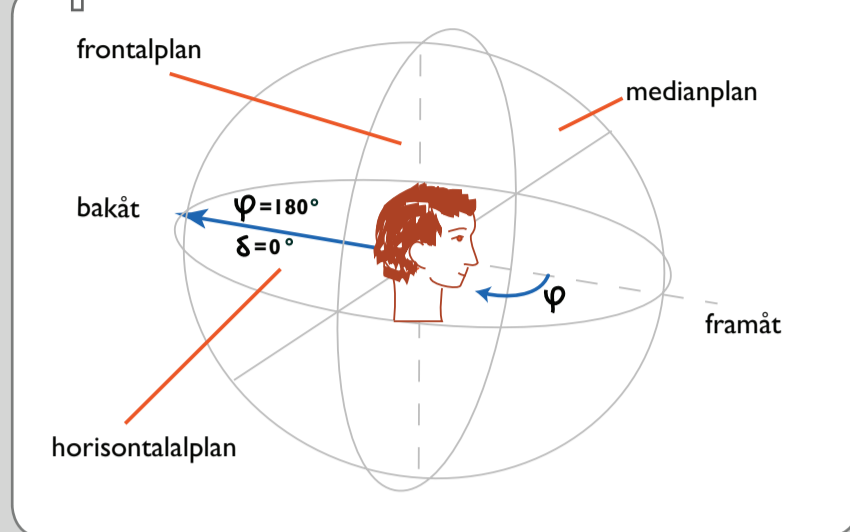


Jag fick hösten 2004 uppgiften att forska kring begreppet 3D-ljud i ett examensarbete på Saab Training Systems. Där används en egenutvecklad grafikmotor, Gizmo3D, som ska utökas med ljudstöd. 3D-ljud i simulerings-sammanhang hämtar vi från fenomenet Virtual Reality VR. Syntetiskt återskapande av vår riktiga verklighet (VR) skulle kunna vara bild och ljud-upplevelsen i filmsammanhang, men bristen på interaktivitet gör VR och 3D-ljuds hemmadomän till datorindustrin. Så i takt med att man lyckas simulera både verklighetstroga grafikvärldar och fysikaliska

modeller i realtid ställs allt större krav på ljudupplevelsen. Men vad är tredimensionellt ljud, 3D-ljud och hur kan man involvera det i Gizmo3D? Rapporten försöker svara på den frågan men även belysa 3D-ljudforskningens hela begreppsvärld, då den utgör grund för att kunna implementera ett Gizmo3D-ljudsystem. Arbetet innebär främst studier av befintliga verktyg och teknologier. Dessa mynnar ut i rekommendationer och egna förslag till lösningar, på viss problematik, för implementering i ett scenografdrivet grafiksystem, som Gizmo3D utgör.

3D-ljud

Spatialt hörande



3D-ljud motsvarar det vi hör. Fysikaliska, fysiologiska och psykologiska fenomen gör att vi upplever vår ljudliga omgivning som en tredimensionell ljudvärld. Vi har "spatialt hörande". Enskilda ljudkomponenter kan exempelvis ha positioner och hastigheter som upplevs relativt positionen på den som lyssnar, användaren.

Fysik

ITD (Interaural Time Difference)
tid- & fasskillnad för ljudvågor som anländer respektive öra för frekvenser under $\sim 1,6$ kHz

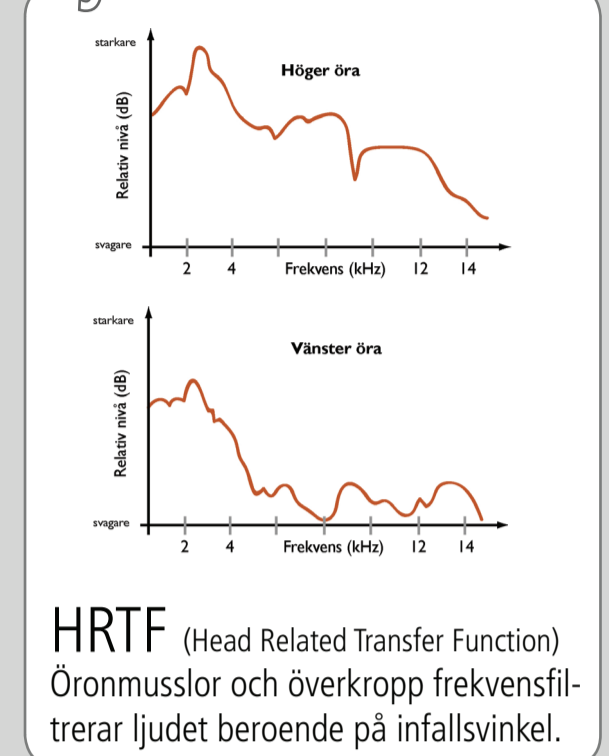
Ljudkälla på medianplanet
Samma sträcka till öronen => ITD = 0

- 34 cm skillnad ger ~ 1 ms skillnad i ankomsttid
- Fasskillnad: läget på ljudvägen är inte samma vid resp. öra

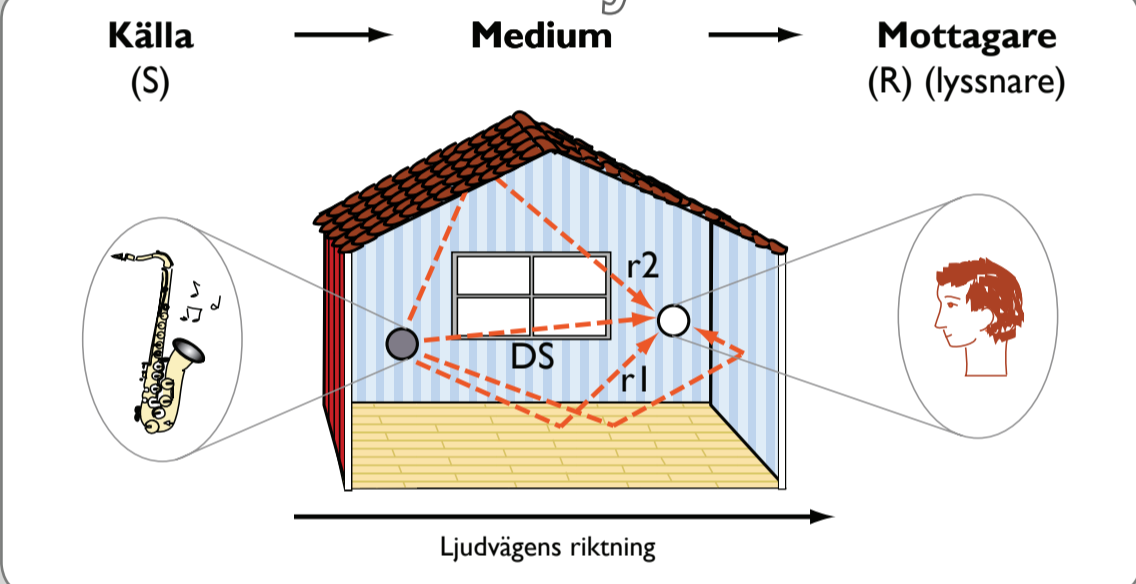
IID (Interaural Intensity Difference)
intensitetsskillnad för frekvenser över $\sim 1,6$ kHz

Vissa fysikaliska skillnader föreligger för en ljudvåg som anländer respektive öra på ett sfäriskt huvud utan öronmusslor. Den s.k. Duplex-teorin ger viss information men är alltför generaliserad.

Fysionomi



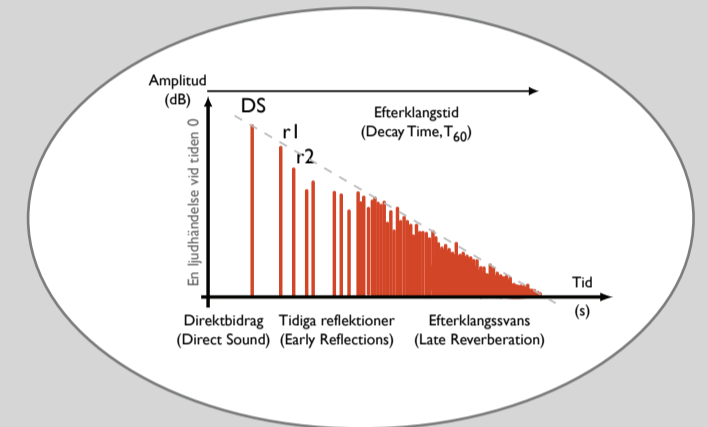
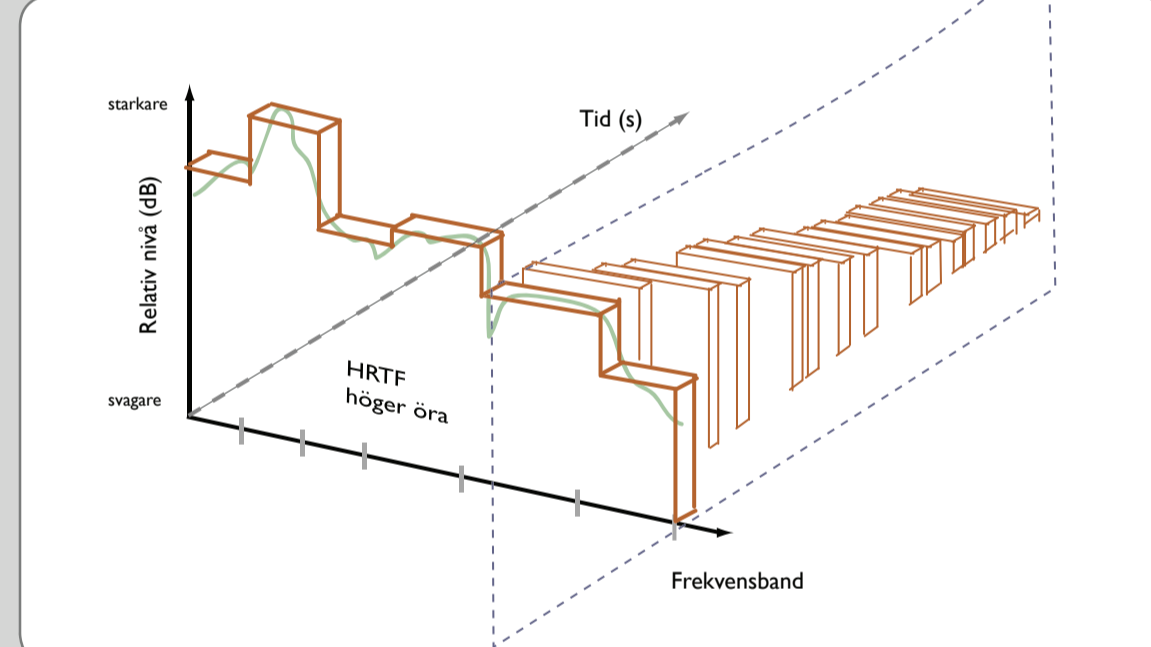
Akustik + Psykoakustik



Ljudet ger i vanliga miljöer upphov till en första direkt vågfront (DS) som är avgörande för rikttningsbedömningen. Dessutom ger reflektioner eller transmissioner upphov till instanser (r1, r2) av samma ljudhändelse, försenade i tid, energiavklingade på grund av längre ljudväg och frekvensmässigt dämpade beroende av absorption. Ger upphov till akustikens fingeravtryck: akustiksvaret.



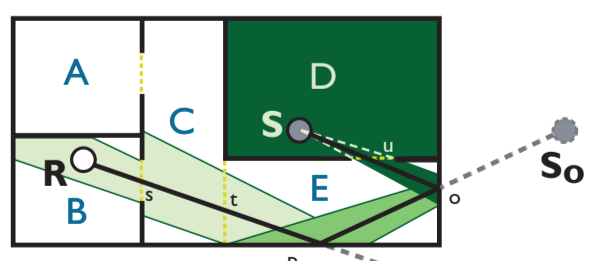
Akustiksvaret



Modellering

I ett VR-sammanhang med geometrisk information vore det önskvärdt att syntetiskt kunna modellera upp akustiksvaret. De flesta för ändamålet användbara metoderna antar att ljudvägen kan beskrivas som en stråle. Fysikaliskt inkorrekt, men godtagbart som approximation.

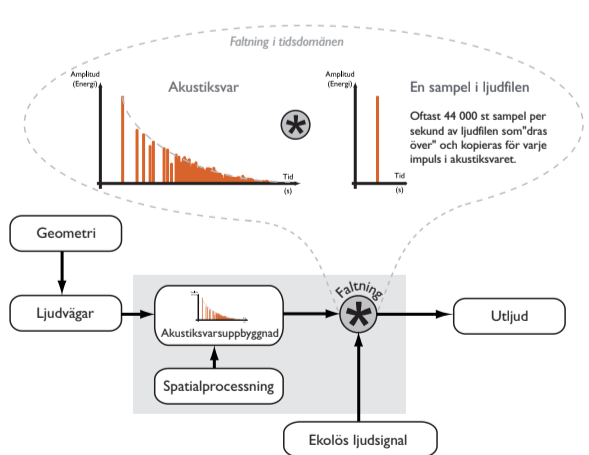
Teorier hämtas delvis från datorgrafikens ljussättningsmodeller med namn som ray- och beam-tracing och image source. Beam-tracing-algoritmen är trots vissa brister en kompetent metod. Förenar en förberäkningsfas (strålnippen, gröna ytor i bilden nedan) med realtidsframställandet av spegelkällor (image sources S_0, S_{op}):



Ovan beskrivna geometriska akustiska modellering lämpar sig för tidiga, diskreta reflektioner (se akustik-svar ovan). För senare delen på avklingningsförloppet kan man statistiskt räkna fram hur mycket ljudet ska dämpas och hur lång tid det tar tills man nått "total avklingning". Tiden är standardiserad och kallas för T_{60} . Vissa antaganden och mycket information om geometrin krävs för att stuligen komma fram till ett modellerat akustiksvaret.

Reproduktion

Oavsett om man syntetiskt framställt eller mätt upp ett akustiksvaret används samma metoder för att "auralisera", att tillsätta efterklang, ett ljud, faltning. Samtidigt måste man "spatialisera", positionera ljudet. Då appliceras matematiskt framställda ITD & IID men framför allt HRTF-filtreras ljudet. Individuellt uppmätta HRTF-värden borde användas men man är ofta tvungen att förlita sig till databaser med generella HRTF:er.



Olika signalteoretiska verktyg och specialanpassad hårdvara finns för att snabba upp faltningsprocessen.

Ovan angivna reproduktionsmetod används för det mest rättframma sättet att återkropa 3D-ljud, tvåkanaligt. Lättast då med hörlurar. Vill man däremot använda flerhögtalaruppsättningar kommer andra spatialiseringsmetoder behövas, vilka inbegriper olika avancerade panoreringar. Dessa ger oftast tydligare resultat men kräver en fast lyssnarposition.

Verktyg

För realtidsapplikationer såsom olika typer av visualisering och spel finns färdigutvecklade programmeringsbibliotek (API:er) för ljud. Jag har specialstuderat två stycken alternativ. OpenAL och FMOD. Båda två stöder flera plattformar, däribland viktiga spelkonsoler. Hela reproduktionskedjan ovan, åtminstone för spatialiseringen, sköts av underliggande rutiner med hårdvarustöd, till vilka API:erna skickar kommandon.

Det som krävs är att man matar ljudmotorn (API:t) med geometrisk information. Däremot saknas generella, dynamiska modeller för akustik. Ofta appliceras s.k. artificiellt reverber, godtyckliga justerbara fasta rumsakustiker, då används företrädesvis Creative's EAX, dock läst till Windows och Xbox.

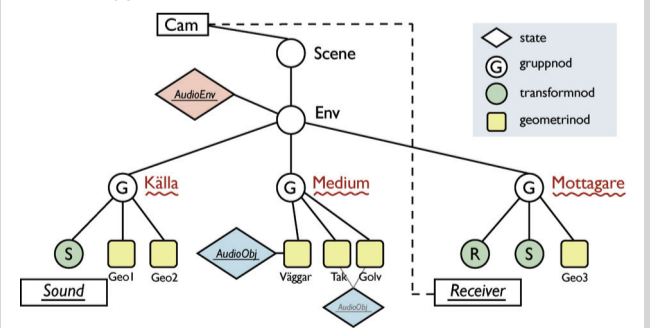
Resultat

Som jag beskrev inledningsvis är mycket av den förstudie och analys av olika tekniker ett delresultat av mitt arbete. Däremot ställer jag avslutningsvis upp en kravspecifikation för en skarp implementering av ett ljud-API i Gizmo3D.

Det kommer krävas ganska stora ansträngningar för att skapa ett väldigt generellt och dynamiskt ljud-API. Men det behövs, för i dagsläget är ljudhanteringen i VR-sammanhang eftersatt och ofta direkt dålig. Det ljudverktyg att utgå ifrån som är mest kompetent är i mina ögon FMOD.

De olika steg som krävs för att innefatta så mycket av hur verkligt ljud beter sig, eller snarare hur man ska approximera verkligt ljud, i en ljudmotor beskrivs och prioriteras i slutet av min rapport. Viktiga aspekter i rangordning är:

- Scenografkoppling - För en fungerande ljudmotor krävs att man på något sätt beskriver och kopplar ihop ljudvärldens geometri med den visuella informationen. På så vis kommer det mest väsentliga för ett 3D-ljudsystem, positionering, fungera. I detta skede måste man således ta ställning till hur man vill lagra ljudinformationen. Scenrafen som datastruktur känns logisk och man kan antingen hänga på ljuddata på befintligt ljudträd eller bygga en ny struktur. Man kommer oavsett kunna dra nytta av GIZMO3D:s underliggande struktur.



- Interaktivitet (rör det sig, låter det)
- Skuggning (täcks ljudvägen, måste ljudet dämpas)
- Bakgrundsmusik (stöd för händelsestyrd musik)
- Enkel rumsakustik (färdiga "reverber-presets")
- Hastighet och Doppler
- Avancerad rumsakustik (generiskt geometriskt drivet)
- Direktivitet (ljudkällors spridningsmönster)

Mer ingående förklaringar, egna uppslag till lösningar, kodkopplingar och slutsatser finns att läsa i min rapport:
<http://3dsound.kastel.org>

Text & Bild: David Kästel
0705 597525 & david@kastel.org