

Horen in de ruimte

Een onderzoeksverslag van: Lucien Immink

Rotterdam, 3 juni 2002

Voorwoord

Dit document is geschreven in het kader van het onderzoek van Lucien Immink wat hij schreef in het derde jaar van de opleiding Media Technologie. Van de student wordt verwacht een literair onderzoek te doen naar een relevant onderdeel van de opleiding, dit kan iets zijn wat de student is tegen gekomen tijdens de lessen, maar mag ook iets zijn wat de student interesseert. De totale duur van het onderzoek bedraagt 80 uur, wat gelijk staat aan 2 studiepunten.

Ik heb voor dit specifieke onderzoek gekozen omdat het onderwerp me al enkele jaren meer dan boeit. Alleen had ik nooit door hoe precies de verschillende onderdelen, die nodig zijn om 3D geluid uit een computer te halen, onderling verband hielden. Dit onderzoek heeft me een veel beter beeld gegeven van het hoe, wat en waarom van 3D audio.

Verder wil ik de volgende personen bedanken die op hun eigen manier hebben gezorgd dat dit onderzoek tot stand is gekomen:

- Mijn moeder, voor de plek, de tijd en de rust die ik nodig had voor het onderzoeken en schrijven
- Mijn stereotoren voor het onderbreken van de rust en het testen van de theorie in de praktijk.
- Paul Go voor het begeleiden en het wijzen op het feit dat er mensen bedankt moeten worden.
- René, Bart en Michiel voor het 'peerreviewen'.

Samenvatting

Dit onderzoek is gericht op de vraag of de akoestiek van de ene ruimte over te brengen is in een willekeurige andere ruimte. Om het onderzoek in te perken worden verschillende factoren weggelaten. Zo wordt niet bekeken wat de menselijke psychie voor invloed heeft op de ervaring. Andere zintuigelijke informatie dan die van het oor worden dus achterwege gelaten. Dit onderzoek is dus puur en alleen gericht op het menselijk oor. Het is dan van belang om te begrijpen hoe de ontvanger het geluid hoort. Daarom is het van belang een inzicht te hebben in de werking van het menselijk oor. Het buitenoor omvat de oorschelp en de gehoorgang. De gehoorgang functioneert als een resonator die frequenties tussen de 1 en de 4 kHz versterkt. Het middenoor bevindt zich tussen het trommelvlies aan de buitenkant en het ovale venster meer hoofd inwaarts. Het binnenoor is voor de waarneming van spraak en muziek het belangrijkste deel van het perifere gehoor. Omdat het menselijk hoofd ruwweg gezien een bol is wordt er vaak gebruik gemaakt van een bolcoördinaten stelsel. Hier zijn de standaard coördinaten *azimut*(hoek met horizon), *elevation*(verhoging) en *range*(afstand). Er zijn meerdere manieren om deze coördinaten te definiëren. Het verticaalpolair systeem wordt het meest gebruikt. In dit systeem wordt eerst azimut θ berekend als een hoek tussen het verticale vlak en een verticaal vlak waarin de bron en de Z-as zich begeven. Daarna wordt de hoogte ϕ (elevation) gemeten als de hoek tot het horizontale vlak. Om het geluid te horen zijn verschillende systemen bedacht die in staat zijn het geluid in de ruimte te plaatsen. Het eerste systeem wat geluid in de ruimte kon plaatsen was Stereo. Het basis idee is simpel, om een geluid uit de linkerbox te krijgen wordt het geluid naar het linkerkanaal gestuurd, om een geluid uit de rechterbox te krijgen wordt het geluid naar het rechterkanaal gestuurd. Een andere manier om geluid in de ruimte te plaatsen is het gebruik maken van meerdere boxen. Een box voor elk mogelijke hoek is natuurlijk het meest ideale in dit geval, al is dit veel te duur en zal dit in de nabije toekomst geen oplossing zijn. In de film industrie wordt wel van dit verschijnsel gebruik gemaakt, met bijvoorbeeld Dolby Pro Logic of Dolby Digital wordt het plaatsen in de ruimte mogelijk gemaakt door het over een bepaalde box te laten klinken. Het is al geruime tijd bekend dat meerdere kanalen niet nodig zijn om voldoende 3D geluid te creëren, 2 kanalen zijn voldoende, immers de mens heeft maar 2 oren. Binaural is het nabootsen van geluidsdruk in de afzonderlijke oren zoals het zou zijn als de luisteraar er live bij aanwezig zou zijn. Om de geluidsdruk die een willekeurige bron $x(t)$ produceert op het trommelvlies is alleen de impuls reactie $h(t)$ van de bron tot het trommelvlies nodig. Dit wordt Head Related Impulse Response (HRIR) genoemd. De Fourier transformatie $H(f)$ die hier bij hoort wordt het Head Related Transform Function (HRTF) genoemd. De HRTF vangt alles af wat nodig voor het vinden van een geluidsbron. Wanneer eenmaal de HRTF van zowel het linker- als het rechteroor bekend zijn, kan er een Binaural signaal gemaakt worden vanuit een monorale bron. Naast deze losstaande systemen is het goed mogelijk ruimtelijk geluid te creëren op een personal computer. De computer hoeft slechts de juiste HRTF's te berekenen en door middel van transaural stereo door de sturen naar de boxen. Dit wordt tegenwoordig vaak toegepast in computerspelen.

Inhoudsopgave

1	INLEIDING	5
2	HET MENSELIJK OOR	6
2.1	OVERZICHT	6
2.2	SAMENVATTING	7
3	3D GELUID	8
3.1	PHYCHO AKOESTIEK	8
3.1.1	Coördinaten systeem	8
3.1.2	Azimut	9
3.1.3	Elevation	10
3.1.4	Range	11
3.1.5	Echo's	11
3.2	SYSTEMEN	12
3.2.1	Stereo	12
3.2.2	Multi-kanaal	13
3.2.3	Binaural	13
3.3	HRTF	14
3.3.1	HRIR : Horizontaal Vlak	15
3.3.2	HRIR : Verticaal Vlak	16
3.3.3	HRTF : Horizontaal Vlak	16
3.3.4	HRTF : Verticaal Vlak	17
3.4	SYSTEMEN GEBASEERD OP HRTF	18
3.4.1	Convolvotron	18
3.4.2	Koptelefoon of Luidsprekers ?	19
3.4.3	Head Tracking	20
3.4.4	Gemeten of Model?	20
3.5	COMPUTER SYSTEMEN GEBASEERD OP HRTF	21
4	APPENDIX A – LITERATUURLIJST	23

1 Inleiding

Dit onderzoek is verricht voor de opleiding MediaTechnologie van de Hogeschool van Utrecht te Amersfoort. Gesteld is dat elke student een literair onderzoek moet verrichten binnen het vakgebied van de opleiding. Met onderzoek wordt dan bedoeld: het verzamelen, bewerken en analyseren van gegevens om meer te weten te komen. Onderzoek is het combineren van gegevens afkomstig uit verschillende bronnen om tot nieuwe inzichten te komen.

Binnen dit kader is er voor dit specifieke onderzoek gekozen om het volgende te onderzoeken:

“Is de computer in staat om de phycho akoestische ervaring van ruimte A op te wekken in ruimte B?”

Met ander woorden; “Kan ik thuis(B) in mijn eigen stoel de ervaring opwekken alsof ik bij een concert aanwezig ben, in een concertgebouw(A)?”

Om de omvang van dit onderzoek te beperken wordt alleen onderzocht of het technisch mogelijk is. Dit geldt alleen voor het gehoor, andere zintuigen die bij een ervaring aanwezig zijn worden binnen dit onderzoek niet behandeld.

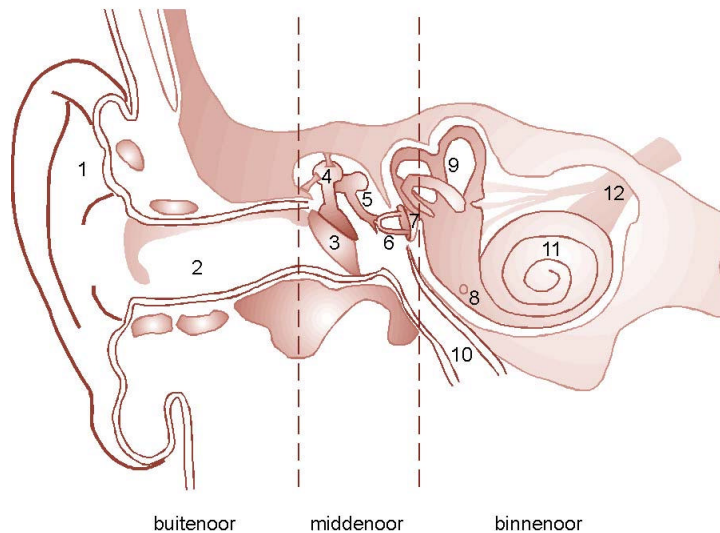
2 Het menselijk oor

Om goed te begrijpen hoe een computer 3D geluid kan produceren is het noodzakelijk het menselijk oor te kennen. In het volgende hoofdstuk wordt uitgelegd hoe het menselijk oor in elkaar zit. Dit hoofdstuk zal slechts een overzicht bieden van het oor, aangezien het gehoor al reeds behandeld is binnen de opleiding. Voor geïnteresseerden staat er achter in een literatuurlijst met links naar websites die geheel over het gehoor gaan.

2.1 Overzicht

Het is gebruikelijk het oor onder te verdelen in drie stukken, het buitenoor, het middenoor en het binnenoor, zoals aangegeven in fig. 2.1. tezamen ook wel genoemd het *perifere gehoor*.

Het *buitenoor* omvat de oorschelp en de gehoorgang, die loopt tot aan het trommelvlies. De oorschelp helpt bij het richting horen. De gehoorgang functioneert als een resonator die frequenties tussen de 1 en de 4 kHz versterkt; dit is precies het frequentiegebied dat voor spraak van belang is.



Figuur 2.1

Het middenoor bevindt zich tussen het trommelvlies aan de buitenkant en het ovale venster meer hoof dinwaarts. De luchtdrukverschillen duwen het trommelvlies beurtelings naar binnen en zuigen het naar buiten. Via drie minuscule botjes, die - naar hun uiterlijk - hamer, aambeel en stijgbeugel heten, wordt de beweging van het trommelvlies doorgegeven aan het ovale venster. Omdat het ovale venster een veel kleiner oppervlak heeft dan het trommelvlies en omdat de botjes als een hefboom werken, worden de zwakke bewegingen van het trommelvlies mechanisch versterkt. De botjes kunnen in hun bewegingen geremd worden door kleine spiertjes die reflexmatig worden aangespannen wanneer het oor getroffen wordt door heel luide, lage trillingen. Na een rockconcert blijven we enige tijd hardhorend omdat deze zelfbeschermingreflex zich niet meteen ontspant. Helaas

komt de reflex te langzaam op gang om ook gehoorbeschadiging als gevolg van explosies te voorkomen.

Het binnenoor is voor de waarneming van spraak en muziek het belangrijkste deel van het perifere gehoor. Hier worden de fysieke geluidstrillingen omgezet in elektrische trillingen voor verdere verwerking in het centrale gehoor (d.w.z. in de hersenen).

2.2 Samenvatting

Dit onderzoek is gericht op de vraag of de akoestiek van de ene ruimte over te brengen is in een willekeurige andere ruimte. Het is dan van belang om te begrijpen hoe de ontvanger het geluid hoort. Daarom is het van belang een inzicht te hebben in de werking van het menselijk oor. Het buitenoor omvat de oorschelp en de gehoorgang. De gehoorgang functioneert als een resonator die frequenties tussen de 1 en de 4 kHz versterkt. Het middenoor bevindt zich tussen het trommelvlies aan de buitenkant en het ovale venster meer hoofd inwaarts. Het binnenoor is voor de waarneming van spraak en muziek het belangrijkste deel van het perifere gehoor.

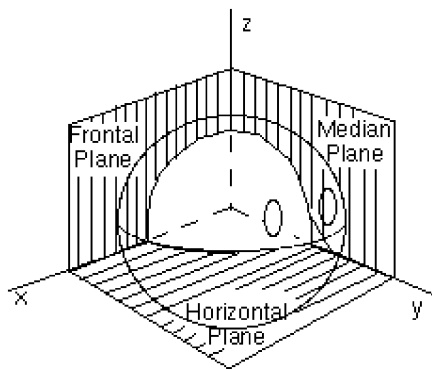
3 3D geluid

Hoe het menselijk oor in elkaar zit is in het vorige hoofdstuk behandeld. In het volgende hoofdstuk wordt behandeld hoe geluid in de ruimte geplaatst wordt.

3.1 Phycho akoestiek

Om te begrijpen hoe geluid in de ruimte geplaatst kan worden, wordt eerst behandeld hoe een mens geluid in de ruimte waarneemt, de phycho akoestiek.

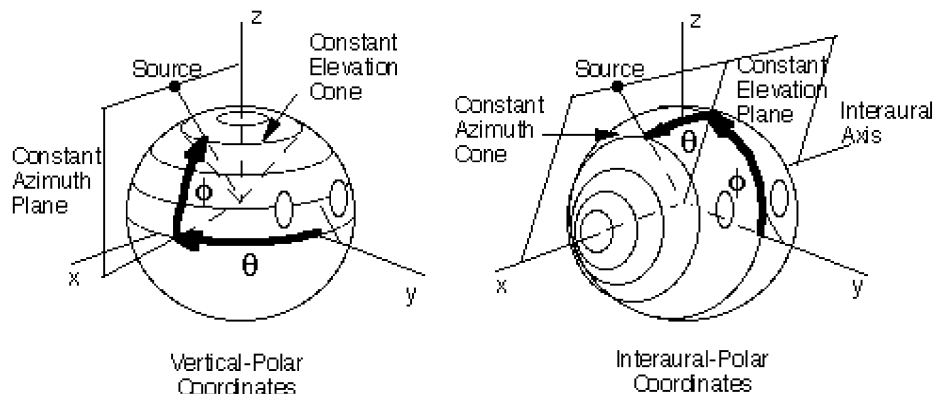
3.1.1 Coördinaten systeem



figuur 3.1

Om de locatie van een geluidsbron te plaatsen ten opzichte van de luisteraar wordt er gebruik gemaakt van een coördinaten systeem. Een van de meest natuurlijke keuzes is het plaatsen van de bron met als oorsprong het menselijk hoofd (figuur 3.1). Hierin gaat de X-as door het rechteroor, de Y-as rechtdoor en de Z-as vertikaal. Dit zorgt voor drie vlakken, het XY vlak, horizontaal vlak (eng. Horizontal Plane) het XZ vlak, het frontale vlak (eng. Frontal Plane) en het YZ vlak, verticale vlak (eng. Median Plane). Het horizontale vlak zorgt voor onderscheiding tussen boven en onder, het frontale vlak voor onderscheiding tussen voor en achter en het verticale vlak voor onderscheiding tussen links en rechts.

Omdat het menselijk hoofd ruwweg gezien een bol is wordt er vaak gebruik gemaakt van een bolcoördinaten stelsel. Hier zijn de standaard coördinaten *azimut* (hoek met horizon), *elevation* (verhoging) en *range* (afstand). Er zijn meerdere manieren om deze coördinaten te definiëren. Het verticaalpolair systeem (figuur 3.2a) wordt het meest gebruikt. In dit systeem wordt eerst azimut θ berekend als een hoek tussen het verticale vlak en een verticaal vlak waarin de bron en de Z-as zich begeven. Daarna wordt de hoogte ϕ (elevation) gemeten als de hoek tot het horizontale vlak.



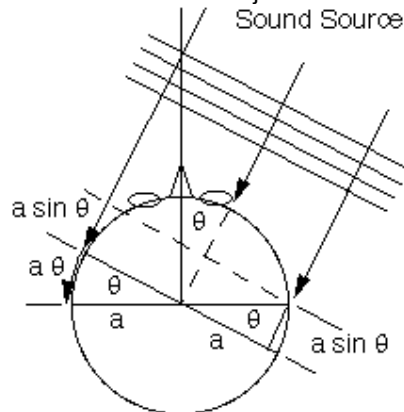
figuur 3.2a en figuur 3.2b

Een belangrijk alternatief is gegeven in figuur 3.2b. In dit model wordt eerst de elevation berekend als de hoek van het horizontale vlak tot een vlak wat door de X-as van de bron gaat (eng. Interaural axis). Het azimuth wordt dan gemeten als de hoek tot de bron Y-coördinaat.

Het verticaal polair systeem is makkelijker om bronnen te beschrijven die vastzitten aan het horizontale vlak, aangezien dan alleen het azimuth als een hoek tussen -180° en $+180^\circ$ beschreven dient te worden. Bij het interaural polair systeem is het azimuth altijd een hoek tussen de -90° en $+90^\circ$. Daarbij dient in dit systeem ook d.m.v. de elevation het verschil tussen voor en achter aan gegeven te worden, in het Verticaal Polair systeem zit dit al in het azimuth.

3.1.2 Azimut

Een van de pioniers in ruimtelijk horen onderzoek van John Strutt. Zo'n 100 jaar geleden ontwikkelde hij zijn zogenoemde Duplex Theorie. Volgens deze theorie zijn er 2 belangrijke tekens voor de hoek met de horizon namelijk Interaural Time Difference (ITD) en Interaural Intensity Difference (IID).



Figuur 3.3

John Strutt had een simpele verklaring voor ITD. Geluid beweegt met een snelheid 'c' van ongeveer 343 m/s. Stel dat een geluidsgolf het menselijk hoofd raakt onder een hoek die bepaald wordt door het azimuth hoek θ . In figuur 3.3 is met dit voorbeeld duidelijk te zien dat het geluid eerder aankomt bij het rechteroor dan bij het linkeroor, omdat het een extra afstand moet afleggen ($a \theta + a \sin \theta$) om bij

het linkeroor te komen. Als dat gedeeld wordt door de snelheid van het geluid komt er een simpele maar wel zeer accurate formule voor het ITD :

$$ITD = \frac{a}{c} (\theta + \sin \theta) \quad , \quad -90^\circ \leq \theta \leq +90^\circ$$

Het ITD is 0 als de bron direct voor de luisteraar staat, en heeft een maximum van $(a/c)(\pi/2+1)$ als de bron aan de linker- of rechterkant van het hoofd staat.

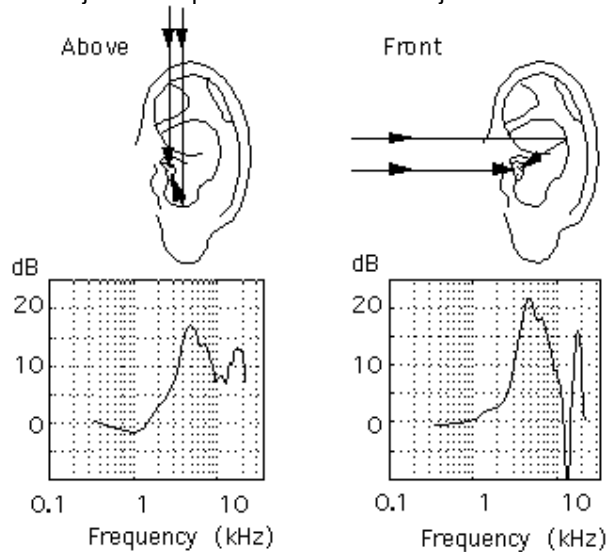
John Strutt observeerde ook dat geluid wordt afgebogen door het hoofd. Hieruit volgde zijn IID.

Het IID is zeer frequentie afhankelijk. Bij lage frequenties, wanneer de golflengte dus relatief lang is ten opzichte van de diameter van het hoofd is, is er nauwelijks een verschil waar te nemen in geluidsdruk tussen de twee oren. Daar tegenover is het bij hoge frequenties mogelijk een verschil waar te nemen tot wel 20 dB. Dit wordt het hoofdschaduw effect genoemd, waarbij het oor wat het verst van de bron afstaat in de schaduw zit van het hoofd.

De Duplex theorie stelt dat de IID en de ITD complementair zijn. Bij lage frequenties (onder de 1.5 kHz) is er weinig IID informatie, maar de ITD verschuift de geluidsgolf een stukje van zijn cyclus, welke snel waarneembaar is. Bij hoge frequenties (boven de 1.5 kHz) zit er dubbelzinnigheid in de ITD omdat er verschillende verschuivingen zijn, maar het IID vangt deze dubbelzinnigheid op. Strutt's Duplex Theorie zegt dat IID en IDT samen de plaatsing van geluid over het complete hoorbare gebied kunnen bepalen.

3.1.3 Elevation

Daar waar de primaire tekens voor azimuth betrekking hebben op 2 oren, hebben de primaire tekens voor elevation(hoogte) betrekking hebben op 1 oor. Ze stemmen af van het feit dat het buitenoor een akoestische antenne is. Door de vorm versterkt het sommige frequenties terwijl het andere verzwakt. Daarbij is de frequentie curve afhankelijk van de richting van het geluid (figuur 3.4)



Figuur 3.4

In figuur 3.4 wordt de frequentie curve getoond voor 2 signalen vanuit verschillende hoeken. In beide gevallen zijn 2 paden getekend, 1 met een direct pad en 1 met een langer pad wat reflecteert op het

buitenoor. Bij lage frequenties komen de 2 signalen op ongeveer dezelfde tijd binnen in het oor. Bij hogere frequenties komt het gereflecteerde signaal later aan dan het directe signaal en volgt er deprivatie van het signaal. Het grootste verlies treedt op als het gereflecteerde signaal op een half van de frequentie binnenkomt. ($f = c / 2d$). Dit verschijnsel wordt een "pinna notch" (oor inkeping) genoemd. In dit voorbeeld ligt het ruwweg op 10 kHz, normale inkepingfrequenties liggen tussen de 6 en 16 kHz.

Het oor is een betere reflector voor geluiden die van voor het hoofd komen dan voor geluiden die van boven het hoofd komen. Dit resulteert in een aanzienlijk grotere inkeping voor geluiden die van voren komen dan voor geluiden die van boven komen. Tevens verschilt de inkeping naarmate de elevation (hoogte) van het geluid veranderd. Hierdoor dient het buitenoor als primaire teken voor elevation.

3.1.4 Range

Het menselijk oor is zeer goed in het bepalen van het azimuth, minder goed in het bepalen van de elevation en zeer slecht in het bepalen van de range (afstand). Zo is het ook met de modellen van deze drie, die voor het azimuth zijn tot in detail uitgewerkt en bewezen, die voor elevation in mindere maten en die voor range in nog mindere maten. Voor de afstand zijn de volgende dingen van belang:

Hardheid

Beweging parallax

Overdreven IID

De primaire basis voor hardheid is dat het afstamt van de bron en naar mate de afstand groter wordt steeds verder afneemt, met andere woorden het geluid wordt zachter naarmate de bron verder van het oor verwijderd is. Hierbij is geen 1 op 1 relatie te ontdekken, een zacht geluid is per definitie niet ver van het oor verwijderd, het kán maar hoeft niet. Om te weten of een bron dichtbij of ver weg staat zijn de eigenschappen van de bron nodig. Een voorbeeld maakt dit duidelijker.

Iedereen weet hoe fluisteren, normaal praten en schreeuwen klinkt, samen met de kennis van de hardheid kan er dan wel bepaald worden hoever de bron van het oor afstaat.

Beweging parallax verwijst naar het feit dat luisteraars het hoofd bewegen. Het verschil in azimuth is afstandsafhankelijk. Voor een bron dichtbij de luisteraar zorgt de beweging van het hoofd in een groot verschil van azimuth, terwijl dit voor een bron ver van de luisteraar af weinig uitmaakt.

Tevens zal de IID toenemen naarmate de afstand van de bron tot de luisteraar afneemt. Deze toename wordt duidelijk bij geluiden die op minder dan 1 meter van het oor ontstaan. Een extreem geval is bijvoorbeeld dat er een insect rond het ene oor vliegt terwijl er iemand in het andere oor aan het fluisteren is. Geluiden die in 1 oor hoorbaar zijn worden als eng ervaren.

3.1.5 Echo's

Alles reflecteert geluid, maar pas als het later dan 30 tot 50 ms dan het originele geluid aankomt wordt het door de hersenen geregistreerd. Dit wordt de echo drempel genoemd. Speciale kamers zijn gemaakt om geluid te absorberen zodat alleen het directe geluid het oor benadert. Deze kamers worden zeker in het begin ervaren als stil, tevens wordt het geluid als saai ervaren. Echo's komen overal voor en toch hebben ze weinig tot geen invloed op het lokaliseren van geluid. Dit komt omdat de mens zich snel aanpast aan een nieuwe omgeving, tevens zijn er systemen in het gehoor wat echo's onderdrukt, al is er nog niet veel bekend over deze systemen. Dat de mens een signaal lokaliseert op het eerste geluid wat de oren binnenkomt staat bekend als de Wet van Eerste Geluidsgolf (Law of the First Wavefront, Blauert). Dit houdt niet in dat er geen rekening wordt

gehouden met de reflecties van het geluid, doch worden deze alleen in acht genomen als de vertraging dermate hoog is dat het voor het oor waarneembaar is. De reflecties hebben weinig tot geen invloed op het lokaliseren van het geluid.

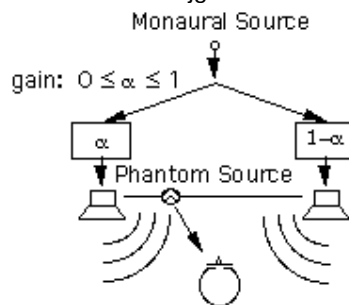
Voorgaande genoodzaakt wel dat de Duplex Theorie aangepast dient te worden. In een normale kamer komt een reflectie een paar milliseconde na het originele geluid aan bij het oor. Voor lage frequentie geluiden (onder de 250 Hz) komen de reflecties aan voordat er een nieuwe sinusbeweging wordt ingezet. Hierdoor zijn lage frequenties nauwelijks te plaatsen en niet nodig voor het lokaliseren van geluid. Van dit effect wordt veel gebruik gemaakt, zo is in veel thuis opstellingen tegenwoordig een sub-woofer aanwezig. Het maakt voor de beleving niet uit waar deze sub-woofer geplaatst wordt, de luisteraar kan toch niet zeggen waarin de kamer dit geluid zich produceert. Dit wordt het Franssen effect genoemd

3.2 Systemen

Om geluid in de ruimte te plaatsen zijn er verschillende systemen beschikbaar, Stereo, Multi-kanaal of Binaural, in de volgende paragraaf worden deze systemen behandeld.

3.2.1 Stereo

Het eerste systeem wat geluid in de ruimte kon plaatsen was Stereo. Het basis idee is simpel, om een geluid uit de linkerbox te krijgen wordt het geluid naar het linkerkanaal gestuurd, om een geluid uit de rechterbox te krijgen wordt het geluid naar het rechterkanaal gestuurd.

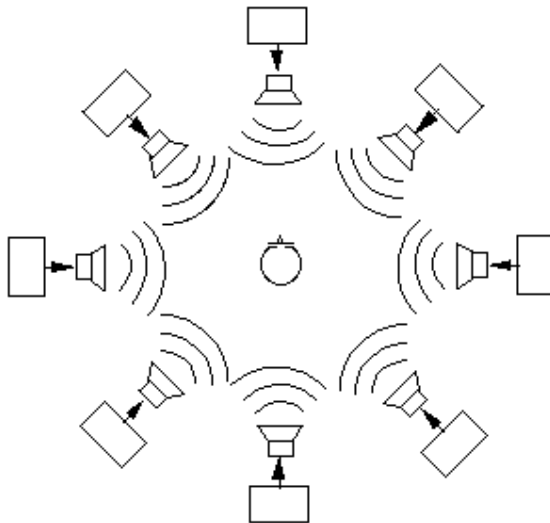


Figuur 3.5: stereo

Als hetzelfde geluid uit beide kanalen even hard en in dezelfde fase wordt gestuurd en als de luisteraar ongeveer in het midden tussen de twee boxen zit lijkt het geluid uit het midden te komen. Door de sterkte van het geluid te verschillen tussen beide boxen kan het geluid zich begeven op een lijn tussen de twee boxen. Het is met deze opstelling niet mogelijk een geluid te doen laten klinken buiten deze lijn. In de paragraaf "Luidspreker of Koptelefoon?" zal een manier behandeld worden die dit wel mogelijk maakt.

Zoals in de paragraaf "Echo's" al naar voren kwam is het ook mogelijk het geluid te positioneren door het geluid uit de ene box een fractie later te laten komen. De luisteraar zal dan alsnog de bron plaatsen bij de box waar het geluid als eerst vandaan komt. Natuurlijk moet dit effect niet overdreven worden, anders zal het overkomen dan een hinderlijke echo.

3.2.2 Multi-kanaal



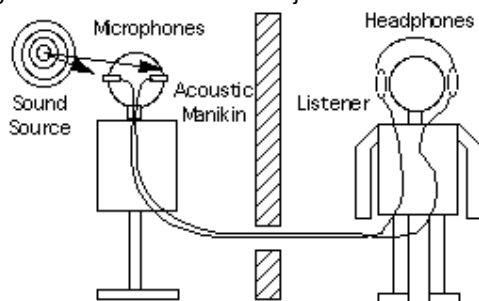
Figuur 3.6: Multi-kanaal

Een andere manier om geluid in de ruimte te plaatsen is het gebruik maken van meerdere boxen. Een box voor elk mogelijke hoek is natuurlijk het meest ideale in dit geval, al is dit veel te duur en zal dit in de nabije toekomst geen oplossing zijn. In de film industrie wordt wel van dit verschijnsel gebruik gemaakt, met bijvoorbeeld Dolby Pro Logic of Dolby Digital wordt het plaatsen in de ruimte mogelijk gemaakt door het over een bepaalde box te laten klinken. Met gebruik van het Franssen effect (zie 3.1.5) is het mogelijk kleine boxen te gebruiken voor de hoog frequente geluiden en slechts 1 grote box voor alle laag frequente geluiden. Hiermee zijn namen als 4.1 of 5.1 te verklaren.

Deze systemen zijn zeer goed in staat geluid in de ruimte te plaatsen maar zijn stukken duurder dan 2 kanaal geluid. Hoewel deze systemen steeds meer gemeengoed zijn geworden zijn ze nog steeds redelijk onhandig in de ontwikkeling en gebruik.

3.2.3 Binaural

Het is al geruime tijd bekend dat meerdere kanalen niet nodig zijn om voldoende 3D geluid te creëren, 2 kanalen zijn voldoende, immers de mens heeft maar 2 oren. De 'truc' is het nabootsen van geluidsdruk in de afzonderlijke oren zoals het zou zijn als de luisteraar er live bij aanwezig zou zijn.



Figuur 3.7: binaural

Een simpel concept om dit te verkrijgen is het plaatsen van 2 microfoons in de oren van een pop (manikin). Deze nemen precies op wat er binnenkomt. Als deze signalen dan via een koptelefoon worden versterkt door 2 speakers lijkt het net alsof de luisteraar aanwezig was tijdens het opnemen van het geluid. Dit principe wordt versterkt als de pop en de luisteraar dezelfde karakteristieken heeft qua omvang en model van het hoofd. In dit geval zullen exact dezelfde ITD en IID gegevens in de pop en de luisteraar aanwezig zijn. Dit principe wordt Binaural (beide oren) genoemd. Binaural opgenomen geluiden zijn zeer goed in de ruimte te plaatsen, de luisteraar ervaart dit als echt 3D geluid.

Ondanks de effectiviteit en kostprijs (zeker ten opzichte van Multi-kanaal geluid) zijn er toch nadelen aan Binaural opnames:

Er moet gebruik gemaakt worden van een koptelefoon, al kan het effect wel gelden als er gebruik wordt gemaakt van "cross-talk cancelled stereo" (zie "Koptelefoon of luidsprekers?").

Niet interactief, het moet van tevoren zijn opgenomen.

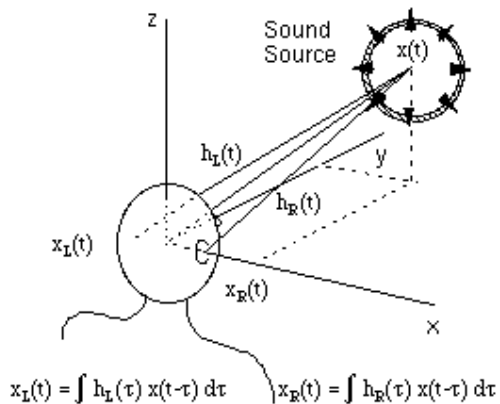
Als de luisteraar zijn/haar hoofd beweegt, beweegt het geluid mee.

Geluiden die direct van voren komen *lijken* veel te hard te klinken.

Omdat de karakteristieken per mens verschillen, zijn de verkregen effecten met Binaural opnames niet betrouwbaar.

Om Binaural opnames te verbeteren, en om het Binaural effect te verkrijgen zonder Binaural opgenomen materiaal is het nodig HRTF (Head Related Transfer Functions) te begrijpen.

3.3 HRTF



figuur 3.8

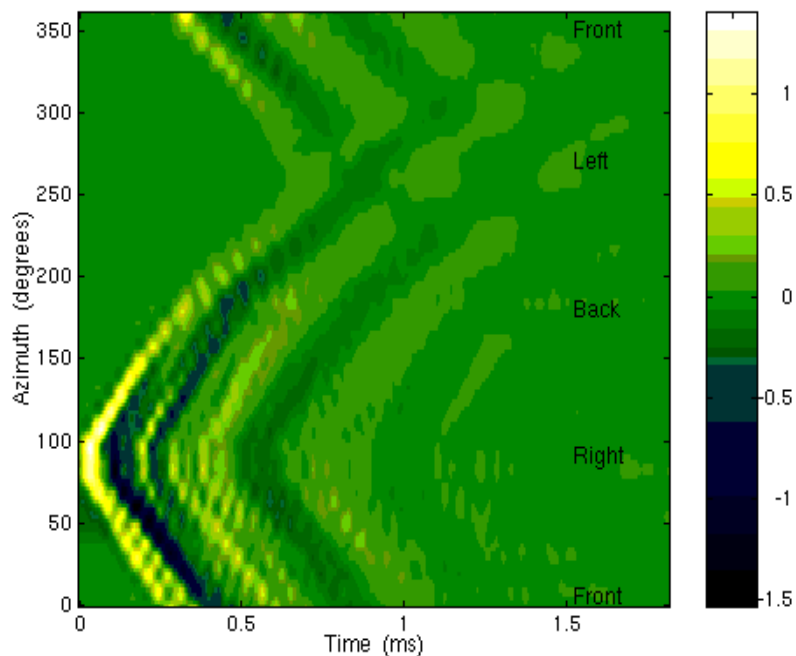
Om de geluidsdruk die een willekeurige bron $x(t)$ produceert op het trommelvlies is alleen de impuls reactie $h(t)$ van de bron tot het trommelvlies nodig. Dit wordt Head Related Impulse Response (HRIR) genoemd. De Fourier transformatie $H(f)$ die hier bij hoort wordt het Head Related Transform Function (HRTF) genoemd. De HRTF vangt alles af wat nodig voor het vinden van een geluidsbron. Wanneer eenmaal de HRTF van zowel het linker- als het rechteroor bekend zijn, kan er een Binaural signaal gemaakt worden vanuit een monorale bron.

De HRTF is een bijzonder gecompliceerde functie met 4 variabelen, 3 ruimte coördinaten en 1 frequentie. Vaak bestaat de HRTF slecht uit een functie van het azimuth, elevation en frequentie.

Er zijn verschillende onderzoeken gedaan op een akoestische manikin die bekend staat als KEMAR (Knowles Electronics Manikin for Auditory Research). Deze KEMAR heeft dezelfde eigenschappen als een doorsnee mens. Om een idee te krijgen hoe KEMAR reageert op verschillende azimuth en

elevation waarde wordt er in de volgende paragrafen een grafische representatie gegeven van de HRIR en HRTF.

3.3.1 HRIR : Horizontaal Vlak

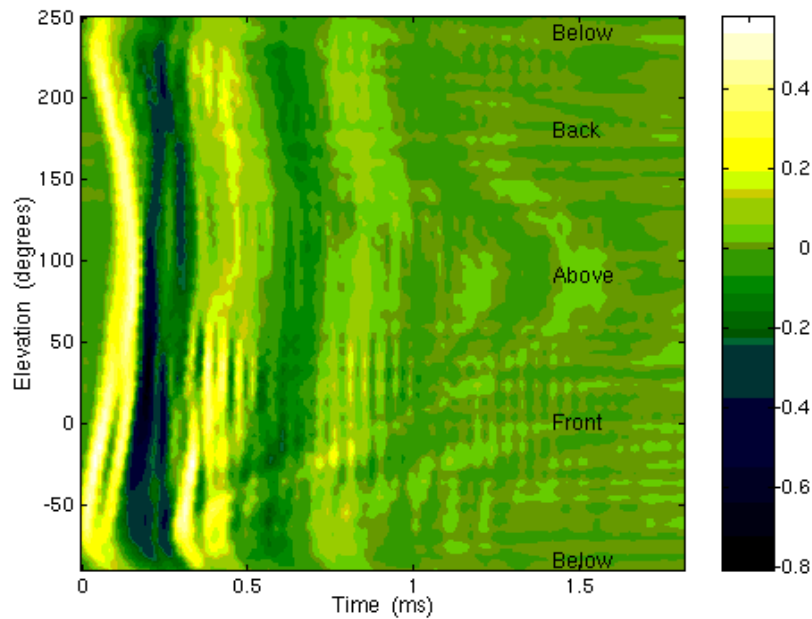


figuur 3.9

Figuur 3.9 geeft een beeld van KEMAR's gemeten HRIR, dit figuur geeft de karakteristieken van het rechteroor. De sterkte van het binnengekomen signaal wordt weergegeven door de helderheid. Zo komt het geluid het hardste binnen als het vanaf de rechterkant komt (azimut = 90°). Zo is het, het zwakste als het vanaf de linkerkant komt (azimut = 270°). Tevens valt op dat het geluid sinusvormig binnenkomt, zoals Strutt al voorspelde met zijn ITD formule. Het verschil tussen een bron aan de rechterkant (minimale vertraging) en een bron aan de linkerkant (maximale vertraging) bedraagt ongeveer 0.7ms, precies datgene wat de theorie verwachtte.

Uit de grafiek blijkt wel dat er slechts weinig verschil waarneembaar is tussen een bron voor en een bron achter het oor. Mensen kunnen dus slecht het verschil tussen voor en achter horen.

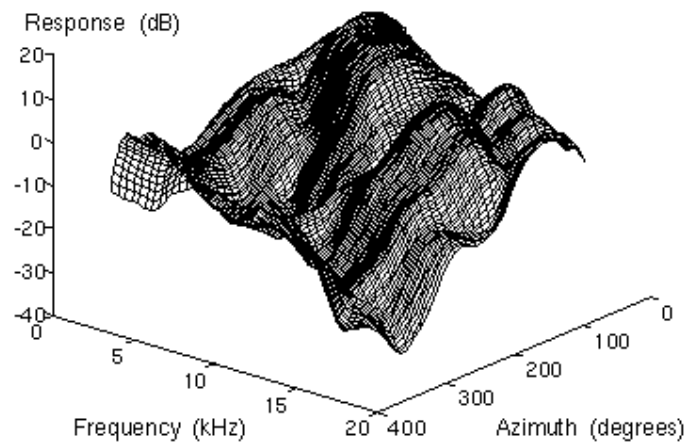
3.3.2 HRIR : Verticaal Vlak



figuur 3.10

Als het geluid verticaal om de luisteraar draait (Boven, voor, onder achter) is het verschil stukken kleiner, zie figuur 3.10.

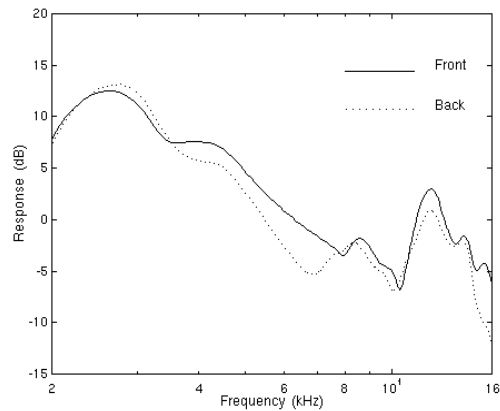
3.3.3 HRTF : Horizontaal Vlak



figuur 3.11

Deze plot geeft de frequentie reactie van KEMAR's rechteroor als de bron zich door het horizontale vlak(links,achter,rechts,voor) begeeft. Hoewel deze plot er zeer gebult uitziet, bestaat er voor elke

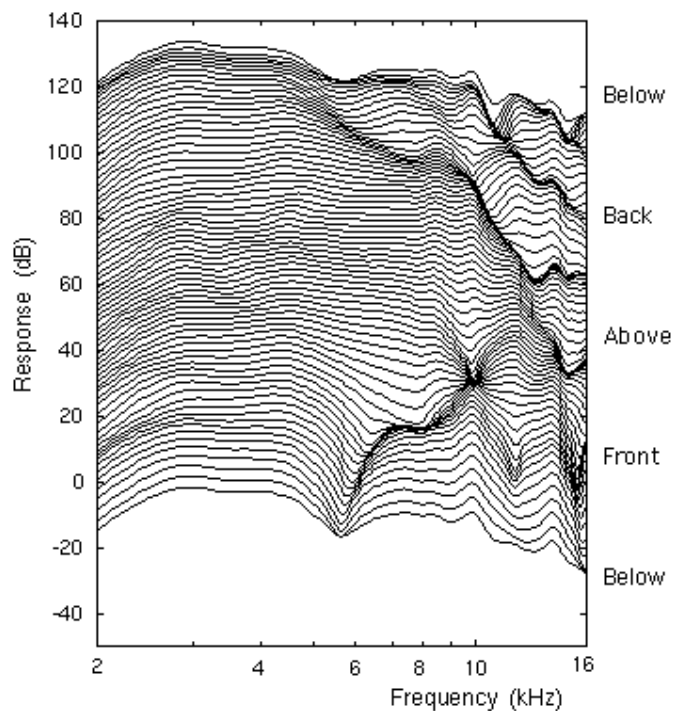
frequentie een sinusfunctie met de Azimut. Zoals verwacht is de reactie het grootst als de hoek 90° is (direct in het oor) en het kleinst als de hoek 270° is (bron vanaf links).



figuur 3.12

Wederom lijken voor en achter (0° en 180°) erg veel op elkaar. Figuur 3.12 geeft het verschil tussen voor en achter weer, het blijkt dat geluid wat van voren komt iets harder binnenkomt (4dB), tevens is de "pinna notch" (oor inkeping) duidelijk zichtbaar rond de 10kHz.

3.3.4 HRTF : Verticaal Vlak



figuur 3.13

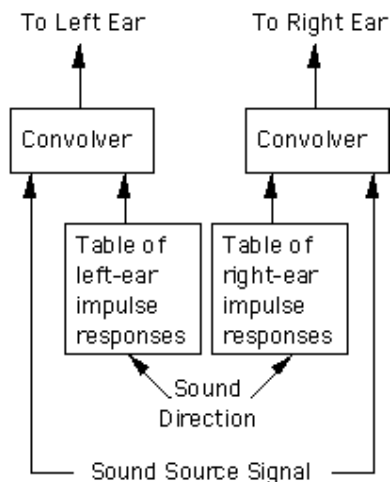
Deze serie curve's geeft de frequentie reacties van KEMAR aan wanneer er een bron beweegt in het verticale vlak. Merk op dat de "pinna notch" (oor inkeping) sterk veranderd met de hoogte mee. Het beweegt mee van 6kHz bij lage hoogte tot een goeie 10 kHz bij bronnen die zich nagenoeg boven het hoofd bevinden. Wanneer de bron zich direct boven het oor bevindt is de "pinna notch" nagenoeg verdwenen en is de gehele curve nagenoeg vlak. De "pinna notch" komt terug zodra de bron zich weer verder beweegt over het verticale vlak. Dit verschijnsel kan perfect gebruikt worden om geluid boven of onder de luisteraar te positioneren. Helaas verschillen deze curve's sterk van mens tot mens, van oor tot oor. Bij grotere oren is de verschuiving minder heftig.

3.4 Systemen gebaseerd op HRTF

Systemen gebaseerd op HRTF's zijn in staat elevation, range en azimuth effecten te produceren. Dat betekent dat ze in staat zijn de illusie te wekken dat een geluid van waar dan ook in het 3D vlak (links, rechts, voor, achter, boven, onder, dichtbij of ver weg) komt.

Als eerste volgt een beschrijving van de Convolvotron, een bekend en effectief systeem. Daarna volgt de vraag of een koptelefoon nodig is om kwalitatief 3D geluid waar te nemen. Wat dient er gedaan te worden aan de verschillen die bestaan tussen de oren van verschillende mensen? De antwoorden op deze vragen worden gegeven in de volgende paragrafen.

3.4.1 Convolvotron



figuur 3.14

De Convolvotron werd gemaakt voor de NASA en is ontwikkeld door Crystal River Engineering. Deze geeft een relatief simpele manier om HRTF's te gebruiken voor 3D geluid. In essentie bestaat het uit 2 "convolution engines" (Winding motoren) elk kunnen dezelfde audio invoer stroom een winding geven met een bepaald aantal HRIR's, die gehaald worden uit een tabel met gemeten resultaten. De uitkomst gaat door versterkers naar de koptelefoon, gedragen door de luisteraar. Als de HRIR's van de luisteraar dichtbij genoeg de HRIR's gebruikt door de winding motoren liggen zal het voor de luisteraar overkomen of het geluid werkelijk 3D is.

Dit basis idee kan op verschillende manieren worden uitgebreid:

Meerdere bronnen kunnen samengenomen worden, Head Motion kan gebruikt worden om de fysieke plaats van het geluid overeen te laten komen met de plaats binnen de Convolvotron. De tabellen van de Convolvotron kunnen worden aangepast aan de HRIR's van een bepaalde luisteraar. Echo's en galm kunnen worden toegevoegd door een Kamer Model toe te voegen. Hoe dan ook, de basis van dit systeem blijft het creëren van juiste linker- en rechteroor signalen, door monorale invoer de winden met vooraf bepaalde HRIR tabellen.

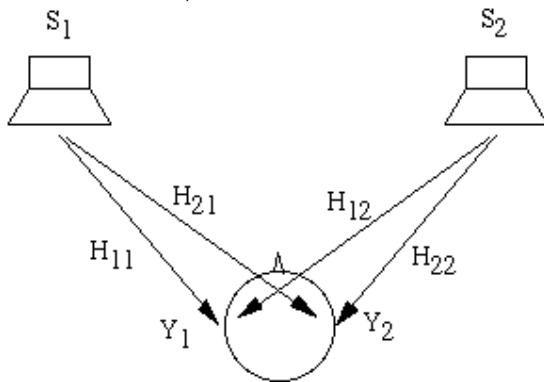
3.4.2 Koptelefoon of Luidsprekers ?

Koptelefoons zorgen er voor dat er geen probleem bestaat tussen het scheiden van wat er naar het ene oor toe moet en wat er naar het andere oor toe moet, maar helaas zijn koptelefoons niet geheel zonder nadelen.

Veel mensen vinden het dragen van koptelefoons nogal onprettig. De akoestisch beste koptelefoons zijn zeer oncomfortabel om voor een langere periode op te hebben. Daarbij zijn deze gesloten en blokkeren dus omgevingsgeluiden wat de luisteraar in een isolement brengt.

Geluiden via de koptelefoon klinken al snel te dichtbij. Uiteraard is de fysieke bron van het geluid ook zeer dichtbij het oor. Om dit verschijnsel op te heffen zijn er compensaties nodig.

Luidsprekers lossen deze problemen op, maar nemen natuurlijk weer andere problemen met zich mee. Hoe kan een luidspreker Binaural geluid geven? Een oplossing hiervoor heet "Cross-talk-cancelld Stereo", ook wel transaural stereo.



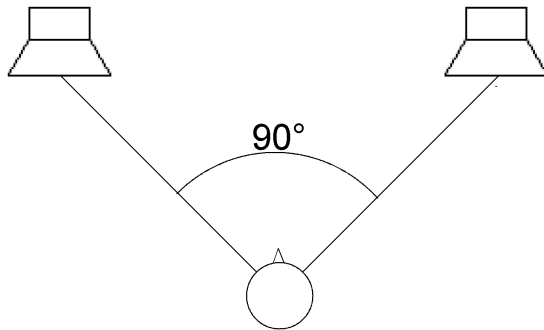
figuur 3.15

Het basis idee is zeer simpel, In figuur 3.15 zijn S_1 en S_2 de 2 signalen die de luidsprekers voeden. Y_1 is het signaal wat bij het linkeroor aankomt. Bij Y_1 komen H_{11} en H_{12} aan. $Y_1 = H_{11} S_1 + H_{12} S_2$, met H_{11} de HRTF tussen de linkerluidspreker en het linkeroor, en H_{12} de HRTF tussen de rechterluidspreker en het linkeroor. Dit geldt op voor het rechteroor, $Y_2 = H_{22} S_2 + H_{21} S_1$. Mits er een koptelefoon gebruikt mag worden zijn de signalen Y_1 en Y_2 bekend. Het probleem schuilt in het vinden van de signalen S_1 en S_2 om tot Y_1 en Y_2 te komen. Vanuit rekenkundig opzicht is dit te vertalen in de vergelijking:

$$\begin{bmatrix} Y_1 \\ Y_2 \end{bmatrix} = \begin{bmatrix} H_{11} & H_{12} \\ H_{21} & H_{22} \end{bmatrix} \begin{bmatrix} S_1 \\ S_2 \end{bmatrix} \quad \begin{bmatrix} S_1 \\ S_2 \end{bmatrix} = \begin{bmatrix} H_{11} & H_{12} \\ H_{21} & H_{22} \end{bmatrix}^{-1} \begin{bmatrix} Y_1 \\ Y_2 \end{bmatrix}$$

In praktijk is het inverteren van een matrix niet voldoende.

Deze vergelijking gaat alleen op dat de luisteraar binnen de zogenaamde “sweet spot” valt, dit is een gebied precies in het midden van de boxen, waarbij de afstand een functie is van de hoek. De hoek waaronder de boxen het hoofd moeten kruisen is 45° (zie figuur 3.16).



figuur 3.16

Mits het goed uitgevoerd wordt is het mogelijk een zeer realistisch signaal te maken, zowel elevation als azimuth effecten kunnen gemaakt worden, de bron kan zich virtueel buiten het veld van de luidsprekers begeven, maar nog steeds zijn Binaural signalen nodig om mee te beginnen.

3.4.3 Head Tracking

Wanneer een koptelefoon gebruikt beweegt het geluid met de luisteraar mee. Dit is natuurlijk niet de bedoeling. Het kan zelfs leiden tot het gehele verlies van de opgewekte illusies (het menselijk brein heeft door dat het niet klopt en compenseert).

Een oplossing voor dit probleem is een apparaat wat een “Head Tracker” wordt genoemd. Het apparaat meet de locatie en oriëntatie van het hoofd en berekent de relatieve positie van elke bron, het apparaat past dus de HRIR's aan. Naast problemen zoals de hoge kosten en de haalbaarheid komen bij dit systeem extra problemen boven:

Vertraging. Wanneer de vertraging tussen de beweging van het hoofd en het bijbehorend verschil in geluid groter is dan 50ms is de ervaring weg, het kan zelfs leiden tot misselijkheid van de luisteraar.

Overgang. Een overschakeling tussen de ene en de andere HRIR kan een hoorbaar verschil met zich meebrengen. Dit is natuurlijk zeer ongewenst.

3.4.4 Gemeten of Model?

Omdat HRTF's zo complex zijn werken meeste systemen met gemeten HRTF's, die van experimenten zoals KEMAR afkomstig zijn, in plaats van Persoonlijk gemeten HRTF's. Op dit moment zijn er 4 verschillende HRTF standaarden.

Standaard HRTF: deze HRTF's zijn afkomstig van metingen over een klein deel van de menselijke bevolking, maar zijn gering in aantal, het behaalde effect is zeer matig te noemen. De grotere standaardisatie organisaties hebben deze HRTF's nog niet erkend dan standaard, maar binnen de computer industrie worden deze veel gebruikt door Microsoft (DirectX), Intel, Creative.

Een set van standaard HRTF's: deze HRTF's zijn afkomstig van meting over een deel van de beoogde gebruikers. Alhoewel deze HRTF's een stuk beter resultaat geven worden ze nagenoeg niet gebruikt. Er is op dit moment nog geen standaard voor deze HRTF's.

Individuele HRTF's: gemeten HRTF's per persoon. Deze geven uitstekende resultaten maar zijn duur om te maken en kosten erg veel tijd.

Model HRTF: een basis model HRTF die aangepast kan worden per gebruiker. De resultaten zijn stukken beter dan de standaard HRTF, zijn makkelijker te verkrijgen dan de Individuele HRTF en zeer goedkoop. Sensaura is een van de bedrijven die een Model HRTF gebruikt voor zijn toepassingen (Virtual Ear, MacroFX, InfinityFX).

3.5 Computer Systemen gebaseerd op HRTF

Naast deze losstaande systemen is het goed mogelijk ruimtelijk geluid te creëren op een personal computer. De computer hoeft slechts de juiste HRTF's te berekenen en door middel van transaural stereo door de sturen naar de boxen. Dit wordt tegenwoordig vaak toegepast in computerspelen. Voor meer professionele doeleinde zijn de meeste van de huidige HRTF modellen voor de computer te onnauwkeurig. Bedrijven zoals Microsoft, Sun, Sensaura en het vroegere Aureal (nu Creative) hebben allen HRTF systemen bedacht die in meer of mindere maten werken. Hoe deze systemen werken is na te lezen op de websites van de bedrijven. Vooral Sensaura heeft een zeer uitgebreide site over 3D geluid.

3.6 Samenvatting

Omdat het menselijk hoofd ruwweg gezien een bol is wordt er vaak gebruik gemaakt van een bolcoördinaten stelsel. Hier zijn de standaard coördinaten *azimut(hoek met horizon)*, *elevation(verhoging)* en *range(afstand)*. Er zijn meerdere manieren om deze coördinaten te definiëren. Het verticaalpolair systeem wordt het meest gebruikt. In dit systeem wordt eerst azimut θ berekend als een hoek tussen het verticale vlak en een verticaal vlak waarin de bron en de Z-as zich begeven. Daarna wordt de hoogte ϕ (elevation) gemeten als de hoek tot het horizontale vlak. Om het geluid te horen zijn verschillende systemen bedacht die in staat zijn het geluid in de ruimte te plaatsen. Het eerste systeem wat geluid in de ruimte kon plaatsen was Stereo. Het basis idee is simpel, om een geluid uit de linkerbox te krijgen wordt het geluid naar het linkerkanaal gestuurd, om een geluid uit de rechterbox te krijgen wordt het geluid naar het rechterkanaal gestuurd. Een andere manier om geluid in de ruimte te plaatsen is het gebruik maken van meerdere boxen. Een box voor elk mogelijke hoek is natuurlijk het meest ideale in dit geval, al is dit veel te duur en zal dit in de nabije toekomst geen oplossing zijn. In de film industrie wordt wel van dit verschijnsel gebruik gemaakt, met bijvoorbeeld Dolby Pro Logic of Dolby Digital wordt het plaatsen in de ruimte mogelijk gemaakt door het over een bepaalde box te laten klinken. Het is al geruime tijd bekend dat meerdere kanalen niet nodig zijn om voldoende 3D geluid te creëren, 2 kanalen zijn voldoende, immers de mens heeft maar 2 oren. Binaural is het nabootsen van geluidsdruk in de afzonderlijke oren zoals het zou zijn als de luisteraar er live bij aanwezig zou zijn.

Om de geluidsdruk die een willekeurige bron $x(t)$ produceert op het trommelvlies is alleen de impuls reactie $h(t)$ van de bron tot het trommelvlies nodig. Dit wordt Head Related Impulse Response (HRIR) genoemd. De Fourier transformatie $H(f)$ die hier bij hoort wordt het Head Related Transform Function (HRTF) genoemd. De HRTF vangt alles af wat nodig voor het vinden van een geluidsbron. Wanneer eenmaal de HRTF van zowel het linker- als het rechteroor bekend zijn, kan er een Binaural signaal gemaakt worden vanuit een monorale bron.

Naast deze losstaande systemen is het goed mogelijk ruimtelijk geluid te creëren op een personal computer. De computer hoeft slechts de juiste HRTF's te berekenen en door middel van transaural stereo door de sturen naar de boxen. Dit wordt tegenwoordig vaak toegepast in computerspelen.

4 Conclusie

“Is de computer in staat om de psycho akoestische ervaring van ruimte A op te wekken in ruimte B?”

Om het geluid te horen zijn verschillende systemen bedacht die in staat zijn het geluid in de ruimte te plaatsen. Het is al geruime tijd bekend dat meerdere kanalen niet nodig zijn om voldoende 3D geluid te creëren, 2 kanalen zijn voldoende, immers de mens heeft maar 2 oren. Binaural is het nabootsen van geluidsdruk in de afzonderlijke oren zoals het zou zijn als de luisteraar er live bij aanwezig zou zijn.

Om de geluidsdruk die een willekeurige bron $x(t)$ produceert op het trommelvlies is alleen de impuls reactie $h(t)$ van de bron tot het trommelvlies nodig. Dit wordt Head Related Impulse Response (HRIR) genoemd. De Fourier transformatie $H(f)$ die hier bij hoort wordt het Head Related Transform Function (HRTF) genoemd. De HRTF vangt alles af wat nodig voor het vinden van een geluidsbron. Wanneer eenmaal de HRTF van zowel het linker- als het rechteroor bekend zijn, kan er een Binaural signaal gemaakt worden vanuit een monorale bron.

De computer hoeft slechts de juiste HRTF's te berekenen en door middel van transaural stereo door de sturen naar de boxen. Samen met de Duplex theorie is het zeer goed mogelijk de acoustiek en plaats, die een luisteraar zou innemen in ruimte A, over te brengen in ruimte B.

Appendix A – Literatuurlijst

- Begault, D. (1994). *3-D Sound for Virtual Reality and Multimedia* (Academic Press, Boston, MA, 1994).
- Berg, R. E. and D. G. Stork (1982). *The Physics of Sound* (Prentice-Hall, Englewood Cliffs, NJ). Blauert, J. (1983). *Spatial Hearing* (MIT Press, Cambridge, MA
- Bracewell, R. N. (1986). *The Fourier Transform and its Applications, 2nd ed.* (McGraw-Hill, New York).
- Duda, R. O. (1993). "Modeling head related transfer functions," in *Proc. Twenty-Seventh Annual Asilomar Conference on Signals, Systems and Computers* (Asilomar, CA).
- Morse, P. M. and K. U. Ingard (1968). *Theoretical Acoustics* (Princeton University Press, Princeton, NJ).
- Wright, M. (1996). "3-D Audio: Above and behind you or moving low left to right"
- <http://xenia.media.mit.edu/~mkc/micArray/node10.html> (Binaural references)
- <http://xenia.media.mit.edu/~kdm/hrtf.html> (KEMAR)
- <http://www.sensaura.com> (Informatie over HRTF systemen voor de pc, tip!)
- <http://www.3dss.com/> (algemene informatie over 3D audio op de pc)
- <http://www.users.dircon.co.uk/~wareing/3daudio.htm> (zeer veel links naar sites over 3D sound)
- <http://www.w3.org/TR/REC-CSS2/aural.html> (Azimuth, Elevation, Range uitleg)