



3D med hörtelefoner

Sedan en lång tid är hemmabio ett "folkhemsmåste". Helt plötsligt är utrustning för att skapa artificiella ljudmiljöer i var mans hylla. Men signalbehandling och forskning på spatialisering av ljud har pågått i många decennier redan. Förvånansvärt mycket finns att läsa och studera på webben om allehanda projekt. Jag har valt att studera en ytterst liten del av detta vidsträckta område inom ljudeffektindustrin: 3D med hörlurar.

Det finns många anledningar att inte använda just hörlurar för att kliva in i en artificiell ljudmiljö. En ofta nämnd anledning är att det är tröttsamt att bära. Samtidigt finns det fördelar framför högtalarsystem som kan bli kostsamma och som dessutom ofta är fysiskt begränsade med olika rums problematiska ljudreflektioner m.m.

Mycket av det som följer nedan gäller för spatialisering i stort. Däremot är bokens implementeringsexempel inriktat för hörlurar.

Människans hörapparat använder sig av olika typer av information för att rumsbestämma en ljudkälla. Förvisso kan ögat hjälpa oss att lokalisera ljud men det är andra fysiska förutsättningar som främst avgör riktning. Undersökningar har visat att två olika faktorer, s.k. nycklar (eng. cues), är extra viktiga för uppskattandet av riktning: tidsförskjutning (ITD, Interaural Time Difference) och intensitetsskillnad (IID, Interaural Intensity Difference) mellan öronen.

- ITD utgör alltså tidsskillnaden för när ljudet anländer till höger resp. vänster öra. *Se fig. 1.* Den ger oss i första hand en horisontell lokaliserings-nyckel. Den beror på avståndet till källan och dess frekvens.
- IID beror mycket på huvudformen och dess skuggning. Vid låga frekvenser är IID-påverkan inte stor, p.g.a. att dessa ljud bryts i huvudet, dvs. går rakt igenom. Den är alltså främst markant vid höga frekvenser, över 1500 Hz där skillnader på ung. 20dB kan mätas.

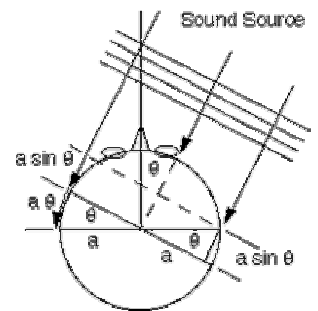


Fig. 1
ITD

Trots dessa två nycklars betydelse för hörseln kan de bara delvis förklara ljudriktning. Faktum är att de ibland ger ganska knapphändig information. Värden för ITD och IID skiljer sig nämligen inte alls (dvs. vi märker ingen skillnad) om ljudkällan rör sig i en cirkelrörelse vid sidan av huvudet framför, bakom, ovan och nedanför oss. Detta ger upphov till en missförstådd ljudriktning. *Se fig 2.*

Ytterligare några parametrar måste till för att kunna modellera en virtuell ljudkälla med någorlunda precision.

Mest känd är de s.k. Head Related Transfer Functions (HRTFs): Bakom denna titel står en annan förkortning HRIR, nämligen impulssvaret för överkroppens olika delar, i tidsplanet. Det innebär en ljudvågs komplexa interaktion, dvs. filtrering med bröstkorget, axlarna, skuldrorna, huvudet och speciellt ytterörona. Dessa ger upphov till signaler med vilka hjärnan hjälps uppskatta positionering.

Den enklaste modelleringen är att använda uppmätta värden på HRIR direkt som koefficienter för FIR-filer. Däremot har det gjorts mycket forskning på olika filterdesigntechniker, i frekvensdomänen för motsvarigheten HRTF. Bl.a. har man försökt ta fram en generell HRTF som inte är individuell, vilket är fallet för flertalet implementationer av HRTF. Överlag vill man hitta

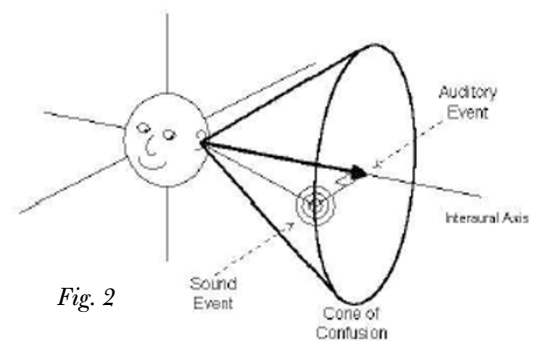


Fig. 2

modelleringsmetoder med så få parametrar som möjligt som dessutom är så generella som möjligt.

Författaren av kursboken DAFX föreslår en uppdelning av 3D-syntes för hörlurslyssning (binaural) med tre huvudkomponenter vars matematiska "huvudberoende överföringsfunktioner" finns beskrivna i boken.:

- Huvudskuggning & ITD
- "Axel- och skuldereko" (bröstkorg)
- Ytteröra-reflektioner.

Den förstnämnda finns implementerad i Matlab i form av en m-fil. Den ger i princip den viktigaste lokaliseringkaraktistikan. Nämligen den 360-gradiga cirkel (azimuth, θ) på ett horisontellt plan på vilken en virtuell ljudkälla vandrar. Denna aspekt är nog den mest generella av de tre. Dvs. den kommer inte att degraderas så mycket beroende på lyssnare.

Axel- och skuldereko ger "reflexer" för signaler med frekvenser från 1 till 3 KHz. Detta eko är ingen signifikant nyckel men påverkar måttligt till höjdlägesbestämning.

Något som absolut avgör höjdbeskrivning är vårt ytteröra (eng. pinna). *Se fig. 3.* Dess komplexa geometri orsakar interferens mellan direkt ljud och reflekterat ljud.

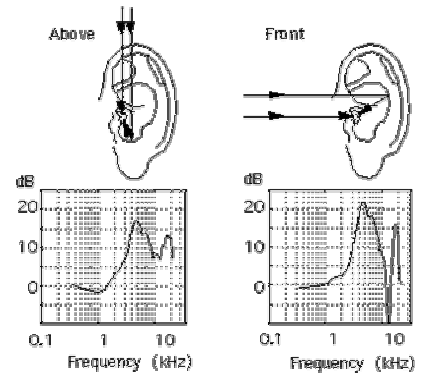


Fig. 3 Pinna=ytterörat

Jag tittat alltså närmare på filen hsfiler.m för att återgå till bokens exempel. Ett script som utifrån en monoljudfil skapar en rumskänsla med hjälp av två kanaler (monaural->binaural). Jag skapade ytterligare två m-script för att kunna testa denna funktion. En, sweeping.m, som har som argument: startläge, slutläge, antal segment och wav-fil. Den ljudfilen spelas sedan upp för varje nytt läge 'hsfilter' räknat ut för höger och vänster kanal på den båge från start- θ till slut- θ .

Den andra, the_sweeping.m, delar istället upp en och samma ljudfil och panorererar den över ett givet fält, återigen från start- θ till slut- θ . En ljudeffekt vanligt förekommande i filmer, där man i diverse actionfilmer får höra kulor vina, bilar väja och fara kors och tvärs över ens tänkta upplevelserum.

Båda dessa m-filer ger en hint om att grunden, hs-filtret, för en 3D-modellering för hörlurar i princip fungerar. Svagheten som i många system är just bak- och framsidmodellering. Lättare tycks vara att fejka bakomvarande ljudkällor än framförvarande dito.

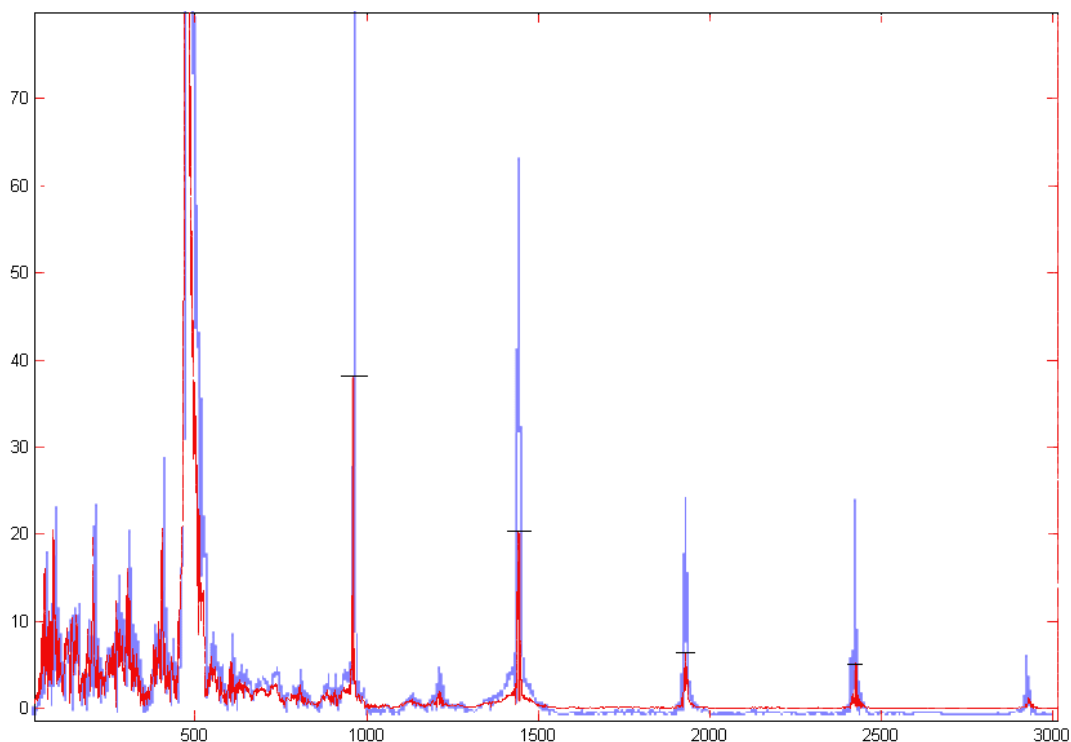


Fig. 4. Spektrum (FFT) över en pianoton (B4) spatialierad rakt ut från höger öra. Ljusblå=höger kanal, Röd (med horisonetlla streck för maxamplitud) = vänsterkanal.

Ett exempel på att huvudmodellen fungerat kan vara fig. 4. Där har en pianoton, H4, 494Hz, lagts rakt ut från högersidan av huvudet. På spektrat över ljudet ser man höger hörtelefon i ljusblått med kraftiga amplituder för grundton och flera övertoner. Medan den vänstra luren (i rött med horisontella streck för maxamplitud) visserligen klingar relativt starkt i grundtonen men sedan tappar mark i övertonerna och klingar av snabbare för övertonerna i högre frekvenser. Huvudet har således sållat bort en del höga frekvenser.

Problem jag märkte med min svepning i planet med samma ljudfil är diskontinuitet i segmenteringen av sampels i den nya ljudfilen. Små clip-ljud för varje skarv skapas och är tydligt framför allt för tonande ljud. Detta åtgärdade jag snöpligt nog, p.g.a. tidspress med en clip-pop-restaurering i programmet CoolEdit. Överlag verkar otonade ljud vara att föredra vid modellering av 3D-miljöer.

Denna modell är tyvärr inte fullständig och skulle behöva utökas med bröstorgs- och ytterörmodellerna för att ta hänsyn till höjdskillnader m.m.

TVå andra faktorer ofta diskuterade i 3D-simulering av ljud är naturligtvis rumseko. Något som markant påverkar riktningssänetecknen.

Den andra är problematiken med att många modeller lider av huvudinstängdhet, i synnerhet de som bara baseras på ITD och IID. Instängdheten består i att människan ofta upplever artificiella modeller som just artificiella p.g.a. att allt ljud känns som det kommer inifrån huvudet. Man hör att det inte handlar om en naturligt avlägsen ljudkälla.

Flera angripssätt finns tillhanda. Däribland en tidsfördröjning (mellan kanalerna) om är uträknad utifrån det faktum att de ljud de båda öronen får till sig är starkt korrelerade. Denna normaliseras sedan. Man kan även avlägsna viss instängdhetskänsla genom att påverka ljudet beroende på hur huvudet rör sig. Dock beräkningsmässigt kostsamt.

Många fler tillvägagångssätt för att skapa 3D-miljöer finns beskrivna i detta kapitel, men relativt kondenserat. För, efter vad jag erfarit, med olika rapporter och projekt, runt om i världen så finns de spaltkilometer inom detta ämne. Intressant och kul!

Referenser:

- Bill Kapralos, Michael R. M. Jenkin, Evangelos Milios :
Auditory Perception and Spatial (3D) Auditory Systems,
www.cs.yorku.ca/techreports/2003/CS-2003-07.pdf
- Kaibo Liu & Syed Hassan :
Matlab Implementation of 3D Synthetic Environment
- Nick Zacharov : *Perceptual Studies on Spatial Sound Reproduction Systems*,
lib.hut.fi/Diss/2000/isbn9512252104/isbn9512252104.pdf
- Zölzer Udo, DAFX