
LABORGESPRÄCHE

MARTIN RUMORI
im Gespräch mit UTE HOLL

PARISFLANEUR: SPAZIERGÄNGE IN BINAURALEN HÖRRÄUMEN

Martin Rumori hat am Institut für Elektronische Musik und Akustik (IEM) der Universität für Musik und darstellende Kunst in Graz das Projekt «Parisflaneur» entwickelt und realisiert. Mithilfe eines optischen Trackingsystems bewegen sich Hörflaneure in einem gerechneten akustischen Raum, der jedoch nicht nur ein einziges homogenes System darstellt, sondern in dem sich akustische Inseln befinden, die Rumori «Hüte» nennt. Diese Hüte sind statische, nicht reaktive Hörräume. Flaneure können in die akustischen «Hörhüte» hineinkriechen, diese Hüte durch ihre Bewegungen verschieben und dann wieder verlassen. Auf diese Weise ändert sich die Szenerie des akustischen Gesamtraums. Die Anordnung experimentiert mit Interaktionen in virtuellen akustischen Räumen und stellt sensomotorisches (audiomotorisches) Verhalten als Orientierung und zugleich Transformation von Räumen und Räumlichkeit auf die Probe. Allerdings stellt sich nicht zuletzt das Problem, wie akustische Objekte und akustische Räume zueinander zu konstellieren sind. Martin Rumori, Musikwissenschaftler und Informatiker, ist künstlerisch-wissenschaftlicher Mitarbeiter an der Kunsthochschule für Medien, Köln.

I. Ohren-Orientierung in fremden Räumen

U.H. Der Clou bei *Parisflaneur* ist, dass sich innerhalb eines größeren, virtuellen und durch ein Trackingsystem interaktiv gehaltenen akustischen Raumes sieben kleinere akustische Räume befinden, die sogenannten «Hüte». Diese sind statisch, keine errechneten Hörräume, sondern wurden auf den Straßen von Paris mit Mikrophon aufgezeichnet.

M.R. Ich möchte damit das Verhältnis zwischen virtuell gerenderten Räumen, hier also dem im Computer erzeugten künstlichen binauralen Raum, und jenen Aufnahmen untersuchen, die ich in Paris gemacht habe. Auch diese sind binaural ... aber nicht reaktiv, sondern so, wie ich sie in Paris aufgenommen habe, statisch.

U.H. Binaural ist nicht einfach stereophon.

M.R. Es gibt raumbezogene Stereophonie und es gibt die kopfbezogene Stereophonie, die in den siebziger Jahren Kunstkopfstereophonie hieß. Bei einem Kunstkopfmikrofon sind Ohren und Kopfverhältnisse genau nachgebildet, was sehr eindrucksvolle räumliche Aufnahmen erzeugt. Bei meinen Aufnahmen in Paris habe ich im Grunde dieselbe Technik verwendet, nur dass ich keinen Kunstkopf genommen habe, sondern meinen eigenen. Ich mache das für die meisten Aufnahmen so, dass ich in eine Richtung schaue und die Szene als stummer Beobachter aufnehme, der sich nicht bewegt und der nicht richtig da ist, der nur hineinlauscht, unbeteiligt. So versuche ich die Szenen einzufangen.

U.H. In der Installation Parisflaneur gibt es innerhalb des größeren virtuellen Hörraums mehrere solche akustischen Paris-Inseln.

M.R. Es sind sechs oder sieben, nicht nur aus Paris, sondern auch aus Giverny, wo Claude Monet seinen Garten hatte, oder aus einem Park; eine Aufnahme ist nachts entstanden am Place de Dauphine, hinter dem Pont Neuf, und eine an meinem Lieblingsplatz, der Passage des Abbesses, ein kleiner Durchgang, wo jemand Harfe geübt hat. Eigentlich sind diese Klänge gar nicht eigens für die Studie *Parisflaneur* gemacht, sondern ich habe genommen, was ich eben hatte. Es gibt also in *Parisflaneur* diese Klanginseln, man kann in diese originalen binauralen Aufnahmen hineingehen, und man kann sie auch wieder verlassen.

U.H. Was heißt «reingehen und verlassen» – ist das metaphorisch gemeint?

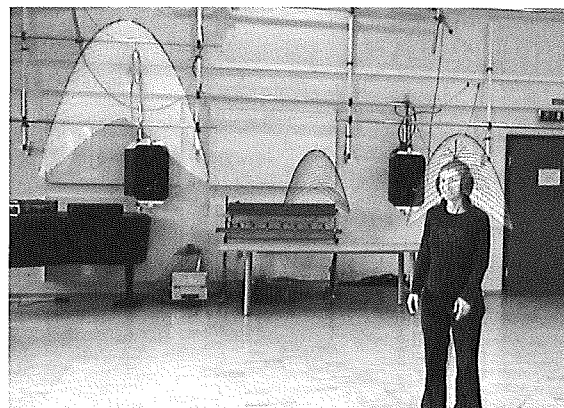
M.R. Die «Klanghüte» sieht man nicht, und das «Hineingehen» geschieht durch Aufsetzen eines Hutes ...

U.H. ... auch das metaphorisch ...

M.R. ... völlig metaphorisch: wie man einen Hut, der irgendwo hängt, sich aufsetzen würde, indem man darunter kriecht und dann den Kopf hebt, ihn aufzieht ... metaphorisch, im übertragenen Sinne. Man sieht überhaupt gar nichts.

U.H. Der Raum insgesamt ist transparent. Wie findet man heraus, wo diese «Hüte» sind?

M.R. Der Raum ist komplett transparent, ein experimenteller Raum, in dem viele Lautsprecher installiert sind und das Trackingsystem, in diesem Fall ein optisches, mit mehreren Infrarotkameras. Die Flaneure tragen einen so genannten Tracking-Target auf dem Kopfhörer, ein Gerüst aus mehreren reflektierenden kleinen Kugeln, den rigid body, auf den die Tracking-Kameras im Raum von allen Seiten gerichtet sind. Damit lassen sich dann im Rechner sehr akkurat und auch sehr schnell die Daten berechnen, die angeben,



Virtuelle «Klanghüte» in
Parisflaneur

wo der Flaneur sich befindet und wie er sich bewegt: Richtung und Orientierung, also Blickrichtung. Entsprechende Informationen werden auf den Kopfhörer gespielt.

Den Inseln im Raum nähert der Flaneur sich schrittweise, merkt, dass er dem Klang näher kommt, wenn es lauter wird, so wie das in der Natur auch wäre.

Durch Kopfdrehen und Herumgehen merke ich, dass ich mich dem Objekt «Hut» annähere. Um mir den Hut aufzuziehen, muss ich mich bücken und druntergehen. Die Hüte haben eine bestimmte Höhe; ich habe um die 1,70, 1,80 m benutzt. Wenn man direkt darunter hockt, kommt diese originale Paris-Aufnahme schon zur virtuellen Szene hinzu. Wenn man sich dann wieder aufrichtet, verschwindet die virtuelle Szene, verschwinden alle anderen Klänge, die drumherum sind, und der Klang verändert sich, weil jetzt in diese binaurale Paris-Aufnahme überblendet wird.

Das Problem ist, eine virtuelle Szene zu rendern. Ich habe Einzelquellen, die ordne ich im Raum an, die eine kommt von rechts, die andere kommt von links, eine richtige 3D-Szene, genau wie in einer Game-Engine, nur sind es bei mir akustische Objekte, keine optischen. Ich muss die Möglichkeit haben, dieses binaurale Bild so zu erzeugen, als hätte ich es aufgenommen und die Szene wirklich nachgestellt.

Dazu gibt es verschiedene Möglichkeiten. Grundsätzlich geht es so, dass man einen Satz von Filtern hat, die das Richtungshören des menschlichen Gehörs beschreiben, sogenannte HRTFs, Head Related Transfer Functions. Darin ist kodiert, wie der Schall aus unterschiedlichen Richtungen von der Ohrmuschel verändert wird. Technisch wird das so gemacht, dass man ein Kunstkopfmikrofon nimmt, es in einen schalltoten Raum stellt und dann Messtöne rundherum von verschiedenen Seiten, z. B. in 5°-Schritten, aufspielt und aufnimmt.

U.H. Das heißt, man misst den menschlichen Hörapparat in eine Matrix ein.

M.R. Ganz genau, man denkt sich ein Raster aus, eine bestimmte Matrix; oben und unten ist man nicht so genau, da geht man nur in 10°- oder 15°-Schritten vor, weil das menschliche Gehör in vertikaler Richtung nicht so gut auflösen kann wie in der Horizontalen.

Impulsantwort ist der physikalische oder technische Ausdruck für den Fingerabdruck eines akustischen Systems. Im Fall von *Parisflaneur* besteht das System aus Lautsprechern und einem Kunstkopfmikrofon und dem Raum und der Luft darin. Dieser Fingerabdruck ist nicht an einen bestimmten Klang gekoppelt. Er umfasst nur die Eigenschaften dieses Systems.

Wenn man den Fingerabdruck einsetzen will, benutzt man ein Verfahren, das sich «Faltung» nennt, auf Englisch *convolution*, eine mathematische Operation aus der Signalverarbeitung. Der Fingerabdruck, die Impulsantwort, wird auf ein beliebiges Signal aufgefaltet – zum Beispiel auf eine meiner Paris-Aufnahmen – und wenn ich den Fingerabdruck von 5° vorne kombiniere mit meiner Paris-Aufnahme, das für das linke und rechte Ohr getrennt mache, und dann den

Kopfhörer aufsetze, klingt das so, als würde meine Parisaufnahme von 5° vorne kommen. Das ist das ganze Geheimnis der Binauralsynthese.

U.H. Das ist speziell für das menschliche Gehör errechnet? Wenn Du das für eine Katze auffalten wolltest, müsstest Du ganz anders vorgehen.

M.R. Dann müsste ich einen Katzenkopf nehmen, in die Ohren Mikrofone setzen, und das genauso vermessen, aber eben mit Katzenohren, oder meinstest Du den Katzenklang? Nein, für eine Katze das ganze Paris.

U.H. Ja.

M.R. Da wird es spannend. Wenn ich naiv davon ausgehe, die Katze hört so ähnlich wie ein Mensch, nur dass sie anders geformte Ohren und einen anderen Ohrenabstand hat – all das ist ja in dem Fingerabdruck kodiert – dann würde es reichen, die Katze oder den Katzenkopf als Nachbildung in den schalltoten Raum zu setzen und genau so zu vermessen und dann die Katzenabdrücke anstelle der menschlichen Fingerabdrücke zu benutzen.

U.H. Man müsste die Katze gar nicht fragen, wie sie hört ...

M.R. Man müsste sie eigentlich fragen, weil man nicht weiß, ob bei ihr derselbe Eindruck entsteht. Ingenieurstechnisch reicht es, das Schallfeld, die Schalldruckänderungen, die man beim natürlichen Hören an den Ohren hat, wieder hervorzurufen. Damit kommt man ja relativ weit. Aber nicht beliebig weit.

U.H. Dann kommt die Psychophysik ins Spiel.

M.R. Die Psychoakustik. Auch wenn ich den genauen Höreindruck habe, fühle ich mich doch nicht wie auf den Straßen von Paris. Ich sehe die Lautsprecher und habe Kopfhörer auf, die ich in den Straßen von Paris nicht trage. Ingenieursmäßig kann ich das Verfahren auch bei der Katze anwenden. Ob ich die Katze genauso gut austricksen kann wie den Menschen oder nicht, das weiß ich nicht, da müsste man die Katze dann fragen.

U.H. Und die würde nur grinsen, wie wir wissen.

M.R. Das ist die Grundlage. Jetzt ist das Problem, dass ich viele Quellen um mich herum habe, oder eine Quelle, die sich auch noch bewegt, oder einen Kopf, der sich dreht. Das sind dann ja ständig unterschiedliche Richtungen, aus denen der Schall kommt. Das heißt, ich muss eigentlich, während das läuft, immer wieder einen anderen Fingerabdruck benutzen. Einmal den für 5° , und wenn ich mich drehe, ist das, vereinfacht, als ob die Quelle weiter nach rechts rutschte. Das heißt, ich muss meine Quelle drehen auf 5° , 6° , 7° , 8° , 9° , 10° . Ich muss ständig herumblenden und einen anderen Fingerabdruck wählen, damit ich dann den Eindruck habe, dass sich die Quelle bewegt.

U.H. ... oder dass sich der Kopf dreht.

M.R. ... entweder dass sich die Quelle bewegt oder dass sich der Kopf bewegt.

U.H. Und wie unterscheidet man das in Deinem Raum?

M.R. In meinem Raum wende ich einen Trick an, in dem sich das tatsächlich nicht unterscheidet. Das ist eine Vereinfachung, eine nicht zulässige, aber die wollte ich eben ausprobieren. Im völlig schalltoten Raum, in dem akustische Raumabdrücke gemessen werden, gibt es, idealerweise, keinen Hall, kein Echo, also im Prinzip gar keinen Raum. Der Akustiker spricht dann vom «Freifeld», wenn der Schall nirgends aneckt, sondern sich in alle Richtungen ausbreitet und schließlich in der Entropie verschwindet. In jedem normalen Raum haben wir jedoch Reflektionen, und das ist für unseren Höreindruck ganz entscheidend. Im schalltoten Raum macht es tatsächlich keinen Unterschied, ob sich die Quelle dreht oder der Kopf. In einem guten Binauralsystem oder einem, das dem ingenieursmäßigen nahe kommen soll, muss man die beiden Fälle jedoch unterscheiden können. Ich habe selber solche Impulsantworten, solche Fingerabdrücke gemessen, und zwar gezielt nicht im schalltoten Raum, sondern in einem realen Raum, und zwar in genau jenem Raum, in dem ich forsche, dem IEM Cube in Graz. Das heißt, ich habe Rauminformationen mit aufgenommen. Ich habe daher Fingerabdrücke, die sich so benehmen, als wäre ich in diesem Raum und würde aus einem der

Lautsprecher, die dort hängen, ein Signal hören mit den gesamten Rauminformationen. Ich habe allerdings nicht die Möglichkeit gehabt, in vielen kleinen Gradschritten zu messen, sondern ich habe einfach die 24 Lautsprecher, die im IEM Cube hängen, und das sogenannte Ambisonic-System für das Setup von 24 Lautsprechern, das da entwickelt wurde, vermessen. Ambisonic ist ein Verfahren, mit dem ein beliebiges Schallfeld wieder resynthetisiert werden kann. Der Vorteil ist, dass es universell ist: Ein in Ambisonic kodiertes Format kann ich auch mit anderen Lautsprecherkonfigurationen wiedergeben. Ich muss es dann nur einfach anders dekodieren, also anders aufbereiten. In meinem Fall habe ich eben nicht 24 Lautsprecher, die im Raum hängen, sondern ich habe jetzt nur 24 Fingerabdrücke, die in diesem Raum genommen wurden. Das heißt, ich dekodiere die Ambisonic-Signale so, als wollte ich sie im Cube abspielen, schalte dann aber nicht die Lautsprecher und den Raum dahinter, sondern eine Faltungseingabe, also die virtuellen 24 Lautsprecher und die Fingerabdrücke von diesem virtuellen Raum. Spiele ich das dann über Kopfhörer ab, so habe ich den Eindruck, ich wäre in der Mitte dieses Raums und hörte die 24

Parisflaneur



Lautsprecher inklusive des Raums um mich herum. Das ist ein vereinfachendes Verfahren, über den Umweg von Ambisonic zwischen den Fingerabdrücken zu interpolieren.

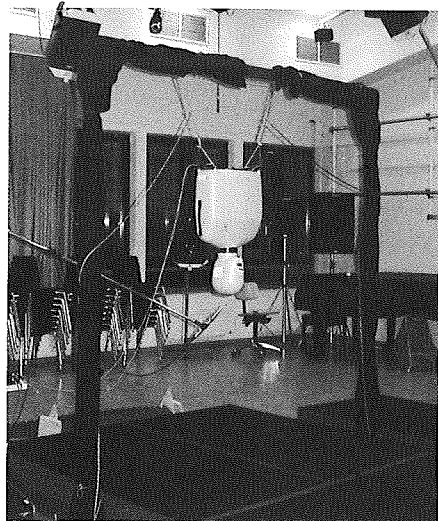
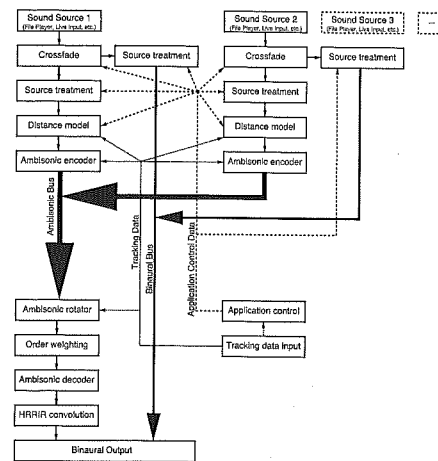
II. Die nicht-euklidische Biene und der Bienen-Raum

M.R. Wenn ich nicht nur eine starre Konfiguration von der Schallquelle und dem Hörer herstellen, sondern das Verhältnis dynamisch halten will, so dass sich entweder die Schallquellen bewegen oder der Hörer sich dreht oder bewegt, dann brauche ich irgendein Verfahren, um zwischen den Fingerabdrücken zu interpolieren. Ich benutze ein virtuelles Zwischenformat, eine Repräsentation von einem Raum, um zwischen diesen einzelnen Fingerabdrücken zu überblenden. Ich habe mich für Ambisonic entschieden, weil man im IEM Cube damit forscht, und weil es mich interessiert hat, den Realraum, so wie er ist, noch einmal abzubilden – denn dieser Raum ist kein virtueller Raum, dieser Raum ist nicht simuliert, sondern basiert auf echten Messungen. Aber der Fehler bei mir ist, dass ich die Messungen mit dem Kunstkopf in der Mitte des Raumes durchführe und die 24 Lautsprecher messe, um dazwischen dann die Quellen zu bewegen. Das heißt, ich habe eigentlich eine Messung unabhängig von der Rotation des Zuhörerkopfes, als wäre es egal, ob sich die Quelle um den Hörer herum bewegt oder aber dessen Kopf sich dreht. Der Kopf ist während der Messung starr an den Raum gebunden, und das bleibt er während der Simulation auch. Das heißt, wenn ich in der Installation den Kopf drehe, dann dreht sich eigentlich der ganze Raum mit ...

U.H. ... wenn Du den Raum im Kopfhörer drehst, ist das eine Übersetzung, als drehst Du den ganzen Raum um Dich herum?

M.R. Ganz genau. Erzähle ich das einem Ingenieur, merkt der bald, dass da was faul ist: Wenn ich in meinem virtuellen Raum sitze, den Blick nach vorne, so wie der Kunstkopf gestanden hat, stimmt das Ganze. Wenn sich eine Schallquelle bewegt, so wie ich sie in jenem realen Raum, dem Ambisonic-System mit Lautsprechern, in Bewegung setzen könnte, stimmt das immer noch. Wenn ich dann den Kopf drehe, müsste sich eigentlich der Kopf relativ zum Raum drehen. Wenn ich das aber in meiner Simulation mache, tue ich so, als würde sich nur die Schallquelle bewegen. Ich korrigiere also eigentlich nur den Fehlwinkel: Angenommen, ich gucke geradeaus und die Schallquelle kommt von rechts. Wenn ich nun meinen Kopf um 90° nach links drehe, nehme ich an, es ist das gleiche, als ob sich die Schallquelle um 90° nach rechts dreht.

Parisflaneur:
 ▼ Funktionsprinzip des binauralen Renderings
 ▼▼ Kopfhörer mit Tracking-Target



U.H. Das heißt, Du hast die Relation zwischen der Biene, die Dir um den Kopf saust, und den Ohren immer nur als einfache Differenz. Und das Verhältnis der Ohren zum Raum, in dem sich die sekundären Geräusche ausbreiten, lässt Du weg.

M.R. Das Problem ist, den Raum lasse ich an den Kopf drangeklebt. Das heißt, die Akustik des Raums dreht sich mit. Die Quelle bleibt da, wo sie akustisch ist ...

U.H. ... die Biene, die um Deinen Kopf schwirrt, darf machen, was sie will.

M.R. Das funktioniert, weil ich die Bewegung der Biene relativ zu meiner Kopfbewegung korrigiere, aber alle Reflektionen dieses Geräusches, die drehen sich mit, alles, was der Raum akustisch noch zu der einen Biene, die da herumschwirrt, hinzufügt ...

U.H. ... das bleibt am Kopf kleben.

M.R. Deswegen stimmt die virtuelle Szene eigentlich nicht. Der Direktschall, das, was von der Biene direkt zu meinem Ohr geht, ist das einzige, was dann noch stimmt, genauer: die Richtung dieses Schalls. Bei den Reflektionen stimmt es aber nicht mehr, die kommen immer so, wie sie kommen würden, wenn ich geradeaus schauen würde.

U.H. Der Sound der Biene macht sich selbständig gegenüber der Biene.

M.R. Richtig, ich habe dann im Prinzip zwei Bienen im Raum, eine an dem Ort «wie wenn ich geradeaus schaute», und eine, die da ist, wo sie wäre, wenn ich den Kopf drehte. Von der einen Biene gibt's nur den direkten Schallanteil und von der anderen gibt's diesen gerade nicht, nur den Rest, den Schwanz sozusagen, den der Raum draufpackt. Ein Ingenieur käme nie darauf, so etwas zu machen, weil das eben völlig falsch ist, und man weiß vom Raumhören, dass diese Reflektionen extrem wichtig sind, auch um Richtungen zu bestimmen. Das ist ein großes Problem bei den virtuellen Akustiksystemen, dass die Richtungsbestimmung und das natürliche Hören so schlecht darin funktionieren. Das heißt, ich mache das Falscheste, was man machen kann; ich führe diesem sehr komplexen kognitiven Prozess des Raumhörens noch mehr Verwirrung hinzu. Aber was mir daran aufgefallen ist: Es funktioniert eben doch erstaunlich gut. Ich wollte wissen, was passiert, wenn ich richtige Raummessungen [aus dem richtigen Grazer Experimentalraum, U.H.] verwende, denn die klingen einfach besser als jeder im Rechner simulierte Raum. Ich habe zumindest bei diesem Versuch festgestellt, dass jeder mir den Raum abgekauft hat und die Leute fanden, ein so gutes synthetisiertes binaurales System hätten sie noch nie gehört. Das funktioniert wunderbar. Ich habe allerdings noch keine systematischen Hörtests gemacht, die mir auch ein Ingenieur abkaufen würde. Aber ich habe den Eindruck, dass es dadurch funktioniert, dass der Raum glaubhaft ist, obwohl es physikalisch nicht funktioniert.

III. Zerstreute Wahrnehmung für Einzelgänger

U.H. Gegen die Idee einer kollektiven Rezeption im technisch-medialen Raum implementierst Du in diesem Projekt stärker das Individuelle und bindest individuelles Verhalten an eine akustische Situation.

M.R. Wenn man sich vorstellt, dass viele Leute in einem Kino sitzen und teilhaben an so etwas wie kollektivem Hören, ist es eigentlich nur ein anderer Grad von Gewissheit, ob der neben mir das gleiche hört wie ich. Dass man sich vorstellt, viele Leute haben Kopfhörer auf und bewegen sich in demselben realen Raum und vielleicht auch im gleichen virtuellen Raum, ist eine andere Form des Glaubens. Da muss ich natürlich darauf vertrauen, dass die Technik denen etwas reinspielt, was sie mir reinspielt, allerdings anders angepasst an ihre Position, dass aber der Raum konsistent ist. Und im Kino glaube ich das ja eigentlich auch nur. Vielleicht hört der neben mir ja gar nicht das gleiche. Im Moment arbeite ich noch an grundlegenden Sachen, schraube nur an den technologischen Parametern und denke dabei nur an eine Person. Im Prinzip geht das aber auch für mehrere Personen.

U.H. Ich habe Deine Arbeiten auch als Trainingsprogramme für eine akustisch zerstückelte Welt verstanden.

M.R. Trainingsprogramm nicht in dem Sinne, dass ich einen bestimmten Trainingsplan hätte oder einen bestimmten Bildungsauftrag, und auch gar keine Anhaltspunkte habe, was man psychoakustisch können müsste, um in dieser Welt nicht durchzudrehen.

U.H. Warum Paris, warum Flaneur?

M.R. Paris wegen meiner Aufnahmen, Flaneur wegen des Spazierengehens. Ich nehme gerne auf, ich habe erst im Zuge dieses Aufnehmens angefangen zu hören, mit diesem Binauralmikrophon höre ich ganz anders, im Soundscape-Modus. Dann ist Hören kein Orientierungssinn mehr, sondern ein Genusssinn.

U.H. Der Flaneur widersetzt sich der Ökonomie.

M.R. Der Flaneur ist zweckfrei unterwegs.

BESPRECHUNGEN

