Realistisk formidling af virtuelle lydkilder

En eksperimentelt funderet udforskning af binaural syntese og *head tracking*, og af hvordan denne kombination kan benyttes i et system til formidling af virkelighedstro 3D-lyd.

Specialeprojekt i Audiodesign
2011-2012

Jakob Hougaard Andersen
Årskortnr. 20042689

Indholdsfortegnelse

Indledning

Når vi lytter til vores omgivelser, er vi i stand til at afkode en imponerende mængde informationer ud fra de to lydsignaler, vi modtager i vores ører. Dette er eksempelvis informationer om, hvad der har forårsaget de enkelte lyde – lydkilderne. Vi er enormt gode til at danne os et billede af, hvilke fænomener der er i spil, også selvom vi ikke nødvendigvis kan se dem. I denne forbindelse er det interessant at bemærke, hvordan vi med synssansen altid befinder os i periferien af det sansede, mens vi med høresansen befinder os i midten[1]. Denne egenskab ved høresansen har uden tvivl øget vores overlevelseschancer i stenalder-junglen og fortsætter med at gøre det i den urbane jungle. Men vi er ikke blot gode til at afkode, *hvad* der forårsagede lyden. Vi er også overraskende gode til at afkode, *hvor* den enkelte lydkilde befinder sig. Hvis man har afkodet, at der befinder sig en sabeltiger eller en bil i ens nærvær, er det naturligvis overlevelsesfremmende også at have en fornemmelse af, *hvor* den befinder sig. Og netop vores evne til at afkode lydkilders position, er helt central i denne opgave.

Siden de første systemer til lagring og reproduktion af lyd blev opfundet i anden halvdel af 1800-tallet, er vi blevet stadig bedre til at gengive lydlige fænomener på en mere og mere naturtro måde. Med avancerede mikrofoner, digital lagring og sofistikerede højttalere med mere kan vi i dag gengive den nær eksakte klang af et hvilket som helst lydfænomen. Og dog er det stadig vanskeligt for os realistisk at gengive et lydbilledes spatiale karakter – oplevelsen af de enkelte lydkilders specifikke position og lydens udfoldelse i rummet.

Denne opgave er en udforskning af specifikke teknologiers/teknikkers potentiale, netop med henblik på at kunne gengive et lydbilledes rumlige karakter på realistisk og troværdig vis. De teknologier, opgaven tager udgangspunkt i, er binaural lyd, herunder specifikt binaural syntese, og *head tracking* (automatisk sporing af lytterens hoveds orientering). Opgaven præsenterer et projekt der, på baggrund af viden fra lignende projekter og forskning, har beskæftiget sig med en eksperimentel udvikling af et funktionsdygtigt system til realistisk formidling af 3D-lyd bygget på de nævnte teknologier. Således har jeg allerede afsløret, at det ikke er det første system af sin art. Binaural syntese i kombination med *head tracking* er blevet udforsket før og er blevet implementeret i diverse eksperimentelle systemer. Eksempelvis er det blevet udforsket med henblik på at give jægerpiloter præcis lydlig information om, hvor der eventuelt befinder sig andre fly, missiler med mere. Andre har beskæftiget sig med teknologien i en mere oplevelsesorienteret sammenhæng, eventuelt med henblik på at formidle virtuelle eller augmenterede verdener. Og desuden er (aspekter af) teknologien ofte blevet udforsket uden erklæret brugssammenhæng. På trods af at teknologien allerede har været genstand for en del forskningsmæssig interesse, vil jeg påstå, at der

[1] Breinbjerg

stadig er et potentiale for at udforske området yderligere. Der er stadig mange udfordringer, der kan angribes på forskellige måder, og hvortil der kan udvikles nye metoder og teknikker. Denne opgave vil forsøge at kaste lys på nogle af disse udfordringer samt præsentere løsningsforslag, som forhåbentlig kan inspirere fremtidigt arbejde på området.
Systemet i dette projekt er udviklet med henblik på at formidle en oplevelse til lytteren. Og netop teknologiens potentiale i en oplevelsesorienteret sammenhæng er et aspekt, dette projekt har sigtet mod at skabe nogle erfaringer med. Således vil opgaven præsentere nogle af de erfaringer, jeg har gjort mig med systemet i oplevelsesorienteret brug. Disse erfaringer vil ligeledes forhåbentlig kunne inspirere og informere fremtidigt arbejde.

God læselyst.

Noter til bilag

Jeg vil gerne henlede læserens opmærksomhed på opgavens bilag, der er vedlagt på en data-DVD (som skal åbnes på en computer). Herunder i særdeleshed bilag 1, som er video-dokumentation, hvori jeg forsøger at præsentere systemet. I fald læseren ikke har mulighed for at afspille videofilen på den vedlagte DVD, kan videoen også ses på *youtube* på adressen:
http://www.youtube.com/watch?v=kH4mVUgk9Do .
I forbindelse med bilag 2 og 3, som er henholdsvis *Max*-kode og Java-kode, vil jeg knytte en enkelt kommentar, nemlig at jeg har valgt at benytte engelsk i alle kommentarer, variabelnavne med mere i forbindelse med al programmeringen. Dette for at gøre det nemmere at dele det udviklede med andre interesserede, som jo ikke nødvendigvis er danskere.

Hvad er binaural lyd?

> ***binaural,*** *(lat. bin- + aur- af auris øre + -al),*
> *med begge ører; det modsatte af monaural.*
> [http://www.denstoredanske.dk]

I forlængelse af ovenstående definition kan vi fristes til at konkludere, at binaural lyd er *lyd, der høres med begge ører*. En sådan definition af binaural lyd er imidlertid relativt meningsløs, idet den ikke siger noget om lyden som sådan men derimod om perceptionen af den – en perception som jo faktisk altid foregår med begge ører. Jeg vil i stedet foreslå denne, måske lettere spidsfindige, definition: Binaural lyd er *lyd, der er hørt med begge ører*. Men hvordan kan man tale om lyd, der allerede er hørt? Det kan man måske heller ikke, hvis vi taler om den fulde lydlige perception, der omfatter hele processen med, hvordan vi omsætter det varierende lufttryk i rummet omkring os til en registrering og forståelse af et elektronisk signal i hjernen. Men hvis vi lægger et snit ned netop det sted i processen, hvor vores ydre ører har

'hørt' lyden, giver det mere mening. Hvorfor det er interessant/relevant at beskæftige sig med lyden på dette perceptionsstadie, kommer jeg ind på om lidt. Binaural lyd kan altså siges at være lyd, der allerede har foretaget en lille del af 'den perceptoriske rejse' og er nået til det sted, hvor den er klar til at fortsætte ind i vores øregange. Det vil sige, at lyden på dette stadie har rejst fra lydkilde(r) til vores øregange, og på denne rejse er den blandt andet blevet 'farvet' af vores ydre anatomi. Denne farvning resulterer i to, mere eller mindre forskellige, lydsignaler – et fra venstre øre og et fra højre.

Men hvad er det for en farvning, der sker, og hvad bruger vi den til?

Allerede i 1907 formulerede Lord Rayleigh (J. W. Strutt) sin såkaldte *Duplex Theory*[2], der beskrev de basale principper i forbindelse med spatial lytning – principper som siden er blevet udbygget, men som essentielt set stadig anses for at være gældende. Han var kommet frem til, at vores evne til at lytte spatialt var baseret på afkodningen af små forskelle mellem den lyd, der kommer ind i henholdsvis venstre og højre øregang. Ifølge Rayleigh kunne man opdele disse lydlige forskelle i to fysiske egenskaber: *Interaural Time Difference*[3] (ITD) og *Interaural Intensity Difference* (IID) – se *Figur 1*.

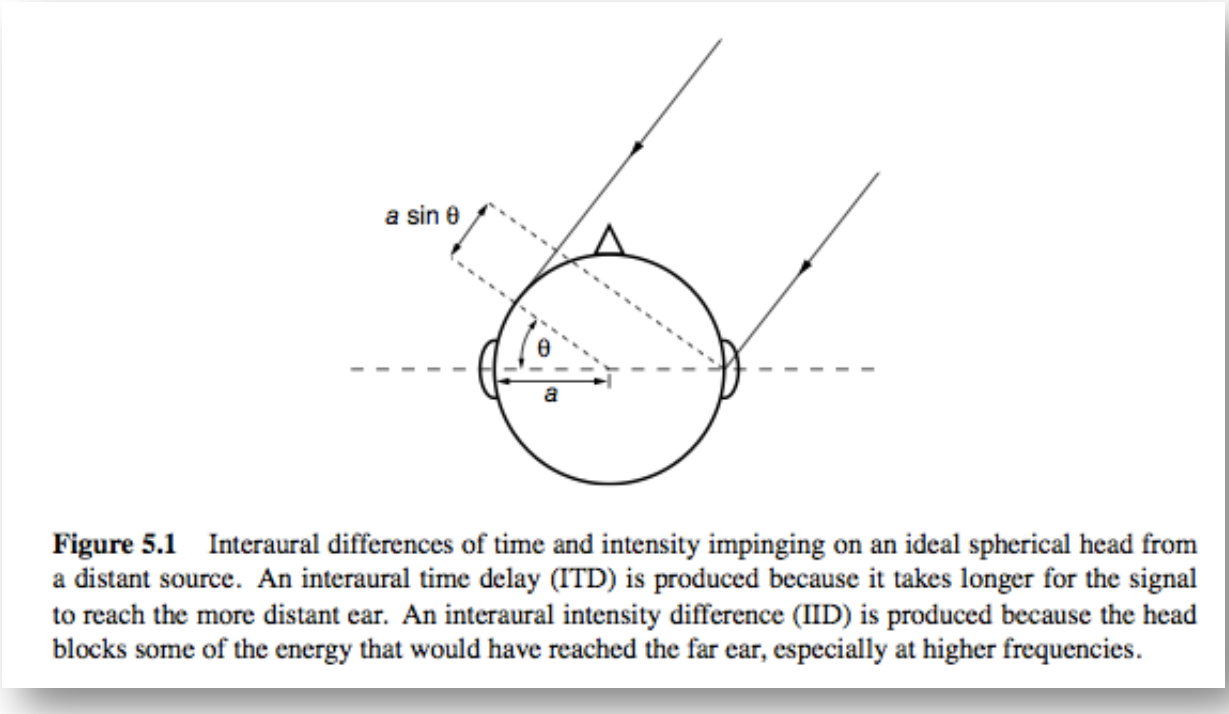


Figure 5.1 Interaural differences of time and intensity impinging on an ideal spherical head from a distant source. An interaural time delay (ITD) is produced because it takes longer for the signal to reach the more distant ear. An interaural intensity difference (IID) is produced because the head blocks some of the energy that would have reached the far ear, especially at higher frequencies.

Figur 1. Figur taget fra Stern, R. M.; Wang, DeL; Brown, G. : Binaural Sound Localization. Kapitel 5 i: *Computational Auditory Scene Analysis*, s. 2.

[2] Stern, Wang og Brown – s. 2

[3] På figuren er fænomenet godt nok benævnt *Interaural Time Delay*, men dette må være en fejl, idet den gængse betegnelse er *Interaural Time Difference*, og idet sidstnævnte betegnelse er benyttet andetsteds i Stern, Wang og Browns kapitel.

ITD er den forskel i tid, der er mellem signalet i det ene og det andet øre. Hvis lydkilden er tættere på det ene øre, vil lyden naturligvis ramme dette øre først, og signalet i det andet øre vil dermed være en anelse forsinket. Denne anelse er maksimalt 0,66 millisekunder for et menneske med et hoved af gennemsnitlig størrelse[4], så det er altså meget små tidsforskydninger, vi taler om. I fald lydkilden befinder sig på lytterens medianplan (og altså dermed har samme afstand til begge ører) vil ITD være lig nul, og omvendt vil ITD være maksimal, hvis lydkilden befinder i retningen stik ud for et af lytterens ører.
IID, som også ofte benævnes ILD (*Interaural Level Difference*), er forskellen i lydstyrken mellem signalerne i de to øregange. Hvis lydkilden er tættere på det ene øre, vil lydstyrken naturligvis være kraftigere i dette øre – denne indvirkning af afstandsforskelle er størst, når lydkilden er meget tæt på det ene øre (hvorved den relative forskel i afstand er meget stor – dette fænomen beskrives yderligere i afsnittet *Rumgeometri – retning og afstand*). Forskellen i lydstyrken skyldes også vores hoveds 'skyggende' effekt på lydbølgerne samt disses komplekse møde med hele vores ydre anatomi (ører, torso, hoved…). Det er vigtigt at pointere, at denne IID eller ILD, er særdeles frekvens-afhængig. Det vil sige, at der er forskel på, i hvilken grad og hvordan forskelligt lydligt frekvens-indhold 'skygges' og formes af vores ydre anatomi. I denne sammenhæng er den generelle regel/tendens, at højfrekvent lyd har sværere ved at bevæge sig rundt om hovedet og ind i øret længst væk fra lydkilden end lavfrekvent lyd – højfrekvent lyd kan betragtes som værende mere retningsbestemt end lavfrekvent lyd. Dermed er den skyggende effekt i IID størst for højfrekvent lyd. Vi kan altså ikke betragte IID som blot et enkelt tal (som vi tilnærmelsesvis kan med ITD[5]), der indikerer lydstyrkeforholdet mellem signalet i de to ører. Rettere er der tale om en slags frekvens-afhængigt filter, der dæmper (og eventuelt forstærker) nogle frekvenser mere end andre. Præcis hvordan vi formår at omsætte disse små forskelle mellem det lydlige signal i højre og venstre øre til en afkodning af de enkelte lydkilders rumlige position, er imidlertid en mere kompleks og omdiskuteret sag. Der findes forskellige modeller, der blandt andet beskæftiger sig med, hvordan vi sammenligner de elektriske signaler på neuron-plan. Den mest kendte og benyttede af disse er nok den såkaldte *Jeffress-model*[6], udviklet af Lloyd Jeffress i 1948. Det er en model, der har dannet basis for mange senere modeller. Jeg vil dog ikke her beskæftige mig mere med, hvordan dennae afkodning foregår – simpelthen fordi det ikke er nødvendigt i konteksten, binaural lyd. Binaural lyd bygger i bund og grund på en meget simpel antagelse: hvis vi indfanger/syntetiserer lyden, som den lyder på vej ind i vores øregange, og af-

[4] Stern, Wang og Brown – s. 3
[5] Der hersker lidt tvivl om, hvorvidt ITD også er frekvens-afhængig. Generelt er det ikke noget, der er antaget i de fleste artikler på området, og dog er det nævnt her: http://www.faqs.org/docs/sp/sp-73.html . Muligvis skyldes forvirringen, at de fleste artikler beskæftiger sig med målinger på bredspektrede lydimpulser, og at ITD i denne sammenhæng simpelthen defineres ud fra ansatsen impulsen. Og desuden, at eventuelle forskelle i forskellige frekvensers tidsforskydning i disse artikler betragtes som en del af IID.
[6] Stern, Wang og Brown – s. 11

spiller den på selv samme sted, må vi få samme slutresultat, som hvis vi blot lod lyden passere direkte – altså realistisk, som om vi var der, osv.… Og så kan vi sådan set være ligeglade med, hvad der sker på lydens videre vej i dens perceptoriske rejse.

Det essentielle og interessante i binaural lyd er, at lyden *i sig selv* 'indeholder' den rumlige konfiguration mellem lytter og lydkilder. Den rumlige konfiguration skal altså ikke søges genskabt via en snedig opstilling af højttalere – dette ville resultere i, at lyden bliver 'indkodet' *to* gange. Derimod skal den afspilles der, hvor den blev indfanget (/ hvortil den blev syntetiseret), altså via hovedtelefoner.

Optagelsen vs. syntesen

Binaural lyd kan opnås enten ved at optage lyd med binauralt optageudstyr eller ved at syntetisere lyd, således at den opnår de binaurale karakteristika. En binaural optagelse kan eksempelvis foretages ved at sætte små mikrofoner i ørerne eller ved at benytte et såkaldt *dummy head* med indbyggede mikrofoner.

Figur 2. Binaurale mikrofoner i ørerne på en virkelig person.
Billedet er taget fra web-adressen:
http://christopherbaker.net/projects/27-may-2003/

Figur 3. Et *dummy head* til binaurale optagelser
Billedet er taget fra web-adressen:
http://www.dv247.com/microphones/neumann-ku-100-dummy-head-binaural-stereo-microphone--21004

En af fordelene ved en binaural optagelse frem for syntesen er, at det er en let og meget overbevisende måde at indfange alle lydens spatiale karakteristika på. Lydkilderne skal ikke efterfølgende processeres via kompliceret software, og endda rummets akustik bliver indfanget meget realistisk. Ulempen ved den binaurale optagelse er, at den, som enhver anden lydoptagelse, er statisk. Det vil sige, at optagelsen gen-

giver netop *den* konfiguration mellem lytter og lydkilder, hvorved optagelsen blev foretaget. Det er altså ikke med den almindelige binaurale optagelse muligt at (gen)skabe dynamiske lydlige miljøer, hvor eksempelvis lytterens kropslige bevægelse har indflydelse på det hørte, som det jo er tilfældet med lyd, der ikke høres via hovedtelefoner. Det er her, den binaurale *syntese* har sin force. Med syntesen er det muligt at indkode en lydkilde, så den lyder, som om den er placeret i en hvilken som helst position i forhold til lytteren (det er i hvert fald tesen). Og på denne måde giver syntesen en større frihed til at skabe mere dynamiske, og dermed paradoksalt nok potentielt mere realistiske[7], lydlige miljøer.

3D-lyd

3D-lyd er et almindeligt benyttet begreb for lydgengivelse, der formidler en oplevelse af realistisk rumligt placerede lydkilder. Således er 3D-lyd et relativt bredt begreb som, foruden binaural lyd også omfatter andre teknikker og bestræbelser, der sigter mod en sådan oplevelse. Af sådanne andre teknikker kan blandt andet nævnes diverse multikanals-setups – altså setups, der benytter mange højttalere rundt om lytteren til at skabe rumligt realistisk lyd. *Ambisonics* er en teknik, der ligeledes arbejder med flere højttalere (fire eller mere), men som desuden via avanceret indkodning/afkodning sigter mod at (gen)skabe et realistisk lydfelt i *en* specifik lytter-position mellem højttalerne[8]. Der findes også systemer, der sigter mod at skabe 3D-lyd via blot *to* højttalere. Sådanne systemer er ret nært beslægtede med binaural lyd[9], idet der her er tale om to lydsignaler - et til hvert øre. Da lyden fra hver højttaler dog imidlertid kan høres af begge ører (og altså ikke er isoleret som i hovedtelefonerne) benytter sådanne systemer sig af såkaldt *crosstalk-cancellation*, som er en måde at håndtere dette kryds-overhør på. Overordnet kan man sige, at de højttaler-baserede systemer har den fordel, at lytteren ikke skal iføre sig hovedtelefoner, for at opleve lyden – man oplever lyden som man er, frit, som i virkeligheden…. Og der er også et socialt aspekt, der umiddelbart kunne synes at tale til fordel for højttalerne. Her tænker jeg på det, at lyden fra højttalere fylder et rum, som kan rumme flere personer, der altså oplever denne lyd sammen. Dog har de fleste højttaler-baserede systemer en sådan natur, at der kun er *et* meget specifikt sted i rummet, hvor oplevelsen er optimal – det såkaldte *sweet spot* – og dette sætter sine begrænsninger både for lytterens færden og for det sociale aspekt. Iført hovedtelefoner befinder lytteren/lytterne sig altid i *the sweet spot*, uanset hvor i rummet vedkommende måtte befinde sig. Desuden er det med hovedtelefoner ikke nødvendigt at forholde sig til rummets akustik. Netop rummets akustiske refleksioner og farvning af lyden, kan gøre det svært i praksis at realisere simuleringen af et virtuelt rum i de højttaler-baserede systemer. Det faktum at de hovedtelefon-baserede systemer er uafhængige af rummet og lytterens placering i dette, kan, paradoksalt nok, også betragtes som disse systemers største ulempe. I den virkelige,

[7] Dette forstået på den måde, at syntesen åbner op for muligheden for at tilpasse lydbilledet i hovedtelefonerne til lytterens bevægelser, hvilket vel, i hvert fald potentielt set, er en tilføjelse af realisme.
[8] Toole
[9] Og kan da også muligvis betegnes som værende binaural lyd.

umedierede verden *er* vores oplevelse af lyd jo netop afhængig af vores bevægelser – vores position og orientering i forhold til de enkelte lydkilder. Hvis vi vil (gen)implementere bevægelsens betydning i et hovedtelefon-baseret system, er det derfor nødvendigt, at vi implementerer en *real-time tracking* af lytterens bevægelser og tilpasser lyden efter disse.

Vision og systemkrav

Forestil dig, at du befinder dig i et rum, hvor du, iført hovedtelefoner, kan gå rundt og lytte til lydkilder, der er placeret/bevæger sig rundt omkring dig. Lydkilderne er muligvis koblet til det konkrete rum, du befinder dig i, forstået på den måde, at lyden er en slags augmentering af det erfarede virkelige rum. Eller måske formidles der i lyden et helt andet, og dermed virtuelt, rum, end det du fysisk befinder sig i. I begge tilfælde oplever du, at du har en klar fornemmelse af, hvor den enkelte lydkilde befinder sig. Og når du drejer hovedet, kan du høre, hvordan lydbilledet ændrer sig, og den enkelte lydkilde bevarer sin oplevede position, som det ville ske i virkeligheden. Ligeledes kan du i denne vision gå rundt i rummet og høre, hvordan du kommer nærmere nogle af lydkilderne og fjerner dig fra andre. Du kan også høre, at lyden udfolder sig realistisk i rummet (det virtuelle eller augmenterede). Dette med hensyn til rummets akustik. Alt i alt opleves det hele meget realistisk – som om der ikke var tale om en mediering. Transparent, som om lydene rent faktisk var i rummet / som om du rent faktisk befandt dig i det formidlede rum. Denne formidlingsform gør, at du lever dig meget ind i den rumlige situation, det lydlige indhold 'beskriver'. Hvad enten der er tale om en slags lydlig fortælling, en musikfremførelse eller noget tredje.

Ovenstående er et forsøg på at formulere den vision, projektet er bygget op omkring. Visionen kan betragtes som det mål, jeg har sigtet mod, og delvist også som min motivation for at give mig i kast med projektet. *Men hvis vi skal realisere denne vision, hvilke krav stiller det så til systemet?* Nedenstående er en skitsering af de overordnede krav.

- **Binaural syntese**
  Den binaurale syntese skal sørge for at indkode lydkilderne, således at de opleves som værende realistisk positioneret i en specifik retning i forhold til lytteren. Da lytteren kan ændre position og orientering, skal den binaurale syntese være dynamisk.
- ***Head tracking* (orientering)**
  For at kunne tilpasse lydbilledet i henhold til lytterens orientering, skal computeren kunne spore dennes hoveds orientering på en robust og stabil måde.
- **Positions-*tracking***
  Hvis vi vil realisere visionen, skal computeren også kunne spore lytterens *position*, foruden orientering. Da positions-*tracking* imidlertid er en temmelig kompliceret og tidskrævende sag, har jeg

dog valgt at undlade at forsøge at realisere det i denne prototype. Dog er det vigtigt at pointere, at positions-*tracking* stadig indgår i min vision for projektet.

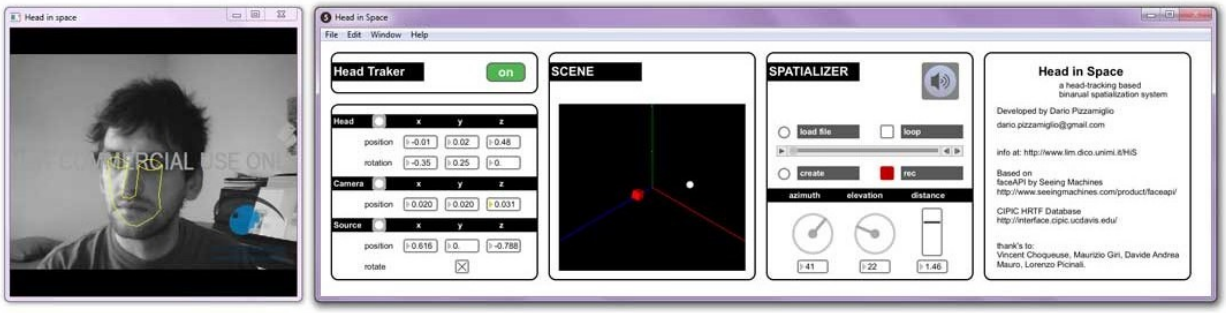
- **Komposition i tid og rum**
  Det er vigtigt, at det i systemet er muligt at strukturere lydligt indhold, således at de enkelte lydkilder har en position både i rummet og i tiden. Dette, hvad enten der er tale om en prædeterminеret eller mere dynamisk komposition.
- **Håndtering af rumgeometriske aspekter**
  Systemet skal kunne udregne forskellige parametre på baggrund af lydkildernes positioner samt lytterens position og orientering. Disse parametre er eksempelvis afstanden fra lytter til lydkilde (som skal benyttes til at bestemme styrken af den enkelte lydkildes signal) og lydkildens relative retning (som skal benyttes i forbindelse med den binaurale syntese).
- **Simulering af rummets akustik**
  For at give et realistisk indtryk af lydens udfoldelse i rummet samt for at understøtte retnings- og eksternaliseringsfornemmelserne hos lytteren er det vigtigt, at systemet kan foretage en simulering af det formidlede rums akustik.
- **Lydligt indhold**
  Det lydlige indhold er ikke en del af systemet som sådan. Grunden til at det alligevel er med på denne liste er, at vi er nødt til at have noget lydligt indhold for at kunne vurdere systemets potentiale i en oplevelses-/formidlings-sammenhæng. Dette lydlige indhold skal være relativt 'færdigt' i sit udtryk, således at lytteren kan få et indtryk af oplevelsespotentialet. Derfor har jeg sat det som et krav til dette projekt, at der skal produceres et par eksempler på lydligt indhold, som kan danne basis for erfaringer med systemet i brug.

Så vidt en overordnet liste over krav til systemet. Dykker vi nærmere ned i de enkelte aspekter, hvilket vi gør senere, dukker der mere specifikke krav og problematikker op.

Inspiration fra lignende systemer og forskning

Dette afsnit beskriver nogle af de relaterede projekter – systemer og forskning – jeg har hentet inspiration i. Det er vigtigt at pointere, at dette afsnit ikke skal ses som en udtømmende redegørelse for, hvad der er *state of the art* på området, men rettere som en række udvalgte nedslag. Nogle af de her skitserede projekter vil også optræde senere i opgaven i forbindelse med gennemgangen af de specifikke udfordringer og løsningsmuligheder.

Head In Space[10]



Figur 4. Screenshot af *Head in Space* -interfacet. Taget fra http://sites.google.com/site/dariopizzamiglio/projects/head-in-space .

Head In Space er et projekt fra 2010, skabt af Dario Pizzamiglio. I projektet har han skabt en *real-time* binaural syntese kombineret med *head tracking*. Lyddelen og interfacet er bygget op i *Max*[11], og *head tracking* -delen bygger på ansigtsgenkendelse via et (web)kamera og den kommercielt udviklede algoritme, *faceAPI*[12]. I den frit tilgængelige demonstration af *Head in Space* er det således muligt at placere en lydkilde i et virtuelt tredimensionelt rum – en lydkilde, der bliver binauralt syntetiseret i henhold til lytterens hoveds orientering. Projektets formål synes ikke helt klart defineret, men det lader til at være motiveret af en lyst til at udforske teknologiens potentiale i en oplevelsesorienteret sammenhæng. I det hele taget minder *Head in Space* en hel del om det projekt, jeg har sat i gang, og jeg har da også hentet inspiration herfra. Dog mener jeg, i lighed med Pizzamiglio, at der er aspekter, der kan udfoldes yderligere, og udfordringer, der muligvis kan løses på en mere hensigtsmæssig måde. Her tænker jeg bl.a. på, at projektet ikke tager skridtet fuldt ud med hensyn til at undersøge, hvordan/hvorvidt et sådant system kan formidle en 'færdig' oplevelse til lytteren. Der er blot fokus på, hvordan en enkelt lydkilde kan placeres i et virtuelt univers. Dette hænger muligvis sammen med den temmelig processor-tunge metode til binaural syntese, man har valgt i projektet (mere om dette i afsnittet *Metode til binaural syntese*), og en deraf følgende begrænsning i antal lydkilder…(?). Derudover beskæftiger projektet sig ikke med simulering af rummets akustik, hvilket angiveligt begrænser fornemmelsen af immersion i det virtuelle rum. Projektets *head tracking* -funktionalitet er ganske smart, idet lytteren kan spores uden at iføre sig nogen form for teknisk aggregat. Dog har den også åbenlyse ulemper, fordi lytterens ansigt skal kunne 'ses' på relativt nært hold af et kamera. Når lytteren har drejet hovedet tilstrækkeligt langt ud til siden, kan kameraet ikke længere se lytterens ansigt, og dermed kan dennes orientering ikke længere bestemmes

[10] Mine kilder i forbindelse med *Head in Space* er henholdsvis artiklen *Head in Space: A Head-tracking Based Binaural Spatialization System* af Pizzamiglio m.fl. og projekt-siden på D. Pizzamiglios hjemmeside: http://sites.google.com/site/dariopizzamiglio/projects/head-in-space

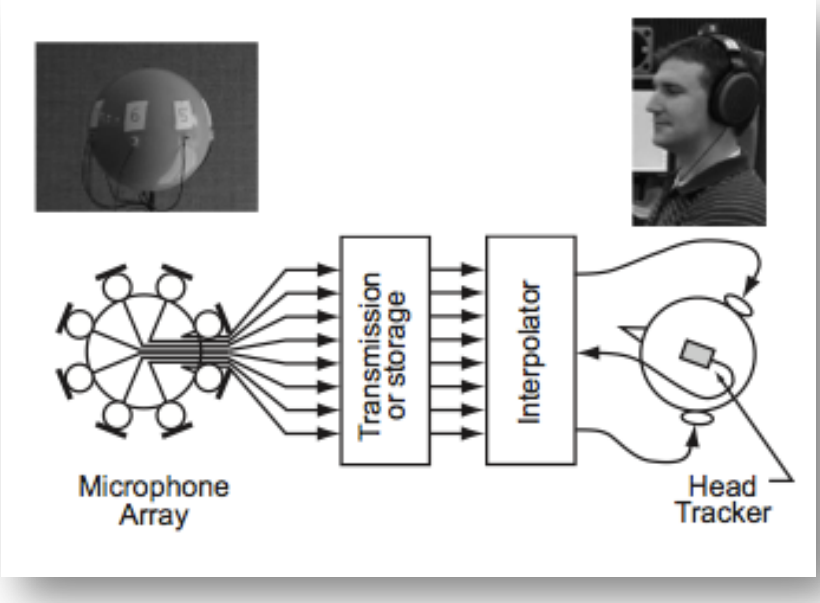
[11] http://cycling74.com/

[12] http://www.seeingmachines.com/product/faceapi/

(dette kunne dog muligvis løses ved brug af flere kameraer). I mit projekt har jeg derfor valgt at arbejde med en anden teknisk realisering af *head tracking*'en såvel som andre aspekter af systemets udformning.

MTB

I artiklen *Immersive Spatial Sound for Mobile Multimedia*[13] fra 2005 beskriver V. Ralph Algazi og Richard O. Duda systemet MTB (*Motion Tracked Binaural*). Her er der ligeledes tale om et projekt, der beskæftiger sig med binaural lyd og *head tracking*. Projektet er fokuseret i en oplevelsesorienteret retning og mod at skabe en metode til at formidle 3D-lyd på, især mobile enheder (mobiltelefoner osv.). Da de mobile enheder ofte har noget større begrænsninger med hensyn til processorkraft, hukommelse og dataoverførsel end den gængse computer[14], koncentrer Algazi og Duda sig om at udvikle en metode, der er realiserbar indenfor disse begrænsninger. Dette betyder en relativt stor grad af pragmatik og forenkling i forbindelse med håndteringen af de fysiske forhold. Centralt i systemets grundudformning er en kugle i hovedstørrelse, hvorpå mikrofoner kan monteres i forskellige positioner – se *Figur 5*.



Figur 5. Figur taget fra Duda og Algazis artikel *Immersive Spatial Sound for Mobile Multimedia.*

Afhængig af hvor fin en opløsning der ønskes, kan flere eller færre mikrofoner benyttes, og de kan også monteres andetsteds på kuglen end blot på 'ækvator', som det er tilfældet på *Figur 5*. Princippet er enkelt: et, potentielt uendeligt komplekst, lydbillede indfanges af mikrofonerne på kuglen, og i afviklingssituationen modtager lytteren et interpoleret signal i forhold til sin orientering. På denne måde er det muligt at formidle et meget komplekst lydbillede på en meget simpel og 'processor-let' måde. Her er der

[13] Algazi og Duda, 2005
[14] Dog er skellet mellem *mobil enhed* og *computer* efterhånden mere og mere udvisket.

ikke tale om binaural syntese[15], men rettere en form for dynamisk multikanals- binaural optagelse. En af metodens helt store fordele er, i min optik, at den let er i stand til at gengive et virkeligt lydbillede i al sin kompleksitet – altså eksempelvis indfange og gengive en symfoniorkester-koncert inklusiv rummets akustik. Mange 3D-lyd-systemer, inklusiv mit eget, er bygget op omkring en modellering af den enkelte lydkilde samt dennes position og udfoldelse i det 3dimensionelle rum. At implementere en live-optagelse af et symfoniorkester i sådanne systemer vil være en meget kompleks opgave, og svær at realisere i praksis – blandt andet på grund af vanskeligheden i at indfange hver lydkilde separat og isoleret.
MTB-systemet har dog også sine begrænsninger. Blandt andet bygger det på nogle pragmatiske tilnærmelser. Her tænker jeg på kuglen, som må siges at være en temmelig kraftig forenkling af vores ydre anatomi, samt det at der interpoleres imellem en temmelig grovkornet sampling af signaler, der indeholder indbyrdes tidsforskydninger. Algazi og Duda er dog udmærket godt klar over, at der er tale om en tilnærmelse:

> *Although MTB produces highly-realistic, well externalized spatial sound, the signals produced by this method only approximate the exact experience, and critical listening tests have revealed various audible defects.*
> [Duda og Algazi: *Immersive Spatial Sound for Mobile Multimedia*]

Ud over at kuglen benyttes som direkte indfangnings-værktøj, beskriver artiklen også, hvordan den kan benyttes som en model, der kan danne basis for binaural syntese baseret på beregnede eller målte impuls-responser. I denne sammenhæng fremgår teknikkens potentiale dog, i mine øjne, ikke så klart i sammenligning med HRTF-baserede teknikker, som den minder om. Fænomet HRTF forklares i afsnittet *Metode til binaural syntese*.
Nogle af Duda og Algazis visioner i forbindelse med teknikkens anvendelsesmuligheder går i retningen af augmented reality. Blandt andet beskriver de nogle idéer, hvor 3D-lyd bruges til at augmentere virkeligheden i navigations-applikationer og i mere narrative lyd-guides. I disse visioner bliver forskellige lyde positioneret i henhold til virkelige fænomeners position i forhold til lytteren. På denne måde ser de nogle muligheder for, hvordan 3D-lyd i sammenspil med den mobile enheds skærm kan fremme en immersiv oplevelse – de mobile enheders skærme, som, i kraft af deres begrænsede størrelse, ikke i sig selv fordrer nogen særlig grad af immersion. De beskrevne visioner om 3D-lyd i en oplevelsesorienteret augmented reality -sammenhæng har meget til fælles med visionerne for mit system og udforskningsprojekt. Dog er mit fokus ikke på mobile enheder, og mit valg af teknisk realisering af den lydlige processering er da også et andet.

[15] I hvert fald ikke i den gængse forståelse af hvad binaural syntese er: convolution via HRTF

Rendering Localized Spatial Audio in a Virtual Auditory Space[16]

I artiklen med ovenstående titel beskrives et forskningsprojekt, som nok er det enkeltprojekt, jeg har hentet mest inspiration og viden fra. I artiklen fra 2004 præsenterer forfatterne, Zotkin, Duraiswami, og Davis, en række metoder, algoritmer og tekniske løsninger, som de har fundet hensigtsmæssige til deres formål.

> *A goal of our work is to create rich auditory environments that can be used as user-interfaces for both the visually-impaired and the sighted.*
> [Zotkin m.fl. : *Rendering Localized Spatial Audio in a Virtual Auditory Space*]

Projektet er altså i højere grad end mit rettet mod en form for konkret, nyttig funktionalitet. Og dog har forfatterne stadig et vist fokus på oplevelses-dimensionen, hvilket også afspejler sig i formuleringen …***rich*** *auditory environments*…. På trods af at vores fokus ikke er helt det samme, har jeg kunnet hente megen inspiration i de tekniske løsninger, artiklen præsenterer. I projektet arbejdes med både binaural syntese, *head tracking* og simulering af rummets akustik – elementer som også er essentielle i mit projekt. Med hensyn til valget af mange af de grundlæggende tekniske metoder og strategier minder vores projekter således en hel del om hinanden. Dog er den konkrete, tekniske implementering af løsningerne forskellig ligesom det lydlige indhold, med hvilket systemets potentiale afprøves.
Men *Rendering Localized Spatial Audio in a Virtual Auditory Space* er en artikel, der præsenterer nogle gode løsningsforslag, og som kommer rundt om forskellige aspekter ved opbygningen af et sådant system.

Og meget mere

Ovenfor har jeg kort beskrevet tre projekter, jeg har hentet inspiration i. Det er dog blot tre ud af mange, i hvilke jeg har hentet viden og inspiration. For et mere komplet overblik over kilder henviser jeg til litteraturlisten. Her vil læseren finde referencer til artikler og ressourcer vedrørende flere lignende systemer. Desuden indeholder litteraturlisten referencer til artikler og ressourcer, der beskæftiger sig isoleret med nogle af de mere specifikke teknologier, teknikker og metoder, der indgår i systemet. F.eks. er der artikler, der beskriver metoder til simulering af et rums akustik, artikler vedrørende interpolering af HRTF, web-ressourcer der forklarer aspekter af arbejdet med rumlig rotation … med mere. I løbet af opgaven vil læseren naturligvis også møde specifikke referencer til disse ressourcer.

[16] Zotkin, Duraiswami, og Davis.

Udviklingen af systemet – valg og udfordringer

Dette afsnit i opgaven beskriver nogle af de udfordringer, hovedsagligt af teknisk art, jeg har mødt i forbindelse med udviklingen af systemet, og de valg jeg har truffet i forbindelse med udfordringerne.
Det skal nævnes, at jeg har valgt at bygge systemet op i programmeringsmiljøet *Max*[17] (også kendt som *Max/MSP*) som er et godt værktøj til at bygge eksperimentelle, og eventuelt interaktive, systemer, installationer, artefakter med mere.

Metode til binaural syntese

En helt fundamental udfordring i forbindelse med konstruktionen af systemet har været at finde eller udvikle en passende metode til processering af lydene, således at de opleves som kommende fra en given retning – med andre ord: en metode til binaural syntese. Den grundlæggende strategi i forbindelse med binaural syntese er at foretage en såkaldt *convolution* mellem målte impuls-responser og de lydkilder, der ønskes indkodet/syntetiseret. Dette kan faktisk siges at være selve definitionen på binaural syntese i den gængse opfattelse af fænomenet. Convolution er en meget anvendt teknik indenfor processeringen af såvel lydsignaler som andre typer signaler.

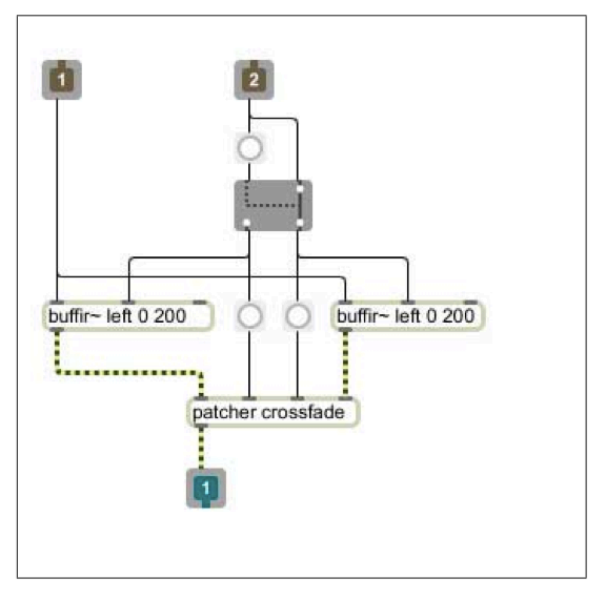
> *Convolution is a mathematical way of combining two signals to form a third signal. It is the single most important technique in Digital Signal Processing. Using the strategy of impulse decomposition, systems are described by a signal called the impulse response. Convolution is important because it relates the three signals of interest: the input signal, the output signal, and the impulse response.*
> [Steven W. Smith: *The Scientist and Engineer's Guide to Digital Signal Processing* – Chapter 6: Convolution (http://www.dspguide.com/ch6.htm)]

Convolution er altså en proces, der kombinerer to signaler til et tredje signal. I lydlig sammenhæng kan man betragte det således, at de to input-lyde bliver 'rullet' sammen, hvorved outputtet bliver en lyd, der rummer egenskaber fra begge lyde. Ofte vil man benytte et systems[18] respons på en slags enheds-impuls (en impuls-respons) som den ene lyd, hvorved man kan simulere den anden lyds udfoldelse i dette system. I sammenhængen vedrørende binaural syntese måles impuls-responserne med små mikrofoner i ørerne på enten et *dummy head*[19] eller på levende testpersoner, og de 'beskriver' altså, hvordan vores anatomi former/indkoder lyd, der kommer fra forskellige målte retninger. Sådanne impuls-responser går under betegnelsen *Head Related Impulse Responses* (HRIR), mens deres måske mere kendte søstre, *Head Related Transfer Functions* (HRTF), er impuls-responserne repræsenteret i frekvensdomænet[20]. I mit sy-

[17] http://cycling74.com/
[18] Et system kan eventuelt være et rum, en effekt-maskine eller som i denne sammenhæng: vejen fra lydkilde til vores to øregange.
[19] En specialdesignet mannekin-hoved og -torso, bygget med henblik på netop binaurale optagelser.
[20] Stern, Wang og Brown – s. 3

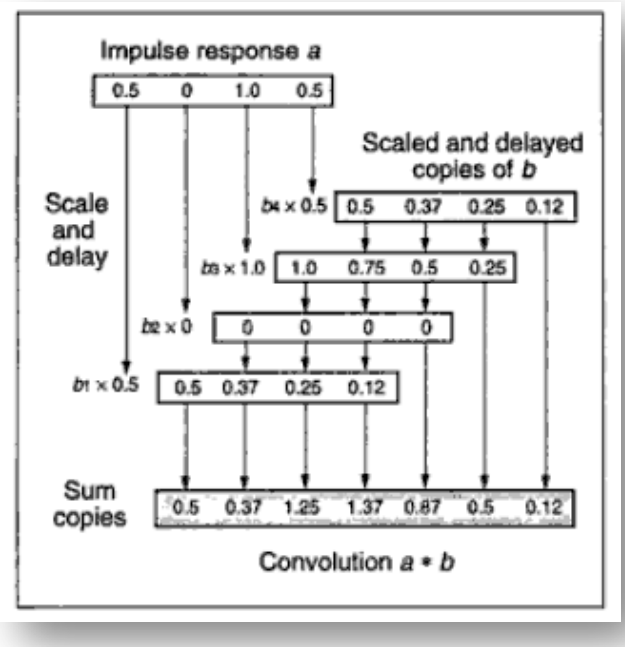
stem har jeg valgt at benytte en frit tilgængelig database af HRIR udviklet af *Center for Image Processing and Integrated Computing* (CIPIC) ved *University of California, Davis*[21]. Databasen indeholder målinger fra 43 forskellige levende test-personer samt af ét *KEMAR dummy head*, og på trods af at den kaldes *The CIPIC HRTF Database,* består den af impuls-responser i tidsdomænet – altså egentlig HRIR og ikke HRTF[22]. Opsummerende kan vi altså konstatere, at vi både har en grundlæggende opskrift (convolution mellem impuls-responser og lydkilder) og tilgængelige ingredienser (impuls-responserne fra CIPIC-databasen) til den binaurale syntese. Hvad vi imidlertid mangler er at føre opskriften ud i livet, hvilket kan gøres på mange måder.

Den oplagte, og lettest tilgængelige, måde at realisere opskriften på er (med de forhåndenværende ingredienser) simpelthen, at foretage convolution mellem impuls-respons og lydkilde i tidsdomænet, da vi jo både har impuls-responserne og vores lydkilder repræsenteret i dette domæne. Det er denne tilgang, vi ser i relaterede projekter som eksempelvis *Head in Space*[23] og *Binaural Tools*[24]. Disse projekter, der ligeledes er bygget i *Max*, benytter *Max*-objektet *buffir~* til at foretage convolution mellem lydkilde og de 200 samples lange impuls-responser fra CIPIC-databasen.



Figur 6. Figur, der viser hvordan den binaurale syntese realiseres i *Head in Space* -projektet. Taget fra artiklen *Head in Space - a Head tracking Based Binaural Spatialization System*[25]. Bemærk at man i projektet har fundet det nødvendigt, at arbejde med to parallelle convolutions (imellem hvilke der *cross-fades*) for hver kanal (venstre og højre), med henblik på at opnå glidende overgange fra én impuls-respons til den næste.

[21] http://interface.cipic.ucdavis.edu/sound/hrtf.html
[22] HRIR og HRTF er naturligvis meget nært beslægtede, og ofte benyttes betegnelsen HRTF om dem begge, hvilket gør sig gældende her.
[23] Ludovico, Mauro og Pizzamiglio samt http://sites.google.com/site/dariopizzamiglio/projects/head-in-space
[24] http://www.ece.ucdavis.edu/binaural/binaural_tools.html
[25] Ludovico, Mauro og Pizzamiglio.



Figur 7. Illustration af convolution-beregning i det digitale tidsdomæne (af Roads kaldet *direct convolution*), taget fra Roads: *The Computer Music Tutorial*, s. 423.

Den, i øjeblikket, nok største problematik vedrørende denne tilgang er, at convolution i tidsdomænet er en proces, der kræver mange beregninger, og derfor beslaglægger en stor del af computerens processorkraft. *Figur 7* viser, hvordan denne beregning foregår i det digitale, tidslige domæne. I vores tilfælde kommer udregningen til at se således ud: *hver* indkommende sample (44100 i sekundet) ganges med *hver* sample i impuls-responsen (200) og lægges sammen med (199 andre) forsinkede værdier. Denne udregning foretages *to* gange (for at opnå glidende overgange) for hvert øre, dvs. *fire* gange i alt. Dette løber op i godt og vel 70 millioner (!) beregninger i sekundet pr. lydkilde, der skal processeres. Dette er, måske overraskende nok, ikke det store problem for nutidens computer – i hvert fald så længe vi arbejder med én eller få lydkilder ad gangen. Men hvis vi ønsker friheden til at have mange (eller i hvert fald flere) lydkilder i spil, sætter denne strategi sine begrænsninger. Derfor er det ikke den oplagte løsning i til mit system.

Artiklen *Real-time, Head-tracked 3D Audio with Unlimited Simultaneous Sounds*[26] beskriver et lignende projekt, hvor man, ganske smart, har løst problematikken ved at processere alle lydene *offline* – altså på forhånd. Således bliver hele lydbilledet *renderet* ned til ét stereospor for hver potentiel lytter-orientering. Dette åbner på den ene side muligheden for et ubegrænset antal lydkilder, men samtidig lukker det muligheden for et mere dynamisk/interaktivt lydmiljø, herunder det at lydmiljøet tilpasser sig varierende lytter-*positioner*. Desuden stiller det store krav til læsehastigheden på computerens harddisk / begrænser potentielle lytter-orienteringer. Disse temmelig væsentlige begrænsninger gør, at jeg ligeledes har fravalgt denne metode med *offline* processering i mit projekt.

Men hvad så?

En løsning på convolution-processens beregningstyngde kan være at foretage beregningen i frekvensdomænet. Dvs., at lydkilder og impulsresponser først konverteres fra en repræsentation i tid (én sample = lydtrykket på det pågældende tidspunkt) til en beskrivelse af frekvens-indhold i små tidsbidder (én sample = mængden af energi i det pågældende frekvensbånd samt frekvensens forskydning i fase) – en sådan konvertering kaldes en *Fourier transformation* (opkaldt efter Jean Baptiste Jeseph, Baron de Fourier, 1768-

[26] Jin, Tan, Leung, Kan, Lin, Van Schaik, Smith, m.fl

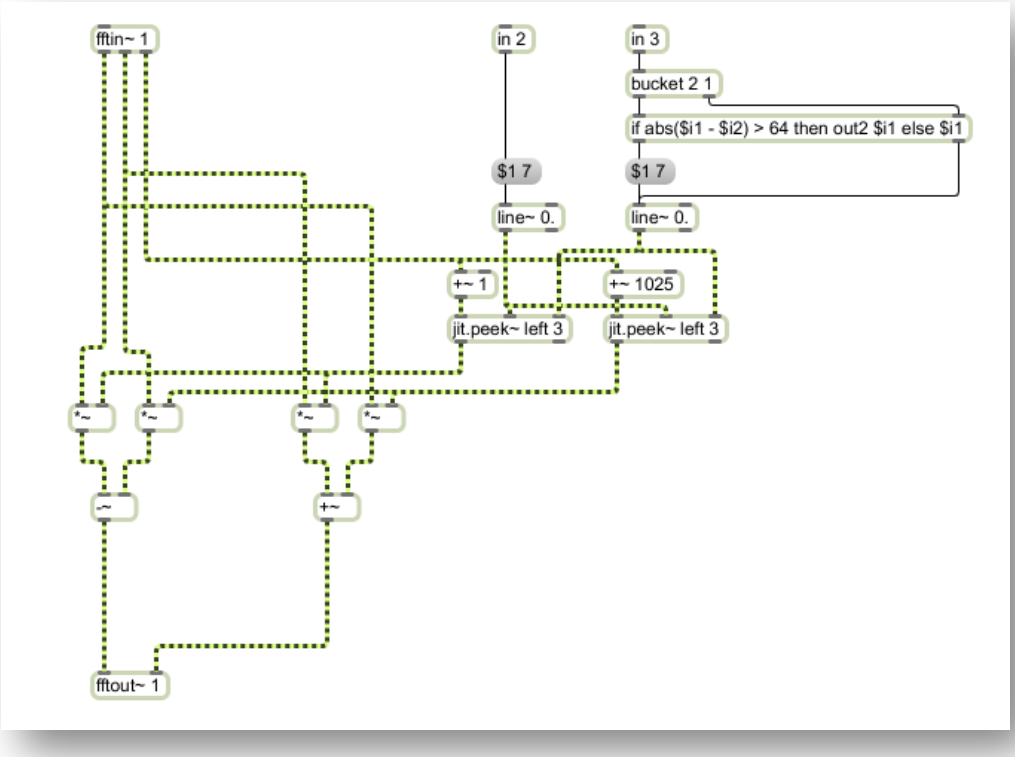
1830)[27]. Når lyden er repræsenteret i denne form, svarer en simpel gange-operation til convolution i tidsdomænet.

> *Convolution in the time domain is equal to multiplication in the frequency domain and vice versa.*
> [Curtis Roads: The Computer Music Tutorial, s. 424]

Dermed kan man ofte spare en masse beregninger ved først at konvertere lyden til frekvens-domænet via algoritmen *fast Fourier transform* (FFT), siden udføre gange-operationen (= convolution) og til sidst konvertere lyden tilbage til tidsdomænet. Derfor er det ofte denne tilgang, man ser i forbindelse med *real-time* convolution som i eksempelvis rumklangseffekter. Det tidligere beskrevne, lignende projekt, *Rendering Localized Spatial Audio in a Virtual Auditory Space*[28], benytter netop den FFT-baserede strategi.
Der er dog et par ting, man bør have *in mente* i forbindelse med et eventuelt valg af denne strategi. For det første skal det nævnes, at strategien kun 'betaler sig', når impuls-responserne er længere end et vist antal samples [29]. For det andet ligger der i FFT-algoritmen et iboende *trade-off* imellem tidslig præcision og frekvensmæssig præcision. Dette i kraft af at algoritmens parameter, *FFT size*, både er målet for, hvor store tids-bidder (antal samples) vi analyserer ad gangen, og samtidig er målet for analysens opløsning i frekvensbånd (antal såkaldte *FFT bins*)[30]. Altså, med en lille *FFT size* er den tidslige opløsning stor (mange bidder pr. tid), men den frekvensmæssige opløsning lille (få *FFT bins*), og omvendt med en stor *FFT size*.
En sidste ting man bør have sig for øje er, at der med FFT-processeringen introduceres en lille forsinkelse af signalet, igen fordi FFT-algoritmen arbejder med opdeling af lyden i små bidder.
Disse forbehold til trods var det alligevel denne FFT-baserede strategi, jeg valgte til projektet, i en forventning om, at det ville åbne muligheden for at processere relativt mange lydkilder ad gangen med en tilfredsstillende lydlig og oplevelsesmæssig kvalitet (= realistisk/virkelighedstro).
Overordnet set kan jeg beskrive min metode til binaural syntese som følgende:
Impuls-responserne konverteres på forhånd i programmet *Matlab*[31] til frekvensdomænet (via *fast Fourier transform* med en *FFT size* på 2048), således at dette ikke skal gøres realtime, hvilket sparer processorkraft. Afviklingen og processeringen foregår i *Max,* hvor lyden fra hver lydkilde konverteres, *real-time* via FFT, til frekvens-domænet og ganges med de relevante HRTF'er (i henhold til lydkildens placering i forhold til lytterens position og orientering). Herefter konverteres resultatet tilbage til tidsdomænet.
I *Max* benytter jeg objektet *pfft~,* som selv håndterer overlapning og såkaldt *windowing* af de små analy-

[27] Roads – s. 1075
[28] Zotkin, Duraiswami, og Davis
[29] Se eksempelvis: https://ccrma.stanford.edu/~jos/sasp/FFT_versus_Direct_Convolution.html
[30] Se eksempelvis http://www.cycling74.com/docs/max5/tutorials/msp-tut/mspchapter26.html
[31] CIPIC HRTF -databasen kommer som *Matlab*-filer. *Matlab*: http://www.mathworks.se/products/matlab/index.html

serede/syntetiserede tidsbidder (*FFT size*). Jeg benytter et overlap på 4 vinduer, i håbet om at sikre en tilfredsstillende høj kvalitet af syntesen i frekvensdomænet.



Figur 8. Indholdet af den *pfft*~-patch, jeg bruger i *Max* (i dette tilfælde for venstre øre). Her ses den relativt simple gange-operation i frekvens-domænet, der erstatter convolution i tidsdomænet. Det kan godt være at det *ser* mindre simpelt ud end *Figur 6*, men fra computerens synspunkt er det lettere at udføre processeringen i denne figur end den i *Figur 6*. *Jit.peek*~-objekterne henter relevante HRTF'er fra de globale *jit.matrix*-objekter *left* og *right* (i dette tilfælde, *left*), som indeholder HRTF'er for alle retninger for et givent *subject*.

Pfft~-objektet suppleres af et *delay*~-objekt, som (gen)implementerer den tidsforskydning, som impuls-responserne er blevet 'frarøvet' i forbindelse med interpolering. Mere om dette i næste afsnit, *Interpolering mellem impuls-responserne*. HRTF'erne for alle potentielle retninger (for et givent *subject* i CIPIC-databasen) indlæses i to globale *jit.matrix*-objekter – *left* og *right*. I processeringen af de enkelte lydkilder (dvs. i den enkelte *pfft*~-instans) 'peges' blot på den relevante data i disse *jit.matrix*-objekter.

For at finde ud af, hvorvidt der rent faktisk *er* en fordel, hvad angår processorkraft, i at processere lyden i frekvensdomænet, har jeg foretaget en sammenligning mellem den binaurale syntese i mit system (som foretager convolution i frekvensdomænet) og den binaurale syntese i projektet *Binaural Tools*[32] (som foretager convolution i tidsdomænet). Ifølge *Max*' måling af benyttet processorkraft til digital signalbehandling benytter mit system ca. 3% af computerens processorkraft til at foretage binaural syntese på *en* lydkilde, mens processeringen i *Binaural Tools* benytter ca. 9%. Der lader altså til at være en klar fordel,

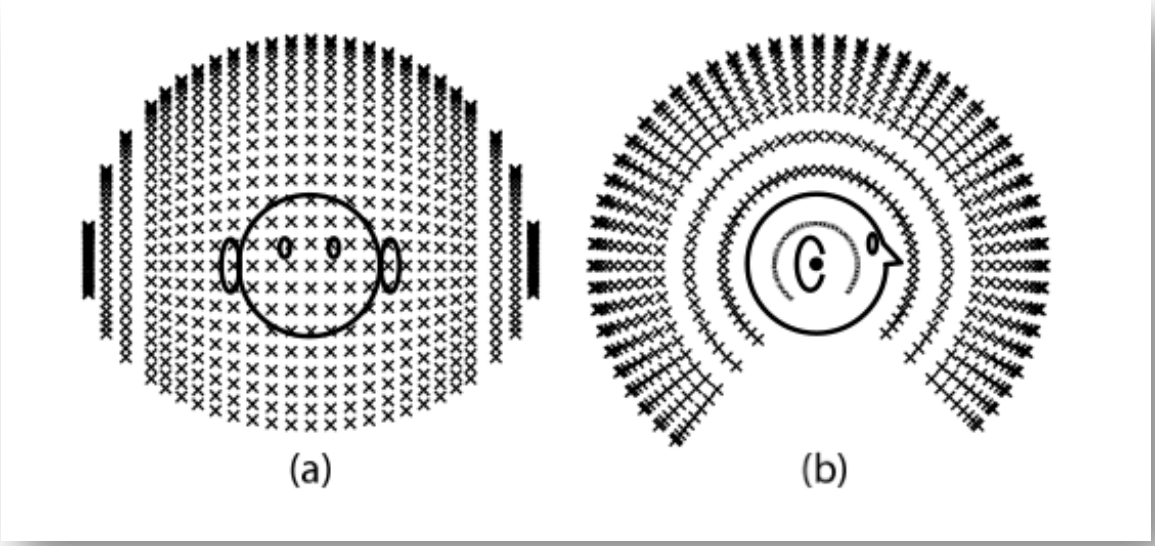
[32] http://www.ece.ucdavis.edu/binaural/binaural_tools.html

hvad angår processorkraft, ved at foretage processeringen i frekvensdomænet. Hvorvidt der er en *kvalitativ* forskel i den oplevelse de to processeringsmetoder skaber hos lytteren, har jeg ikke undersøgt med nogen videnskabelig gyldighed. Jeg har ikke selv bemærket nogen kvalitativ forringelse ved den FFT-baserede teknik, men jeg har heller ikke fortaget en direkte og systematisk sammenligning. Man kunne måske have en forventning om, at den FFT-baserede strategi, i kraft af sine *trade-offs*, medfører en vis forringelse, som måske, måske ikke, har en mærkbar negativ indflydelse på oplevelsen af realisme. En sådan sammenligning hører ind under kategorien relevant fremtidigt arbejde.

For at andre også eventuelt kan få gavn af /arbejde videre på de udviklinger, jeg har lavet mht. den binaurale syntese, har jeg lagt en version af denne grundfunktionalitet op på nettet til fri afbenyttelse og videreudvikling. Dette, på *cycling'74*s *tool*-side: http://cycling74.com/toolbox/fft-based-binaural-panner/

Interpolering mellem impuls-responserne

Som nævnt benytter jeg CIPIC-databasen, som altså består af impuls-responser (lydfiler, om man vil), der angiver, hvordan vores øregange 'hører' lyde kommende fra forskellige retninger. Disse responser er optaget på baggrund af lydimpulser afspillet fra 1250 forskellige positioner rundt om hver test-person (43 personer).



Figur 9. Viser de 1250 måle-positioner i CIPIC-databasen. Som det ses, er der tale om halvcirkel, der er flyttet gradvist rundt om lytteren. Der er dog en vinkel under lytteren, hvori der ikke findes målinger.

Selv om 1250 positioner lyder af mange, så er det ikke flere, end at jeg har haft behov for at lave en mere finkornet version. Dette behov er i høj grad dikteret af behovet for at kunne bevæge lydkilder gliden-

de i henhold til lytterens glidende hovedbevægelser[33]. I denne sammenhæng er problematikken: hvordan finder man frem til en mellemting mellem to lydfiler, som godt nok har meget til fælles, men som også kan siges at rumme forskelle i både tid og amplitude? Problematikken vedrørende interpolering af netop disse HRIR/HRTF er et emne, man kan finde overraskende mange videnskabelige artikler om. (Eksempelvis *Perceptual Consequences of Interpolating Head-Related Transfer Functions During Spatial Synthesis*[34], *On The Minimum-Phase Approximation of Head-Related Transfer Functions*[35] og *Frequency-Domain Interpolation of Empirical HRTF Data*[36]). Omfanget af videnskabelig interesse afspejler også, at det ikke er en helt simpel problematik med en entydig løsning. Tværtimod kan man 'løse problemet' på mange forskellige måder – nogle mere komplekse end andre. En del af artiklerne beskæftiger sig med en såkaldt *minimum phase* repræsentation af impuls-responserne. En sådan *miminum phase* repræsentation skulle være frugtbar med henblik på at interpolere mellem disse HRIR/HRTF. De omtalte artikler er dog som regel henvendt til et ingeniør-segment og bruger ofte en meget teknisk terminologi, der forudsætter en del viden om signal-processering og den matematiske beskrivelse af denne. Derfor er det, for at være ærlig, også begrænset, hvor meget jeg forstår. For at få en hjælpende hånd med at vælge en metode til interpolering i min konkrete sammenhæng, skrev jeg til V. Ralph Algazi, som er *Research Professor* ved *University of California, Davis* og har været med til at lave CIPIC HRTF -databasen, som jeg benytter. Ifølge Algazi er en acceptabel interpoleringsmetode i min sammenhæng følgende[37]:

Først *align'*er man impuls-responserne i tid (og gemmer den enkelte respons' tidsforskydning), således at de starter samtidigt. Dernæst interpoleres hver sample lineært, og til sidst kan de enkelte impuls-responser flyttes tilbage til der, hvor de var, før de blev *align'*et i tid.

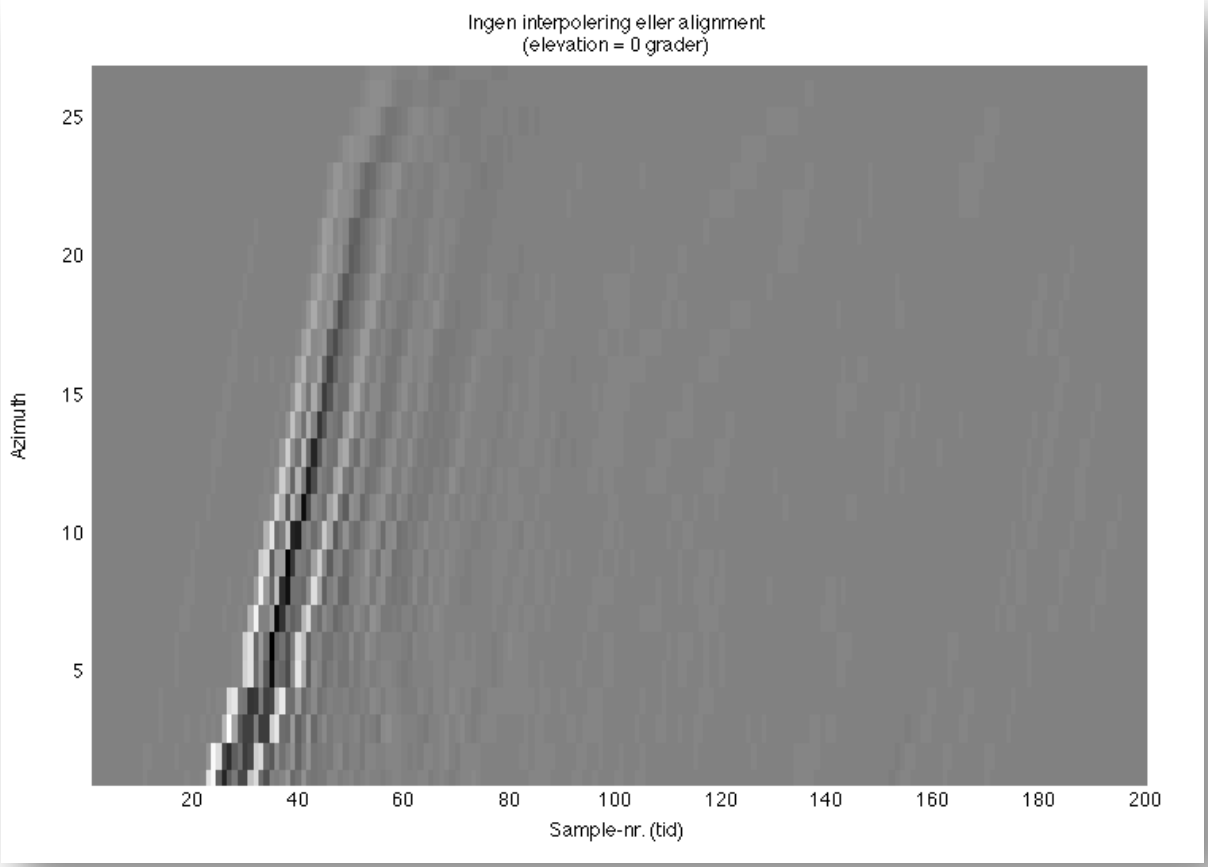
Siden denne mail-korrespondance har jeg faktisk fundet et *Matlab-script*, der medfølger CIPIC-databasen, som udfører denne nævnte interpolering. Dermed har jeg blot skullet implementere dette script samt tilpasse det en anelse til mit specifikke behov[38].

Inspireret af artiklen *Head-Related Impulse Response Interpolation in Virtual Sound System*[39] har jeg valgt at lade være med at flytte impuls-responserne tilbage til deres 'præ-timealign-position' og i stedet gemme tidsforskydningerne separat. Det smarte i at have tidsforskydningerne separat er, at det giver mulighed for at interpolere *real-time* på dette parameter. Hvis tidsforskydningen for et givent signal eksempelvis er 14 samples ved én hovedposition og 20 samples ved den næste, kan man lade sit implementerede delay

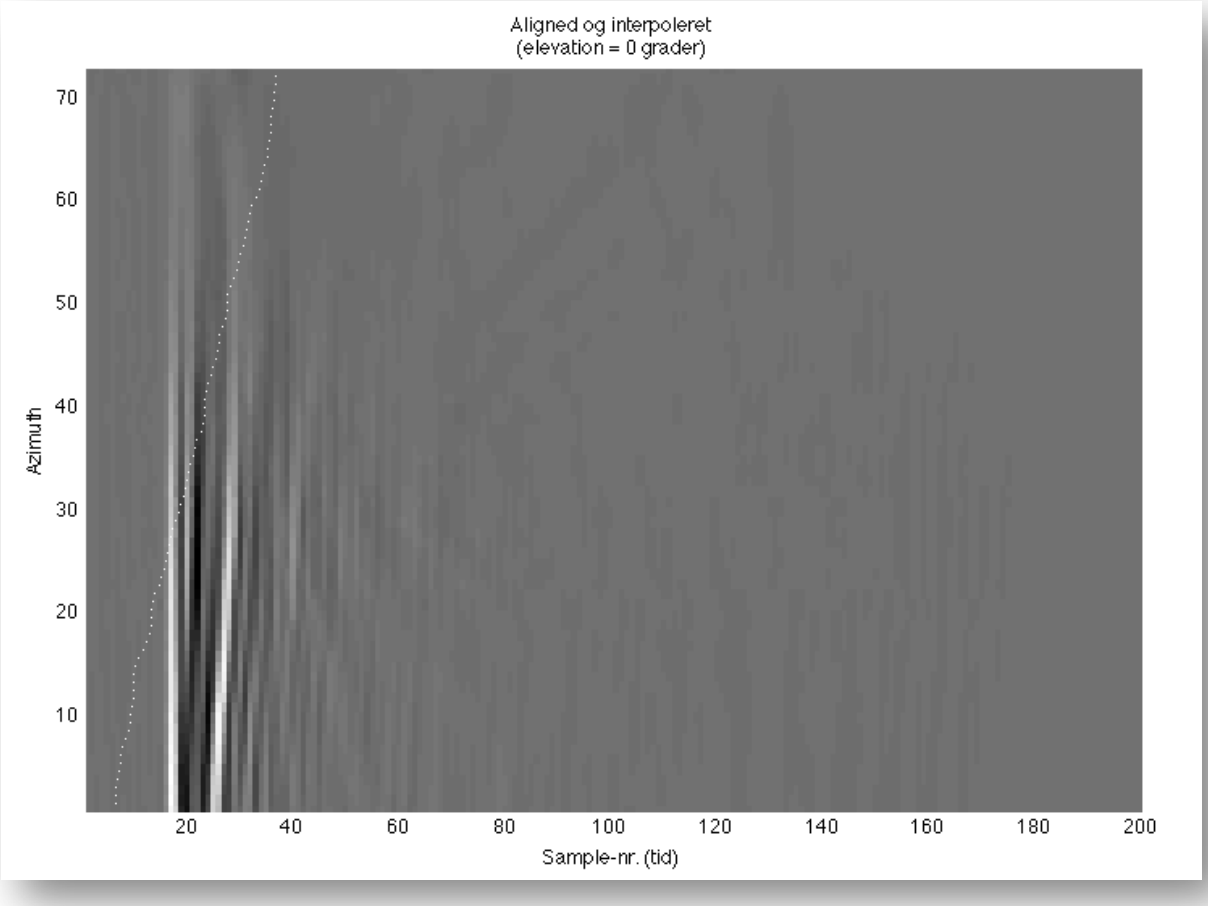
---

[33] Det skal nævnes at *pfft~*-objektet i sig selv foretager en form for udglatning mellem responserne i kraft af sin *windowing*-funktionalitet. Men dette skaber ikke flere statiske punkter.
[34] Wenzel og Foster
[35] Kulkarni, Isabelle og Colburn
[36] Carty og Lazzarini
[37] Se evt. bilag 6 – *Mailkorrespondance med Algazi*
[38] Se evt. bilag 4 – *Matlab-script (interpolation, fft, and export)*, som er det script der både sørger for interpolering, konvertering via FFT og eksport til tekstfiler.
[39] Chen, Hu og Wu

(i mit tilfælde *delay~*-objektet i *Max*) foretage dette skift i en glidende bevægelse, frem for at springe direkte mellem impuls-responer indeholdende statiske delays. Dette mindsker også risikoen for hørbare *clicks* ved hurtige hovedbevægelser, og måske mere vigtigt: det sikrer, at disse perceptuelt meget betydningsfulde tidsforskydninger (*Interaural Time Difference* eller *ITD*) ikke bliver 'tværet ud' i FFT-konverteringen – jf. at jeg benytter en relativt stor *FFT size*.



Figur 10. Grafisk repræsentation af en række af impulsresponserne (altså stadig i tidsdomænet) for *subject 3* i CIPIC-databasen, lavet i *Matlab*. Der er her hverken foretaget interpolering eller *algnment* – dog har jeg tilføjet 2 responser, som er estimeret til at repræsentere stik vestre og stik højre (jf. at databasen kun indeholder målinger fra -80 til + 80 grader *azimuth*). I dette eksempel ses hvordan, responserne for venstre øre ændrer sig gradvist, når lydkilden bevæges, på en vandret halvcirkel, over mod højre øre. *Azimuth* 1 betyder lige ud for venstre øre, mens *Azimuth* 27 er lige ud for højre. Denne repræsentation giver et meget godt billede af, hvordan både tidsforskydningen og lydtrykket/amplituden ændrer sig, når lydkilden flyttes over mod det modsatte øre. Lydtrykket er angivet ved gråtone-skalaen.



Figur 11. Her er impuls-responserne fra *Figur 10* først blevet *alignet* i tid og siden interpoleret – bemærk at de 27 responser fra *Figur 10* er blevet til 73, altså en mere finmasket version. Tidsforskydningerne, som impulserne er blevet frarøvet, gemmes separat til genimplementering i afviklingssituationen. På denne figur er tidsforskydningerne, som også er blevet interpoleret, repræsenteret ved den stiplede hvide linje. Impuls-responserne interpoleres både *azimuth-wise* og *elevation-wise*, og herefter er impuls-responserne klar til at blive konverteret til frekvens-domænet.

En mindre udfordring i forbindelse med disse HRTF har været at få disse data over i *Max* fra *Matlab*. Dette har jeg gjort ved først at eksportere al interpoleret og FFT-konverteret data til tekstfiler – over 400mb tekst pr *subject* i CIPIC-databasen! – ikke så elegant, men det virker. Dernæst læses disse tekstfiler ind i *Max*, hvor impuls-responserne (i frekvens-domænet) og deres tilhørende overordnede tidsforskydning, som tidligere nævnt, skrives ind i to *Jitter*-matricer (*jit.matrix*), *left* og *right*. I *Max kan Jitter*-matricerne gemmes til en fil, så tekstfilerne ikke længere er nødvendige. I afviklingssituationen benyttes objektet *jit.peek~* til at pege på den relevante HRTF for en given lydkildes position i forhold til lytteren.

Interpoleringen lader generelt til at virke efter hensigten. Dog er der stadig en problematik vedrørende den vinkel (se *Figur 9*), hvori databasen ikke indeholder målinger. Jeg har forsøgt at lave en interpolering mellem yderpunkterne i dette område, men der er tydeligvis nogle problemer med udfasninger og lydlige artefakter hvilket egentlig ikke er underligt. I fremtidige udviklinger af systemer som mit, er det der-

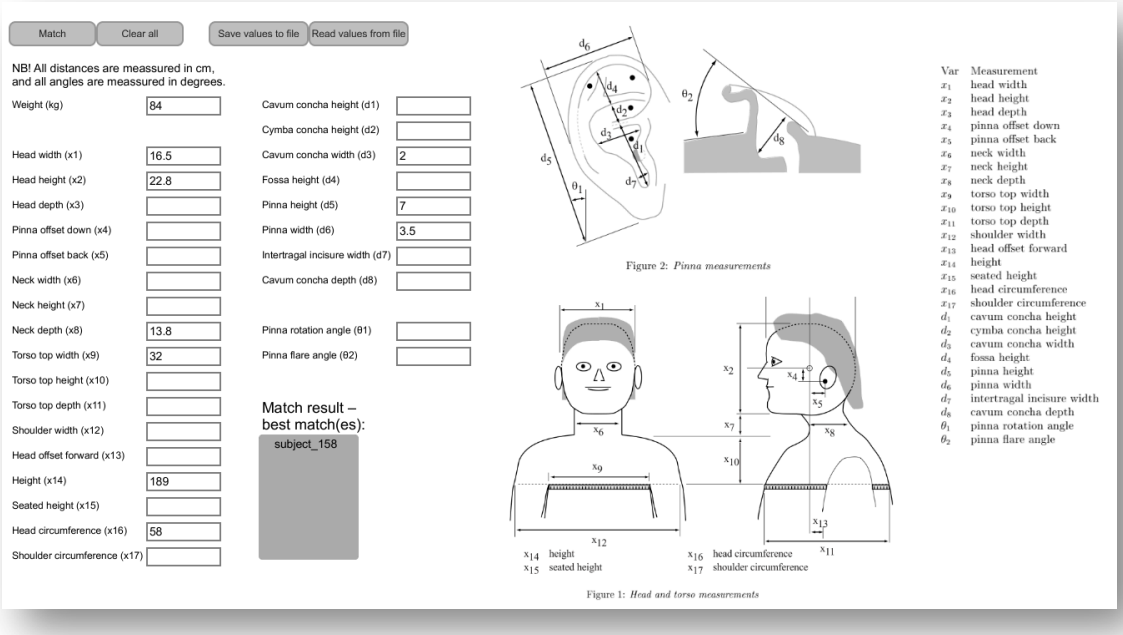
for ønskværdigt at benytte HRTF-databaser, som indeholder målinger i den del af den manglende vinkel, hvor det er fysisk mulig at foretage målinger. Her tænker jeg især på området ud mod siderne, skråt ned for lytterens ører. Når dette er sagt, er jeg dog meget taknemmelig for, at CIPIC har gjort deres database frit tilgængelig til gratis afbenyttelse.

At finde den bedst egnede HRTF

Som nævnt indeholder CIPIC-databasen impuls-responser optaget med små mikrofoner i ørerne på 43 forskellige personer (*subjects*). Grunden til at man har valgt at optage så mange er, at der angiveligt er ret stor forskel på forskellige personers HRIR/HRTF (i det følgende, blot kaldet HRTF). Dette skyldes forskellene i vores fysiske udformning – vi er alle forskellige.

> *It is known* […] *that differences in ear shape and geometry strongly distort perception and that the high-quality synthesis of a virtual audio scene requires personalization of the HRTF for the particular individual for good virtual source localization.*
>
> [Zotkin m.fl. : Rendering Localized Spatial Audio in a Virtual Auditory Space]

I CIPIC-databasen er også inkluderet fysiske mål på alle *subjects* – både på øret og på hoved og torso. Præcis hvilken indflydelse de forskellige enkelte anatomiske variationer har på HRTF, vides ikke. Dog kan man rimeligvis antage, at jo større fysisk, anatomisk lighed der er mellem to personer, jo mere vil deres HRTF ligne hinanden. Derfor gælder det, i min sammenhæng, om at finde det/den *subject* i CIPIC-databasen, der minder mest om lytteren. Til dette formål har jeg forsøgt at lave et lille program, som mathcer indtastede lytter-mål med databasens *subjects*.



Figur 12. Screenshot af programmet (*Max*-patchen), der matcher lytter-mål med *subjects* i CIPIC HRTF -databasen.

Programmet kommer altså med et bud på, hvilke(n) HRTF det vil være mest hensigtsmæssigt for lytteren at bruge. Ideelt set burde lytteren naturligvis benytte et sæt af HRTF, der er optaget på lytteren selv, men den næstbedste løsning må være at benytte et sæt HRTF fra en person, der minder om lytteren. Det er ikke en helt enkel og entydig beregning at vurdere hvilket *subject*, der minder mest om lytteren. For eksempel er der sandsynligvis forskel på, hvor stor betydning de enkelte anatomiske mål har for HRTF. Derfor har jeg forsøgt at lave en vægtning for hvert mål. Denne vægtning er på ingen måde videnskabeligt begrundet, men er blot et skøn fra min side. Dog håber jeg, at programmet kan være en hjælp til at finde et egnet match.

Programmets udregning foretages som følger:

For hvert *subject* i databasen matches hvert indtastet lytter-mål med *subjectets* mål og en afvigelseskvotient bestemmes. Denne afvigelseskvotient beregnes som forskellen mellem lytter-mål og *subject*tets mål divideret med målets standardafvigelse (standardafvigelserne er angivet i databasens dokumentation) og til sidst ganget med en vægtning af dette måls vigtighed. For hver database-*subject* summeres alle gyldige afvigelseskvotienter og herefter divideres med summen af gyldige vægtninger. På denne måde findes frem til et overordnet afvigelsesmål for hvert *subject* i databasen. *Subjectet* med det laveste afvigelsesmål er programmets bud på det bedste match. Programmet er essentielt set en *Max*-patch, som foretager udregningen ved hjælp af en *Java external*, jeg har kaldt for *HRTFSubjectMatcher*[40]. I forbindelse med udarbejdelsen af programmet er jeg blevet inspireret af artiklen, *Rendering Localized Spatial Audio in a Virtual Auditory Space*, som skitserer, hvordan man i det projekt har valgt at udregne det bedste match.

Realisering af *head tracking*

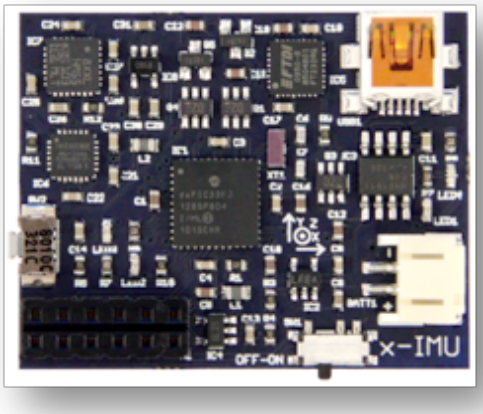
Der findes forskellige metoder til at bestemme lytterens hoveds orientering. *Head in Space* -projektet benytter, som beskrevet, en form for kamera-*tracking* af lytterens ansigt. Jeg har dog valgt at basere *head tracking*en i mit projekt på sensorteknologi. Dette med henblik på at opnå stabil og robust *tracking*.

Der findes forskellige sensorer på markedet. I projekter, der ligner mit, benyttes ofte sensorer produceret af firmaet *Polhemus*[41] [42]. Hvor disse sensorer fra *Polhemus* sandsynligvis er gode og præcise, er de samtidig temmelig bekostelige – et system ligger gerne på den gode side af 20000kr! I min research stødte jeg også på den såkaldte *AmmSensor*[43], som kan kommunikere trådløst og er noget billigere, men dog stadig udenfor min økonomiske ramme. Jeg skrev derfor til *Amm Technologies* med henblik på at overtale dem til at sponsorere en sensor til mit projekt[44]. Da de imidlertid ikke var interesserede i dette, mål-

[40] Se evt. bilag 3 – Java-kode
[41] http://polhemus.com/
[42] Eksempelvis Zotkin, Duraiswami, og Davis samt Sandvad
[43] http://www.ammsensor.com/
[44] Se evt. bilag 5 – Project description and application for Amm.

te jeg finde en anden løsning. I min videre research stødte jeg så på *x-IMU* [45], der ligeledes er i stand til at kommunikere trådløst, og som koster i omegnen af 2250kr. Den bliver produceret af det lille, nyopstartede, engelske firma *x-io Technologies*. Foruden den betydeligt lavere pris gav produktet og firmaet et indtryk af åbenhed og en eksperimentel tilgang, som tiltalte mig. På trods af at Seb Madgwick, Ph.D-studerende ved Bristol Universitet og manden bag firmaet, bestemt ikke var afvisende overfor idéen om at låne mig en *x-IMU*, valgte jeg at investere i en – bl.a. fordi jeg ikke følte, at jeg kunne love ham at producere materiale til brug på firmaets hjemmeside.

En udfordring i denne forbindelse har været at få *Max* til at kommunikere med *x-IMU*. I virkeligheden er *x-IMU* ikke *en* sensor, men rettere en slags *device* (en *Inertial Measurement Unit* eller *IMU*), som kombinerer data fra forskellige sensorer (gyroskop, accelerometer, magnetometer og termometer), og som på baggrund af denne kombination er i stand til at angive sin egen orientering i det tredimensionelle rum.



Figur 13. *x-IMU*

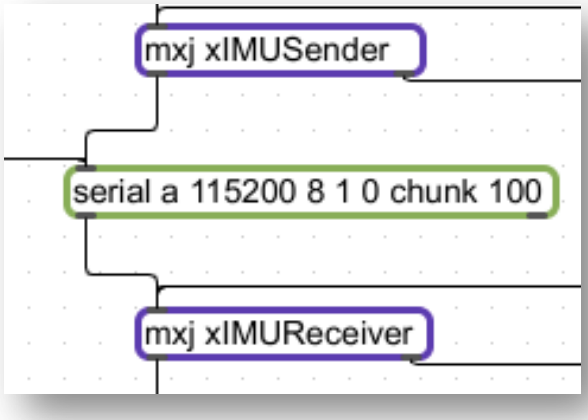
X-IMU kommunikerer med computeren via en seriel port og kan forbindes enten via et usb-kabel eller trådløst via bluetooth. Til *x-IMU* hører et program (*x-IMU GUI*), med hvilket man kan kommunikere med *devicen*, men dette hjælper mig ikke med at få data ind i *Max*. *Max* indeholder et objekt kaldet *serial,* som er i stand til at modtage den strøm af bytes, som *x-IMU* sender, og det kan ligeledes sende bytes til *x-IMU*. Problematikken i denne sammenhæng ligger i at trække en mening ud af de bytes, som *x-IMU* sender, samt at konvertere en given instruks til bytes, som *x-IMU* forstår. Eksempelvis kan det umiddelbart være svært at trække mening ud af *255 6 25 18 127 55 10 235 9 87 16 1 220 7 123 0* – måske er det en angivelse af temperatur, måske orienteringsdata eller muligvis en fejlmeddelelse af en eller anden art. Heldigvis er der hjælp at hente i *x-IMU API*'en[46], som er en samling af C#-kode, der viser, hvordan man kan kommunikere med *x-IMU*. På baggrund af *x-IMU API*'en har jeg konstrueret to Java-klasser til brug som objekter i *Max* (*Java externals*): *xIMUReceiver,* som tolker data sendt fra *x-IMU, og xIMUSender,* som er i stand til at indkode givne instrukser til det byte-sprog, *x-IMU* forstår.

Det har dog ikke været helt problemfrit at konvertere C#-koden til Java. Dels fordi jeg ikke er programmør, og dels fordi der åbenbart er forskel på datatypen *byte* i C#, *Max*, og Java. Jeg vil ikke påstå, at jeg forstår denne forskel til fulde, men en grov forståelse kan formidles således: en byte i C# og *Max* er *unsigned*, dvs. at den kan repræsentere en værdi fra 0 til og med 255 (jf. at 1 byte = 8 bit = 2^8 mulighe-

[45] http://www.x-io.co.uk/node/9
[46] http://www.x-io.co.uk/res/sw/ximu_api_13_1.zip

der = 256), mens en byte i Java er *signed*, dvs. at den første bit benyttes til at angive positiv eller negativ. Dermed kan en byte i Java repræsentere en værdi fra -128 til og med 127. Man skulle ikke synes, at denne forskel var så vanskelig at håndtere, men ikke desto mindre har jeg bøvlet en del med det. Dog er det lykkedes mig at få Java-objekterne til at virke, om end det sandsynligvis ikke er særlig 'kønt' rent programmeringsmæssigt. Jeg har en løs aftale med Seb Madgwick, om at jeg sender ham mine *Max*-objekter med tilhørende Java-kode[47], så andre muligvis kan få gavn af dem. Det vil jeg gøre i den nærmeste fremtid.

Figur 14. *Java-externals xIMUSender* og *xIMUReceiver* i brug i *Max*.



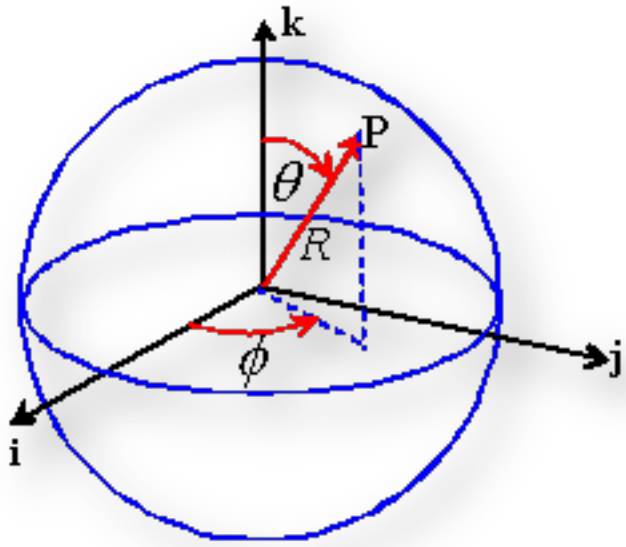
Figur 15. Hovedtelefoner med *x-IMU* i en lille plastik-kasse, elegant monteret med gaffa-tape og strips. Selvom *x-IMU* i princippet kan kommunikere trådløst via bluetooth, har jeg ikke kunnet få dette til at virke stabilt, og har derfor forbundet via USB. Ideelt set burde hovedtelefonerne også være trådløse. Måske ikke just det, man vil betegne som et transparent interface, men det fungerer ok som prototype.

[47] Se evt. bilag 3 – Java-kode

Rumgeometri – retning og afstand

Hvilken HRTF er det lige nu relevant at benytte til processeringen af en given lydkilde?

For at kunne besvare dette spørgsmål, må vi se på tre parametre: lydkildens position, lytterens position og lytterens orientering. Alle disse tre parametre har en afgørende betydning i denne sammenhæng, og systemet må derfor inddrage dem i dets *real-time*-beregning af relevant HRTF. Lad os i første omgang ser bort fra lytterens orientering og betragte forholdet mellem lytter og lydkilde som en vektor gående fra lytter til lydkilde. Hvis lytteren eksempelvis befinder sig i punktet *a*(2,4,1) og lydkilden i punktet *b*(5,2,2), har vi en lytter-lydkilde-vektor *v*(3,-2,1), *b* minus *a*. Vektor *v* peger altså på lydkilden ud fra lytterens perspektiv, og det er jo netop dette perspektiv, vi er interesserede i. For at finde frem til den relevante HRTF blandt de interpolerede og FFT-konverterede impuls-responser fra CIPIC-databasen må vi omtolke denne vektor *v* til polære koordinater[48] – jf. at målingerne i databasen er struktureret via sfæriske polære koordinater (*azimuth*, *elevation*). Bemærk, at vi her er ligeglade med, hvor lang vektoren er – vi er i første omgang blot interesserede i, i hvilken retning den peger. Når vektorens retning er omtolket til sfæriske polære koordinater, skal disse blot skaleres og afrundes for at pege på en given kombination af de 73 mulige *azimuths* og 128 mulige *elevations* i de interpolerede (og FFT-konverterede) versioner af databasen.



Figur 16. Figur taget fra http://www.engin.brown.edu/courses/en3/Notes/Vector_Web2/Vectors6a/Vectors6a.htm , der viser forholdet mellem en vektor, *P*, og de polære koordinater, *θ* og *ϕ*. De polære koordinater er her ikke specificeret helt på samme måde som i CIPIC-databasen, men det grundlæggende princip er det samme. Man kan måske forestille sig, hvordan vektoren peger på en af impuls-responserne illustreret i Figur 9.

I denne udregning har vi imidlertid ikke taget højde for lytterens orientering. Heldigvis behøver vi ikke starte helt forfra for at medregne denne. Vi skal blot, inden vi omtolker vektoren til polære koordinater, rotere vektoren med det omvendte af lytterens rotation – lytterens rotation forstået som den rotation, der skal til for at rotere lytteren fra en 'udgangspunkt-orientering' til lytterens nuværende orientering. Lytterens rotation får vi fra *x-IMU*, og den er specificeret i det, der hedder kvaternioner (på engelsk,

[48] Jeg benytter metoden skitseret her: http://www.engin.brown.edu/courses/en3/Notes/Vector_Web2/Vectors6a/Vectors6a.htm

quaternions). Kvarternioner er egentlig et matematisk talsystem, der er en slags udvidelse af de komplekse tal, og de blev første gang beskrevet i år 1843 af den irske matematiker Sir William Rowan Hamilton[49]. Kvaternioner, eller mere præcist de såkaldte enheds-kvaternioner, er dog også en meget anvendelig måde at beskrive rumlig rotation på, og de har angiveligt en række fordele sammenlignet med eksempelvis *Euler angles* og rotationsmatricer[50]. Dette betyder også, at de er flittigt brugt indenfor mange områder – eksempelvis i computerspil, 3D-visualiseringer, fysik, rumforskning osv.. I min konkrete sammenhæng møder jeg kvaternionerne som en liste bestående af fire tal, som *x-IMU* spytter ud omtrent 128 gange i sekundet[51]. Jeg vil ikke foregive en komplet forståelse for disse fire tals kobling til rumlig rotation, men den, sandsynligvis stærkt forenklede, forståelse, jeg har *googlet* mig frem til, lyder som følger: De tre af tallene angiver en rumlig vektor, som kan betragtes som den akse, omkring hvilken rotationen er foretaget, mens det fjerde tal angiver, hvor meget objektet er roteret omkring den angivne vektor. På denne måde kan alle tænkelige rotationer repræsenteres. Heldigvis behøver man ikke forstå alting for at kunne høste frugten af andres forståelse.

$$\vec{v'} = q\vec{v}q^{-1}$$ [engelsk wikipedia: *Quaternions and spatial rotation*]

Med ovenstående matematiske formel, som beskriver rotation af vektor *v* via kvaternion *q*, kan vi rotere vores lytter-lydkilde-vektor, således at vi tager højde for lytterens orientering. Jf. at vi skal rotere med den *omvendte* rotation af *x-IMU*, hvilket i ovenstående formel resulterer i, at *q* og q^{-1} skal byttes om[52]. Alle disse udregninger varetages i mit system af endnu en *Java-external*, jeg har kaldt for *spaceHandler*. Hver lydkilde har tilknyttet en *spaceHandler*, der, blandt andet, har til opgave at pege på den relevante HRTF for en given kombination af lydkilde-position, lytter-position og lytter-orientering. Derudover specificerer *spaceHandler*'en også hvor meget en lydkilde skal dæmpes/forstærkes og forsinkes i henhold til dens afstand til lytteren.

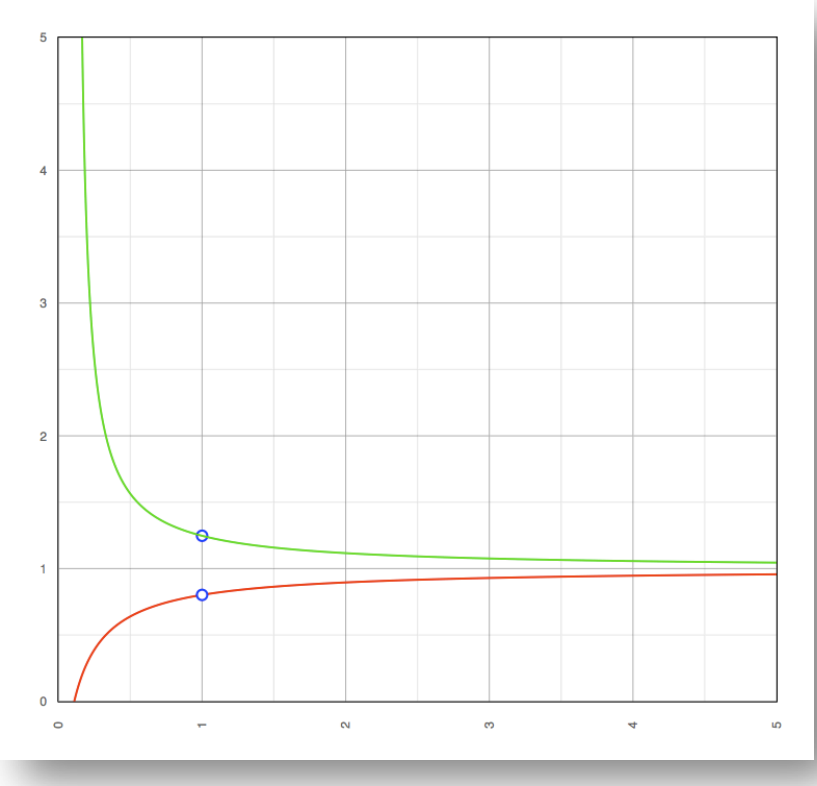
> *The sound pressure decreses with the ratio 1/r to the distance.*
> [http://www.sengpielaudio.com/calculator-distance.htm]

Ovenstående lovmæssighed beskriver, hvordan forholdet er mellem ændring i lytter-lydkilde-afstand og ændring i lydtryk (*Sound Pressure Level, SPL*) – når afstanden fordobles halveres lydtrykket (ganges med faktor 0.5 eller sænkes med 6dB). Det er en almindelig misforståelse at antage, at dette forhold er $1/r^2$,

[49] http://en.wikipedia.org/wiki/Quaternion
[50] http://en.wikipedia.org/wiki/Quaternions_and_spatial_rotation
[51] *x-IMU* kan indstilles til at udsende sin data med rater gående fra 1Hz til 512Hz. Jeg har fundet 128Hz tilstrækkeligt til dette projekt.
[52] Jf. også at kvaternioner er ikke-kommutative. Dvs. at operationen *q1*q2* ikke giver samme resultat som *q2*q1*. Jf. også at q^{-1} (*q* inverteret) er det samme som *q* konjugeret (en simplere udregning), når der er tale om enheds-kvaternioner.

som beskriver lydens *intensitet*, men denne intensitet kan ikke kobles direkte til den hørbare forandring af lydens styrke og ej heller til den faktor, lydsignalet skal ganges med for at forårsage denne forandring. I denne sammenhæng er det altså *Sound Pressure Level*, vi skal have fat i og ikke *Sound Intensity*[53].

I forbindelse med simuleringen af lydens aftagende styrke over afstand er jeg stødt på en særlig problematik, der især gør sig gældende, når lydkilderne er meget tæt på lytteren. Denne problematik vedrører det, at forholdet mellem afstandene til henholdsvis venstre og højre øre varierer i henhold til lydkildens afstand til lytteren. Altså, den maksimale forskel i afstand er konstant (= afstanden fra øre til øre), og denne forskel har naturligvis større betydning for lydstyrken, når den overordnede afstand til lydkilden er lille. Tænk eksempelvis på en, der hvisker dig i øret – her høres lyden temmelig kraftigt i det pågældende øre, mens lyden stort set ikke høres i det andet. Impuls-responserne i CIPIC-databasen er optaget på baggrund af lydkilder med en afstand på en meter til lytteren (centrum af lytterens hoved). Det vil sige, at disse responser sørger for at 'indkode' de forhold mellem øre-lydstyrkerne (IID/ILD), der optræder ved denne overordnede afstand.



Figur 17. På denne figur har jeg forsøgt at illustrere, hvordan forholdet mellem afstanden til henholdsvis venstre og højre øre ændrer sig med den overordnede afstand mellem lytter og lydkilde – der er her tale om en situation, hvor afstandsforskellen er størst mulig, dvs. at lydkilden befinder sig i retningen stik ud for det ene øre (i dette tilfælde, venstre øre). X-aksen er afstanden (meter) fra lydkilden til centrum af lytterens hoved, og y-aksen er ratio-forholdet mellem lydkildens afstand til henholdsvis det ene og det andet øre. Afstanden mellem lytterens to ører har jeg angivet til 22cm. Den grønne graf er afstanden til højre øre divideret med afstanden til venstre. Den røde graf viser det samme forhold, men med brøken vendt på hovedet – altså venstre divideret med højre. De blå cirkler viser hvor i denne forholdsudvikling CIPIC-databasens målinger er foretaget – ved *en* meters afstand. Det er tydeligt at se, at der sker meget med graferne på venstre side af de blå cirkler. Altså, det potentielle indbyrdes forhold mellem afstandene til de to ører bliver kraftigt påvirket, når lydkilden bevæger sig tættere på lytteren end en meter. Jo længere væk lydkilden kommer, jo mindre en ændring i dette forhold (pr. ændring i afstand) ser vi.

[53] http://www.sengpielaudio.com/calculator-distance.htm

Men hvis vi vil simulere en lydkilde, der er tættere på lytteren (og egentlig også en, der er længere væk fra lytteren, om end problematikken er mest betydningsfuld for helt nære lydkilder – jf. *Figur 17*), må vi altså kompensere for, at impuls-responserne er optaget ved en specifik afstand med et specifikt øre-afstandsforhold. For at kunne gøre dette må vi i systemet arbejde med afstandene fra lydkilde til lytterens, henholdsvis, venstre og højre øre, og ikke blot *en* afstand til centrum af lytterens hoved. I mit system er det *SpaceHandler*'en der også står for at udregne disse afstande – dette på baggrund af afstanden mellem lydkilde og lytterens hovedcentrum samt oplysninger fra *x-IMU* vedrørende lytterens orientering. *SpaceHandler*'en udregner ligeledes parametre til brug i simuleringen af rummets refleksioner – mere om dette i næste afsnit, *Simulering af rummets akustik*.
Vedrørende simulering af den karakteristiske Doppler-effekt henviser jeg til afsnittet, der beskriver konstruktionen af det lydlige indhold til lydfortællingen.

Simulering af rummets akustik

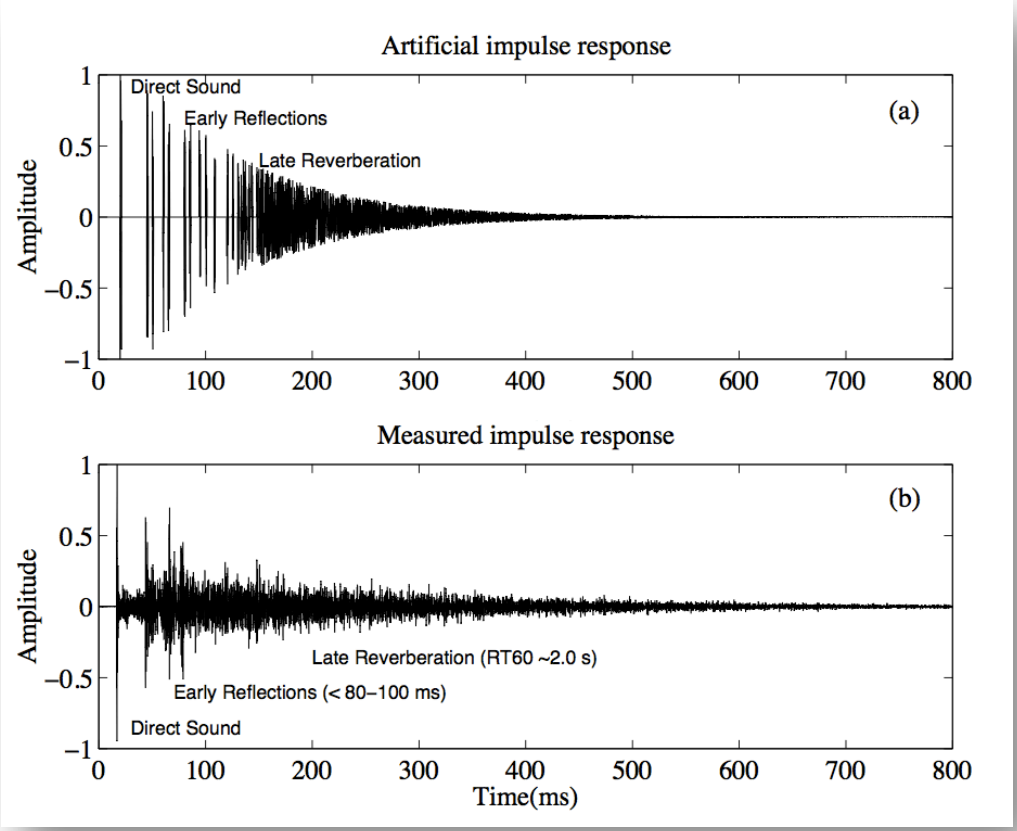
> *A review of literature states that incorporating room simulation in binaural sound reproduction systems is important to improve localisation capabilities as well as out of head localisation.*
> [Falch m.fl.: *Room Simulation for Binaural Sound Reproduction Using Measured Spatiotemporal Impulse Responses*]

Ovenstående citat indikerer vigtigheden af at arbejde med rummets akustik i forbindelse med binaural syntese. Og det virker da også intuitivt som et vigtigt område at beskæftige sig med i bestræbelserne på at simulere et lydligt miljø.
Simuleringen af et rums akustiske egenskaber er dog imidlertid et helt videnskabeligt område for sig, og jeg har ikke gjort mig forhåbninger om at nå i dybden på dette felt. Mit mål har blot været at finde en praktisk løsning, der kan supplere den binaurale syntese i en søgen mod rumlig auditiv realisme. I denne forbindelse har jeg ladet mig inspirere af Lauri Saviojas doktorafhandling *Modeling Techniques for Virtual Acoustics*[54], som beskriver nogle forskellige strategier på området.
En ofte brugt, og meget effektfuld, måde at simulere et rum på, er via målte *Room Impulse Responses* (RIR) – altså rummets reaktion på en impuls, fuldstændig i lighed med HRIR/HRTF. Disse målte impuls-responser kan også optages med binauralt optageudstyr, og i sådanne tilfælde taler man om *Binaural Room Impulse Responses* (BRIR). Hvis der foretages en convolution mellem den 'tørre' lyd fra en given lydkilde og en sådan rum-impuls-respons, opnås en effekt af, at lyden afspilles på det pågældende sted i det pågældende rum. Problemet er bare, at denne rum-impuls-respons naturligt nok angiver rummets respons ved netop den specifikke, statiske placering af lytter (mikrofoner) og lydkilde, hvorved den blev optaget. I mit projekt arbejder jeg med lydkilder og en lytter, der kan placeres i alle tænkelige positioner i rummet, og derfor er det ønskværdigt at have et rumsimuleringssystem, der er mere dynamisk adaptivt, end hvad der er tilfældet med denne strategi. Muligvis kunne man få et tilfredsstillende resultat ud af

[54] Savioja

at optage rum-impuls-responser i nogle grund-positioner, og efterfølgende, i afviklingssituationen, lave et dynamisk mix mellem disse. Dog tror jeg også, at man ville kunne risikere uønskede udfasninger og lydlige artefakter i forbindelse med en sådan metode. Dette på grund af de tidslige forskydninger mellem refleksionerne i de forskellige optagede impuls-responser.

Andre strategier går ud på at foretage en rent beregnet simulering af rummet på baggrund af viden om rummets geometri og overflade-karakteristika. I sådanne strategier kan man også tale om rummets impuls-responser, men så fald er de beregnet, ikke målt.

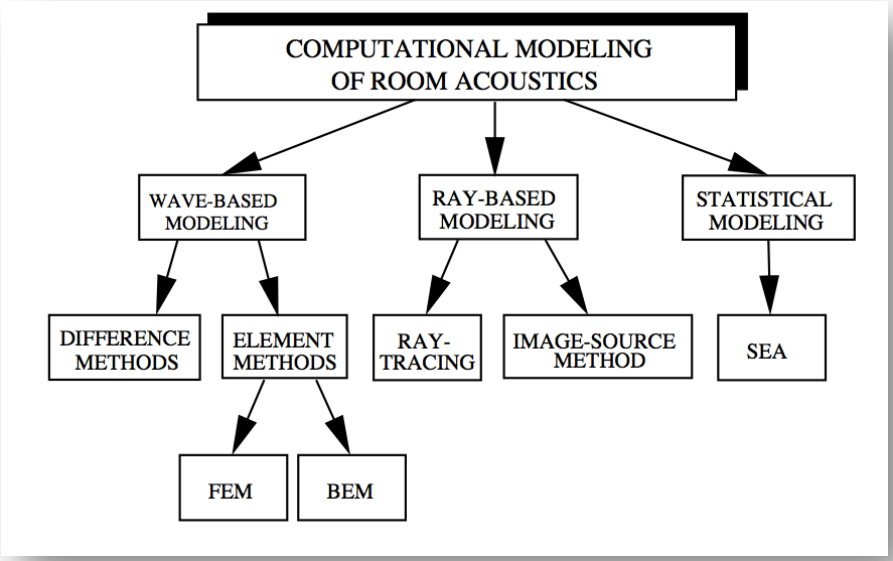


Figur 18. Figur taget fra Lauri Saviojas doktorafhandling *Modeling Techniques for Virtual Acoustics*. Figuren viser nederst en målt impuls-respons, og øverst en beregnet/simuleret.

Som det også fremgår af *Figur 18*, taler man i rumklangsterminologi ofte om, at den hørte lyd i et rum består af tre dele: den direkte lyd fra lydkilden, rummets tidlige refleksioner, og rumklangens 'hale'. Den direkte lyd kan ikke betragtes som en del af rummets akustik, da lyden ikke har haft mulighed for at interagere med rummet. Af rummets refleksioner er det angiveligt[55] de tidlige af slagsen, der er vigtigst i forbindelse med perceptionen af en lydkildes position, idet det hovedsageligt er disse, der ændrer sig i

[55] Zotkin, Duraiswami, og Davis – s. 12

forbindelse med ændring af lydkildens og lytterens position[56]. Rumklangens hale er en, efterhånden, mere fortættet og diffus ophobning af refleksioner, og den er tilnærmelsesvis statisk på tværs af variation i lytter- og lydkilde-placering. Hvis vi vælger en af strategierne, der er baseret på en beregnet simulering af rummet, er vi i stand til at lave en rumklang, der tilpasser sig dynamisk og *real-time*. Det er dog en kompleks og tung beregning at udregne hele rumklangen – alle refleksioner. Derfor har jeg valgt at basere de tidlige refleksioner i rumsimuleringen på *real-time* udregninger, mens rummets hale håndteres via convolution med en statisk, optaget impuls-respons. Jf. at halen er tilnærmelsesvis statisk. Der findes imidlertid en del forskellige metoder til at lave en beregnet simulering af rummets akustik. *Figur 19* er en oversigt over sådanne metoder.

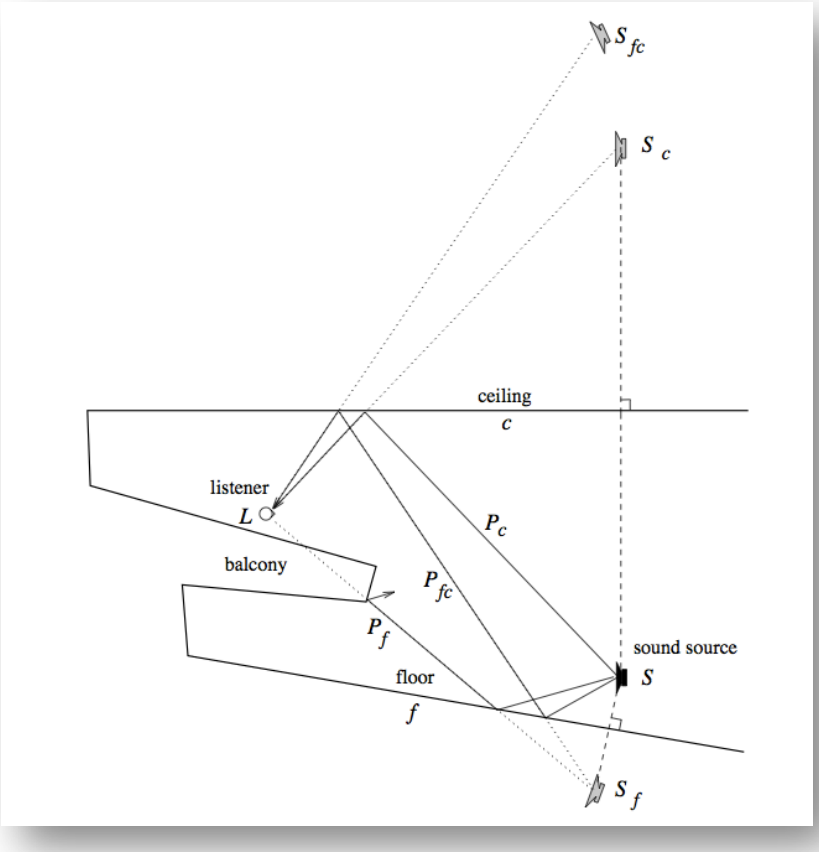


Figur 19. Figur taget fra Lauri Saviojas doktorafhandling *Modeling Techniques for Virtual Acoustics*. Figuren viser forskellige metoder til beregnet simulering af et rums akustik. Savioja pointerer, at metoderne evt. kan kombineres til at forme ganske gyldige hybrid-metoder. I mit system har jeg valgt at kombinere en af disse beregningsmetoder (*image-source-metoden*) med metoden, der baserer sig på empirisk målte impuls-responser.

Jeg vil ikke beskrive alle disse metoder, men blot den metode, *jeg* har valgt til beregning af de tidlige refleksioner i systemet. Det drejer sig om den såkaldte *image-source-metode*, som jeg, pragmatisk, har valgt, fordi den er let forståelig og relativt let at implementere. Metoden består i, at de lydlige refleksioner beregnes ved simpelthen at spejle lydkilden i rummets forskellige flader. Metoden arbejder med lydbølger, som var de en stråle af, eksempelvis, lys, der spejler sig i fladerne med præmissen: indgangsvinkel er lig udgangsvinkel. Dette er naturligvis en stærk forenkling af det faktiske fysiske fænomen med lydbølger,

[56] Det er dog vigtigt at pointere at det er den direkte lyd, der spiller den *største* rolle i bestemmelsen af en lydkildes position i forhold til lytteren – i hvert fald med hensyn til den retningsmæssige bestemmelse.

der kastes rundt i rummet. Men det er ikke desto mindre sådan, fænomenet modelleres i denne metode.



Figur 20. Figur taget fra Lauri Saviojas doktorafhandling *Modeling Techniques for Virtual Acoustics*. Figuren viser, hvordan de forskellige lydlige refleksioner findes i *image-source-metoden*.

Vi kan betragte refleksionerne som værende forsinkede og dæmpede versioner af den direkte lyd fra lydkilden. Refleksionernes forsinkelser bestemmes ud fra afstanden til disse spejlinger, og deres lydstyrke er ligeledes afhængig af afstanden samt af den absorption og diffraktion, der sker i mødet med rummets flader – denne absorption og diffraktion er frekvens-afhængig. Når en refleksion har ramt to flader, før den når lytteren, taler man om en refleksion af 2. orden, ved tre flader 3. orden osv. . I mit system har jeg nøjedes med at arbejde med refleksioner op til og med 2. orden. Dette giver 36 refleksioner for hver lydkilde i et almindeligt kasseformet rum. Hver af disse refleksioner kan betragtes som en lydkilde, der er positioneret på det sted, den sidst har ramt en flade. Ideelt set bør hver af disse 36 tidlige refleksioner, pr. lydkilde, processeres med hver sit sæt HRTF i henhold til deres position i forhold til lytteren, men dette bliver alt for processor-tungt i denne sammenhæng. I forsøget på at konstruere en realiserbar og samtidig perceptuelt acceptabel løsning har jeg ladet hver af de seks flader (gulv, loft og fire vægge) have sin egen globale kanal, som bliver processeret med *et* sæt HRTF. Altså benytter rum-simuleringen 'kun' seks HRTF-processeringer i alt i stedet for 36 pr. lydkilde. Hver flades HRTF vælges ud fra en slags

gennemsnit for fladens position i forhold til lytteren. Dette er naturligvis ikke fuldstændig optimalt i forhold til en oplevelse af retning, men som nævnt er det et kompromis mellem det realiserbare og det ønskværdige.

I praksis sker der følgende i systemets rumsimulering:

Hver lydkildes signal splittes ud i 36 yderligere signaler, som skal udgøre de tidlige refleksioner for denne lydkilde. *SpaceHandler*'en udregner forsinkelser og afstandsforårsaget dæmpning af de enkelte refleksioner, hvilket herefter implementeres. For hver flade i rummet summeres alle signaler af første og anden orden for sig, således at hver flade har to signaler. For at simulere fladernes frekvens-afhængige absorption og diffraktion løber signalerne igennem et *equalizer*-filter, som dæmper nogle frekvenser mere end andre. Refleksionerne af anden orden løber igennem filtret to gange, i henhold til at de har ramt to flader. I denne sammenhæng har jeg hentet inspiration i oversigter over forskellige bygningsmaterialers frekvensafhængige absorption. Nu lægges de to signaler pr. flade sammen til *et* signal pr. flade, som herefter processeres med et sæt HRTF udvalgt af Java-*external*'en *surfacePointer*. Så vidt de tidlige refleksioner. Rumklangens hale simuleres, ved at de tidlige refleksioner sendes ud til en global VST-rumklang, som foretager convolution med impuls-responser, hvor de tidlige refleksioner er skåret fra. For at få halen til at virke så realistisk som muligt har jeg i udarbejdningen af lydfortællingen som lydligt indhold optaget min egen impuls-respons[57] i det rum, hvori fortællingen foregår og opleves, mit eget køkken.

Komposition i fire dimensioner

I et almindeligt sequencer-program, som eksempelvis *Cubase* eller *Logic,* er en given lyd-events tidslige placering indikeret ved en grafisk placering fra venstre mod højre – præcis som et nodebillede, hvor man spiller fra venstre mod højre, eller for den sags skyld almindelig tekst[58]. Det er her interessant at bemærke, at vi bruger *det spatiale* til at udtrykke *det tidslige*. Denne spatiale repræsentation af det tidslige er ikke unik for sequencer-programmet, tværtimod bliver tid oftest repræsenteret (og forstået?) sådan – jf. tidslinjen, urskiven og vores sproglige metaforer som *frem* og *tilbage* i tiden. I sequencer-programmet klares den *rumlige* placering oftest med en kombination af de uafhængige parametre: panorering, lydstyrke, rumklang og eq. Man kan altså sige, at i sequencer-programmet er det *tiden*, der får lov til at tegne det primære spatiale rum, mens *rummet*, som jo ellers er spatialt i sin natur, er henvist til en kombination af drejeknapper, *fadere* og diverse *plugins*. Årsagerne til dette er sandsynligvis flere. For det første vil jeg påstå, at vi i vores oplevelse af et udtryk, der udfolder sig i både tid og rum (musik, fortællinger,…, vores virkelighed) vil tillægge det tidslige en større betydningsværdi end det rumlige. Altså, vi synes, at *hvornår* (underforstået en form for rækkefølge) er vigtigere end *hvor*. Dette gør sig i hvert fald gældende i vestlig kultur, hvor vi traditionelt set er særligt glade for det lineært tidslige og en form for kausalitets-

[57] Rummets respons på en ballon, der springes – en tilnærmelse til en slags enheds-impuls (en uendelig kort impuls, der indeholder alle hørbare frekvenser). De tidlige refleksioner er efterfølgende skåret fra.
[58] I nogle kulturer er læseretningen en anden, men det er sådan set ligegyldigt i denne sammenhæng.

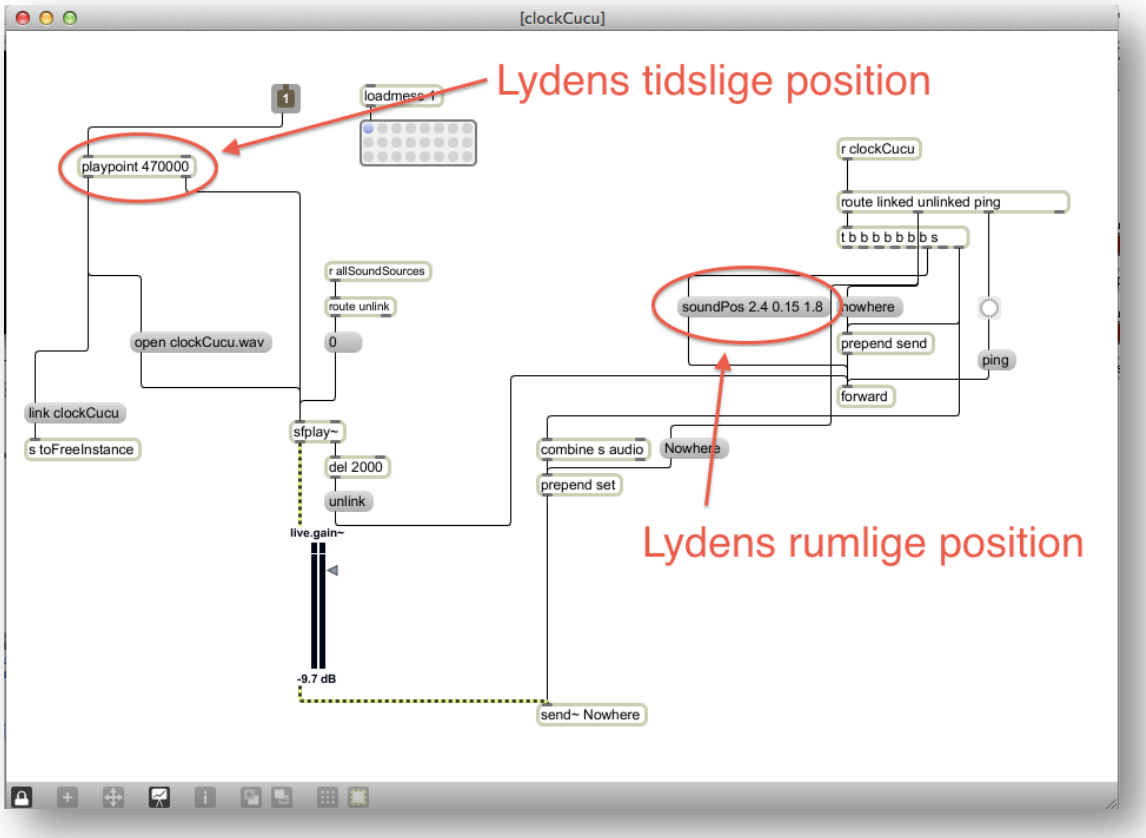
tankegang. Dette afspejler sig både i vores lineære historiesyn og kompositoriske traditioner. I mere cykliske eller tilstands-udfoldende udtryk, som ofte tilskrives andre kulturer end de vestlige, kan det måske diskuteres, hvorvidt der er denne forrang for *hvornår* i forhold til *hvor.* På den ene side kan man argumentere for, at de ofte meget rytmiske, groove-orienterede elementer i sådanne udtryk jo netop er meget tidsbundne – blot på et cyklisk mikro-plan. På den anden side kan man også betragte sådanne udtryk som værende mere statiske stemningsbilleder, der udfolder sig i, og karakteriserer, rummet – for eksempel kan rækkefølgen af ordene i visse digte måske være ligegyldig, og det samme kan måske rækkefølgen af formled i en reggae-sang. Her er *rummet* dog forstået som en mere abstrakt størrelse og ikke blot: *hvor i rummet befinder den enkelte lydkilde sig?*. *Hvad* er naturligvis også af stor vigtighed – altså hvad er det for en lyd, tone, hvad er det for et ord… – men dette *hvad* kæmper sædvanligvis ikke om retten til at repræsentere sig spatialt (måske lige med undtagelse af nodebilledet, hvor tonehøjden repræsenteres med nodens højde-placering).

Navnet *sequencer* indikerer naturligvis også, at dette værktøjs primære formål er at håndtere lydlige sekvenser – altså tidsligt strukturerede lyde. En anden årsag til at det rumlige ikke har fået tildelt en entydig spatial repræsentation i sequencer-programmet er, at der ikke findes en entydig standard for afvikling af lyde i henhold til en rumlig placering. I min sammenhæng er det dog vigtigt, at jeg både kan angive en lydkildes tidslige placering og dens præcise, konkrete, rumlige placering på et givent tidspunkt. Man kan måske sige, at jeg skal kunne komponere entydigt i fire dimensioner – de tre rumlige og i tidens dimension. De rumlige dimensioner er koblet til tiden, idet en given lydkilde kan bevæge sig over tid. Jeg har tænkt en del over, hvordan man bedst kunne tilgå dette med at komponere/placere lyde i tiden og rummet, sådan at det bliver en overskuelig og intuitiv proces. Jeg er kommet frem til, at det kunne være smart, hvis man byggede et kompositionsværktøj, der virker på samme måde, som når man animerer i 3D-programmer. I 3D-programmerne placerer man jo netop objekter i rummet, samtidig med at de er koblet til en tidslinje – både rum og tid er repræsenteret spatialt. Jeg har dog ikke kunnet (nå at) konstruere et sådant værktøj til dette projekt. På længere sigt kunne det helt bestemt være frugtbart at lave et interface/ en struktur for komposition i et eksisterende 3D-program som eksempelvis *Unity*[59] – jf. at *Unity* og *Max* er i stand til at 'tale' sammen . En sådan kobling ville gøre kompositionsprocessen nemmere, hurtigere, mere overskuelig og intuitiv.

Min nuværende løsning må betegnes som værende temmelig *lo-fi* og hverken særligt overskuelig eller intuitiv. Der er ikke en overordnet tidslinje (som sequencer-programmerne har), som giver et overblik over lydkildernes tidslige placering. Der er heller ikke en overordnet *scene* (som 3D programmerne har), som giver et overblik over lydkildernes rumlige placering. I stedet er der tale om, at den enkelte lydkilde selv holder styr på, hvor den befinder sig i rummet, og ligeledes 'hvornår' den befinder sig. Det-

[59] http://unity3d.com/

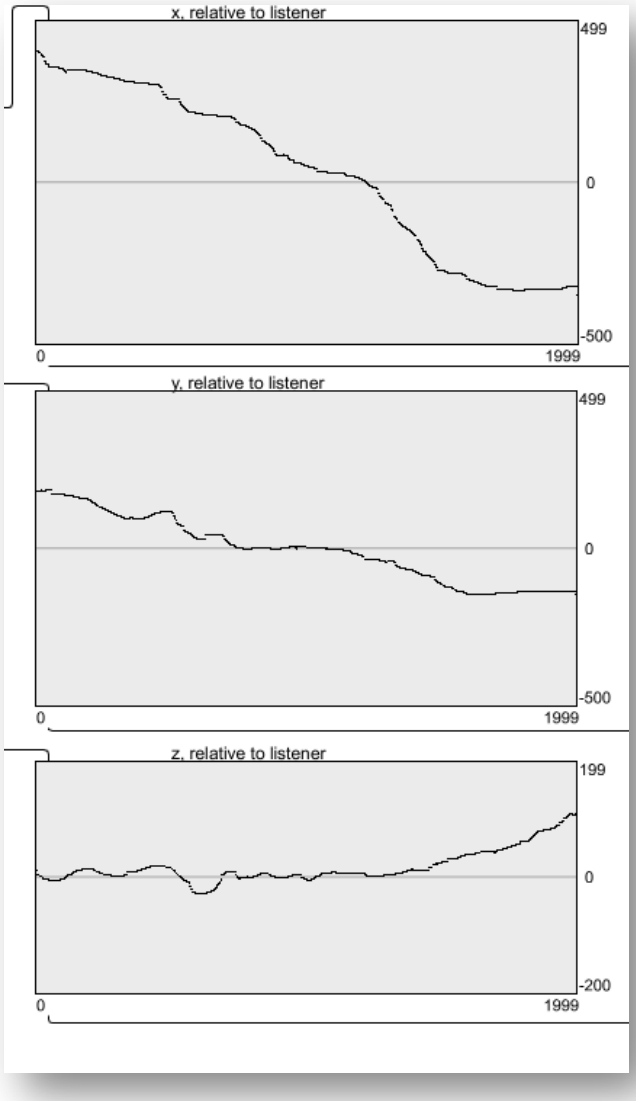
te sker essentielt set via *et* tal, der definerer, hvornår lydkilden skal afspilles (antal millisekunder fra start) og *tre* andre tal, der definerer x-, y- og z-koordinaterne for lydkildens rumlige placering.



Figur 21. Viser lydkilden *clockCucu* (et kuk-ur, der kukker). Her ses hvordan lydkildens tidslige placering (470 sekunder fra start) og lydkildens rumlige placering (2,4m; 0,15m; 1,8m) angives i systemet.

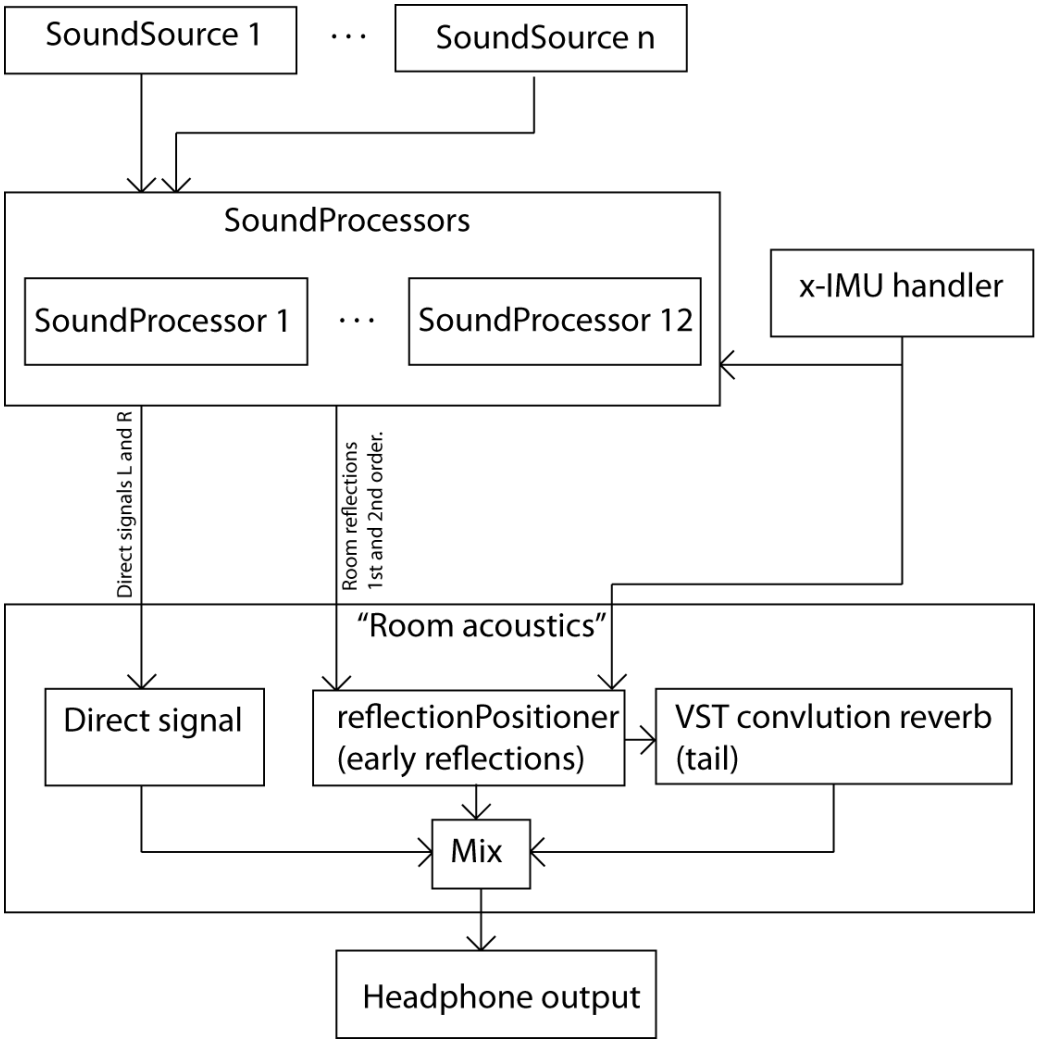
Der er dog også lydkilder, der ikke har en konstant rumlig placering, men bevæger sig rundt i rummet over tid. I sådanne tilfælde kan man naturligvis ikke nøjes med et enkelt sæt koordinater til at repræsentere lydkildens rumlige position – vi er nødt til at koble positionen til tiden. Dette har jeg gjort ved simpelthen at tegne grafer for lydkildens position i forhold til tiden – *en* graf for hver akse (x, y og z). Se *Figur 22*. Dette er naturligvis ikke en særligt overskuelig og intuitiv måde at arbejde med rumlig placering

på, men selve det lydlige resultat har fungeret ganske fint.



Figur 22. Eksempel på repræsentationen af en bevægelig lydkildes position. Repræsentationen er fordelt på tre akser, x, y og z. I dette tilfælde er der tale om en flues position i forhold til lytteren. Normalt er lydkildernes position angivet absolut og ikke relativt i forhold til lytteren, men netop i tilfældet med fluen har jeg valgt, at den skal lægge sin bane i forhold til lytteren. På graferne er x-akserne tid i 100.dele sekunder og y-akserne er position i centimeter.

Systemets struktur



Figur 23. Systemets struktur i meget overordnede træk.

Figur 23 viser i grove træk, hvordan systemet er opbygget i *Max*[60]. Som det ses, består systemet blandt andet af en række *SoundSources*. Hvor mange *SoundSources,* der er, afhænger af, hvor mange selvstændige lydkilder der er i det specifikke lydlige indhold – i princippet kan der være uendeligt mange (så længe de ikke er aktive samtidig)[61]. Jf. her, at jeg skelner mellem, hvad der er *systemet*, og hvad der er *det lydlige indhold*. I denne skelnen befinder disse *SoundSources* sig på grænsen mellem system og indhold, idet de både indeholder fast funktionalitet, der kræves for integrationen med resten af systemet, og samtidig indeholder referencer til de specifikke lydfiler samt en eventuelt meget individuel håndtering af disse – *SoundSource*'ene er altså ikke baseret på en klasse i datalogisk forstand, men er derimod potentielt set alle forskellige. I udviklingen af et programmeringsbaseret system er det altid vanskeligt at vurdere, hvornår det er hensigtsmæssigt at arbejde med generaliseringer/skabeloner (= klasser), og hvornår det er mere hensigtsmæssigt at programmere hver enhed for sig. I tilfælde, hvor man arbejder med mange

[60] Se evt. bilag 2 – Max-kode
[61] Se evt. Figur 26

enheder 'af samme slags', er det ofte smart at programmere en klasse. Og selvom man kunne argumentere for, at lydkilderne langt hen ad vejen er 'samme slags', så har jeg alligevel valgt, at holde muligheden åben for helt individuel og specifik afvikling af den enkelte lydkilde. Dette valg er ikke mindst baseret på erkendelsen af, at jeg befinder mig i en meget eksperimenterende fase, og her kan det være svært, og endda uhensigtsmæssigt, at forsøge at konstruere en skabelon, der kan håndtere alle ønskværdige muligheder. Der er dog, som nævnt, nogle faste opgaver, som disse *SoundSources* skal varetage. Hver *SoundSource* sørger for at afspille den pågældende lydkildes lydfil på det relevante tidspunkt og desuden at skalere dette lydsignal til den lydstyrke, lydsignalet skal høres ved på *en* meters afstand. Derudover holder hver *SoundSource* styr på den pågældende lydkildes placering i det tredimensionelle rum – en placering, som eventuelt kan variere over tid.

Umiddelbart inden en *SoundSource* skal til at afspille lyd, 'beder' den om at blive linket til en *SoundProcessor*. *SoundSource*'ene er altså ikke koblet permanent til nogen *SoundProcessor*, men der skabes *ad hoc* - koblinger mellem *SoundSource* og en tilfældig 'ledig' *SoundProcessor*[62]. Når en *SoundProcessor* er koblet sammen med en *SoundSource*, aktiveres *DSP*-processeringen i *SoundProcessor*'en, og den markeres som værende 'optaget' (og omvendt, når koblingen brydes). *SoundProcessor*'nes *DSP*-processering er temmelig tung at trække for computeren, og derfor bør den naturligvis kun være aktiv i instanser, der er koblet til en *SoundSource*. *SoundProcessor*'ne, som er instanser af en klasse, har som hovedopgave at foretage den binaurale syntese på de enkelte lydkilders signaler. Dette foregår, som tidligere nævnt, ved at konvertere lydkilde-signalet til frekvens-domænet, foretage en gange-operation med den relevante HRTF og til sidst konvertere signalet tilbage til tids-domænet. I *SoundProcessor*'en befinder sig også den ligeledes tidligere nævnte *SpaceHandler*, som sørger for at udpege lydkildens i øjeblikket relevante HRTF samt at udregne parametre til lydkildens *early reflections*. *Poly~*-objektet hvori *SoundProcessor*'ne er indkapslet sender i alt 14 lydsignaler[63] videre til systemets næste stadie, rum-simuleringen, bestående af *patch*'en "*Room acoustics*". "*Room acoustics*" håndterer, som beskrevet i afsnittet *Simulering af rummets akustik*, tre forbundne akustisk rumlige fænomener: den direkte lyd fra lydkilden, rummets tidlige refleksioner og den mere diffuse hale af refleksioner. Den direkte lyd, der af *SoundProcessoren* er blevet binauralt syntetiseret, passerer sådan set blot igennem *patchen*. De tidlige refleksioner, der, som tidligere nævnt, er blevet summeret pr. flade i rummet, bliver i patchen *reflectionPositioner* binauralt syntetiseret (pr. flade, dvs. seks syntetiseringer i et almindeligt kasseformet rum) i henhold til fladens position i forhold til lytteren. VST-convolution-rumklangen, som håndterer den diffuse hale, får sit input fra de tidlige refleksioner. Disse

[62] I systemets nuværende udformning indeholder det 12 *SoundProcessors* indkapslet i objektet *poly~,* som er i stand til at håndtere aktiveringen/deaktiveringen samt tildelingen af enkelte 'stemmer' (i dette tilfælde, *SoundProcessors*)

[63] Disse 14 lydsignaler er: to direkte signaler (venstre og højre øre), første-ordens-refleksioner for hver flade i rummet (seks stk.) og anden-ordens-refleksioner for hver flade i rummet (også seks stk.)

tre akustiske komponenter mikses til sidst sammen til det endelige signal (*de* endelige signal*er* til henholdsvis venstre og højre øre), der bliver sendt til lytterens hovedtelefoner.

Det lydlige indhold

Som nævnt i afsnittet *Systemets struktur* arbejder jeg med en skelnen imellem systemet og systemets indhold. I dette afsnit vil jeg beskæftige mig med indholdet.

I projektet kunne jeg godt have valgt at have 'nøjedes' med at konstruere systemet og herefter teste dets evne til at positionere et par, mere eller mindre tilfældige, lyde. Dog har min ambition været at få et indblik i denne teknologis potentiale i en oplevelsesorienteret sammenhæng, og derfor har jeg forsøgt at konstruere et par bud/eksempler på konkret lydligt indhold til systemet. Lydligt indhold, som, i samspil med systemet, formidler en form for oplevelse til lytteren. Det drejer sig om to overordnede bud på indhold: det ene er en slags lydfortælling, og det andet er musik, nærmere bestemt firestemmig korsang. I de følgende underafsnit vil jeg beskrive (arbejdet med) disse to indholdseksempler.

Lydfortællingen

Idéen til lydfortællingen som indhold udspringer af min deltagelse i en workshop, som gik ud på at brainstorme over idéer til formidling af ældre menneskers livshistorie og -situation. Initiativet til workshoppen var taget af en lille gruppe på tre unge aarhusianere, som har til hensigt at lave en sådan formidling på Kvindemuseet i Aarhus i løbet af 2012. Projektet har arbejdstitlen *Mit livs historie*. Gruppen havde selv en idé, som involverede binaural lyd. Denne idé gik ud på at sætte den oplevende person i den ældres sted, forstået på den måde, at man befinder sig i den ældres rum (evt. en rekonstruktion af vedkommendes stue) og samtidig også lydligt er nedsænket i en meget livagtig gengivelse ved hjælp af binaurale optagelser. Det kunne eventuelt være, at man som lytter 'er' en af flere (virkelige) ældre mennesker, der snakker sammen – en samtale, der er optaget i netop det pågældende rum med binaurale mikrofoner i ørerne på disse ældre personer. Personerne er naturligvis ikke fysisk til stede i oplevelsessituationen, kun deres lyd. Deres fysiske placering / deres rolle er udfyldt af, sandsynligvis, yngre medmennesker, der besøger museet. En stærk symbolik – vi er ikke så forskellige, selv de ældre har været unge en gang, vi bliver alle ældre … Det var faktisk disse idéer, der inspirerede mig til at lave hele dette projekt vedrørende binaural syntese. Gruppens idé drejede sig godt nok om binaurale *optagelser*, men jeg syntes, at det kunne være spændende at udforske den binaurale *synteses* muligheder for at skabe mere dynamiske fortællinger, der, blandt andet, tager højde for lytterens bevægelser. Da det var dette formidlingsprojekt, der havde inspireret mig, syntes jeg også, at det var på sin plads at forsøge at konstruere lydligt indhold, der kunne passe med gruppens idé om at formidle ældre menneskers livshistorie og -situation.

Nedenstående er nogle tanker vedrørende min vision for oplevelsen af dette lydlige indhold:

Man træder ind i et rum, som er et ældre menneskes køkken eller stue (enten en rekonstruktion eller

det faktiske rum). Her iører man sig hovedtelefoner med indbyggede sensorer (disse er potentielt trådløse, men i min prototype har de ledning). Nu får rummet liv i form af en masse lyde. Man kan høre køleskabets summen, og man kan høre, at lyden kommer fra det køleskab, der rent faktisk står i hjørnet, men før var lydløst. Når man bevæger hovedet, positionerer lydene sig, som de ville gøre 'i virkeligheden'. I denne vision kan man som lytter/deltager også bevæge sig rundt i rummet og lytte nærmere til de forskellige ting (i min prototype er lytteren dog fikseret til én position i rummet (evt. siddende ved et bord), da jeg ikke arbejder med positions-*tracking*). Lydene kan også bevæge sig rundt i forhold til lytteren – eksempelvis kan det være, at der flyver en flue rundt i rummet. I dette rum befinder der sig også en ældre person (ikke faktisk, men lydligt). Denne person kommer måske til lytteren med en kop kaffe, imens personen snakker om løst og fast. Måske begynder personen at fortælle en historie fra sit liv. Denne historie udspiller sig måske i selv samme rum, måske ikke. Pludselig befinder vi os (lydligt set) inde i den historie, som den ældre person fortæller. Altså en slags lydligt flashback, hvor den ældre persons stemme nu ikke længere er diegetisk placeret i rummet (det nye rum), men har karakter af en ophævet fortæller.
På denne måde kan vi i fortællingen skifte mellem at befinde os i 'nutidens rum' og det rum, hvori den ældre persons fortælling foregår.

Det er værd at bemærke, at der i gruppens idé og også i min version af den, ligger et element af *Augmented Reality*, fordi den fysiske virkelighed 'udvides' med et medieret, lydligt lag, der forholder sig / er koblet til netop den fysiske virkelighed[64]. Det er tanken i min lydfortælling, at den oplevende person vil opleve lydene som værende koblet til rummet, man befinder sig i. Altså vil lyden af eksempelvis en kaffemaskine opleves som kommende fra den kaffemaskine, der er synlig i rummet, men som dog måske ikke er tændt. Selv når lytteren, der er iført hovedtelefoner, drejer sit hoved, vil lyden stadig opleves som kommende fra kaffemaskinen. De hørte lydlige refleksioner vil også passe med rummets fysisk visuelle fremtoning samt med de erfaringer, lytteren har med rummets akustik, fra før vedkommende iførte sig hovedtelefoner.

Jeg besluttede mig for at konstruere/indfange en lydlig fortælling med min mormor som den ældre person, om/af hvem fortællingen fortælles. Min mormor har blandt andet nogle spændende oplevelser, fra da hun og min morfar var unge i København under krigen. Hendes nuværende dagligdag i Hjørring er på mange måder temmelig langt fra tiden i København. Hun har mistet min morfar og mange af sine gamle venner, og samtidig bøvler hun med mange fysiske skavanker. Dog forsøger hun at holde sig i gang, så godt hun kan. Jeg var godt klar over, at det ikke var muligt, at indfange hele hendes livshistorie og -

[64] Det kan naturligvis diskuteres, om der er tale om en fysisk virkelighed i det tilfælde, at det er en rekonstruktion af et virkeligt rum. Men ikke desto mindre kan man vel konstatere, at *rekonstruktionen* fremtræder fysisk virkeligt for den besøgende.

situation i en kort lydfortælling, men mit mål var alligevel at forsøge at give et indtryk af den udvikling og de kontraster, der er i hendes/et liv.

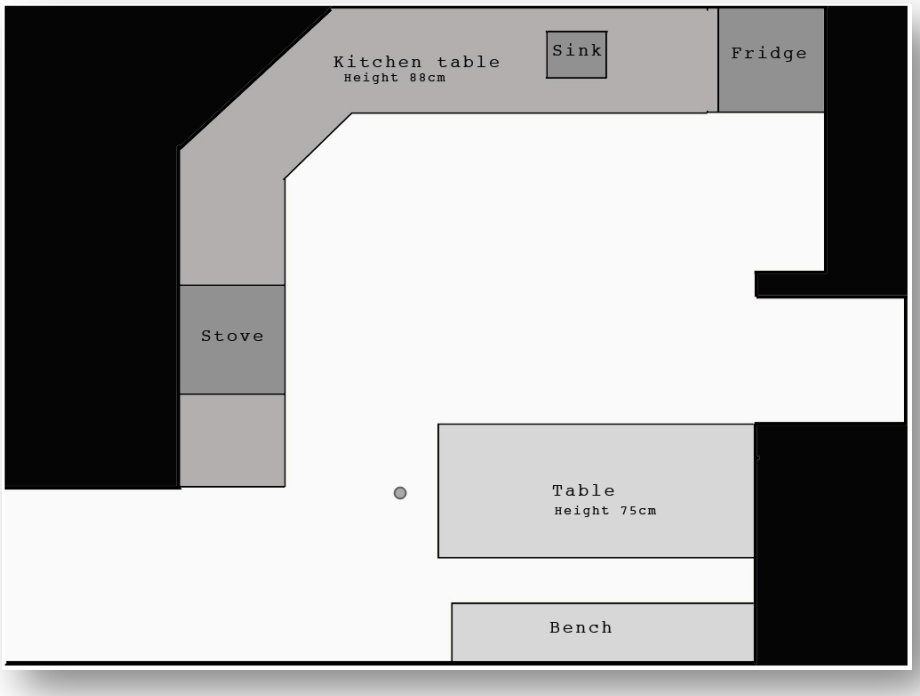
Ideelt set burde alle lydkilder i systemet optages i ellers lydløse og lyddøde omgivelser[65], men jeg optog hende i hendes hjem i Hjørring for at gøre det så trygt som muligt med henblik på, at hun kunne tale naturligt. Derudover forsøgte jeg at gøre det til en samtale mellem hende og mig – igen for at fremme naturligheden i denne ellers temmelig opstillede og uvante situation. Jeg fik hende til at tale både om sin ungdom og sin nuværende situation. Desuden instruerede jeg hende i at sige ting som "vil du ikke have en kop kaffe?", så jeg senere kunne konstruere en lydlig kaffe-hentnings-situation.

Figur 24. Min mormor fortæller.

Selvom jeg optog min mormor i hendes hjem i Hjørring, har jeg konstrueret lydfortællingen til at blive oplevet i mit køkken i Aarhus. Køkkenet fordi det er et uformelt, hyggeligt rum, der åbner muligheden for en del lydkilder, og mit eget fordi det er let tilgængeligt for mig. Det at fortællingen er konstrueret til

[65] Det er vigtigt, at hver lydkilde så vidt muligt optages isoleret, da de i systemet bliver positioneret som et entydigt punkt. Ydermere er det ønskværdigt at undgå rummets refleksioner på optagelserne, da systemet selv simulerer disse.

dette rum betyder altså, at lydkilderne skal placeres i henhold til deres (potentielle) placering i mit køkken – lyden af vandhanen skal placeres i henhold til vandhanens position i mit køkken osv..



Figur 25 Et screenshot fra *Max*. Her ses en grafisk oversigt over mit køkken. Den lille cirkel angiver lytterens x-y -position på en stol for enden af bordet – en position, der, potentielt, kan ændres ved at trække cirklen rundt med musen (lytterens z-position angives andetsteds). Jeg har benyttet denne grafiske oversigt til også at bestemme placeringen af de enkelte lydkilder i lydfortællingen.

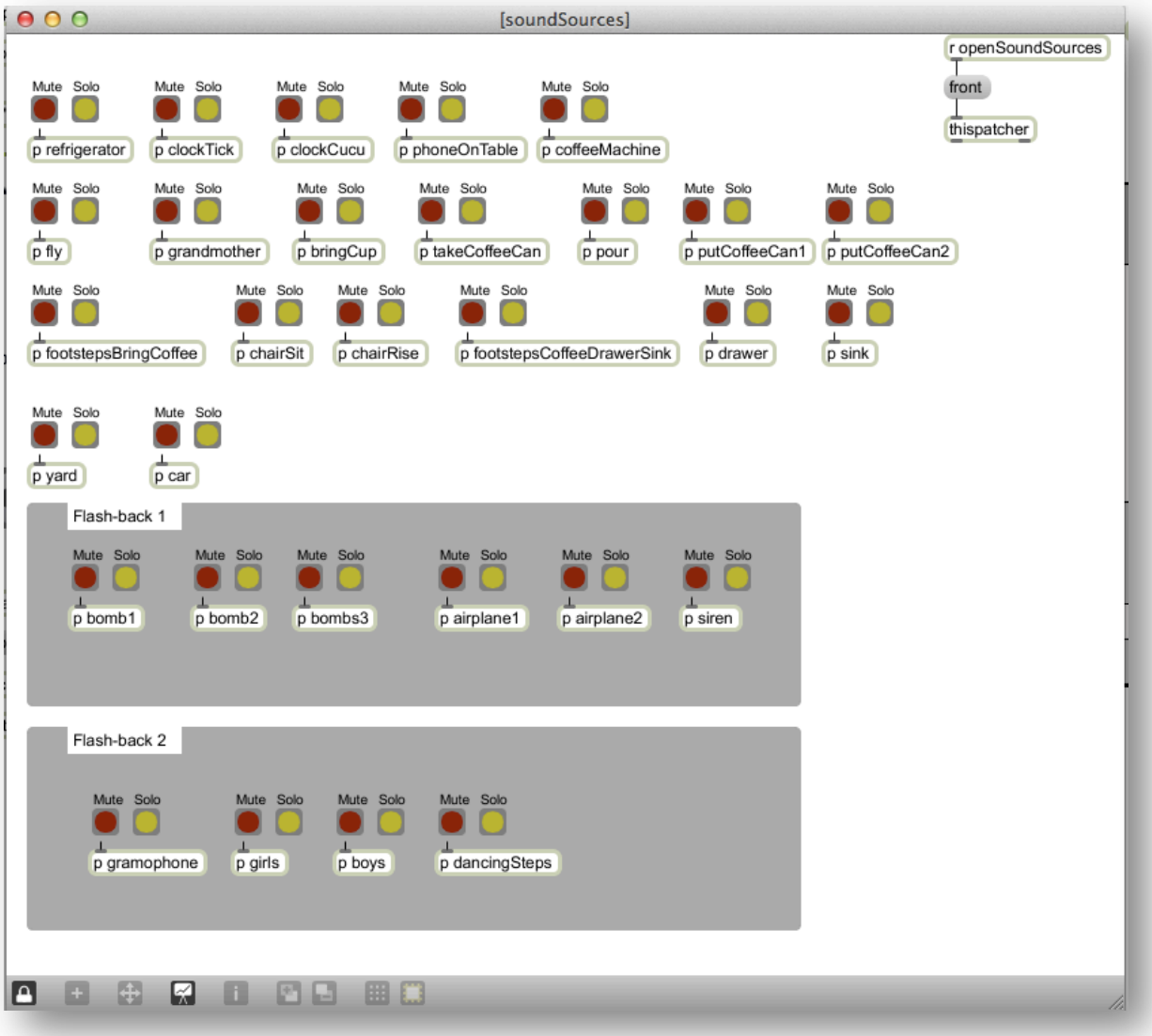
For at tilføre en ekstra dimension til lydfortællingen har jeg arbejdet med at konstruere lydlige flashbacks som beskrevet i visionen. Det er blevet til to flashbacks i løbet af den ca. otte minutter lange lydfortælling – man kunne sagtens lave flere, men to er rigeligt til at vurdere effekten. Under disse flashbacks skifter vi fokus fra fortællerens/nutidens rum til det fortaltes/fortidens rum. Med mindre det fortalte tilfældigvis foregår i det samme fysiske rum, bryder vi altså her med *det augmenterede* til fordel for det rent virtuelle/forestillede. De to situationer, jeg har konstrueret som flashbacks, er henholdsvis min mormors fortælling om bombningen af Shell-huset i 1945, hvor hun bl.a. så flyvemaskinerne komme ind over byen, og hendes fortælling om hendes ungdomsliv under krigen, med festlige komsammener på små klubværelser i København. Under disse flashbacks bliver den binaurale syntese og den rumlige simulering på min mormors stemme slået fra. Det vil sige, at hendes stemme skifter fra at blive oplevet som kommende fra et sted i rummet til nærmest at befinde sig inde i lytterens hoved. Det er jo netop det karakteristiske ved binaural lyd (og rum-simulering), at den sigter mod og slår på sin evne til at skabe denne eksternalisering af de oplevede lyde i hovedtelefoner – ellers opleves lyd i hovedtelefoner ofte(st?) som værende inde i lytterens hoved. Men altså, dette skift i oplevelsen af min mormors stemme har jeg konstrueret

med henblik på at underbygge skiftet i fortællingens fokus. Vi skifter fra at være nedsunket i nutidens rum, hvor lytteren og min mormor befinder sig i rummet sammen, til at være i et andet rum/i en anden tid, hvor min mormors stemme ikke længere kan betragtes som værende en diegetisk lydkilde, men rettere en slags ophævet fortæller. Hun er dog måske nok stadig diegetisk, i fald vi oplever fortællingens forskellige tider som parallelle, sideløbende spor. Under disse flashbacks forsvinder også nutids-rummets (køkkenets) lyde til fordel for de lyde, der hører til i flashback'ets rum. I flashback nr. 1, bombningen af Shell-huset, hører vi flyvemaskinerne komme ind over os og bomberne falde og ramme tæt på. Jeg har tilladt mig at placere lytteren en hel del tættere på begivenhederne, end min mormor i virkeligheden var. Dette for at forstærke den dramatiske effekt. Flash-back-lydene er placeret i systemet på samme måde som de andre lyde – det vil sige, at lytteren (forhåbentlig) oplever, at lydene er realistisk rumligt positioneret, og at de, som de er det i virkeligheden, er afhængige af lytterens hovedbevægelser. I dette tilfælde har vi at gøre med lydkilder, der bevæger sig forholdsvist hurtigt – flyvemaskiner og bomber. Arbejdet med implementeringen af disse har budt på både udfordringer og interessante opdagelser. En interessant opdagelse var, at jeg fandt ud af, at den karakteristiske Doppler-effekt, som vi kender fra lydkilder, der passerer os hurtigt, faktisk kunne håndteres meget nemt ved at implementere et delay med variabel forsinkelse. Da jeg jo i forvejen arbejder med lytterens afstand til lydkilden, er det meget nemt at omtolke denne afstand til en forsinkelse (antal samples) i henhold til lydens hastighed og systemets *sample rate*:

$$Delay\ (antal\ samples) = \frac{afstand}{lydens\ hastighed} \cdot sample\ rate$$

Nogle af os kender effekten af at variere forsinkelsen på en delay-maskine, imens der spilles lyd igennem – lydens *pitch* og afspilningshastighed varierer. På denne måde kan vi simulere Doppler-effekten på en måde, som er lettere og mere direkte koblet til det naturlige fænomen, end hvis vi skal ind og arbejde med afspilningshastigheden af lydfilen. Det er dog vigtigt, at det varierende delay ændres glidende i henhold til den simulerede lydkildes glidende bevægelse og ikke i de, godt nok fint inddelte, hak, som systemet opdateres med. Derfor er det hensigtsmæssigt at foretage en udglatning mellem de udregnede forsinkelser. En mindre udfordring i forbindelse med disse hurtigt-bevægende lydkilder er, at deres grundlyd, den afspillede lydfil, hverken må være influeret af Doppler-effekt eller lyddæmpning i henhold til afstand, idet systemet selv håndterer disse fænomener. Det vil sige, at lydfilen skal rumme en meget konstant og relativt nær lyd af, eksempelvis, en flyvemaskine, der flyver, men ikke flyver *forbi* – der er her en åbenlys problematik vedrørende optagelse af sådanne lyde. Jeg har derfor konstrueret konstante *loops* sammensat af forskellige optagelser i et forsøg på at skabe sådanne lyde.

Flashback nr. 2 er mere enkelt og rummer lyde som en gammel grammofon, der spiller et hit fra krigens tid, piger der snakker, drenge der skåler og snakker samt lyden af dansende trin.



Figur 26. Oversigt over alle lydkilderne, der indgår i lydfortællingen. Denne *Max-patch* er blot en samling af lydkilderne, dvs., at deres spatiale placering på den todimensionelle flade ikke har nogen betydning for afviklingen af dem. I kompositionsfasen fandt jeg det nyttigt at have nem adgang til at bestemme, hvilke lydkilder der skal høres. Til dette formå har jeg udviklet et *solo-mute*-system, som man kender det fra lydmixere.

I forbindelse med konstruktionen af lydfortællingen er jeg stødt på en generel problematik i forbindelse med systemet. Denne drejer sig om sammensatte lyde – altså lydoptagelser, som rummer lyd fra flere forskellige lydkilder. I udarbejdelsen af et lydligt univers er det ofte ønskværdigt at arbejde med sådan sammensatte, ofte ambiente, lyde. Det kan eksempelvis være byens summen, rummets 'tone', fuglenes kvidren osv.. Det ville selv sagt være et meget stort arbejde at skabe sådanne lydfænomener ud fra de enkelte lydkilder, der indgår i dem – den enkelte fugl, den enkelte bil…. For ikke at tale om den betydelige belastning et sådant hav af lydkilder ville lægge på computeren i et system som mit. Men der opstår, som nævnt, en problematik, når vi vil implementere en sammensat lyd i systemet: *hvor skal vi placere den, rumligt set?* Systemet arbejder med lydkilder som et punkt i rummet og har derfor vanskeligt ved at håndtere lydkilder med en væsentlig rumlig udstrækning. Og de sammensatte lyde har oftest en betydelig

rumlig udstrækning – ja ofte kan de siges at fylde hele 'rummet' og dermed omslutte lytteren. Helt konkret stødte jeg på dette problem, da jeg ville indfange og implementere lydene fra byen udenfor køkkenets vinduer. I et forsøg på at komme udenom systemets punkt-håndtering af lydkilder optog jeg disse lyde med tre spredte mikrofoner samtidigt med henblik på at implementere lyden fra disse som tre separate lydkilder i systemet. Dette altså for at få en 'bredere' fornemmelse af den ambiente lyd fra byen udenfor. Jeg havde imidlertid ikke gennemtænkt alle implikationer ved denne metode og har siden forkastet den, blandt andet fordi, mikrofonerne, på trods af at være relativt retningsbestemte, indfanger lyd fra de samme kilder. Og dette kryds-overhør, med dets indbyrdes forsinkelser og dæmpninger, kommer til at karambolere med den binaurale synteses indkodning af ITD og IID. Jeg har ikke fundet frem til en ideel løsning på denne problematik, men det ville bestemt være relevant i fremtidigt arbejde.

Koret

Ud over at benytte systemets teknologi til formidling af lydfortællinger, ser jeg et muligt potentiale i at benytte musik som lydligt indhold – altså at man kan formidle musik, så lytteren oplever, at de enkelte musikere/instrumenter er placeret meget realistisk i det oplevede, lydlige rum. Den altdominerende formidlingsform for optaget musik er, og har i mange år været, stereo – altså afviklingen af to separate, statiske lydkanaler. Stereo var en revolution, da det først blev almindeligt tilgængeligt i 50'erne[66], og det bliver stadig betragtet som værende en ganske god måde at formidle musik på. De bestræbelser, man ser i retning af multikanalslyd og andre udforskninger af det lydspatiale aspekt, er oftest rettet mod andre kunst- og underholdningsformer end musikken – eksempelvis film og virtuelle verdener som computerspil. Den manglende interesse på musikkens område skyldes nok i høj grad de praksisformer, der omgiver, og den rolle, vi tilskriver den populærmusikalske[67] udgivelse – populærmusikken, som udgør langt størstedelen af musikindustrien. I 60'erne begyndte populærmusikere at omfavne lydstudiets efterhånden mange muligheder, og den populærmusikalske udgivelse, med albummet som det kunstneriske format, blev løsrevet fra et ideal om 'blot' at skulle være en (transparent/realistisk/autentisk) gengivelse af en, mere eller mindre, live performance. Den blev en selvstændig og mere abstrakt udtryksform uden noget krav om, at vi som lyttere har et indtryk af, at musikerne står og spiller samtidigt et sted i samme rum. Da populærmusikken hermed alligevel har forkastet tanken om, at udgivelsen skal formidle noget realistisk/transparent, har der fra denne side ikke været noget stort behov for at søge mod potentielt mere realistiske gengivelsesformer. Filmen, derimod, har i mere udstrakt grad omfavnet de løbende teknologiske bestræbelser i retningen mod en større grad af realisme. Diverse udviklinger af multikanalslyd og det, for tiden meget populære, visuelt tredimensionelle er eksempler på sådanne bestræbelser. Nu kan filmen dog bestemt også siges at være en udtryksform, der langt hen ad vejen bygger på meget di-

[66] Toole
[67] Populærmusik, her forstået som den store kategori af musik, vi også betegner *rytmisk musik*, dog minus jazzen.

rekte og entydige referencer til virkeligheden. I tilfældet musik kan man måske tale om, at der er en dobbelthed: på den ene side kan musik anskues som værende abstrakt lyd – det abstrakte, som jo netop er løsrevet fra entydige virkelighedsreferencer. På den anden side kan vi også betragte musik som værende lyd frembragt af musikere og instrumenter – det kan godt være, at det musikalske *indhold* er abstrakt, men den lydlige frembringelsesproces er oftest meget virkelig og entydig.

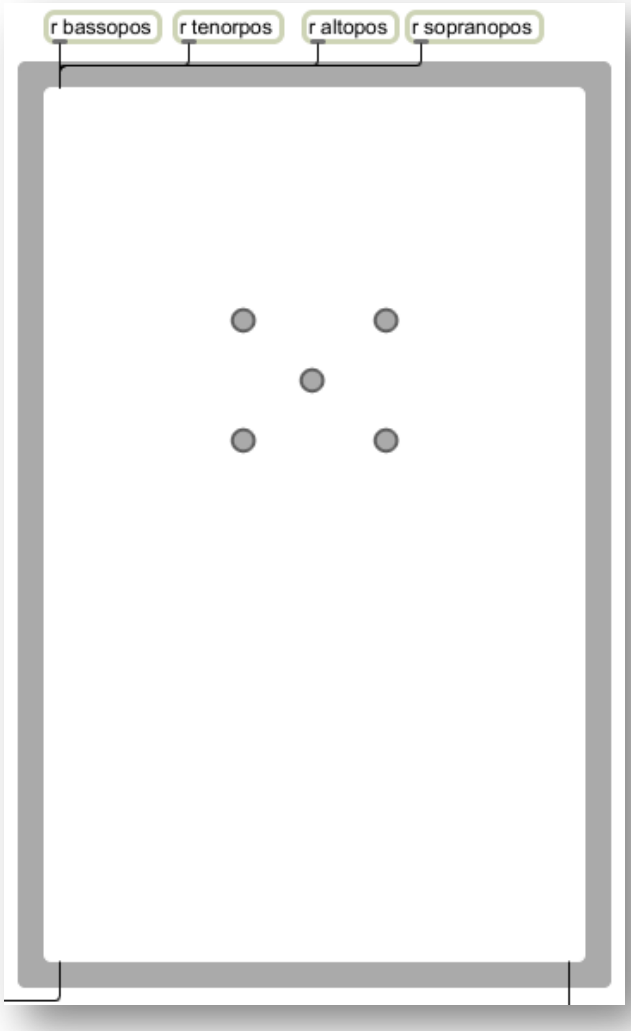
Selv om jeg tror, at det kunne være frugtbart at benytte populærmusik i mit binaurale system, har jeg valgt i første omgang at forsøge mig med musik, hvor det er mere oplagt at placere lydkilder entydigt i rummet. I denne sammenhæng tænker jeg på musik som jazz, klassisk musik[68], akustisk musik[69]… musik, hvor det æstetiske ideal i forbindelse med den optagede udgivelse, i modsætning til populærmusikken, er en form for transparens – den naturtro gengivelse af en performance. Jeg forsøgte bl.a. at finde sporopdelte[70] optagelser af akustiske ensembler, som eksempelvis en jazz-trio. Dog fandt jeg det mest oplagte på en hjemmeside, som sælger kor-arrangementer[71]. På denne hjemmeside er det også muligt og gratis at downloade lydspor, hvor hver korstemme (sopran, alt, tenor og bas) i de pågældende arrangementer er indspillet på sit eget spor. Godt nok er der en smule overhør mellem de stemme-specifikke mikrofoner på indspilningerne, og godt nok er der tale om mp3-filer (128 Kbit/sek), men sangene er meget musikalsk og professionelt fremført. Dermed skulle jeg blot placere fire statiske lydkilder i systemet – en for hver korstemme – og selvfølgelig sørge for, at de begynder at synge den samme sang samtidigt.

[68] Klassisk musik, her forstået som det dagligdagsbegreb, der dækker over alt fra middelalderlig munkesang til det vi kalder moderne kompositionsmusik. Egentlig en broget samling af tider, former og traditioner, som vi alligevel formår at putte ned i *en* kasse.

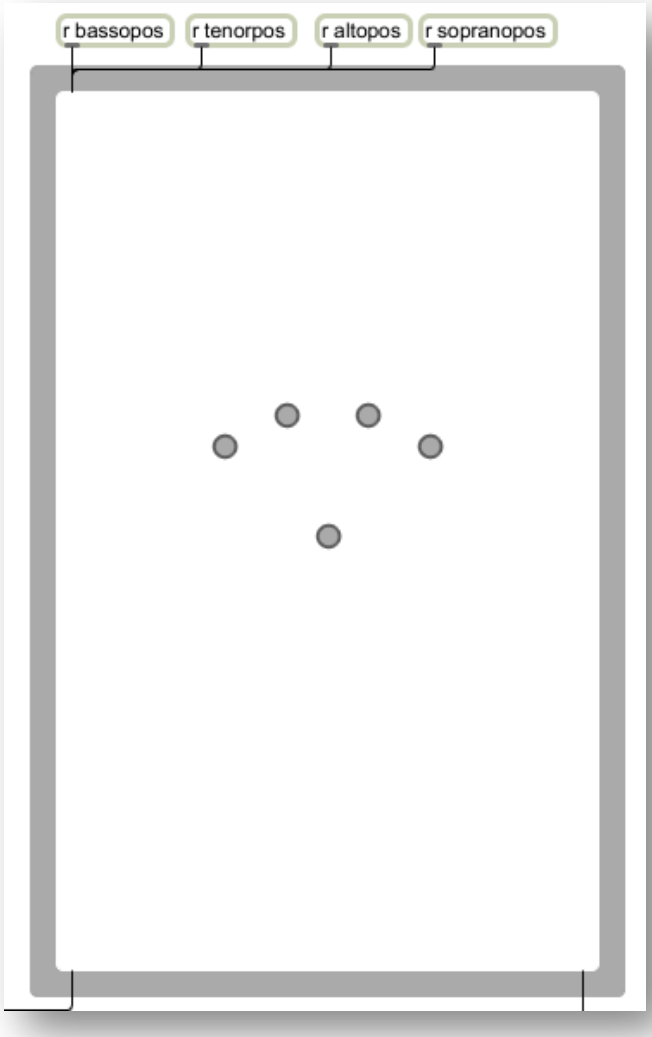
[69] Akustisk musik, forstået som musik der ikke benytter elektricitet som led i lydfrembringelsen – egentlig også et underligt ord for dette fænomen.

[70] Altså hvert instrument / hver stemme på et lydspor for sig.

[71] http://www.carlfischer.com/partbypart/index.html



Figur 27. Screenshot fra *Max*. Her ses en grafisk repræsentation af en mulig positionering af fire korstemmer og lytteren i midten (set ovenfra). Rummet er angivet til at være 25 meter bredt, 50 meter langt og 25 meter højt – eksempelvis en kirke. I denne grafiske repræsentation er det muligt at trække lytterens x,y-position rundt med musen (det er sådan set også muligt i lydfortællingen) og derved opleve hvordan det lyder at stå forskellige steder i rummet. Lydkildernes (altså korstemmernes) position kan ikke manipuleres her. Det er altså blot en grafisk repræsentation af den position, der er angivet med tal i hver *SoundSource*.



Figur 28. Her står lytteren foran et kor i vifte-formation.

Da der her er tale om et ganske andet rum end lydfortællingens, benytter jeg her et andet sæt impuls-responser til VST-rumklangen, der varetager den diffuse hale. I dette tilfælde er der tale om impuls-responser, der fulgte med rumklangen, og som er optaget i en kirke. Udregningen af rummets tidlige refleksioner foregår på præcis samme måde som i lydfortællingen – blot med en anden angivelse af rummets dimensioner.

I konstruktionen af koret som lydligt indhold har jeg ikke tænkt den samme nødvendige kobling til det fysiske rum, hvori det opleves. Altså, jeg har ikke tænkt *augmented reality* som værende en essentiel præmis i denne sammenhæng. Måske fordi vi her har at gøre med musik. Musikken, der i dag er så stærkt bundet op på massedistribution, og som i den grad er løsrevet fra en kobling til et specifikt sted – vi har den med os i ørerne overalt, og den udfolder sig i, eller måske i højere grad former, hvilket som helst rum, vi end måtte befinde os i. Vi er så vant til musikken som den akusmatiske lyd, der kan udfolde sit rum overalt. Men når dette er sagt, vil jeg nævne, at på trods af at jeg ikke ser *augmented reality* som en nødvendig præmis her, så kunne det være spændende at eksperimentere med aspektet. Eksempelvis

kunne man lave en installation i en kirke, hvor lytterne kan gå rundt[72] og lytte til dette (virtuelle) kor, der er placeret i kirken. Igen forudsætter en sådan installation, at man kan *tracke* lytternes position, og således må det gå ind under kategorien fremtidigt arbejde.

Erfaringer med oplevelsen af systemet i funktion

De erfaringer, jeg beskriver i dette afsnit, er baseret dels på mine egne oplevelser med at prøve systemet og dels på mine oplevelser med at afprøve det på andre. Disse andre er hovedsagligt venner og bekendte, som har afprøvet systemet i dets forskellige udviklingsstadier og med forskelligt lydligt indhold. I alt er der i omegnen af 10 personer, der har prøvet systemet – nogle af dem flere gange. Der er altså ikke tale om en systematiseret og videnskabeligt funderet test af veldefinerede brugsaspekter, men rettere improviserede og åbne afprøvninger. Hvor dette kan synes at være et relativt løst grundlag at basere sine erfaringer på, vil jeg alligevel mene, at jeg kan præsentere nogle erfaringer af en vis gyldighed.

Indledningsvis vil jeg sige, at jeg generelt har fået positiv respons på brugen af systemet. Folk har generelt syntes, at det har været en interessant lydlig oplevelse. Selvom jeg er klar over, at sådanne tilbagemeldinger fra venner og bekendte skal tages med et gran salt, har jeg fornemmet en ægte interesse, forundring og måske til tider endda begejstring. Generelt er det min fornemmelse, at folk har følt at systemet har 'virket' – altså at lydene er blevet oplevet som værende realistisk positioneret i det virtuelle/augmenterede rum. Der har naturligvis også været enkelte situationer, hvor lytteren har følt et brud på illusionen – altså situationer hvor lyden opfører sig i strid med lytterens forventninger. Den præcise afdækning af alle årsagerne til sådanne brud kræver yderligere brugertests og udforskninger. Dog er det min tese, at nogle af dem kan forklares med, at lytteren ikke benytter sin egen HRTF. Oftest har testpersonerne faktisk blot benyttet en tilfældig HRTF fra databasen. Dette simpelthen fordi det ikke altid har været praktisk muligt at foretage præcise målinger af lytternes ører og ydre anatomi. I fremtidige udforskninger vil det naturligvis være relevant at finde frem til den mest relevante lytter-HRTF og eventuelt lave en sammenlignende test for at afgøre, om der er kvalitativ forskel. Andre af disse brud på illusionen kan forklares med simple nedbrud i applikationen, og den sidste årsags-tese, jeg vil nævne, er, at simuleringen af rummets akustik ikke i alle tilfælde har været tilstrækkelig – et aspekt, jeg kommer nærmere ind på nedenfor.

En af de mere specifikke erfaringer, jeg har gjort mig, drejer sig om (bogstaveligt talt) det, at lytteren kan interagere med det lydlige indhold ved at ændre orientering. Altså det, at lydbilledet tilpasser sig lytterens hoveds bevægelser, som det gør det i 'den umedierede virkelighed'. Jeg havde, som nævnt, en forventning om, at det at inddrage *head tracking* i den binaurale syntese ville gøre oplevelsen meget mere

[72] Iført trådløse hovedtelefoner.

levende og realistisk for lytteren. Det har imidlertid været slående, hvor lidt de forskellige testpersoner rent faktisk har bevæget hovedet. Ofte har jeg måttet sige til vedkommende: *du kan jo også prøve at dreje hovedet*. Dette på trods af at jeg forinden har informeret testpersonerne om systemets funktionalitet. Det har især været i forbindelse med lydfortællingen som lydligt indhold, at jeg (ikke) har oplevet dette. En af grundene, som jeg ser det, er, at lytteren i lydfortællingen er placeret i en stol. Og når vi sidder sådan fikseret i en stol, er det ikke så oplagt at bevæge hovedet. En anden grund er nok hovedtelefonernes fysiske udformning, som jo, ikke mindst i kraft af mine modifikationer, er/kan føles bevægelsesbegrænsende. Dertil kommer, at vores generelle erfaringer med hovedtelefoner er, at det altså ikke gør nogen forskel, hvilken vej vi vender – det kan være meget svært at omdefinere vores forventninger til et medie / en teknologi, når først vi har lært *et* sæt regler. I forbindelse med koret som lydligt indhold lod jeg testpersonerne stå op, og her var bevægelseslysten større, om end stadig ikke prangende. Når jeg selv har afprøvet systemet, har jeg bevæget hovedet relativt meget, men dette skyldes nok, at jeg som udvikler har et naturligt fokus på, hvorvidt det nu også virker, som det skal. Og derudover har jeg et indgående kendskab til, hvad systemet kan, og vil gerne udforske dette. Men et helt essentielt spørgsmål i forbindelse med bevægelseslysten er: *hvad skal motivere brugerne til at bevæge hovedet?* Der skal være en gevinst for brugerne ved at foretage en bevægelse, og de gevinster, der er i systemets nuværende udformning, står muligvis ikke mål med indsatsen. I hvert fald nok ikke i forbindelse med lydfortællingen. Jeg kan godt hos brugerne opleve en smule fascination over teknologien: *nåh ja, lyden af køleskabet kommer jo fra den samme retning, selvom jeg bevæger hovedet…* – en fascination, som jeg jo også selv deler. Men en umiddelbar fascination er ikke nok til på længere sigt at motivere bevægelse. I forbindelse med koret som lydligt indhold oplever jeg, både hos mig selv og hos andre, en større gevinst ved hovedbevægelse. Som nævnt skyldes dette nok dels brugerens her frit stående position, men jeg tror også, at det skyldes det lydlige indholds karakter. I koret har vi at gøre med et umiddelbart æstetisk tilfredsstillende lydligt udtryk. Vi kan eventuelt placere lytteren i midten af koret, hvorved vedkommende er omsluttet af en, synes jeg, smuk samklang af stemmer. Her føler jeg i hvert fald selv en lyst til at udforske dette musikalske udtryk. I lydfortællingen er vi ikke på samme måde motiveret af et umiddelbart smukt udtryk, og der mangler måske til en vis grad det element, der i givet fald skulle motivere bevægelsen.

Men tilfører head tracking så ikke oplevelsen en øget realisme?

Hvis brugeren slet ikke bevæger hovedet, er *head tracking*'en naturligvis ikke i stand til at tilføre oplevelsen noget som helst. Men det er sjældent, at brugeren slet ikke bevæger hovedet, og jeg vil mene, at der i hvert fald er tale om, at det tilføjer et realisme-*potentiale*. Og blot det at lydbilledet uden *head tracking* forbliver statisk ved eventuelle hovedbevægelser, kan vel betragtes som et brud med realismen. Omvendt kan man også argumentere for, at vi efterhånden er så gode til at indgå på præmisserne for de mange forskellige typer af medierende lag, vi bliver præsenteret for. I sådan en argumentation kunne det pointeres, at normale hovedtelefoner er så velkendt et medierende lag, at det bliver 'transparent', og at

man altså ikke kan tale om noget brud med realismen i deres manglende tilpasning til lytter-orientering. Artiklen *Direct Comparison of the Impact of Head Tracking, Reverberation, and Individualized Head-Related Transfer Functions on the Spatial Perception of a Virtual Speech Source*[73] beskriver en undersøgelse af forskellige elementers betydning for forskellige aspekter af den rumlige oplevelse af virtuelle lydkilder i et system som mit. Disse elementer er blandt andet rumsimulering, individualiserede HRTF og *head tracking*. Med hensyn til *head tracking* konkluderes blandt andet følgende:

> *The inclusion of head tracking will significantly reduce reversal rates, also by a ratio of about 2:1, but does not improve localization accuracy or externalization.* [Beagault m.fl. : *Direct Comparison of the Impact of …*]

Der beskrives altså her, hvordan *head trackingen,* i den pågældende undersøgelse, faktisk halverer de såkaldte *reversal rates*. *Reversal rates* i denne sammenhæng er situationer, hvor lytteren vurderer, at lyden kommer fra *en* retning, men at den i virkeligheden (den virtuelle virkelighed) kommer fra en slags spejling af den vurderede retning. Det er især tilfælde, hvor lyden kommer lige forfra eller lige bagfra, der kan være tvetydige, men også lyd fra andre retninger kan forveksles med positioner, som har, tilnærmelsesvis, tilsvarende ITD og IID. Sådanne forvekslinger/tvetydigheder forekommer også i vores virkelige virkelighed, men i sådanne tvivlssituationer kan en lille hovedbevægelse altså være med til at bestemme hvilken 'retnings-hypotese', man skal stole mest på[74]. Citatet konkluderer dog også, at den generelle nøjagtighed i vurderingen af lydkildernes position, ikke blev forbedret af inddragelsen af *head tracking*, og ej heller blev eksternaliseringsfornemmelsen. Andetsteds i artiklen konkluderes, at der faktisk ikke var nogen af de testede elementer (rumsimulering, individualiserede HRTF og *head tracking*) der havde nogen nævneværdig indflydelse på oplevelsen af realisme (testpersonerne blev adspurgt om deres oplevelse af realisme i de forskellige tests). Dog erkender artiklens forfattere, at dette kan skyldes, at realismebegrebet kan være svært at forholde sig til, og at testpersonerne muligvis ikke har samme opfattelse af, hvad realisme er.

Mine erfaringer med det system, jeg har udviklet, er imidlertid, at både rumsimulering og *head tracking* spiller en forstærkende rolle i oplevelsen af lydens realisme. Igen vil jeg pointere, at mine erfaringer ikke bygger på en systematiseret, sammenlignende undersøgelse, som det er tilfældet med den nævnte artikel. Mine erfaringer med rumsimulering er, at den faktisk spiller en større rolle, end jeg først havde antaget. Et par af mine testpersoner bad mig skrue højere op for rumsimuleringens lydstyrke, idet de fornemmede, at det gav et mere realistisk lydbillede. Rumsimuleringen bør nok indtænkes som en helt

[73] Begault, Lee, Wenzel og Anderson.

[74] Ud over muligheden for at benytte hovedbevægelser til at afkode hvilken position, der er den rigtige (af de mulige hypoteser), benytter vi også de informationer, der ligger i den måde lyden bliver reflekteret i rummets flader og genstande. Dette pointeres af Zotkin m.fl. i artiklen *Rendering localized spatial audio in a virtual auditory space*. En pointe, der igen indikerer vigtigheden af simuleringen af rummets akustik.

grundlæggende præmis for systemet – dette i højere grad, end jeg har gjort. I mit system har jeg taget udgangspunkt i den *head tracking*-tilpassede binaurale syntese, mens mit fokus på rumsimuleringen er kommet lidt i anden række. Mine erfaringer med *head tracking* er omvendt, at den muligvis spiller en lidt *mindre* rolle, end jeg først havde antaget. Denne konklusion drager jeg i høj grad på baggrund af testpersonernes manglende lyst til at bevæge hovedet (hvilket som nævnt var mest udtalt i forbindelse med lydfortællingen). Dog vil jeg vurdere, at selvom dens rolle muligvis er mindre end først antaget, så bidrager *head tracking*'en stadig positivt til oplevelsen af realisme. Dette ikke mindst, fordi jeg, som nævnt, simpelthen anser det for at være mere realistisk, at lydbilledet ændrer sig, når lytteren drejer hovedet, frem for at det forbliver statisk. Det er naturligvis ikke helt ligegyldigt, hvorvidt disse ændringer af lydbilledet forekommer troværdige, men blot det, at der *er* en interaktion/frihed, tilføjer en ekstra dimension, som øger ligheden med lydlig perception i/af 'den umedierede virkelighed'.

Tidligere beskrev jeg, hvordan visionen for projektet omfatter, at lytteren kan bevæge sig rundt i rummet og lytte til forskellige lydkilder. Det har dog ikke været muligt for mig, inden for specialets rammer, at realisere positions-*tracking* af lytteren, idet det er vanskeligt, og et meget stort arbejde at konstruere en stabil, præcis og hurtig *tracking* af menneskers position. Her udtaler jeg mig både på baggrund af egne erfaringer[75], bekendtes erfaringer[76] og på baggrund af en indledende research af teknologiske muligheder i forbindelse med positions-*tracking*. Det har dog været vigtigt for mig at konstruere systemet således, at det er umiddelbart klart til at modtage lytterens dynamiske positions-koordinater fra en eventuel fremtidig positions-*tracking*. Som systemet er nu, kan lytterens position angives manuelt. Mine erfaringer med systemet i brug synes imidlertid at understrege, at det ville være en betydelig udvidelse af det oplevelsesmæssige potentiale, hvis lytterens position blev inddraget som dynamisk input. Da jeg testede systemet med koret som lydligt indhold på en af mine bekendte, begyndte han at gå rundt i rummet for at forsøge at komme tættere på nogle af sangerne. Jeg måtte forklare ham, at systemet altså ikke kunne 'se', at han ændrede position med sin krop. Dog kunne han få lov til at ændre sin position i det virtuelle rum ved at trække en markør (en slags lytter-avatar) rundt på computerskærmen med musen. Hvor dette lod til at give en vis tilfredsstillelse, er der ingen tvivl om, at det ville være mere tilfredsstillende, hvis interaktionen foregik transparent, med lytterens krop. Implementeringen af positions-*tracking*, ville sandsynligvis give en større motivation for lytter-bevægelse end orienterings-*trackingen*. Denne antagelse begrunder jeg bl.a. i, at lydbilledet forandrer sig mere drastisk ved ændring af position end ved ændring af orientering. Ved ændring af orientering 'drejes' lydbilledet godt nok, og vi er måske i stand til at foku-

[75] Mine egne erfaringer i denne sammenhæng stammer hovedsagligt fra projektet SoundPeople, som jeg udviklede i samarbejde med Sune Hede og Helga Rosenfeldt-Olsen. Se evt. http://denevigebeta.blogspot.com/2009/11/sound-people.html

[76] To af mine bekendte, Sune Hede og Lasse Knud Damgaard, udviklede i deres specialeprojekt en positions-*tracking* via infrarøde *leds*. En omfattende og kompliceret sag at sætte op – også selvom jeg ville kunne tage udgangspunkt i deres system. Se evt. http://www.cycling74.com/forums/topic.php?id=28422

sere vores lytning i en given retning, men der sker ikke den helt store forandring i, hvilke lyde vi kan høre. Ved ændring af position er vi i langt højere grad i stand til at komme tæt på specifikke lydkilder, og vi kan måske høre lyde, vi ikke kunne høre før. Det at lytteren ved ændring af position kan opleve nye lyde, vil altså, efter min overbevisning, være en større motivation til lytter-bevægelse end de ændringer, der sker ved ændring af lytter-*orientering*. Dette ligger umiddelbart i forlængelse af mine tanker i afsnittet *Komposition i fire dimensioner*, hvor jeg påstår, at *hvornår* og *hvad* generelt synes vigtigere for os end *hvor*. I denne sammenhæng giver positionsændring mulighed for en større udforskning af *hvad*, mens orienteringsændring i højre grad lægger sig til bestemmelsen af *hvor*.

Perspektiver og fremtidigt arbejde

På baggrund af mine erfaringer fra dette projekt vil jeg konkludere, at der bestemt er et potentiale i kombinationen af binaural syntese og *head tracking* (*tracking* af lytterens hoveds *orientering*) med henblik på at formidle virkelighedstro 3D-lyd. Dog synes det også klart, at potentialet sandsynligvis ville blive betydeligt større, hvis *tracking* af lytterens *position* blev inddraget som input i et sådant system. Den dynamiske tilpasning af lydbilledet til lytterens *orientering* viste sig muligvis mindre interessant for lytteren, end jeg havde forventet. Men i denne forbindelse er det min tese, at tilpasningen til lytter-orientering kombineret med tilpasning til lytter-position er en frugtbar sammensætning, der, så at sige, rummer et større potentiale end summen af enkeltdelene. Dette forstået sådan, at positions-tilpasningen vil gøre orienterings-tilpasningen mere interessant og vice versa. Simpelthen fordi lytteren dermed oplever fuld 'betydningsfuld' bevægelsesfrihed. Et af mange områder for fremtidigt arbejde er altså at inkorporere positions-*tracking* i systemet. Dette indebærer blandt andet research af forskellige mulige tekniske løsninger, implementering af valgt løsning samt at indsamle erfaringer med dette system i brug – gerne i en iterativ proces.

Af andre aspekter af fremtidigt arbejde kan nævnes udviklingen af et mere intuitivt og praktisk kompositionsværktøj, som beskrevet i afsnittet *Komposition i fire dimensioner*. Desuden er det relevant i fremtidige udviklinger, at finde en løsning på problematikken vedrørende håndtering af lydkilder med en væsentlig rumlig udstrækning – herunder ikke mindst sammensatte ambiente lyde.

Som nævnt i afsnittet *Metode til binaural syntese* vil det være interessant, at foretage en kvalitativ sammenligning mellem binaural syntese, der foretages i henholdsvis tidsdomænet og frekvensdomænet.

Endnu et område for fremtidigt arbejde er, at udforske systemets potentiale i andre sammenhænge end i forbindelse med lydfortællinger og musik. Ikke forstået på den måde, at jeg anser systemets potentiale i disse sammenhænge for fuldt afsøgt – på ingen måde. Men det kunne være interessant at udforske tek-

nologien i sammenhæng med eksempelvis film eller computerspil. På computerspilsområdet er der allerede forsket en del i binaural lyd, og der findes allerede kommercielle produkter, der inkorporerer binaural lyd[77]. Dog er det stadig et relativt nyt område, som bestemt kan udforskes yderligere. Jeg er ikke umiddelbart stødt på eksempler på brug af binaural lydprocessering i forbindelse med filmmediet. Der ligger muligvis også en problemstilling i, at film ofte er en social begivenhed – man ser ofte film sammen med andre – og i denne sammenhæng kan brug af hovedtelefoner føles/være isolerende, hvilket reelt set er en problematik i de fleste sammenhænge. Dog vil jeg stadig mene, at det vil være et interessant område at udforske.

I fremtiden ville det være relevant at foretage flere brugstests – disse gerne med henblik på at teste og forfine de specifikke aspekter af systemet. Disse aspekter er eksempelvis rumsimuleringen, selve den binaurale syntese, interpoleringen af HRTF med mere.

Det er netop karakteristisk ved mit og lignende systemer, at de består af mange specifikke teknologier, der kan finpudses hver for sig. Det kan betragtes som en kæde af led, der alle gerne skal være så stærke som muligt for at få illusionsnummeret til at virke. Simuleringen af rummets akustik er et af disse led, som har vist sig at være temmelig essentielt, og som i fremtiden bestemt fortjener mere opmærksomhed, end jeg har givet det indtil nu. Men alle kædens led kan forstærkes og vil uden tvivl blive det i fremtidige projekter i takt med nye landvindinger på de forskellige teknologiske fronter. Den teknologiske søgen mod transparent gengivelse af en fiktiv eller faktisk virkelighed – realisme – er en uendelig rejse. Der kan altid perfektioneres yderligere.
I denne forbindelse kunne man måske også spørge sig selv, om det er disse forsøg på en fuldstændig naturtro gengivelse af det lydlige rum, der resulterer i den størst mulige realisme. Og i forlængelse heraf kan man spørge, hvorvidt det er den realistiske gengivelse, der fordrer den største indlevelse hos lytteren, der giver den mest interessante oplevelse, der rummer det største formidlingspotentiale osv.. Eller rummer subjektivt bearbejdede/overdrevne/æstetiserede/abstrakte lydlige udtryk større potentialer i disse sammenhænge? Man kan måske også spørge sig selv, om der ikke ligger et underligt paradoks i det faktum, at det lydlige indhold i systemer som mit nødvendigvis må være stærkt konstrueret med henblik på at opnå realisme i oplevelsessituationen. Alle sådanne kritiske spørgsmål lægger imidlertid op til en længere udredning og diskussion af begreber som lydlig realisme, objektivitet og subjektivitet, mediering og det medierede, kunst og æstetik med mere. Og det ligger ikke indenfor rammerne af dette speciale, at udfolde en sådan diskussion. Her vil jeg nøjes med at fremlægge det synspunkt at selvom en blind stræben efter en form for lydlig realisme nok ikke er svaret på alle vores problemer, så er det alligevel

[77] Se eksempelvis http://www.actionreactionlabs.com/ghost.php , http://www.papasangre.com/

nyttigt af have teknikker til troværdig formidling af lydens rumlige karakter med i vores værktøjskasse. Det bliver spændende at følge, hvad der sker på området i de kommende år.

English summary

English title: *Realistic mediation of virtual sound sources. An experimental exploration of binaural synthesis and head tracking, and of how this combination can be used in a system for communication of realistic 3D sound.*

Since the first systems for storing and playing back sound were invented in the second half of the 19th century, we have become increasingly better at reproducing sound, resulting in an increasing degree of fidelity. However, we still have difficulties reproducing the spatial properties of a sound phenomenon – the experience of the exact spatial location of the individual sound sources.
This paper describes the development of a system that aims at doing exactly this. It uses binaural synthesis, head tracking and acoustics simulation in an attempt to create a realistic sense of the spatiality of sound played though headphones – this, within an experience-oriented context. It is not the first system of its kind. These technologies have been combined before in different systems for reproducing 3D sound for different purposes. The system described in this paper is based on the knowledge and experiences gained in the development of such similar systems. However, I also hope to be able to present solutions and experiences that might be a help to any future work in this area.

The following is a dense summary of the system's construction and functionality.

The system is built in the programming environment *Max*, also known as *Max/MSP*[78].
The binaural synthesis is based on convolution with impulse responses from *the CIPIC HRTF database*[79]. Inspired by the article, *Rendering Localized Spatial Audio in a Virtual Auditory Space*[80], it was decided to perform the convolution in the frequency domain, since this is 'cheaper' than time domain convolution processor-wise. The impulse responses from the CIPIC database were, beforehand, interpolated in *Matlab*[81], by first aligning the responses in time, and then interpolating (still in the time domain). The cropped time delay of each impulse response is saved separately and re-implemented real-time (after convolution) where an interpolation is performed on this parameter. Then the responses were then converted to the frequency domain via FFT (so that this should not be done real-time) with a FFT size of 2048 and exported to *Max*. In *Max*, the data of a given CIPIC *subject* is 'held' by *jit.matrix* objects. The binaural syn-

[78] http://cycling74.com
[79] http://interface.cipic.ucdavis.edu/sound/hrtf.html
[80] Zotkin, Duraiswami, and Davis.
[81] http://www.mathworks.se/products/matlab/index.html

thesis is performed in *pfft~* objects with an overlap of 4 windows. The *pfft~* objects obtain the HRTFs by pointing to relevant data in the *jit.matrix* objects.

Head tracking of the listener's orientation was implemented in order to compensate for the fact that headphones, and thereby the whole sound environment, move along when the listener rotates his/her head. The *x-IMU*[82] sensor was mounted on top of headphones and special Java externals based on the *x-IMU API*[83] were created in order to be able to communicate with the sensor in *Max*.

Simulation of room acoustics was implemented to achieve a realistic sense of the sound sources being placed in the room outside the listener's head, and also to support localization capabilities. Since it is mostly the room's early reflections that change with changing listener rotation and position[84], only these early reflections were calculated dynamically and real-time. Reflections up to 2nd order were calculated for each sound source using the image-source method as described in *Modeling Techniques for Virtual Acoustics*[85]. As it is not possible (processor-wise) to perform binaural synthesis on each of the 36 reflections from each sound source, all reflections were summed per surface in the room before performing *one* binaural synthesis for each surface in regard to the surface's position in relation to the listener's position and orientation. This resulted in six syntheses overall for room simulation instead of 36 for each sound source in a regular box-shaped room. The 'tail' of the reverberation was handled by a static VST convolution reverb.

Two examples of sound content were created with the aim of investigating the system's (and the technology's) potential ability to communicate an experience to the listener. One of these is a kind of sound narrative, where my grandmother talks about her life – both her present life and experiences from her youth. Occasionally, the narrative 'dives into' specific narrated situations, creating a kind of auditory flashbacks. An important aspect of this narrative as sound content is Augmented Reality. The listener will experience the narrative in the same physical room as the basic 'room of narration'. The idea was that the listener should have an impression of the physical room being sonically augmented in such a way that, for example, my grandmother is perceived as sitting in the empty chair that is physically present in the room.

The other sound content that was created was an example of how music could be mediated in such a system. More precisely, it consisted of a four-voiced choir (SATB) that was positioned as individual sound sources in the system/room.

---

[82] http://www.x-io.co.uk/node/9

[83] http://www.x-io.co.uk/res/sw/ximu_api_13_1.zip

[84] Zotkin, Duraiswami, and Davis.

[85] Savioja

The following is a short summary of some of my experiences with the system in use. The experiences are not based on scientific and systematic user tests of well-defined aspects. Rather, they are based on open and improvised try-outs. Beside myself, approximately 10 others – primarily friends and family – have tested the system. Some of them more than once. I know that this can seem as quite a fragile foundation to base my experiences upon. However, I still believe that I can present experiences with some validity.

Generally, the users seemed to react positively to the tests – the system generally and overall seemed to work as intended. I know that friends and family generally would do so regardless of their actual experience, but I also feel capable of decoding to what extend a positive response is pretended. Of course there were situations with 'breakdowns' in the illusion of realism. One of the reasons for such breakdowns might be that the users did not use personalized HRTFs. To get a more precise knowledge of what caused the breakdowns, more user tests must be conducted.

A more specific experience was a noticeable lack of user head movements. The users did generally not rotate their head, even though that I, beforehand, had explained the system's functionality. One of the reasons for this might be that our general experience with headphones is that rotating your head has no influence on the sound, and another reason is probably that the listener doesn't 'gain' enough by rotating his/her head.

The tests of the system seem to emphasize that an implementation of *position* tracking of the listener would greatly enhance the system's potential in an experience-oriented context. This would, I presume, also make the effects of listener *rotation* more meaningful and appealing to the users.

The tests also pointed to the importance of the simulation of room acoustics in such a system. In fact, it seemed to be more important than I initially assumed, and therefore it deserves more attention than I have given it during the development of the current system.

Concluding this summary, I will list some aspects of relevant future work.

One such aspect is to implement and gain experiences with position tracking of the listener. Furthermore, it could be interesting to investigate the technology in other contexts such as movies and computer games. It will also be relevant to examine the quality of binaural synthesis in the frequency domain compared to synthesis in the time domain, as well as examining the quality of different ways of performing frequency domain synthesis. Generally, it will be relevant to conduct a lot more, and perhaps more systematic, user tests. These maybe focused on specific aspects of the system, for it does consist of a lot of different technologies and techniques that can be improved individually. And future developments in technology will certainly make it possible to create more realistic, more immersive experiences of the spatial aspects in sound reproduction.

Post Scriptum

Jeg vil gerne sige tak til alle, der har støttet og hjulpet mig i denne specialeproces. Herunder særligt familie og venner, som har hevet mig op af det hul, jeg til tider fik gravet mig ned i.

Også tak til alle, der har hjulpet mig med at afprøve systemet, og til folket på speciale-gangen, som gjorde det lidt sjovere at skrive speciale.

Desuden vil jeg gerne takke Sune Hede, som har været en god sparringspartner i udviklingen af systemet, Peter Fjordbak, som agerede 'ung mand under krigen' i lydfortællingen, Frederikke og Nina, som agerede 'unge kvinder', min mormor, som fortalte, CIPIC for at stille deres database til rådighed, Mr. Algazi for hjælp med interpolering og Seb Madgwick for hjælpsomhed i forbindelse med x-IMU.

Litteraturliste

Bøger

- Begault, Durand R. : *3-D Sound for Virtual Reality and Multimedia*. Academic Press Professional, 1994.
- Roads, Curtis : *The Computer Music Tutorial*. The MIT Press, 1996
- Smith, Steven W. : *The Scientist and Engineer's Guide to Digital Signal Processing*. E-bog: http://www.dspguide.com/ Besøgt i perioden marts 2011 til februar 2012.
- Stern, R. M.; Wang, DeL; Brown, G. : Binaural Sound Localization. Kapitel 5 i: *Computational Auditory Scene Analysis*. Redigeret af DeL. Wang og G. Brown. Wiley/IEEE Press, 2006.

Artikler

- Algazi, V. Ralph; Duda, Richard O. : Effective Use of Psychoacoustics in Motion-Tracked Binaural Audio. I: *Tenth IEEE International Symposium on Multimedia, 2008*, s. 562-567.
- Algazi, V. Ralph; Duda, Richard O. : Immersive Spatial Sound for Mobile Multimedia. I: *Proceedings of the Seventh IEEE International Symposium on Multimedia*, 2005.
- Begault, Durand R.; Lee, Alexandra S.; Wenzel, Elizabeth M.; Anderson, Mark R. : Direct Comparison of the Impact of *Head Tracking*, Reverberation, and Individualized Head-Related Transfer Functions on the Spatial Perception of a Virtual Speech Source. I: *AES 108th Convention, Paris, February 2000*.
- Breinbjerg, Morten; The Aesthetic Experience of Sound – Staging of Auditory Spaces in 3D Computer Games. I: *On-line Proceedings of The Aesthetics of Play Conference, Bergen 13. – 14. Oktober 2005*.

- Carty, Brian; Lazzarini, Victor : Frequency-Domain Interpolation of Empirical HRTF Data. I: *126th AES Convention, Munich, May 2009*.
- Chen, Liang; Hu, Hongmei; Wu, Zhenyang : Head-Related Impulse Response Interpolation in Virtual Sound System. I: *Fourth International Conference on Natural Computation, 2008. ICNC '08.*
- Cheung, Steve; Kramer, Lorr; Smyth, Michael; Smyth, Stephen : A Virtual Acoustic Film Dubbing Stage. I: *AES Conference Papers: 40th International Conference: Spatial Audio: Sense the Sound of Space*, 2010.
- Falch, Cornelia; Noisternig, Markus; Warum, Stefan; Ho□ldrich, Robert : Room Simulation for Binaural Sound Reproduction Using Measured Spatiotemporal Impulse Responses. I: *Proc. of the 6th Int. Conference on Digital Audio Effects (DAFX-03), London, UK, September 8-11, 2003*
- Freeland, Fabio P.; Biscainho, Luiz W. P.; Diniz, Paulo S. R. : Efficient HRTF Interpolation in 3D Moving Sound. I: *AES Conference Papers: 22nd International Conference: Virtual, Synthetic, and Entertainment Audio*, 2002.
- Fö□rster, Marius: *Auralization in Room Acoustics*. Bacheloropgave ved Graz University of Technology, 2008.
- Jin, C., Tan, T., Leung, J., Kan, A., Lin, D., Van Schaik, A., Smith, K., m.fl. : *Real-time*, Head-tracked 3D Audio with Unlimited Simultaneous Sounds. I: *International Conference on Auditory Displays (2005).*, s. 1-4
- Kulkarni, A.; Isabelle, S. K.; Colburn, H. S. : On The Minimum-Phase Approximation of Head-Related Transfer Functions. I: *IEEE ASSP Workshop on Applications of Signal Processing to Audio and Acoustics, 15-18 Oct 1995.*
- Laitinen, Mikko-Ville; Pulkki, Ville: Binaural Reproduction for Directional Audio Coding. I: *IEEE Workshop on Applications of Signal Processing to Audio and Acoustics, 2009. WASPAA '09.*, s. 337-340, 2009.
- Lalime, Aimee L. : *Development of a Computationally Efficient Binaural Simulation for the Analysis of Structural Acoustic Data*, Master of Science In Mechanical Engineering ved Virginia Polytechnic Institute and State University, 2002.
- Ludovico, Luca A.; Mauro, Davide A.; Pizzamiglio, Dario : Head in Space: A Head-*tracking* Based Binaural Spatialization System. I: *Sound and Music Computing Conference 2010*
- Menzer, Fritz : *Binaural Audio Signal Processing Using Interaural Coherence Matching*, Doktorafhandling ved E□cole Polythecnique Fedérale de Lausanne, 2010
- Newman, Duncan James Philip; *Head-Related Transfer Functions – And Their Application in: Real-Time Generation of Binaural Spatial Audio*. Bachelor of Engineering ved The University Of Queensland, 2003

- Sandvad, Jesper : Dynamic aspects of auditory virtual environments. I: *100th AES Convention, Copenhagen, May 11-14, 1996*
- Sandvad, Jesper : Dynamic Aspects of Auditory Virtual Environments. I: *100th AES Convention, Copenhagen, May 1996*.
- Savioja, Lauri : *Modeling Techniques for Virtual Acoustics*. Doktorafhandling ved Helsinki University of Technology, december 1999.
- Tikander, Miikka; Karjalainen, Matti; Riikonen, Ville : An Augmented Reality Audio Headset. I: *Proc. of the 11th Int. Conference on Digital Audio Effects (DAFx-08), Espoo, Finland, September 1-4, 2008*.
- Toole, Floyd E. : Direction and Space – the Final Frontiers. Det har ikke været muligt at finde frem til det medium, artiklen er blevet bragt i. Artiklen er fundet på denne webadresse: http://www.harman.com/EN-US/OurCompany/Technologyleadership/Documents/White%20Papers/HowManyChannels.pdf . Webadressen er blevet besøgt i perioden marts 2011 til februar 2012.
- Wenzel, Elisabeth M.; Foster, Scott H. : Perceptual Consequences of Interpolating Head-Related Transfer Functions During Spatial Synthesis**.** I: *Applications of Signal Processing to Audio and Acoustics, 1993. Final Program and Paper Summaries., 1993*, s.102-105.
- Zotkin, D.N.; Duraiswami, R.; Davis, L.S. : Rendering localized spatial audio in a virtual auditory space. I: *Multimedia, IEEE Transactions on* , vol.6, no.4, s. 553- 564, aug. 2004

Web-ressourcer

Det gælder for alle webressourcer, at de er blevet besøgt i perioden marts 2011 til marts 2012.

- *AmmSensor*: http://www.ammsensor.com/ .
- *Binaural Tools*: http://www.ece.ucdavis.edu/binaural/binaural_tools.html .
- *Head in Space*: http://sites.google.com/site/dariopizzamiglio/projects/head-in-space .
- *Brown University->Division of Engineering->Introduction to Engineering->Teach Yourself Vectors->APPENDIX II. ADVANCED TOPIC: Non-Cartesian vector components:* http://www.engin.brown.edu/courses/en3/Notes/Vector_Web2/Vectors6a/Vectors6a.htm
- *Carl Fischer Music* – Part by Part: http://www.carlfischer.com/partbypart/index.html
- *Ghost Dynamic Binaural Audio* : http://www.actionreactionlabs.com/ghost.php
- *Introduction to Sound Processing -> 3.6: Spatial sound processing* : http://www.faqs.org/docs/sp/sp-72.html
- *Max 5 Help and Documentation* (tilsvarer den indbyggede dokumentation i *Max 5*) : http://cycling74.com/docs/max5/vignettes/intro/docintro.html .

- *OpenGL:Tutorials:Using Quaternions to represent rotation - GPWiki* : http://content.gpwiki.org/index.php/OpenGL:Tutorials:Using_Quaternions_to_represent_rotation
- *Papa Sangre* : http://www.papasangre.com/
- *Polhemus*: http://polhemus.com/ .
- *The CIPIC HRTF Database*: http://interface.cipic.ucdavis.edu/sound/hrtf.html .
- *Unity 3D*: http://unity3d.com/
- *Wikipedia->Quaternion* : http://en.wikipedia.org/wiki/Quaternion .
- *Wikipedia->Quaternions and spatial rotation* : http://en.wikipedia.org/wiki/Quaternions_and_spatial_rotation .
- *x-io technologies*: http://www.x-io.co.uk/ .