

Raumklang mit Ambisonics in Max/MSP

Philippe Kocher

ICST Institute for Computer Music and Sound Technology

ZHdK Zürcher Hochschule der Künste

Seit einigen Jahren wird in Zürich im Bereich der räumlichen Klangprojektion geforscht und man hat sich auf die aus Großbritannien stammende Technologie Ambisonics spezialisiert, die nicht nur der Wiedergabe mit dem Soundfield Microphone erstellter, peripherer Aufnahmen dient, sondern mit der es auch möglich ist, Klangquellen im virtuellen Raum zu setzen und diesen über eine beliebige Anzahl Lautsprecher zu reproduzieren. Die am ICST entwickelten External für Max/MSP ermöglichen dem Benutzer einen einfachen und intuitiven Zugriff auf dieses Raumklangverfahren, einerseits durch Module für das Encodieren und Decodieren in Ambisonics bis dritter Ordnung, andererseits durch eine graphische Schnittstelle zur räumlichen Manipulation virtueller Klangquellen in Echtzeit und einem Modul für die algorithmischen Steuerung von Bewegungen im dreidimensionalen Raum.

Ambisonics

Die Raumklangtechnologie Ambisonics wurde in den 1970er Jahren vom britischen Mathematiker Michael Gerzon entwickelt. Ursprünglich handelt es sich um eine besondere Mikrophontechnik, mit der eine mehrkanalige Aufnahme erstellt wird, welche die räumliche Information des Schallfeldes trägt und schließlich über ein Lautsprechersystem derart wiedergegeben werden kann, dass sich der Eindruck dreidimensionalen (periphonen) Hörens einstellt. Das heißt, das Schallfeld wird weitgehend originalgetreu übertragen und virtuelle Schallquellen sind korrekt und präzise lokalisierbar. Für eine solche Aufnahme verwendet man das so genannte Soundfield Microphone, bestehend aus vier Mikrophonkapseln mit nierenförmiger Charakteristik, angeordnet als regelmäßiges Tetraeder. Das Ausgangssignal dieses Mikrophons wird elektronisch aufbereitet, und man erhält das vierkanalige B-Format, das als Standardsignal einer Aufnahme mit dem Soundfield Microphone gilt. Das in der Fachsprache als A-Format bezeichnete direkte Ausgangssignal dieses Mikrophons ist nicht dazu ge-

dacht, ohne weitere Verarbeitung verwendet zu werden.

Das B-Format ist das eigentliche Kernstück von Ambisonics, denn es beschreibt mit seinen vier Kanälen (W, X, Y, Z) die gesamte Dreidimensionalität des Schallfeldes an jenem Punkt, an dem die Aufnahme gemacht wurde. Zum Verständnis des B-Formats stelle man sich vier Mikrophone vor, die sich in einem einzigen Punkt befinden. Drei Mikrophone besitzen Achtercharakteristik und sind rechtwinklig zueinander, den drei Raumkoordinaten gemäß, ausgerichtet. Sie liefern die Signale für die Kanäle X, Y und Z. Ein weiteres Mikrophon schließlich besitzt Kugelcharakteristik und liefert das monophone Signal W. Dieses vierkanalige Format nennt sich B-Format erster Ordnung.

Zur Wiedergabe wird ein Decoder benötigt, der aus dem B-Format die Signale für die einzelnen Lautsprecher errechnet. Bis hierher wurde noch kein Wort verloren über Anzahl und Aufstellung der Lautsprecher. Tatsächlich handelt es sich hier um eine große Stärke, die Ambisonics gegenüber anderen Raumklangtechniken hat: Man ist an keine fixe Lautsprecherkonfiguration gebunden. Bleibt man innerhalb vernünftiger Grenzen, was Minimalzahl und Symmetrie angeht, ist man frei in der Anordnung der Lautsprecher und kann sie den jeweiligen Gegebenheiten und der Form des Raumes anpassen. Der Decoder benötigt lediglich die genauen Koordinaten eines jeden Lautsprechers.

Zu den weiteren vorteilhaften Eigenschaften von Ambisonics zählt die Tatsache, dass keine Raumrichtung bevorzugt wird. Dies im Gegensatz zu Raumklangtechniken, die im Zusammenhang mit dem Film entwickelt wurden und eindeutig nach vorne, zur Bildfläche hin ausgerichtet sind. Ebenfalls ist bei Ambisonics der Bereich, in dem sich das optimale räumliche Klangempfinden einstellt – der "sweet spot" – merklich größer als bei anderen Formaten. In der Tat mag es für den Heimgebrauch reichen, wenn dieser bloß einer Person Platz bietet, für öffentliche Konzerte sollte aber auch einer umfangreicheren Zuhörerschaft ein überall ähnlich

gutes Hörerlebnis geboten werden. Auch außerhalb des Lautsprecherkreises ist es möglich, eine gewisse räumliche Abbildung wahrzunehmen, ein Hineinhören von Außen sozusagen.

Virtuelle Räume

Selbstverständlich kann das B-Format auch synthetisch hergestellt werden. Dem Sounddesigner oder dem Komponisten elektroakustischer Musik eröffnet sich damit die Möglichkeit künstliche Räumlichkeiten zu kreieren. Das Verfahren, das ein monophones Schallsignal mit einer Richtungsangabe in ein B-Format überführt, also eigentlich die Aufnahme mit einem Soundfield Microphone virtuell nachbildet, nennt man Encodierung.

Beim Arbeiten mit Ambisonics ist dadurch die Produktion von der Reproduktion getrennt, mit dem B-Format als zwischengeschaltetes Übertragungsformat (Abb. 1). Kompositionen können im B-For-

mat aufgezeichnet sein und werden erst zum Zeitpunkt der Wiedergabe für die spezifische Lautsprecheraufstellung des Aufführungsortes decodiert. Dadurch wird Ambisonics zu einer ausgesprochen flexiblen Technologie, was dem Kompositionsprozess sehr entgegenkommt, arbeitet man doch während des Komponierens meist mit einem anderen und oft wesentlich bescheideneren Lautsprechersystem, als es der Konzertsaal bietet. Zudem bleibt das fertige Stück unabhängig von irgendeiner spezifischen Lautsprecheraufstellung.

In der Praxis sieht die Produktion wie folgt aus: Der Ort einer virtuellen Klangquelle wird durch seine Raumkoordinaten festgelegt. Um den Effekt der Bewegung zu erzeugen, ändern sich diese Koordinaten in der Zeit. Zuerst wird der monophone Ursprungsklang aufbereitet, indem gemäß seiner Entfernung die Lautstärke angepasst und ebenso mit einem Tiefpassfilter die Luftabsorption simuliert wird. Dann wird der Klang mit den Koordinaten verrechnet und ins B-Format encodiert.

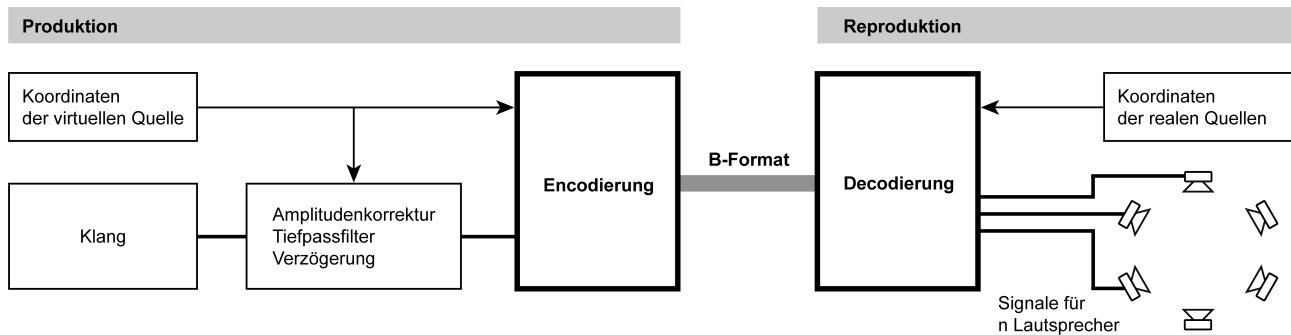


Abbildung 1: In Ambisonics ist die Produktion von der Reproduktion getrennt

Formeln

Wenn man versucht, an einem beliebigen Punkt im Raum die Summe der Schallwellen exakt zu berechnen, treten komplizierte mathematische Funktionen auf. Vereinfachend wird deshalb bei Ambisonics angenommen, dass nur ebene Wellen auftreten und der Hörer sich genau im Mittelpunkt befindet. Schallwellen werden als Summe von Kugelflächenfunktionen (Spherical Harmonics) dargestellt. Dadurch reduziert sich die Komplexität auf eine Multiplikation der monophonen Klänge mit Richtungsvektoren. Die Codierung der Signale

kann unterschiedlich genau erfolgen. Die Genauigkeit wird durch die so genannte Ordnung bestimmt. Die nullte Ordnung ist ein Monosignal, die Aufnahme mit dem Soundfield Microphone ist in erster Ordnung. Aufnahmeverfahren für höhere Ordnungen wurden zwar entwickelt, befinden sich aber noch in einem experimentellen Stadium.

Höhere Ordnungen lassen sich synthetisch berechnen, indem man das Signal mit den Funktionswerten von Kugelflächenfunktionen höherer Ordnung multipliziert. Unsere Implementation verwendet die Koeffizienten des Furse-Malham-Sets.

Ist die virtuelle Richtung eines Schallsignals S gegeben durch den Azimutwinkel φ und den Elevationswinkel δ , werden die Kanäle des B-Formats wie folgt berechnet:

Ordnung	Komponente
Nullte	$W = S \cdot 0.707$
Erste	$X = S \cdot \cos \varphi \cdot \cos \delta$
	$Y = S \cdot \sin \varphi \cdot \cos \delta$
	$Z = S \cdot \sin \delta$
Zweite	$R = S \cdot (1.5 \cdot \sin^2 \delta - 0.5)$
	$S = S \cdot \cos \varphi \cdot \sin (2\delta)$
	$T = S \cdot \sin \varphi \cdot \sin (2\delta)$
	$U = S \cdot \cos (2\varphi) \cdot \cos^2 \delta$
	$V = S \cdot \sin (2\varphi) \cdot \cos^2 \delta$
Dritte	$K = S \cdot \sin \delta \cdot (5 \cdot \sin^2 \delta - 3) \cdot 0.5$
	$L = S \cdot \cos \varphi \cdot \cos \delta \cdot (5 \cdot \sin^2 \delta - 1)$
	$M = S \cdot \sin \varphi \cdot \cos \delta \cdot (5 \cdot \sin^2 \delta - 1)$
	$N = S \cdot \cos (2\varphi) \cdot \sin \delta \cdot \cos^2 \delta$
	$O = S \cdot \sin (2\varphi) \cdot \sin \delta \cdot \cos^2 \delta$
	$P = S \cdot \cos (3\varphi) \cdot \cos^3 \delta$
	$Q = S \cdot \sin (3\varphi) \cdot \cos^3 \delta$

Die Kanäle des B-Formats sind standardisiert mit Grossbuchstaben bezeichnet. Ihre Anzahl ist nur von der Ordnung abhängig, nicht von der Zahl der virtuellen Schallquellen oder der Lautsprecher. Die Anzahl der Kanäle wächst quadratisch mit der Ordnung.

$$\text{Anzahl Kanäle} = (\text{Ordnung} + 1)^2$$

Das B-Format bleibt aber abwärtskompatibel, d.h. es kann auch immer in einer tieferen Ordnung decodiert werden.

Obschon die Theorie suggeriert, dass die Anzahl und die Positionen der Lautsprecher völlig frei gewählt werden können, hat die Praxis gezeigt, dass dennoch eine möglichst symmetrische Aufstellung anzustreben ist. Zudem sollte je nach Ordnung eine Mindestanzahl von Lautsprechern verwendet werden, da durch die engmaschigere räumliche Auflösung höherer Ordnungen sich sonst "Löcher" zwischen zu weit voneinander entfernten benachbarten Lautsprechern ergeben würden. Für eine Lautsprecheraufstellung in der Ebene (2D) gilt:

$$\text{Anzahl Lautsprecher} \geq (\text{Ordnung} + 1) \cdot 2$$

Und für eine dreidimensionale Aufstellung:

$$\text{Anzahl Lautsprecher} \geq (\text{Ordnung} + 1)^2$$

Als Kompromiss zwischen Genauigkeit der räumlichen Auflösung und bewältigbarer Datenmenge und Rechenleistung wird am ICST die dritte Ordnung verwendet.

Entwicklung von Software

Für die ersten Schritte mit Ambisonics wählte man in Zürich die im Feld der Computermusik sehr verbreitete Klangsprache CSound. Die oben genannten mathematischen Formeln wurden implementiert und führten zu guten Resultaten, bloß die Arbeitsweise war weit entfernt von Unmittelbarkeit und Interaktion. Um ein Werkzeug zur Hand zu haben, das auch ermöglicht, in Echtzeit die klangliche Räumlichkeit zu kontrollieren, begannen wir damit, für Ambisonics die Programmierumgebung Max/MSP zu verwenden.

Dass wir uns damit das Feld der Live-Elektronik eröffneten, versteht sich von selbst. Aber auch die Arbeit des Komponisten im Studio veränderte sich dadurch. Es ist eine unermessliche Hilfe für die Vorstellungskraft, wenn man unmittelbar hören kann, wie man seine Klänge am Bildschirm räumlich manipuliert. Mit solchen Experimenten und Studien können wertvolle Erfahrungen gesammelt werden und es wird der Sinn für Wirkung und Ausdruck räumlicher Gesten geschärft, was letztlich dem Kompositionsprozess zugute kommt. Die Frucht unserer Arbeit mit Max/MSP ist eine kleine, aber äußerst leistungsfähige Kollektion von so genannten Externalen, die wir Interessierten im Internet frei zur Verfügung stellen (www.icst.net/downloads). Damit lassen sich in Max/MSP Raumklanganwendungen baukastenartig zusammenstellen. Unser Ziel ist es, dem Benutzer, der ja meist eher Musiker als Mathematiker ist, eine Schnittstelle zu bieten, mit der sich zweckmäßig, unkompliziert und ohne großes Wissen um die theoretischen Hintergründe arbeiten lässt.

DSP-Bausteine

Das Herzstück bilden die beiden Externalen "ambidecode~" und "ambidecode~", die für das Encodieren

ren und das Decodieren zuständig sind, also die ganze digitale Signalverarbeitung besorgen. Mit ihnen lässt sich eine beliebige Anzahl Klangquellen in das B-Format überführen – in erster, zweiter oder dritter Ordnung – und dann für eine beliebige Anzahl Lautsprecher decodieren. Die Trennung in zwei Module ermöglicht den Zugriff auf das B-Format, das somit bearbeitet oder aufgezeichnet werden kann. Ebenfalls kann so die Encodierung und

Decodierung auf zwei Rechnern laufen, sollten praktische Überlegungen dies erfordern.

Um die Positionen der virtuellen Klangquellen zu steuern werden dem Encoder pro Kanal die jeweiligen Koordinaten (Azimut, Elevation und Distanz) eingegeben. Eine weitere Variable legt den dB-Abfall pro Distanzeinheit fest. Der Decoder erwartet die Eingabe einer Liste mit den Koordinaten der Lautsprecher, um die entsprechenden diskreten Signale zu errechnen.

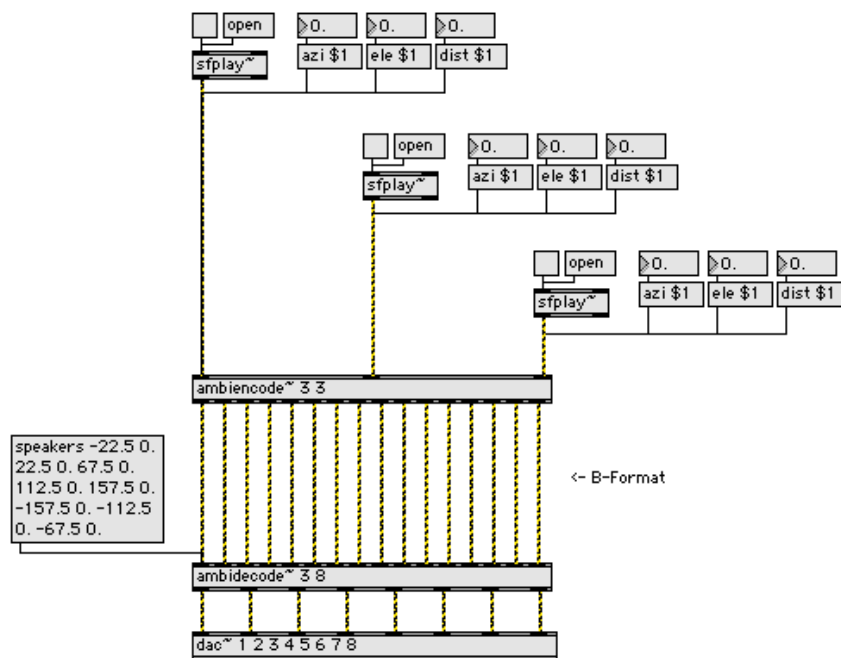


Abbildung 2: Eine einfache Anwendung, die drei Schallquellen auf acht Lautsprecher codiert in dritter Ordnung

Das Tiefpassfilter zur Simulation der distanzabhängigen Frequenzabsorption sowie ein Modul zur Erzeugung des Dopplereffekts, der psychoakustisch für die Wahrnehmung von Bewegung eine wichtige Rolle spielt, sind nicht fest in den Encoder eingebaut, sondern müssen vom Benutzer bei Bedarf selber in den Signalfuss integriert werden. Die Erfahrung zeigt, dass Komponisten hier unterschiedlich vorgehen und individuelle Vorlieben haben, ob und wie solche distanzabhängigen Bearbeitungen vorgenommen werden sollen. Ebenso müssen Verzögerungen, um Lautsprecherpositionen, die von

einer kreisrunden Aufstellung abweichen, von Hand nach dem Decoder in den Signalfuss eingefügt werden.

Graphische Benutzerschnittstelle

Optische Kontrolle über die Koordinaten der virtuellen Klangquellen im Raum bietet ein "ambimonitor" genanntes GUI-Objekt, das jede Quelle mit einem Punkt darstellt, ähnlich einem Radarschirm. Dieses Objekt erzeugt automatisch eine für die DSP-Bausteine verständlich formatierte Liste mit Koordina-

ten und kann somit direkt an jene angeschlossen werden.

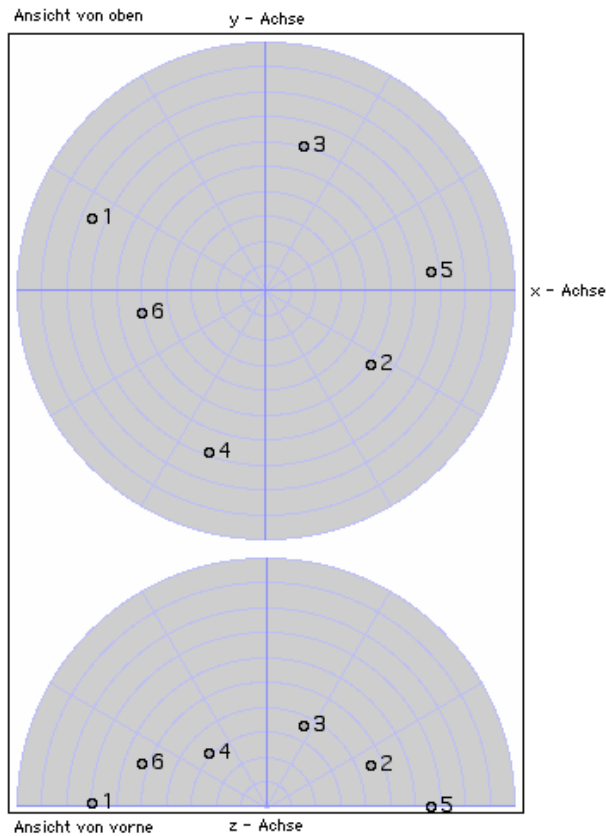


Abbildung 3: Der Ambimonitor

Die Punkte können mit der Maus am Bildschirm bewegt werden und mit Hilfe der Computertastatur können sie an- und abgewählt, sowie neu gesetzt oder gelöscht werden. Falls gewünscht, wird zu jedem Punkt der numerische Index oder ein frei wählbarer Name eingeblendet, sowie die Koordinaten, wahlweise in kartesischem oder polarem Format. Die Anordnung der Punkte auf dem Bildschirm kann zu jeder Zeit gespeichert und später wieder aufgerufen werden und damit lässt sich eine Bibliothek von Raumszenen für eine Komposition erstellen. Solcherart gespeicherte Anordnungen können schließlich in einer XML-formatierten Textdatei gespeichert werden.

Dem Benutzer, der die Bewegungen seiner Klangquellen algorithmisch steuern möchte, wird der Baustein "ambicontrol" gute Dienste leisten. Hier lassen sich automatische Bewegungen auf einzelne oder mehrerer Punkte anwenden, von der schlichten Kreisbahn über die Zufallsbewegung innerhalb eines begrenzten Volumens bis zur benutzerdefinierten Trajektorie. Dadurch, dass man diese Bewegungen in Geschwindigkeit, Ausdehnung und Lage interaktiv steuern kann, verlagert sich die Kontrolle des Benutzers auf eine höhere Ebene. Wenn man diese Kontrollobjekte miteinander verbindet oder kaskadiert, lassen sich Bewegungen beliebiger Komplexität erzeugen. Ein Beispiel hierfür findet sich in Abbildung 4.

Benutzerdefinierte Trajektorien können entweder als Liste von Koordinaten (Zeit, Position) eingegeben oder aber in Echtzeit als Bewegung aufgezeichnet werden. Diese Daten können ebenfalls in einer XML-formatierten Textdatei gespeichert werden.

Anwendungen

In einem Max/MSP-Patch finden sich üblicherweise Klangquellen, wie Mikrophoneingänge, synthetische Klangerzeuger oder auf einem Buffer, beziehungsweise einer Datei basierende Wiedergabemodule, gefolgt von irgendeiner Form von Klangbearbeitung, deren Audiosignale dem Encoder zugeführt werden. Die Koordinaten für den Encoder können vorprogrammiert sein oder werden vom Benutzer interaktiv gesteuert, gegebenenfalls mittels eines angeschlossenen Controllers. Die vom Decoder benötigten Koordinaten erhält man durch das Vermessen der Lautsprecherpositionen am jeweiligen Aufführungsort.

Weitere Elemente, wie distanzabhängige Filterung, Dopplereffekt oder Hall, werden wie bereits erwähnt an den entsprechenden Stellen eingefügt. Im Falle von Hall, können die frühen Reflexionen wiederum der Spatialisation zugeführt werden, um ihr räumliches Eintreffen zu simulieren, während der Nachhall der monophonen Komponente des B-Formats (Kanal W) beigemischt wird.

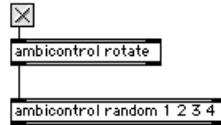
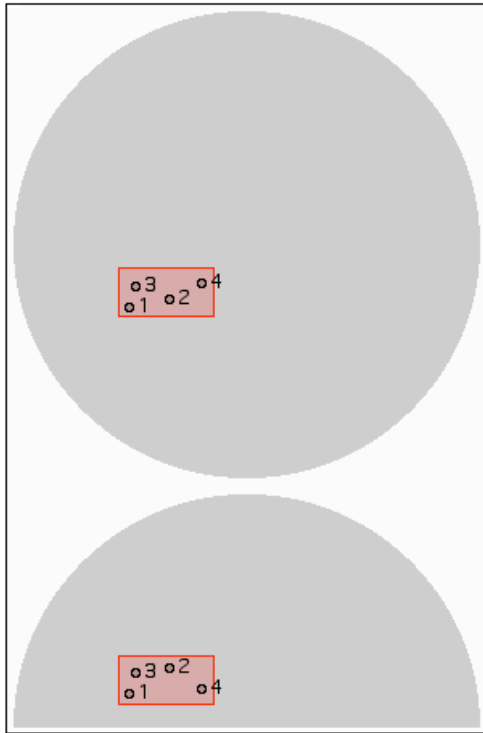
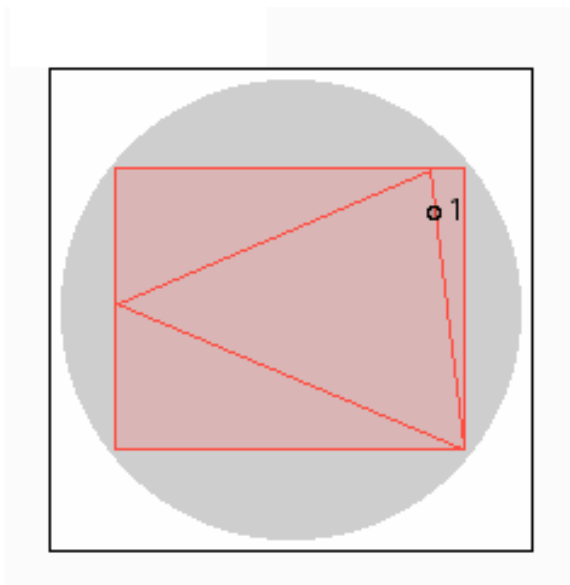


Abbildung 4: Vier Punkte in Zufallsbewegung, deren Begrenzung sich ihrerseits auf einer Kreisbahn um den Mittelpunkt befindet



```
<?xml version="1.0" standalone="yes" encoding="UTF-8"?>
<ambiscore version="1.0">
  <xyz-def handedness="right" x-axis="right" />
  <trajectory>
    <point>
      <time>0</time>
      <xyz>0.605751 0.600000 0.000000</xyz>
    </point>
    <point>
      <time>2000</time>
      <xyz>0.750000 -0.600000 0.000000</xyz>
    </point>
    <point>
      <time>4000</time>
      <xyz>-0.750000 0.030307 0.000000</xyz>
    </point>
    <point>
      <time>6000</time>
      <xyz>0.605751 0.600000 0.000000</xyz>
    </point>
  </trajectory>
</ambiscore>
```

Abbildung 5: Eine benutzerdefinierte Trajektorie und die dazugehörige XML-formatierte Textdatei

Die Max/MSP External des ICST wurden bereits vielfach eingesetzt in Konzerten elektronischer Musik, teils zur Wiedergabe von im B-Format gespei-

cherten Tonbandstücken, teils für live-elektronische Musik, sowie für anspruchsvollere Projekte mit Kombinationen von vorgefertigtem Material, Inter-

aktionen in Echtzeit und Instrumentalisten. Für Installationen mit räumlicher Beschallung wurden die Externalen ebenfalls schon mehrfach verwendet, wobei auch gerade hier sich die Flexibilität von Ambisonics bewährt hat, unabhängig zu sein von ortsspezifischen Bedingungen, ohne die Integrität des Raumklanges zu kompromittieren.

Die Effizienz dieser Technik, was die Rechenleistung angeht, ermöglicht es, die gesamte Spatialisation auf einem einzigen leistungsfähigen Laptop-Rechner auszuführen. Eine stattliche Anzahl virtueller Schallquellen kann in Echtzeit encodiert und decodiert werden, wobei immer noch genügend Prozessorleistung für andere Arten von Signalverarbeitung übrig bleibt.

Ausblick

In einem zukünftigen Schritt soll ein weiteres DSP-External programmiert werden, das die gesamte Encodierung und Decodierung in einem einzigen Modul vornimmt, für die Fälle, wo dies zweckmäßiger ist als zwei einzelne Module. Ebenfalls werden weitere Externalen zur algorithmischen Bewegungssteuerung dazukommen.

Eine Sammlung von "Pluggo" Plug-ins, um die Studioarbeit mit Sequenzerprogrammen zu unterstützen, ist in Arbeit. Ferner soll die Decodierung des B-Formats in ein binaurales Signal mittels HRTF entwickelt werden, um auch die Arbeit mit dem Kopfhörer zu ermöglichen. Schließlich ist geplant, die Externalen nach PD zu portieren und diese damit einer noch größeren Öffentlichkeit verfügbar zu machen.

Philippe Kocher, Komponist und Computermusiker, arbeitet am ICST Zürich als Software-Entwickler.

www.icst.net

philippe.kocher@zhdk.ch