

Sonifikation og spatial interaktion

Sebastian Thinggaard Rostved
Aarhus Universitet
Aarhus, Denmark
201702966@post.au.dk

ABSTRACT

This paper examines how to implement sonification in interactive art installations and how the room can be used as an effect and as something that can be analysed visually with computer vision technology. The interactive art installation *Khaos* address these approaches and this raises, with *Khaos* as an example, a discussion about what interactive art really is and the problems related to that, including technology and definitions.

Keywords

Sonification, space, interactive art, technology, computer vision, visualisation

1. INTRODUKTION

Når man taler om rum i en lydlig spatial kontekst, befinder vi os næsten altid i et rum. Det er ligegyldigt om man står midt på en mark, i et lydstudie eller i et fly; der vil altid være noget som kan reflektere lyd. Der findes dog to undtagelser; det første er anekoiske kamre, som simulerer et uendeligt stort rum, hvor alle lydbølger bliver absorberet af væggene. Den anden undtagelse er vakuum, hvor lyd ikke kan eksistere, da der ikke er noget medie, som det kan transporteres igennem. Hvis man tager udgangspunkt i ikke-lydlig definitioner på rum, kan der dog argumenteres for at der stadig er tale om et rum, da man er omgivet af fire vægge, et gulv og et loft. I forhold til vakuum, er der også tale om et rum, da menneskeskabt vakuum altid foregår i et afgrænset og lufttæt rum. Et naturligt vakuum, som rummet, er i ordets betydning også et rum, og oversat fra engelsk er der også tale om *space*.¹ Ud fra denne tese om, at man altid befinder sig i et lydligt rum, vil jeg i dette projekt undersøge, hvordan man kan bruge et *levende rum* som noget der kan trækkes lydlig betydninger ud af, som kan sonificeres. Med et levende rum, menes der et rum, hvor et sansende subjekt befinder sig i, og subjektet dermed, bevidst eller ubevidst, sanser rummet og interagerer med det. Desuden vil jeg undersøge hvordan subjektets placering i (forhold til) et rum, kan influere hvordan de sanser rummet og dets visuelle og lydlig kilder og installationer.

1.1 Khaos

Med udgangspunkt i denne undersøgelse, har jeg skabt en installation, som hedder *Khaos*. Ordet *kaos* er af mytologisk oprindelse og stammer fra oldgræsk, hvor det betyder "gab" eller "svælg". *Kaos* beskriver en tilstand eller et tomrum, som fandtes før den ordnede verden (kosmos) blev til [1]. Valget af dette navn vil blive behandlet senere i dette afsnit.

Khaos er programmeret i Max MSP, og vil bruge teknologi som computere og webcams. Med *Khaos* vil jeg undersøge hvordan man kan bruge computer vision som en teknologi til at sonificere bevægelsesmønstre og variationer i mængden af bevægelse, samt hvordan sonifikation kan skabe algoritmisk musik.

Når der i de følgende afsnit bliver nævnt sonifikation, er det som en lydlig repræsentation af noget data, som ikke er baseret på tale. Sonifikation bliver brugt i så forskellige ting som filmsoundtracks, hjælpemidler til blinde, videnskabelige forsøg og som lydlig respons i en smartphone eller computer. I dette tilfælde er det ikke meningen af sonifikationen skal være en direkte, praktisk eller videnskabelig lydlig repræsentation af noget data, men et kunsterisk og æstetisk udtryk.

2. BESKRIVELSE

Installationen står i et langt rum, hvor der i hver ende er en stor projekteret skærm, som viser et video-feed fra et webcam. Ved siden af hver skærm er der ophængt et par større stereo-højtalere, som sender lyd og musik ud. Rummet skal være lettere mørkt, for at brugernes opmærksomhed bliver rettet mod skærmene. Den ene skærm viser et live video-feed fra et mikroskop, som forstørret og oplyser indholdet af noget vand fra en sø, potteplante eller lignende. Kilden er underordnet, så længe at vandet indeholder liv i form af protister og flercellede organsimer. Protister er en gruppe af encellede organismer, som indeholder en cellekerne, herunder amøber, bakterier og lignende. Vandets indhold, herunder organismene, bliver vist på stormskærmen, imens videoen-feedet bliver analyseret af en computer, som sonificerer organismernes bevægelser i real-time. Sonificeringen bliver herefter afspillet for publikum. Brugerne kan selv flytte videokameraets placering ved hjælp af et interface, og stoppe det, hvis de ser noget interessant.

Den anden del er en samling af live-webcams, som filmer større gader, specielt vejkryds, hvor der er meget aktivitet. Video-feedet bliver også analyseret og sonificeret af en computer, så brugerne kan se video, samtidig med at de hører sonificeringen. Analysen tracker biler, ud fra bilernes farver, og det samlede aktivitetsniveau i områder på billedet. Med en knap kan brugerne skifte imellem videofeeds, og dermed selv vælge hvad de vil se på.

¹ I Cambridge Dictionary defineres space på følgende måde: "the *area* around everything that *exists*, *continuing* in all *directions*".

Forslag til opstilling

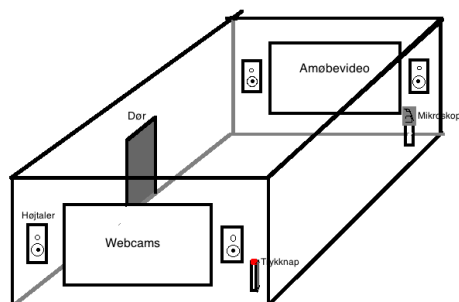


Figure 1. 3D oversigt over rummet.

2.1 Koncept

Idéen med installationen er at sonificere liv og rum på to forskellige planer, og give brugere og publikum en sonificeret oplevelse af de to planer, hvor de kan finde ligheder og forskellige.

På det mikroskopiske plan er det organismernes uforudsigelige bevægelser, som skaber musikken. Myron Kruger beskriver i *Responsive Environments*, hvordan der både er en dirigent, en instrumentalist og en komponist i tilblivelsen af responsive systemer og det kan dermed sammenlignes med forholdene for et musikalsk værk [2]. I et responsivt system er designeren komponisten, som har designet systemet og mulighederne i dette. Brugeren er instrumentalisten som udfører værket, og computeren er dirigenten der styrer udfaldet af brugens interaktioner. På denne måde kan man sige at organismerne er instrumentalisterne, som står for udførelsen af værket. I en musikalsk sammenhæng, kan man også sammenligne dem med komponister, da de også skriver stykket. Organismernes meget forskellige størrelser og bevægelsestempi, er også med til at sonificere dem og skabe et uforudsigeligt, men samlet musikalsk udtryk.

Det andet plan er af en mere relaterbar størrelse, da det indebærer mennesker og menneskeskabt teknologi. Her er trafikken og interaktionen mellem mennesker komponisten. Den komponerede musik ændrer sig i takt med kulturelle fænomener som myldretid, ugedagene, begivenheder, dag og nat og så videre. Folk på gaden interagerer med og skaber stykket ubevidst, da de ikke ved at de bliver filmet oppefra. En af måderne man kan analysere trafik på, er ved hjælp af kaosteori [3]. Måden Max analyserer trafikken på er ikke ved hjælp af kaosteori, men trafik-flowdata er kaotisk, og dette er en af grundene til at jeg har kaldt installationen for *Khaos*. En anden er, at en hel slægt af amøber har det latinske navn *Chaos* og jeg fandt at den mytologiske overgang imellem *Kaos* og *Kosmos*, passer godt sammen med de to planer, set ud fra et evolutionært perspektiv, men også rumligt, da *Kaos* og *Kosmos* er to forskellige rum.

2.2 Teknologi

Kameraerne er hhv. et mikroskop og en håndfuld overvågningskameraer/webcams. Mikroskopet indeholder et

lille videokamera, som kan sende et klart og skarpt billede ud. De to drejknapper på mikroskopet et skiftet ud med to motorer, som styres af et simpelt interface, der kan køre frem og tilbage på x-aksen og y-aksen. Interfacet er udstyret med op/ned/frem/tilbage knapper.

Webcams er opsat rundt omkring ved større trafikale færdselsårer. Hvert kamera sender en video-feed via internettet. Videopløsningen skal være tilpas god til at man kan se hvad der sker på billedet, og så computeren kan analysere det, men personer, biler osv. skal ikke kunne genkendes, og derfor skal kameraerne også hænges tilstrækkeligt højt op.

2.2.1 Max MSP og cv.jit

Video-feedsne bliver analyseret på henholdsvis to computere i Max MSP. I Max bliver videosignalet manipuleret og analyseret ved hjælp af Max's eget videobehandlingsystem kaldet jitter og en samling af objekter til computer vision applikationer kaldet cv.jit [4]. Computer vision betegner algoritmer som giver computere mulighed for at lave intelligente påstande på baggrund af video- og billedmateriale. Computer vision bliver blandt andet brugt til automatisk ansigtsgenkendelse og analyse, analyse af gestik, bevægelser og stedgenkendelse [5]. Eksempler på brug af computer vision er smartphone apps og Xbox Kinect. Analysen af videoen fra de to videokilder, bliver i store træk analyseret på samme måde. Først bliver billedopløsningen nedskaleres, for at spare computerkraft. Herefter sker der to processor. I den første bliver en enkelt farve udvalgt. For eksempel rød for røde biler, eller grøn for grønne biler eller bakterier/plankton. Ved hjælp af jit-objekter bliver den enkelte farve isoleret og billedet gøres sort/hvid. Herefter sørger objekter og indstillinger for, at det kun er ting på billedet i den udvalgte farve som bevæger sig der viser sig på billedet. Til sidst gør et computer vision centroid-objekt, at de får et punkt som følger dem i det koordinatsystem, som er video-billedet. Ud kan man trække punkternes koordinater løbende, og man har dermed information om koordinater og størrelse på henholdsvis røde biler, grønne biler eller bakterier/plankton. Den anden process deler billedet i et matrice med 12 lige store felter, hvor den samlede aktivitet analyseres i hvert felt. Hermed har man både relative tal for aktiviteten i de enkelte felter, men også den samlede aktivitet på hele billedet.

Musik og lyd er også skabt i Max MSP. Bakterier og plankton trackes som punkter og deres koordinater og størrelse/aktivitet bliver mappet til forskellige parametre på en additiv synthesizer i fem stemmer, som består af sinustoner. Deres placering på x-aksen styrer stereopanning, så hvis et punkt er helt ude til højre på x-aksen vil synthesizer lydilden kun være i højre højtaler, hvis den er til venstre på x-aksen vil den være i venstre højtaler og i midten af x-aksen vil dens tilhørende synthesizerstemme være ligeligt fordelt imellem højre og venstre højtaler. Y-aksen styrer tonehøjde, hvor tonehøjden stiger med y-aksens værdi, og størrelsen på punktet styrer synthesizerstemmens amplitude. Fra mikroskopvideofeedet vil der som maksimum være fem af de trackede punkter, som bliver sonificeret på grund af begrænsede mængder computerkraft. St Sonificeringen af de røde og grønne bliver sker tilnærmelsesvist på samme måde, med et par enkelte forbehold. Her er der kun en enkelt stemme for hver bil, som til gengæld er en subtraktiv synthesizer. Værdierne fra y-aksen er vendt om for at skabe en illusion om,

at bilernes lyd stiger i pitch og et lowpassfilter åbner op når de nærmer sig kameraet og falder i pitch og et lowpassfilter lukker, når de bevæger sig væk. Dette skaber en illusion af doppler-effekt.

Aktiviteten i tolv-felts matricen er også sonificeret med synthesizer-stemmer. Videosignalet fra mikroskopet bliver sonificeret igennem tolv subtraktive synthesizerstemmer, hvor aktiviteten i hvert felt i matricen, styrer amplituden på en synthesizerstemme. Alle stemmerne bliver mixet sammen og filteret af et lowpassfilter, som styres af den samlede aktivitet på billedet. Stemmerne er stemt en fis-mol skala. Aktiviteten i videosignalet fra webcams bliver sonificeret af tolv synthesizerloops optaget på en analog synthesizer, med en hel masse rumklang. Aktiviteten i hvert felt styrer amplituden på hver loop. De tolv loops er også stemt i en fis-molskala, så de passer sammen.

2.2.2 Det fysiske rum

Det fysiske rum vil bestå af to computere, som er koblet til to projektorer og to sæt stereohøjtalere i god kvalitet, så man får et godt stereobillede. Højtalere og forstærkere skal indstilles ens, så de ikke unødvendigt spiller højere end hinanden.

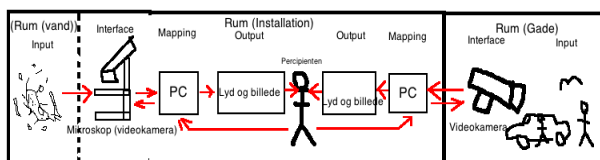


Figure 2. Oversigt over mappings og interaktioner

2.3 Oplevelsen i rummet

Mappingstrategier af sonifikation af rummet i *Khaos* er hovedsageligt metaforiske. Der kan skelnes imellem tre forskellige typer af mappingstrategier [6]:

Symbolisk: Er en abstrakt repræsentation, og skal derfor indlæres. Eksempel: Sirene.

Metaforisk: Er til dels arbitrære, men gør brug af ligheder imellem data og lyd. Eksempel: Nedadgående tone for at repræsentere noget som falder.

Nomisk: Er knyttet til konteksten og refererer til en fysisk aktivitet. Eksempel: For eksempel lyde af regn, hvis det regner eller lyden af et stykke papir som bliver kryllet sammen, når man tømmer papirkurven på en PC.

For som bruger/publikum at kunne følge med i udviklingen på skærmene, uden nogen forudgående kendskab til installationen, og for at ramme et bestemt musikalsk udtryk, er størstedelen af mappingstrategierne i *Khaos* metaforiske og nomiske. For eksempel er bakteriernes placering på y-aksen mappet til pitch (op og ned) og mængden af aktivitet er mappet til amplituden og antallet af synthesizerstemmer som bliver spillet.

Af nomiske mappings, kan der nævnes de røde og grønne biler, hvor det lydige udtryk minder om en doppler-effekt, og placeringen af biler og bakterier på x-aksen, som stemmer overens med deres placering i stereo-feltet.

Alle lydene skaber tilsammen et musikalsk udtryk. Da musikken ikke skabes aktivt af personer, men repræsenterer en form for data, kan den kaldes algoritmisk, og dermed generativ. Hvorvidt den er baseret på stokastiske eller deterministiske principper, kan lægge op til komplekse filosofiske diskussioner, som ikke vil behandles her. Herunder om encellede og flercellede organismer har bevidsthed, og om menneskelig adfærd er deterministisk [3].

Det æstetiske lydige udtryk er præget af synthesizere. Der er en vis form for tonalitet, som stammer fra de mange synthesizer pads fra mappingen af aktiviteter, da alle stemmerne er stemt i en fis-mol. Forskellige bølgeformer på stemmerne som knytter sig til punkterne, gør at man kan skelne imellem de røde og grønne biler, og bakterierne. Selve musikken er meget ambient, da der ikke skabes nogle rytmer, og synthesizerpads og leads, giver også referencer til ambient musik. Rumklang på mikroskopssynthesizeren

Ved at placere de to skærme og højtalere i hver sin ende af rummet, kan der skabes en naturlig overgang imellem de to udtryk, som kan bruges aktivt med fokus på brugere og publikum. Hvis lyd niveauerne mixes rigtigt, kan der skabes en overgang imellem lydtrykket fra højtalerne. Publikum kan derfor bevæge sig imellem de to skærme og rummet kan hermed fungere som en form for cross-fade, hvor man kan skruer op og ned for lyden, men hvor det samlede lydtryk nogenlunde vil være konstant. På denne måde kan rummet bruges aktivt i installationen.

Undervejs kan publikum interagere med de to installationer, med nogle simple interfaces. Hvis der ikke sker noget, eller hvis det keder publikum, kan de ændre billedet, så der sker noget mere. Man kan sammenligne det lidt med at skifte nummer på en CD – det er stadig den samme musik, bare i en lidt anden udgave.

3. DISKUSSION

Når publikum og brugerne interagerer med installationen, vil de gå rundt i rummet og crossfade imellem de to ender, og måske trykke på knapperne, som ændrer videobillederne. Men er det tilstrækkelig interaktion for publikum? Kan det sidestilles med at gå på et traditionelt kunstmuseum, hvor man som passiv tilskuer går rundt og ser på værkerne? Hvis man tager udgangspunkt i Kruegers definition på et responsivt system, kan der sås tvivl om rummet er interaktivt, da publikum og brugere kun i en lille grad har kontrol over systemet. Derfor er brugeren som er bevidst om værket ikke instrumentalist, der skal findes i små organismer og mennesker, som er fuldstændig uvidende om at de bliver filmet [2]. Musikken er derfor fuldstændig styret af henholdsvis organismer og mennesker/maskiner, og deres daglige gøren og laden. Publikum og brugere har dog en vis magt over systemet, da de kan skifte kameraets (interfacets) placering, og dermed det som det filmer, men de kan stadig ikke styre hvad som der sker på kameraet. Ifølge Brian Massumi er forskellen på interaktiv- og

traditionel kunst, at traditionel kunst er envejskommunikation, imens at interaktiv kunst fordrer en bestemt type adfærd hos brugeren, som igen kan ændre adfærden fra systemet [7]. Han ser to problemer ved interaktiv kunst:

1. Både Massumi og Krueger advarer imod, at teknologien kommer til at spille en for central rolle i systemet og interaktionen, så teknologien er i centrum, og ikke interaktionen og kunsten [2] [7].

2. Interaktion skaber en hvis lukkeheds i værket, da det kommer til at minde om et computerspil, hvor man skal løse en opgave, som man lægger fra sig, når den er løst [7].

Massumi mener at løsningen skal findes i at skjule systemet, og dermed lave mystik om det. En anden måde er ved at lave rum for folk opfører sig anderledes [7]. Hvis man tager udgangspunkt i *Khaos*, er systemet hverken skjult, og det får heller ikke folk til at opføre sig anderledes end normalt.

Men måske skal man ikke se på det værket som et levende responsivt system, der fordrer en bestemt type af interaktion, som Krueger skriver. Huhtamo tager udgangspunkt i installationen *Listening Post*, af Mark Hansen og Ben Rubin, som han har svært ved at definere som interaktiv kunst, da brugeren ikke interagerer mere med det, end for eksempel en film eller mere traditionel kunst [8]. Da kunstværket bliver fodret med data fra internettet, foreslår han at man kan kalde det "Database aesthetics", og ser det som en naturlig udvikling i den interaktive kunst, som har udviklet sig til noget andet og nyt. *Khaos* har visse ligheder med *listening post*, da de begge sonificerer data fra internettet, og skaber et musikalsk lydbillede ud fra det. Til forskel fra *Listening Post* har brugeren i *Khaos* mulighed for at bruge rummet aktivt og ændre kameraet.

Et teknologisk problem ved interfacet i *Khaos*, er hvordan brugeren kan interagere med det. Skal der være en form for karenstid over controlleren, så én bruger ikke kan skifte webcam hele tiden, til gene for de andre brugere? Og hvad kommer der til at ske, når der er mange besøgende og kun en enkelt controller til hver skærm. Mange mennesker i rummet vil også påvirke oplevelsen af værket, og ændrer interaktionen mellem mennesker og værk, og skabe en anden atmosfære i rummet.

Computer vision er kompliceret teknologi, og det kræver meget arbejde at få algoritmerne til at fungere efter hensigten []. I forhold til algoritmerne i *Khaos*, er der en stor udfordring i at få systemet til at tracke bevægelserne efter hensigten. Der er et stort arbejde i at tilpasse indstillingerne og kalibrere threshold og lignende, og det vil derfor være en stor udfordring at få systemet til at fungere lige godt, når der sker ændringer i videofeedet, herunder ændringer i lys, opløsning, farve, størrelse på objekter og andet. Derfor kan det også blive problematisk, hvis arbejdet kommer til at handle for meget om teknologien, som både Massumi og Krueger pointerer [2][7].

4. RELATERET ARBEJDE

En af pionererne, som startede med at bruge computer vision, er Myron Krueger, med sit værk *Video Place* fra 1975. I værket interagerer en bruger med systemet og videoteknologi som tracking og hvad Krueger kalder for "artificial reality" [5]. Den amerikanske kunstner Micah Frank, har også arbejdet med

sonification af videodata. I sit projekt kaldet *Junction* tracker han taxier i New York ved hjælp af offentligt opsatte videokameraer. Dataen samlede og analyserede han i *Max/Jitter*, som skabte et musikalsk soundtrack [8], meget lig metoden i *Khaos*. Et tredje eksempel på et relateret værk er Mark Hansen og Ben Rubins *Listening Post*. Ligesom i *Khaos*, er der ikke så meget traditionel interaktion over forholdet mellem bruger og system i *Listening Post*. *Listening Post* er også en æstetisk sonificering af data, og i *Listening Posts* tilfælde er der tale om real-time twitterbeskeder.

5. FREMTIDIGT ARBEJDE

Der er to store problemer i *Khaos*, der skal løses. Det første er den teknologiske del, som indebærer computer vision algoritmerne. Hvis man skal have et velfungerende interface, hvor man kan skifte billeder, er der brug for et system som ikke er for sensitivt overfor ændringer i inputtet. Hvis man kan lave algoritmen avanceret nok, kan bruge video fra over flere tidspunkter af døgnet og i forskellige vejrforhold. Et andet problem er forholdet imellem interaktivitet og passivitet. Måske skal hele forholdet imellem bruger og system omtænkes, eller måske skal der holdes et bedre fokus, som for eksempel hvordan man kan bruge rummet i installationen til andet end cross-fades. Desuden kan man undersøge brugernes adfærd og ændre installationen ud fra dette. En anden måde at inkludere brugerne i en interaktiv rolle, kan ved for eksempel at bruge mikroskopet aktivt, så brugere selv kan fylde væsker på, som de kan se igennem. Dette vil kræve en meget avanceret algoritme.

6. KONKLUSION

Man kan bruge rummet til mange forskellige ting. I *Khaos* bruges rummet som noget der kan analyseres og sonificeres, og selve det fysiske rum som brugeren befinder sig i bruges som en effekt til at crossfade imellem to musikalske udtryk. I *Khaos* bliver bevægelse og farver analyseret og lavet til musik igennem sonification. Analysen foregår i *Max/Jitter* og *cv.jit* som er en samling af algoritmer som kan skabe computer vision. Computer vision teknologi er blevet brugt til interaktiv visuel og auditiv kunst, siden halvfjerdserne, og bliver stadig brugt den dag i dag. *Khaos* rejser også nogle problemstillinger i forhold til definitionen på interaktiv kunst, og hvornår noget er interaktivt eller en anden æstetik.

7. ACKNOWLEDGMENTS

Thanks to Jean-Marc Pelletier for creating *cv.jit* and making it available to the public for free.

8. REFERENCER

[1] Sørensen, J. P., Den Store Danske – Kaos. Accessed 2 June 2018. Available: http://denstoredanske.dk/Sprog_religion_og_filosofi/Religion_og_mystik/Primitiv_religion/kaos

[2] Krueger, M. *Responsive Environments*. National Computer Conference. The University of Wisconsin. Madison, Wisconsin. 1977.

[3] Frazier, C., and Kockelman, K. Chaos Theory and Transportation Systems: An Instructive Example. Transportation Research Record No.1897: 9 17, 2004.

[4] M. Pelletier, Computer Vision for Jitter. Accessed 1 June 2018. [Online]. Available: <http://jmpelletier.com/cvjit/>

[5] Levin, G. "Computer Vision for Artists and Designers: Pedagogic Tools and Techniques for Novice Programmers". *Journal of Artificial Intelligence and Society*, Vol. 20.4. Springer Verlag, 2006. Accessed 2 June 2018. Available: http://www.flong.com/texts/essays/essay_cvad/

[6] Gaver, W. *Auditory Icons: Using Sound in Computer Interfaces*. Human-Computer interaction, Volume 2, pp. 167-177, University of California, San Diego, 1986.

[7] Massumi, B. The Thinking-Feeling of What Happens: Putting the Radical Back in Empiricism

[8] Huhtamo, E. *Trouble at the Interface, or the Identity Crisis of Interactive Art*. Framework, The Finnish Art Review, 2/2004, University of California, Los Angeles, 2004.

[9] <https://cycling74.com/projects/junction-taxi-cabs-generate-a-soundtrack-in-real-time/>