungen wird das Gehör sowohl IIDs, ITDs als auch den Effekten der HRTFs ausgesetzt.

Signale anderer Dauer und/oder anderem **Spektrum** können zu abweichenden Messwerten führen. Bei reinen Zeitdifferenzen der Ohrsignale ist die seitliche Auslenkung des Hörereignisses wie auch die Lateralisationsunschärfe vorwiegend durch die Signalanteile unterhalb von 1500 Hz bestimmt. Unter Freifeldbedingungen (bei gleichzeitigen interauralen Zeit- und Pegeldifferenzen) ist auch die Lokalisationsunschärfe am geringsten, wenn das Signal Anteile unterhalb von 1500 Hz enthält. Die Lateralisation bei Signalen über 1600 Hz erfolgt durch die Hüllkurvenauswertung. Interaurale Zeitdifferenzen sind bei der Lokalisation von Hochfrequenzsignalen mit relative langsamer Hüllkurvenmodulation nützlich [3].

So lassen die Ergebnisse vermuten, dass bis zu einer Signaldauer von etwa 700 ms die Lokalisationsunschärfe mit der Dauer sinkt. Andere Vermutungen konnten das nur mit Einschränkung bestätigen. Jedenfalls ist es sicher, daß bei allen Signalarten die Lokalisationsunschärfe wächst, wenn die Quelle aus der Vorwärtsrichtung ausgelenkt wird. Bei einem Schallpegel zwischen 40 und 80 dB ist die Lokalisationsunschärfe etwa konstant und minimal. Die Lateralisationsunschärfe sinkt mit steigendem **Pegel** und steigender **Signaldauer**, wenn eine interaurale Zeitdifferenz da ist.

Die Filterwirkung des Kopfes und der Ohren kann sich bei Schmalbandsignalen nur auf den Pegel auswirken, nicht aber in einer Veränderung der Spektral-Anteile untereinander. Bei breitbandigen Signalen können sich jedoch sowohl die Frequenzfunktion als auch die Phasenlage ändern. Ein breitbandiges Ohrsignal enthält mehr Information über den Schallquellenort als ein schmalbandiges. Deswegen wird bei **Breitbandsignalen** öfter Richtungskoinzidenz auftreten. **Schmalbandsignale** enthalten kein Information über die Schalleinfallrichtung, die vom Gehör ausgewertet werden könnte. Die Hörereignisrichtung wird aufgrund der Frequenz der Ohrsignale und durch die sog. „richtungsbestimmenden Bänder“ ausgewertet [14]. Sie sind dafür verantwortlich, daß das Hörereignis von „vorne“ kommt, anstatt von hinten. In den Bändern, in denen mehr Signalleistung auftritt, wird die Richtung bestimmt. Wenn die Leistung gleichgroß ist, kann es zur Im-Kopf-Lokalisation oder zu einem diffusen Schallereignis kommen. Gauß-Töne sind solche Signale, deren Zeitfunktion und komplexes Spektrum gauss-verteilt sind.

Weißes Rauschen ist breitbandig, hat einen flachen Frequenzgang, unbegrenzte Energie und kann beliebig lange andauern - im Gegensatz zu Dirac-Impulsen. Deswegen sind solche Signale und deren gefilterte Versionen gut für die Untersuchungen geeignet. Durch lineare Filterung lassen sich aus weißem Rauschen regellose Vorgänge mit beliebigen gewünschten Bandbreiten und Spektren herstellen, wie bei unserem Fall. Damit kann man auch Eigenresonanzen und stehende Wellen vermeiden. Die Hüllkurve spielt auch keine Rolle.

Die Häufigkeit eines Lokalisationsfehlers ist von der Art des Signals abhängig. Bei breitbandigen, insbesondere längeren oder bei wiederholten Signalen wurde eine recht gute Übereinstimmung gefunden und auch eine gute Lokalisation. Die Lokalisationunschärfe ist trotzdem recht groß bei Änderungen der Quellenrichtung (dreimal so groß wie in der Horizontalebene). Wenn ein Signal breitbandig ist und Anteile oberhalb von 7 kHz enthält, sinkt die Lokalisationsunschärfe bei Quellenrichtungsänderungen bis auf 4°. Bei Schmalbandsignalen (1-2 Terzen) hängt die Lokalisation nicht vom Ort der Quelle, sondern nur von der Frequenz ab. Bei Tiefpassrauschen und bei sehr kurzen Signalen (weniger als 1 ms) treten Richtungsinversionen auf, die von der oberen Grenzfrequenz abhängig sind. Ein allgemeiner Fehler ist, dass man Hörereignisse mit größerer Tonhöhe weiter oben lokalisiert, als „tiefe“ Töne.

In der Regel hilft bei der Lokalisation, ob die Versuchsperson das Signal kennt oder es vertraut ist, aber nicht in jedem Fall. Die Lokalisation wird durch *a priori*-Kenntnisse, **Erwartungen und Erfahrungen** der Versuchsperson beeinflusst. Beim Richtungshören in der Medianebene sind Adaptations- und **Lernvorgänge** beobachtbar. Die Lernprozesse sind wichtig, aber es ist noch nicht ganz geklärt, warum. Die Lokalisation ist also zeitabhängig.

3. Frühere Ergebnisse

Viele vorherige Untersuchungen haben sich bereits mit der menschlichen Lokalisation beschäftigt. Der Bereich des „schärfsten Hörens“ liegt in oder nahe der Vorwärtsrichtung. Die absolute untere Grenze für die

Lokalisationsunschärfe liegt bei rund 1° . Sie hängt von Signaldauer, Schallquellenort, Schalldruckpegel und Frequenzbereich ab. Eine Übersicht über die Messungen der Lokalisationsunschärfe zeigt Tabelle 1 und Tabelle 2 in der Horizontalebene bzw. in der Medianebene.

AUTOR	SIGNAL	GRADWERTE	SONSTIGES
Blauert [14]	(alle Signale, vorwärts)	1°	Absolutes Minimum
	Klicks, Klickfolge	0,75°-2°	
	Dauertöne	1°-4°	
	Schmalbandrauschen, Gauss-Töne	1,4°-3,3°	
	Sprache	0,9°-1,5°	
Haustein, Schirmer [1]	Breitbandrauschen	3,2°	
	100 ms weisser Rauschimpuls	± 3,6° vorwärts ± 9 bis 10° seitlich	900 VPn 700 phon
Hartmann [22]		1°-2°	Absolutes Minimum
Heffner, Heffner [25]	Rauschsignal	1,3°-1,8° vorwärts 9°-10° seitlich	MAA Untersuchung
Litovsky, Ashmed [16]		1°	MAA Untersuchung
Kremer [6]		3°-5°	
Barfield, Furness [3]		1°-5° vorwärts 5°-10° seitlich	MAA Untersuchung
		10°-20°	Absolutmessung
Begault [8]	Breitbandsignale	1°	MAA Untersuchung, optimale Bedingungen
Oldfield, Parker [9, 10]	azimuthal mean value	9°	Absolutmessung
	Azimuth-Fehlerwerte mit HRTF Filterung	4°-6°	$\varphi = 0^\circ-80^\circ$
	Azimuth-Fehlerwerte ohne HRTF Filterung	11,9°	
Middlebrooks [12]	Durchschnittfehler von 150 ms breitbandigem Rauschsignal	5,8°	VP zwischen 19 und 36 Jahren
McKinley, Ericson [17]	Durchschnittfehler	5°	
	MAA Untersuchung	5°	
	Durchschnittfehler für Oktavband- Geräusche	4,4°-5,9°	10 VP, absultmessung
	Durchschnittfehler für Rosarauschen	6°-7°	10 VP, absultmessung
	MAA für 500 Hz Sinuston	4°-5° vorwärts	7 VP, absultmessung
Makous, Middlebrooks [24]	Minimaler Durchschnittfehler für 150 ms Störsignalbursts	2° vorwärts	Freifeld, Absolutmessung
Middlebrooks [26]	Durchschnittfehler	17,1° 14,7°	Nicht individuelle HRTFs (other-ear-condition) individuelle HRTFs (own- ear-condition)
Duda [27]	Durchschnittfehler	5,1° 4,5°	Max.Likelihood Schätzungsprozedur von KEMAR-Daten Mit Menschlichen HRTFs

Tabelle 1. Zusammenfassung der Ergebnisse der früheren Untersuchungen in der Horizontalebene.

AUTOR	SIGNAL	GRADWERTE	SONSTIGES
Blauert [14]	Unbekannter Sprecher	17° vorwärts	20 VPn
	Bekannter Sprecher	9° vorwärts ± 10° bei $\delta=36^\circ$ ± 13 bis ± 22° oben	7 VPn
	Weisses Rauschen	4° vorwärts	2 VPn
Wettschurek [4]	MAA für weisses Rauschen	± 4° vorwärts ± 10° oben	8-10° Standard Deviation
	Tiefpassrauschen mit 4 kHz Grenzfrequenz	± 8° vorwärts ± 20° oben	
Oldfield, Parker [9, 10]	elevational mean value	12°	Absolutmessung
	Erhebungsfehlerwerte mit HRTF Filterung	6°-8°	
	Erhebungsfehlerwerte ohne HRTF Filterung	21,9°	
Wenzel, Foster [11]	durchschnittfehler	ca. 25° (niedrigere Elevationen, vorwärts) ca. 22° seitlich	Freifeld 16 VPn
		ca. 24° (niedrigere Elevationen, vorwärts) ca. 23° seitlich	16 VPn virtuelle Quellen, Kopfhörerwiedergabe, non-individuelle HRTFs
Wightman, Kistler [36]	Durchschnittfehler	ca. 20° (niedrigere Elevationen, vorwärts) ca. 18° seitlich	Freifeld
		ca. 21° (niedrigere Elevationen, vorwärts) ca. 20° seitlich	Kopfhörer
Middlebrooks [12]	Durchschnittfehler von 150 ms breitbandigem Rauschsignal	5,7°	VP 19 und 36 Jahre
McKinley, Ericson [17]		30°-35°	MAA von KEMAR HRTFs
Moller [19, 29]	Relativer Lokalisationsfehler	36% 55%	HRTFs von einem zufällig ausgewählten Person KEMAR HRTFs
Makous, Middlebrooks [24]	Minimaler Durchschnittfehler für 150 ms Störsignalbursts	3,5° vorwärts 20° seitlich	Freifeld, Absolutmessung 94% der VPn waren innerhalb der 10° Standard Deviation
Duda [27]	Durchschnittfehler	12° 19,2° 17,2°	Max.Likelihood Schätzungsprozedur von KEMAR-Daten Mit menschlichen HRTFs Breitbandsignale bis 12kHz

Tabelle 2. Zusammenfassung der Ergebnisse der früheren Untersuchungen in der Medianebene.

4. Unsere Untersuchung

Im Folgenden wird die Lokalisation in der Median- und in der Horizontalebene untersucht. Auf Grund früherer Ergebnisse des GUIB-Projektes (Graphical User Interfaces for Blind People) werden wir weiterhin mit einem Kopfhörer arbeiten [34]. Dabei beschränken wir uns auf die zweidimensionale „flache“ Darbietung, und auf einfache Signale. Weiterhin nehmen wir auch an, daß sich die akustisch dargestellte Oberfläche vor dem Hörer befindet (vgl. Abb.1.) und bekannte Fehler, wie Vorn-Hinten-Verwechslung oder axialsymmetrische Richtungsänderungen (die Hörereignisse treten nicht in der Schall-Einfallrichtung auf sondern systematisch in einer Richtung bezüglich der Ohrenachse) durch *a priori* Kenntnisse beeinflusst werden.

4.1. Basis und Ziel der Untersuchung

Das oben genannte GUIB-Projekt hat sich mit einer akustischen Unterstützung für blinde Rechnerbenutzer beschäftigt. Es ging um eine Lösung die „Earcons“ (auditive Repräsentation von Bildschirm-Icons), die richtige Form der Dateneingabe (Tastatur, Touch Pads kombiniert mit Braille Display und Sprache) und die Diskriminationsfähigkeit festzustellen. Dabei hat sich ergeben, daß blinde Versuchspersonen nicht besser lokalisieren können als sehende. Damals nahmen nur wenige VPn bei der Untersuchung teil, so daß repräsentative Ergebnisse fehlen.

Unseres Ziel ist die Diskriminationsfähigkeit und deren Abhängigkeit von den Signalen festzustellen. Diese Signale und das Messverfahren liefern Ergebnisse für spätere GUIB Anwendungen. Wir beschränken uns auf eine rechteckförmige 2D Oberfläche mit einem maximalen Winkel von $\pm 60^\circ$ seitlich und vertikal.

Das Messprogramm läuft auf einem PC. Die Echtzeit DSP Effekte werden durch eine Beachtron™ Karte durch den mitgelieferten Sennheiser HD 540 Kopfhörer und HRTFs realisiert. Diese Karte wurde zu diesem Kopfhörer optimiert: ist diffuse-field-equalised, circumaural, offen und dynamisch. Die HRTFs der Beachtron-Karte stammen aus einer Untersuchung von *Wightman und Kistler* [36]. Es wurde die am besten lokalisierende Person ausgewählt. 72 HRTFs wurden aufgenommen, jeweils für das linke und das rechte Ohr von 6 verschiedenen Elevationen (zwischen -36° und $+54^\circ$) und von allen 30° Positionen horizontal. Die fehlenden HRTFs sind von den vier benachbarten interpoliert [33]. Die HRIRs für die zeitliche Konvolution sind Minimalphasen-FIR-Filter mit 75 Punkten. Minimalphase bedeutet, daß das Filter die kürzeste Phasen-Verzögerung unter allen stabilen Filtern mit der selben Amplitudencharakteristik hat, d.h. dieses Filter hat die kürzeste Impulsantwort. Minimalphasen-HRTFs könnten dabei zu einer reduzierten Lokalisationsschärfe führen [13], [20]. Die HRIRs werden durch Echtzeit-Konvolution mit den monotonischen Schallquellen von der Festplatte (wave-files) konvolviert mit 44.1 kHz und 16 bit Auflösung.

4.2. Vorbereitung und Testergebnisse

Vor der richtigen Untersuchung sind Testmessungen mit 7 Versuchspersonen durchgeführt worden. Dabei wurde die optimale Lautstärke, Länge und Darbietungsdauer der Signale festgestellt, die als Rahmenbedingungen für die repräsentative Untersuchung dienen. Da die Entfernung der Schallquelle bei einer seitlichen Auslenkung immer größer wird, wird sie dabei gedämpft. Die geringste Entfernung vom Hörer ist in der Vorwärtsrichtung (D_0). Bei Bewegungen parallel zur Y-Achse ändert sich diese als $1/\cos$ -Funktion des Seitenwinkels (Abb. 3).

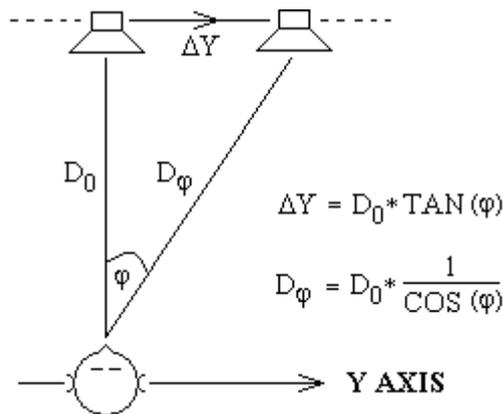


Abb.3. Die Schallquellen-Entfernung in der Horizontalebene ist eine 1/COS-Funktion des Seitenwinkels φ , wenn die Schallquellen sich parallel mit der Y-Achse bewegen. Die Y-Koordinaten sind eine TAN-Funktion des Seitenwinkels.

Es besteht keine Notwendigkeit für einen GUI basierende VAD unbedingt diese 2D Lösung zu wählen. Man kann auch kugel- oder zylinderförmige Oberflächen wählen. Diese Zuordnungen entsprechen jedoch nicht den üblichen Anwendungen am besten, z.B. unter MS Windows, wo Fenster, Icons und alle Ereignisse auf einer rechteckigen Oberfläche erscheinen.

Die individuelle Anpassung der HRTFs an die Testpersonen betrifft die Messung der Entfernung der Gehörgängeingangs. So besteht die Möglichkeit die Größe des Kopfes einzustellen, und damit die HRTFs individueller zu machen (Abb.4).

Die Aufgabe während der MAA-Untersuchung besteht darin zwischen drei möglichen Antworten zu wählen (three-categorie-forced-choice), nämlich: „kein Unterschied“, „unsicher“ und „verschiedene Schallquellenorte“ je nachdem, ob die VP die nacheinander dargebotenen Signalimpulse mit Sicherheit, nur mit Unsicherheit oder überhaupt nicht diskriminieren kann.

Die Diskriminationsfähigkeit wird für drei Signale von etwa gleicher subjektive Lautstärke aufgenommen (siehe Abb.5). **Signal A** ist ein 300 ms digital erzeugtes breitbandiges weißes Rauschen, mit flachem Spektrum zwischen 20 und 22000 Hz. **Signal B** entsteht durch passive Tiefpaß-Filterung von Signal A mit einer oberen Grenzfrequenz von 1500 Hz. **Signal C** entsteht durch Hochpaß-Filterung von Signal A mit einer unteren Grenzfrequenz von 7000 Hz. Signal B ist +10 dB, Signal C ist +6 dB stärker als Signal A.

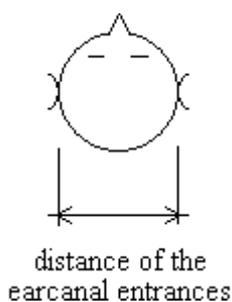
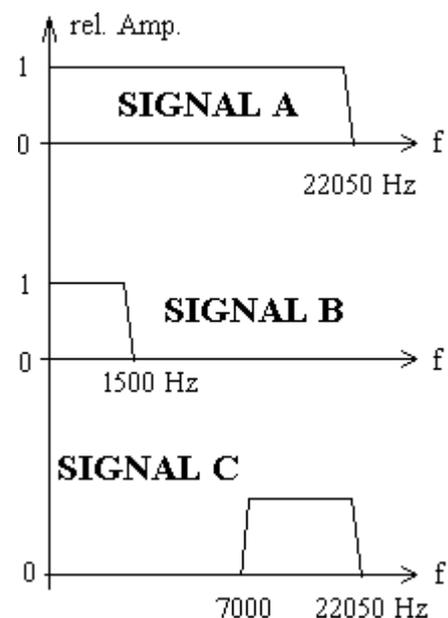


Abb.4. Distanz zwischen den Gehöreingängen der Versuchsperson.

Abb.5. Frequenzgang der Signale. Signal A ist weißes Rauschen mit 16 bit und 44.1 kHz Auflösung. Signal B und C sind werden Tiefpaß- bzw. Hochpaßfilterung mit einer Grenzfrequenz von 1500 und 7000 Hz erzeugt.



Die Diskriminationsfähigkeit wird in vier Schritten (nach rechts, links, oben und unten) in der Horizontalebene bzw. in der Medianebene untersucht. Die Rauschimpuls-paare erfolgen in einem Zwischenzeitraum von 400 ms (Abb.6). Der erste Impuls kommt immer von der Referenzrichtung (Anfangs vom Ursprung), die andere von einer seitlichen Auslenkung oder von einem bestimmten Erhebungswinkel.

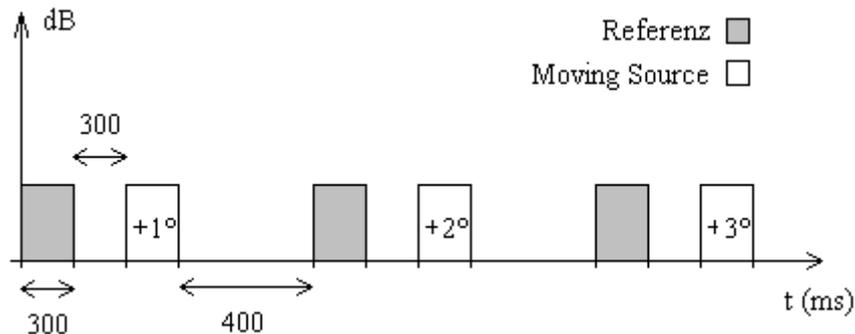


Abb.6. Signaldarbietung während der MAA-Untersuchung. Die 300 ms Rauschbursts-Paare erfolgen in 400 ms takt. Zwischen den Bursts zum MAA-Vergleich sind 300 ms. Der erste Impuls kommt immer von einer Referenzrichtung, der andere von einer seitlichen oder vertikalen Auslenkung (in $+1^\circ$ Schritte).

Der zweite Impuls bewegt sich in 1° Schritten. Erstmal entfernt er sich vom Referenzpunkt bis die VP die Impulse mit Sicherheit diskriminieren kann („Punkt der Sicherheit“). Dann kommt er rückwärts auf die Referenzquelle zu bis die VP die Impulse nicht mehr diskriminieren kann. Zwischen diesen Punkten kann die VP eine Unsicherheits-Region bestimmen. Als neuer Referenzpunkt wird dann der erste Punkt festgestellt, wo die VP in beiden Bewegungsrichtungen des zweiten Impuls mit Sicherheit die Impulse diskriminieren konnte. In der Horizontalebene ist der Ausgangspunkt zur Rückwärtsbewegung 5° , in der Medianebene 10° weiter vom Punkt der Sicherheit entfernt. In den Horizontalrichtungen werden maximal sechs neue Referenzpunkte bestimmt, in den Vertikalrichtungen maximal zwei.

4.3. Auswertung der Ergebnisse

Die Versuchspersonen sind ungeübte Freiwillige zwischen 25 und 50 Jahre alt, durchschnittliche Rechnerbenutzer mit durchschnittlicher Erfahrung in der Kopfhörerwiedergabe. Keiner der 50 Personen (25 männliche und 25 weibliche) hat Gehörschäden oder andere Probleme, welche die Untersuchung oder die Lokalisation selbst beeinflussen könnten. Vor und nach der Untersuchung werden außer den Ergebnissen auch andere Daten im Datenblatt aufgenommen (Name, Alter, Geschlecht, Datum, verbrauchte Zeit für die Untersuchung, tabellenförmige Ergebnisse, allgemeine Meinung über die Untersuchung).

Die Messung findet in einem beleuchteten schalltoten Raum statt. Der Kopf ist nicht fixiert, aber die VP ist angewiesen ihn ruhig zu halten. Um sich an die Umgebung zu gewöhnen, bekommt jede VP eine ausführliche Erklärung über das Verfahren und führt einen Richtungstest durch, wobei sie die Signale und das Messverfahren kennenlernt und sich ihr Gehör den reflexionsarmen Bedingungen anpassen kann.

Die Ergebnisse können sowohl graphisch als auch in Tabellenformat gezeigt werden. Die Tabellen zeigen die durchschnittlichen Ergebnisse von der Horizontalebene und von der Medianebene für alle Signale. Das zeigt, dass Signal A die beste räumliche Auflösung (Diskriminierbarkeit) hat, Signal C ist fast damit identisch, Signal B kann mit etwa $5-6^\circ$ schlechter lokalisiert werden.

Durchschnittsergebnisse

Signal A links

AVG	8,28°	16,71°	25,57°	34,71°	43,42°	47,8°
MAX	12°	22°	36°	48°	58°	70°
MIN	4°	8°	13°	18°	24°	31°

Signal A rechts

AVG	7,28°	15,28°	25,28°	33,85°	43,71°	51,66°
MAX	11	24	35	43	56	69
MIN	3	7	12	18	25	32

Signal A oben

AVG	16,33°	34,16°
MAX	22	44
MIN	11	25

Signal A unten

AVG	17,7°	34°
MAX	29	53
MIN	8	21

Signal B links

AVG	9,71°	20,57°	31,7°	41,85°	49,5°	52,8°
MAX	14	30	44	58	70	65
MIN	6	12	21	28	36	43

Signal B rechts

AVG	7,8°	16,57°	26,43°	37,28°	42,66°	52,66°
MAX	12	27	46	63	57	68
MIN	4	10	16	25	31	38

Signal B oben

AVG	21°	39,16°
MAX	27	53
MIN	15	30

Signal B unten

AVG	18,5°	35,8°
MAX	24	46
MIN	13	25

Signal C links

AVG	9,7°	19,4°	28,0°	38,9°	44,0°	48,6°
MAX	13	28	43	55	56	63
MIN	5	10	16	22	28	33

Signal C rechts

AVG	8,9°	17,9°	30,1°	40,2°	52,0°	60,4°
MAX	11	23	39	53	64	66
MIN	6	10	19	32	44	51

Signal C oben

AVG	16,4°	37,4°
MAX	21	55
MIN	11	26

Signal C unten

AVG	20,6°	37,4°
MAX	30	55
MIN	13	27

5. Zusammenfassung

Das System wir präsentiert haben ist in der Lage Raumeindruck und Schallquellenort-Information zu übertragen nur mit bechrenkter Möglichkeit. Die übliche Fehler die bei der Kopfhörerwiedergabe auftauchen sind vorhanden. Versuchspersonen können vertikale Richtungen sehr schlecht lokalisieren. Horizontal könnte man 5-7 Schallquellen plazieren um eine räumliche Auflösung von 15 Grad zu erhalten.

Im weiteren sollte diese Auflösung mit wirklichen „Earcons“ dargestellt und untersucht werden.

6. References

- [1] B.G. HAUSTEIN, W. SCHIRMER, „Messeinrichtung zur Untersuchung des Richtungslokalisationsvermögens“, Hochfrequenztech. und Elektroakustik **79**, pp. 96-101, 1970.
- [2] R. WETTSCHUREK, „Über Unterschiedsschwellen beim Richtungshören in der Medianebene“, Gemeinschaftstagung für Akustik und Schwingungstechnik, Berlin, pp. 385-388, VDI-Verlag, Düsseldorf, 1970.
- [3] W. BARFIELD, T.A. FURNESS III (Editors), „Virtual Environments and Advanced Interface Design“, Oxford Univ. Press, New York, Oxford, pp.295-302, 1995.
- [4] R. WETTSCHUREK, „Die absoluten Unterschiedswellen der Richtungswahrnehmung in der Medianebene beim natürlichen Hören sowie beim Hören über ein Kunstkopf-Übertragungssystem“, *Acoustica* **28**, pp. 197-208, 1973.
- [5] J.C. MIDDLEBROOKS, D.M. GREEN, „Sound localization by human listeners“, *Ann.Rev.Psychol.* **42**, pp. 135-159, 1991.
- [6] <http://www.dasp.uni-wuppertal.de/audite/psychoak/psychoak26.htm>
- [7] <http://www.cs.stir.ac.uk/~frm/ewns2/schaikabstract.html>
- [8] D.R. BEGAULT, „3-D Sound for Virtual Reality and Multimedia“, Academic Press, London, UK, 1994.
- [9] S.R. OLDFIELD, S.P.A. PARKER, „Acuity of sound localisation: a topography of auditory space I-II.“, *Perception* **13**, pp. 581-617, 1984.
- [10] S.R. OLDFIELD, S.P.A. PARKER, „Acuity of sound localisation: a topography of auditory space III.“, *Perception* **15**, pp. 67-81, 1986.
- [11] E.M. WENZEL, S.H. FOSTER, „Perceptual consequences of interpolating head-related transfer functions during spatial synthesis“, *Proc. of the ASSP Workshop on Applications of Signal Processing to Audio and Acoustics*, New York, IEEE Press, 1993.
- [12] J.C. MIDDLEBROOKS, „Spectral Shape Cues for Sound Localization“, in *Binaural and Spatial Hearing in Real and Virtual Environments* (edited by R.H. Gilkey and T.R. Anderson), Lawrence Erlbaum Ass., Mahwah, New Jersey, pp. 77-97, 1997.
- [13] S.H. FOSTER, E.M. WENZEL, „Virtual Acoustic Environments: The Convolvotron“, CDemos system presentation at SIGGRAPH'91, 18th ACM Conference on Computer Graphics and Interactive Techniques, Las Vegas, NV (ACM Press, New York), 1991.
- [14] J. BLAUERT, *Spatial Hearing*, The MIT Press, MA, 1983.
- [15] H. MOLLER, „Fundamentals of binaural technology“, *Applied Acoustics* **36**, pp. 171-218, 1992.
- [16] R.Y. LITOVSKY, D.H. ASHMEAD, „Development of Binaural and Spatial Hearing in Infants and Children“, in *Binaural and Spatial Hearing in Real and Virtual Environments* (edited by R.H. Gilkey and T.R. Anderson), Lawrence Erlbaum Ass., Mahwah, New Jersey, pp. 571-592, 1997.
- [17] R.L. McKINLEY, M.A. ERICSON, „Flight Demonstration of a 3-D Auditory Display“ in *Binaural and Spatial Hearing in Real and Virtual Environments* (edited by R.H. Gilkey and T.R. Anderson), Lawrence Erlbaum Ass., Mahwah, New Jersey, pp. 683-699, 1997.
- [18] W.M. HARTMANN, B. RAKERD, „On the minimum audible angle – A decision theory approach“, *J. Acoust. Soc. Am.* **85**, pp. 2031-2041, 1989.
- [19] H. MOLLER, D. HAMMERSHOI, C. B. JENSEN, M. F. SORENSEN, „Evaluation of artificial heads in listening tests“, *J. Acoust. Soc. Am.* **47(3)**, pp. 83-100, 1999.
- [20] E. M. WENZEL, M. ARRUDA, D. J. KISTLER, F. L. WIGHTMAN, „Localization using nonindividualized head-related transfer functions“, *J. Acoust. Soc. Am.* **94(1)**, pp. 111-123, 1993 July.
- [21] J. C. MIDDLEBROOKS, „Individual differences in external-ear transfer functions reduced by scaling in frequency“, *J. Acoust. Soc. Am.*, **106(3)**, pp. 1480-1491, 1999.
- [22] W. M. HARTMANN, „How we localize sound“, *Physics Today*, pp. 24-29, 1999 November.

- [23] H. MOLLER, M. F. SORENSEN, C. B. JENSEN, D. HAMMERSHOI, „Binaural Technique: Do We Need Individual Recordings?“, *J. Audio Eng. Soc.* **44(6)**, pp. 451-469, 1996 June.
- [24] J. C. MAKOUS, J. C. MIDDLEBROOKS, „Two-dimensional sound localization by human listeners“, *J. Acoust. Soc. Am.*, **87(5)**, pp. 2188-2200, 1990 May.
- [25] R.S. HEFFNER, H.E. HEFFNER, „Sound localization acuity in the cat: Effect of azimuth, signal duration and test procedure“, *Hear. Res.* **36**, pp. 221-232, 1988.
- [26] J. C. MIDDLEBROOKS, „Virtual localisation improved by scaling nonindividualized external-ear transfer function in frequency“, *J. Acoust. Soc. Am.* **106(3)**, pp. 1493-1510, 1999.
- [27] R.O. DUDA, „Elevation Dependence of the Interaural Transfer Function“, in *Binaural and Spatial Hearing in Real and Virtual Environments* (edited by R.H. Gilkey and T.R. Anderson), Lawrence Erlbaum Ass., Mahwah, New Jersey, pp. 49-75 1997.
- [28] GY. WERSENYI, „Measurement system upgrading for more precise measuring of the Head-Related Transfer Functions“, in *Proceedings of Inter-Noise 2000*, Vol.II., Nice, pp. 1173-1176, 2000.
- [29] H. MOLLER, „On the quality of artificial head recording systems“, in *Proceedings of Inter-Noise 97*, vol. II., pp. 1139-1142, Budapest, 1997.
- [30] M. BODDEN, G. CANAVET, J. GRABKE, K. HARTUNG, T. TAKAHASHI „Räumliches Hören in komplexen akustischen Umgebungen“, *DAGA 94*, pp. 1137-1140, 1994.
- [31] S. HEHRGART, V. MELLERT, „Transformation characteristics of the external human ear“, *J. Acoust. Soc. Am.* **61(6)**, pp.1567-1576, 1977 June.
- [32] H. MOLLER, M. F. SORENSEN, D. HAMMERSHOI, C. B. JENSEN, „Head-Related Transfer Functions of human subjects“, *J. Audio Eng. Soc.* **43(5)**, pp. 300-321, 1995.
- [33] Crystal River Engineering, Inc. : BEACHTRON – Technical Manual, Rev.C., 1993.
- [34] K. CRISPIEN, H. PETRIE, „Providing Access to GUI's Using Multimedia System – Based on Spatial Audio Representation“, *AES 95th Convention Preprint*, New York, 1993.
- [35] D.J. KISTLER, F.L. WIGHTMAN, „Principal Component Analysis of Head-Related Transfer Functions“, *J. Acoust. Soc. Am.* **88**, pp. 98, 1990.
- [36] F.L. WIGHTMAN, D.J. KISTLER, „Headphone Simulation of Free-Field Listening I-II.“, *J. Acoust. Soc. Am.* **85.**, pp. 858-878, 1989 February.