



## Virtual Gamelan Graz (VGG)

Projektleiter: Alois Sontacchi (sontacchi@iem.at)

Mitarbeiter: Gerhard Eckel, Robert Höldrich, Gerhard Nierhaus, Franz Zotter

### 1 Einleitung

Gamelan - „der Ursprung des Gongs“ – hat auch bei westlichen Musiker, Künstler und Musikwissenschaftler schon seit Jahrzehnten für Interesse und Auseinandersetzung gesorgt, wobei eine Reihe von Problemen nach wie vor ungelöst sind. Gamelan ist die Bezeichnung für Ensembles in Südostasien, die vorwiegend aus gestimmten Idiophonen (Gongs, Gongspiele, Metallophone) sowie weiteren Instrumenten bestehen (siehe dazu Abb.1). In Zentraljava (Indonesien) wird mit solchen Orchestern die an lokalen Fürstenhöfen entstandene klassische Musik gespielt, die heute auch an örtlichen Musikhochschulen gelehrt wird.

Virtual Gamelan Graz (VGG) soll im Hinblick auf musikethnologische Forschung neue Möglichkeiten bieten, welche auch den gegenwärtigen Stand der Entwicklung im Bereich der spektralen und spatialen Klangsynthese sowie der algorithmischen Verfahren zur Generierung von musikalischen Strukturen nutzen.

Ziel des Projektes ist ein steuerbares virtuelles Gamelan-Orchester, das selbständig Kompositionen der klassischen Musik Zentraljasas in verschiedenen Stimmungen und modifizierbaren Klangfarben generiert und wodurch das Spiel auf unterschiedlichen Sets von Instrumenten simulieren werden kann.

Dieses interdisziplinäre Forschungsprojekt erfolgt in Zusammenarbeit mit dem Institut für Musikethnologie (IME) der Kunstuniversität Graz (KUG). Die Gesamtprojektleitung obliegt Prof. Gerd Gruppe. Das Projekt VGG startet mit Oktober 2005 und wird für den Zeitraum von 2 Jahren durch Fördermittel des Zukunftsfonds Steiermark ermöglicht.



Abbildung 1: Gamelanorchester (Quelle: IME Homepage)

## 2 Aufgabenstellungen und Vorhaben

Durch die oben formulierte Projektidee ergeben sich einige Arbeitsfelder. Im Folgenden werden hier nur jene Aufgabenbereiche aufgelistet und kurz beschrieben, die durch die Mitarbeiter am IEM hauptsächlich behandelt werden. Zu diesen Aufgabenbereiche zählen die:

- Algorithmische Modellierung musikalischer Strukturen
- Klangmodellierung (Klanganalyse- und Synthese)
- Aufnahmetechnik und Modellierung der Abstrahlcharakteristik
- Softwareimplementierung

### 2.1 Algorithmische Modellierung musikalischer Strukturen

Als Teilbereich des *Virtual-Gamelan* Projektes wird versucht grundlegende musikalische Strukturen der klassischen Stile (Karawitan) der zentraljavanischen Gamelanmusik zu modellieren. Für die unterschiedlichen Problemstellungen, die sich in Zusammenhang mit diesem Teilprojekt ergeben, können prinzipiell zwei unterschiedliche Herangehensweisen verwendet werden:

- Regelbasierte Methoden eignen sich besonders für die Abdeckung von Domänen, bei denen eine Derivierung musikalischer Struktur aufgrund gängiger Musikpraxis eindeutig möglich ist, oder aber ein explizites Regelwerk vorhanden ist. Diese Vorgehensweise wird bei der Umsetzung der Instrumentalparts im Bonangan-Stil<sup>1</sup> verwendet. Für die Erstellung von Regelwerken, aufgrund bekannter Gesetzmäßigkeiten oder aber neuer wissenschaftlicher Erkenntnisse ist Expertenwissen notwendig, welches am IME zur Verfügung steht. Dem Institut für Elektronische Musik und Akustik fällt in diesem Fall jedoch die Aufgabe zu, für das Regelwerk ein programmiertechnisches Umfeld zu adaptieren um dem IME eine selbständige Arbeit zu ermöglichen.
- Eine andere Möglichkeit musikalische Struktur eines bestimmten Idioms zu erzeugen bietet sich durch die Anwendung von Problemlösungsmethoden, welche im Prinzip kein problemspezifisches Wissen über die untersuchte Domäne benötigen. Mögliche Zugänge sind hier unterschiedliche stochastische Verfahren (vgl. *Conklin, Chai et al.*), fallbasiertes Schließen (vgl. *Pereira*) sowie unterschiedliche Ansätze aus dem Gebiet des maschinellen Lernens (vgl. *Conklin et al.*) und einiges mehr. Notwendig für diese Zugänge ist jedoch ein ausreichend großer Bestand von Datenmaterial aus der zu modellierenden Domäne. Das IEM arbeitet mit diesen Methoden und versucht eine algorithmische Modellierung der Kernmelodie. Das IME stellt die notwendigen Daten zur Verfügung und überprüft die gewonnenen Resultate<sup>2</sup>.

---

<sup>1</sup> auch: lauter Stil, welcher ein eingeschränktes Instrumentarium verwendet

<sup>2</sup> eventuell in Zusammenarbeit mit javanischen Musikern

## 2.2 Klangmodellierung – Klangsyntheseverfahren

Zur Klangmodellierung stehen in der digitalen Signalverarbeitung unterschiedliche Verfahren zur Verfügung. Um z.B. ein Instrument durch *Sampling-Synthese* vollständig abzudecken, müssten sämtliche Spielzustände über den gesamten Tonumfang aufgenommen und geeignet gespeichert werden. Neben der vollständigen physikalischen Modellierung (*physical modeling*) von Instrumenten gibt es auch die Möglichkeit einen Klang durch Zusammensetzung von elementaren Signalkomponenten zu realisieren, wobei eine exakte Nachbildung des zeitlichen Verlaufs oder des Amplitudenspektrum angestrebt wird; oder es wird überhaupt nur eine Lösung die weitgehend perzeptiv übereinstimmt gesucht. Während beim *physical modeling* Änderungen der Spielweise z. B. unterschiedliche Anblasstärke eines Horns genau durch einen Parameter bei der Resynthese zur Klangsteuerung eingestellt werden kann, ist bei der *Frequenzmodulations- (FM) Synthese* die Steuerung bei weitem nicht so intuitiv und einfach. Bei der *additiven Klangsynthese* kann durch die geeignete Wahl von Signalkernen eine Lösung zwischen reiner physikalischen Modellierung und reiner Signalresynthesierung gefunden werden.

Für eine echtzeitfähige Umsetzung ist generell die damit verbundene Rechenleistung und der erforderliche Speicheraufwand für die einzelnen Syntheseverfahren zu berücksichtigen. Da in diesem Projekt nicht nur die Umsetzung, sprich Klangsynthesierung, für ein Instrument erfolgen sollte, sondern ein ganzes Orchester, ist eine minimale Anzahl von klangbeschreibenden Parametern erstrebenswert.

Aufbauend auf erste gemeinsame Untersuchungen des IME und IEM an Lamellophonen (vgl. *Grube et al.*) die zur Gruppe der Idiophone zählen, konnte gezeigt werden, dass deren Klang als eine Summe von exponentiell gedämpften Sinusschwingungen aufgefasst werden kann. Daher liegt es nahe, die Analyse der Lamellophon-Klänge mit dem Steiglitz-McBride Ansatz zu lösen. Dieser Ansatz ist ein iteratives Verfahren, basierend auf der Prony-Methode, welches einen Signalausschnitt (Impulsantwort) durch eine Summe von Pole innerhalb des Einheitskreises (bzw. entlang der Frequenzachse) modelliert. Aus den Residuen der Übertragungsfunktion wird die Amplitude und Startwinkel der einzelnen gedämpften Sinusschwingungen bestimmt. Wird bei der Analyse das originale Klangsample durch frequenzselektive Subbändern aufgeteilt, so ist die Verteilung der Pole pro Frequenzband steuerbar. Dadurch kann trotz geringer Gesamtanzahl von Analyseparameter eine weitgehend perzeptiv nicht unterscheidbares Klangabbild erreicht werden.

## 2.3 Abstrahlverhalten – Abstrahlcharakteristik

Untersuchungen von *J. Meyer* an klassischen westlichen Instrumenten zeigen instrumentenspezifische frequenzabhängige Abstrahlverhalten. Während Blechblasinstrumente durch die spezielle Bauform (geführte Welle und Anpassung der Strahlungsimpedanz durch den Trichter) eine relativ einfach beschreibbare Schallausbreitung um die Trichterachse (nur frequenzabhängige Bündelung)

besitzen, ergeben sich bei Holzblasinstrumenten, bedingt durch die Grifflöcher und Mundstück, oder auch bei Streichinstrumenten eine wesentlich komplexeres Abstrahlverhalten.

Neben der Klangmodellierung, die ein möglichst realistisches auditives Abbild der realen Musikinstrumente anstrebt, spielt somit auch die Abstrahlcharakteristik eine essentielle Rolle. Daher ist bei einer möglichst natürlichen Wiedergabe dieser Sachverhalt zu berücksichtigen.

Die Erhebung der Richtcharakteristik impliziert aber damit auch eine speziell geeignete Aufnahmetechnik. Dies kann mittels einer Kartierung der Schallschnelle über der Klangfläche durch eine Nahfeldmessung (2D Mikrofonarray) bzw. mittels einer Kartierung des Schalldrucks durch eine Fernfeldmessung bei gleichzeitiger Unterdrückung des Einflusses des Aufnahme-raums durch akustisches Beamforming (*akustische Kamera*) realisiert werden. Bei diesen Untersuchungen müssen auch unterschiedliche Anregungspunkte (kleine Varianz um den idealen Anschlagpunkt) jeweils berücksichtigt werden.

Bei der Wiedergabe unter Berücksichtigung der unterschiedlichen Richtcharakteristika werden spezielle Lautsprecheranordnungen (-Arrays) erforderlich. Durch Zusammenwirken von mehreren elementaren Einzelquellen mit einfachen Richtcharakteristika ist es möglich, die bei der Schallschnellenkartierung gemessenen komplexere Richtcharakteristika zu approximieren.

## **2.4 Implementierung**

Die Implementierung des virtuellen Gamelanorchester sieht eine digitale Synthese, die durch einen elektronischen Score (algorithmisch generierte musikalische Struktur) und freien Steuerparametern für die Klangsynthese (Intonation, Klangfarbe – als vorliegendes bzw. modifizierbares Preset) angesteuert wird, vor. Darüber hinaus werden die im Vorfeld erhobenen Richtcharakteristika der einzelnen Instrumente zum Teil softwaretechnisch durch die Signalverarbeitung als auch durch entsprechende Lautsprecheranordnungen berücksichtigt und über diese (im Raum) abgestrahlt.

Ein wesentlicher Punkt für die Verwendbarkeit in Feldstudien ist dabei auch die Skalierbarkeit der Software, damit diese nicht nur auf einer Rechnerfarm sondern auch auf einem einzelnen Rechner (Laptop) verwend- und einsetzbar ist.

Die Implementierung erfolgt mit pure data (PD), einer grafischen Computermusik Programmiersprache. Anhand einer grafischen Bedienoberfläche können importierte Scores selektiert werden, einzelne Signalparameter für die Resynthese selektiv modifiziert werden etc.. Die Wiedergabe erfolgt durch die oben genannten Lautsprecheranordnungen oder mittels geeigneter Spatialisierung im IEM CUBE. Dabei werden die Parameter der Oszillatorbank (gedämpfte Sinusschwingungen) durch die Ergebnisse der Klanganalysen festgelegt. Für die Interaktivität bezüglich Intonation und Klangfarbe werden die prominentesten Teilklangkomponenten erstgereiht und können auch zur Spielzeit beliebig variiert werden.

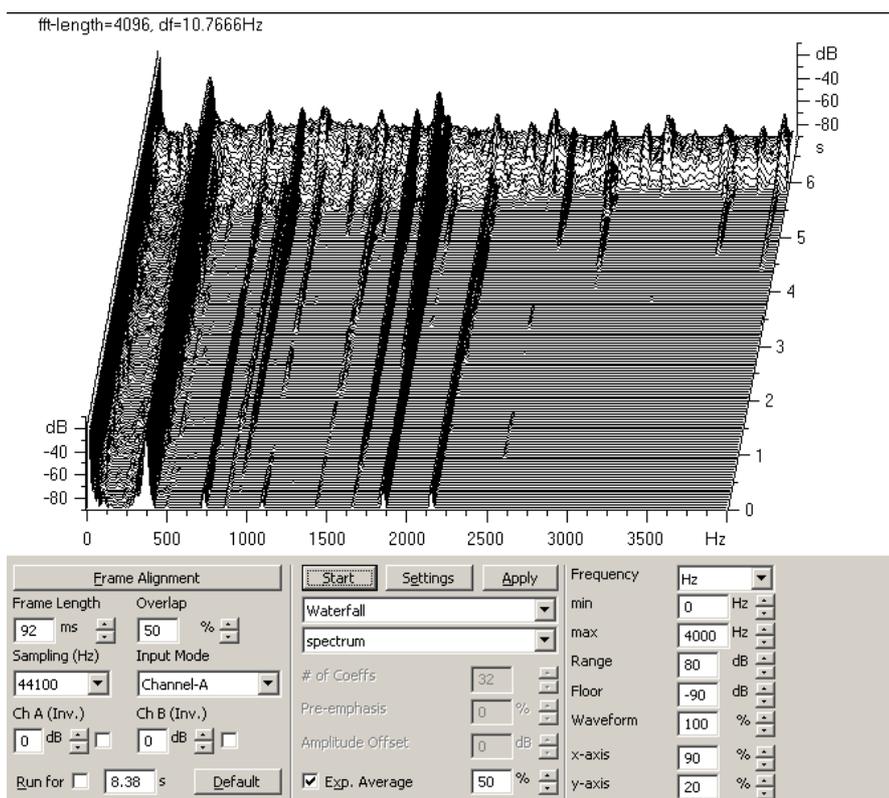
### 3 Erste Ergebnisse

In den ersten Schritten im Projekt wurden einzelne Platten der Orchesterinstrumente und der Gong-Ageng (siehe Abb. 2) aufgenommen. Anhand dieser Aufnahmen können die jeweiligen Eigenschaften der Klangkörper bzgl. Ein- und Ausschwingverhalten (Abb. 3) sowie die spektrale Zusammensetzung der Klangkomponenten als auch prinzipielle Abstrahlverhalten untersucht und mit Ergebnissen aus der Literatur verglichen werden.



**Abbildung 2:** Gong-Ageng an der KUG  
(Quelle: Homepage IME)

In weiterer Folge wurden sowohl des 5- (*Slendro*) als auch des 7-stufige (*Pelog*) Set der KUG aufgenommen - zum einen für den Aufbau einer Samplebank und zum anderen auch für die Klanganalyse und -synthesestudien.



**Abbildung 3:** Wasserfalldiagramm für die Klangplatte 3 des Instrumentes Demung.

Darüber hinaus wurden für eine Dokumentations- und Präsentations-DVD bei einem Workshop im Dezember 2005 des IME eine Surroundaufnahme mit Video erstellt, deren Fertigstellung im Frühjahr 2006 geplant ist.

### 3.1 Ergebnisse algorithmische Modellierung musikalischer Strukturen

Die Softwaregrundlage für die Part-Ableitung des IME wurde in SuperCollider erstellt. Ein Ausschnitt einer Kernmelodie (*balungan*) – erste Zeile - und ihrer zugehörigen Ableitungen – folgende Zeilen - findet sich in der Abbildung 4.

A) balungan (notiert)	3	1	3	2	3	1	3	2	5	6	1	2	1	6	4	5
saron/demung	3	1	3	2	3	1	3	2	5	6	1	2	1	6	4	5
slenthem	.	3	.	2	.	3	.	2	.	1	.	6	.	4	.	5
peking lanc.	3	1	3	2	3	1	3	2	5	6	1	2	1	6	4	5
bon.b. lanc.	<u>2</u>	.	<u>5</u>	.	<u>5</u>	.										
bon.p. lanc.	.	.	<u>2</u>	.	<u>5</u>	.										
kendh. lanc.	.	.	p	.	p	.	p	.	p	.	p	.	p	.	p	.

Abbildung 4: Beispiel zu einer Kernmelodie und deren zugehörige Ableitungen.

### 3.2 Ergebnisse Klangmodellierung – Klangsynthefverfahren

Bezüglich der Klanganalyse und –synthese wurden bisher drei unterschiedliche Ansätze verfolgt, die im Folgenden kurz beschrieben werden.

Der erste und intuitivste Ansatz untersucht das aufgenommene Zeitsignal eines Klangkörpers nach den perzeptiv relevanten Klangkomponenten d.h. es werden nur all jene Komponenten betrachtet, die über der Ruhehörschwelle liegen und die auch durch weitere in dem Signal innewohnenden Komponenten nicht maskiert werden (vgl. Abb. 5). Deren spektrale Lage und deren zeitlicher Verlauf wird durch einfache Funktionsformen approximiert und einer somit **gehörgerechten Additiven Klangsynthese** zugeführt.

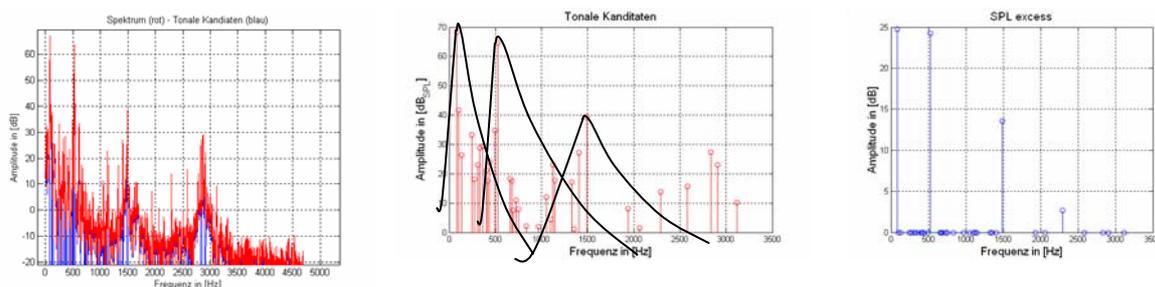


Abbildung 5: Extrahieren von perzeptiv relevanten Klangkomponenten.

Beim zweiten Ansatz handelt es sich um eine, ähnlich der bereits in Abschnitt 2.2 angesprochene, Modellannahme - das Gesamtsignal als Superposition einzelner ausklingender Exponentialfunktionen zu beschreiben. Im vorliegenden Ansatz wird jedoch aufgrund des perzeptiv relevanten transienten Klangbeginns, der durch dieses reine Polstellen-Modell mit geringer Ordnung nur unzureichend nachgebildet werden kann, zusätzliche Nullstellen hinzugezogen. Somit handelt es sich um einen Ansatz der in der Literatur auch als „**Frequency-Zooming ARMA Modeling**“ bekannt ist. In der ersten Phase werden zuerst die relevanten Signalkomponenten durch eine *peak-detection* aus dem Langzeitspektrum des Originalsignals ermittelt. Danach erfolgt eine Modulation anhand der relevanten Frequenzkomponenten in das Basisband (*heterodyning*) und eine Abtastratenreduktion um einen Faktor K. Im

Basisband werden die Pole durch den *Steiglitz-McBride* Algorithmus berechnet. Abschließend werden die bestimmten Pole aufgrund der Abstratenreduktion und Modulation wieder in die ursprüngliche Lage rotierte und die Radien entsprechend korrigiert. Von den evaluierten Pol- u. Nullstellen wird jedoch nur das reine Polstellenfilter herangezogen. Im folgenden Schritt wird mit dem erhaltenen inversen Filter das Originalsignal gefiltert. Das dadurch erhaltene Ausgangssignal wird als Anregungssignal (Nullstellenfilter) für das oben bestimmte Polstellenfilter verwendet. Mit diesem Ansatz ergeben sich immense Datenkompressionen. Als Beispiel für einen Zahlenvergleich: ein 8sek Klang eines Gong-Ageng kann mit 200 Anregungssamples und 6 komplexen Polstellen klanglich weitgehend ausreichend nachgebildet werden – bei direkten Sampling (und zusätzlicher Bandbegrenzung [fs@16kHz](#)) ergeben sich immerhin 128.000 Datensamples.

Der dritte Ansatz, der bisher behandelt wurde, folgt der Theorie der ***Empirical Mode Decomposition*** (EMD), wobei hier noch einige Modifikation für eine Verbesserung der spektralen Auflösung unter recheneffizienter Umsetzung erfolgte. Bei EMD handelt es sich ebenso um einen physikalisch motivierten Ansatz. Dabei wird versucht aus dem Originalsignal alle vorhandenen Moden, die durch Amplituden und Frequenzmodulationen beschrieben werden können, zu extrahieren. Diese werden als *Intrinsische Moden* bezeichnet. Diese Intrinsische Moden werden durch einen iterativen Prozess gewonnen. Ausgangspunkt ist das vorliegende Originalsignal. Es werden die Einhüllende für positive und negative lokale Signalmaxima bestimmt und daraus die mittlere Einhüllende berechnet. Diese wird vom Ausgangssignal abgezogen und das resultierende Signal als Residuum bezeichnet. In Folge wird der beschriebene Prozess, wobei als Ausgangssignal dient nun jeweils das Residuum der vorhergegangenen Iteration, solange wiederholt, bis die mittlere Einhüllende den Mittelwert identisch Null und nahezu keine Varianz besitzt. Bei Unterschreitung des Abbruchkriteriums wird das letzte Residuum als erster Intrinsischer Mode (IMF) bezeichnet. Wird nun das erste IMF vom Originalsignal abgezogen kann obiger Algorithmus auf ein neues Ausgangssignal angewandt werden, wobei nun nach durchlaufen der Iterationen abschließend die zweite IMF erhalten wird usw.

Um eine höhere Anzahl von schmalbandigeren IMFs zu erhalten, werden diese nicht sukzessive aus dem Signal bzw. den Residuen ermittelt, sondern jeweils nur die erste IMF aus einem hochpassgefilterten und danach zu tieferen Frequenzen verschobenen Signal berechnet. So wird für die erste IMF das Signal mit z.B.  $f_c=1600$  Hz hochpassgefiltert (das komplementäre TP-Signal wird gespeichert), das Hochpasssignal wird durch Einseitenbandmodulation (SSB) um z.B. 1200 Hz gegen DC verschoben. Aus dem demodulierten Hochpasssignal wird die erste IMF\_SSB und ein Residuum\_SSB, welche durch Einseitenband-Demodulation wiederum um z.B. 1200 Hz in das Originalband geschoben werden. Dies ergibt die erste IMF, das Residuum wird mit dem TP-Signal addiert und fungiert als Ausgangspunkt für die nächste Hochpassfilterung mit z.B.  $f_c= 1200$  Hz usw.

Die dadurch bestimmten IMFs ergeben in Summe das ursprüngliche Originalsignal. Anhand dieser Methode kann eine sehr gute Signalrekonstruktion erreicht werden. Da sich nun die IMFs durch niederfrequente FM- und AM-Signale darstellen lassen,

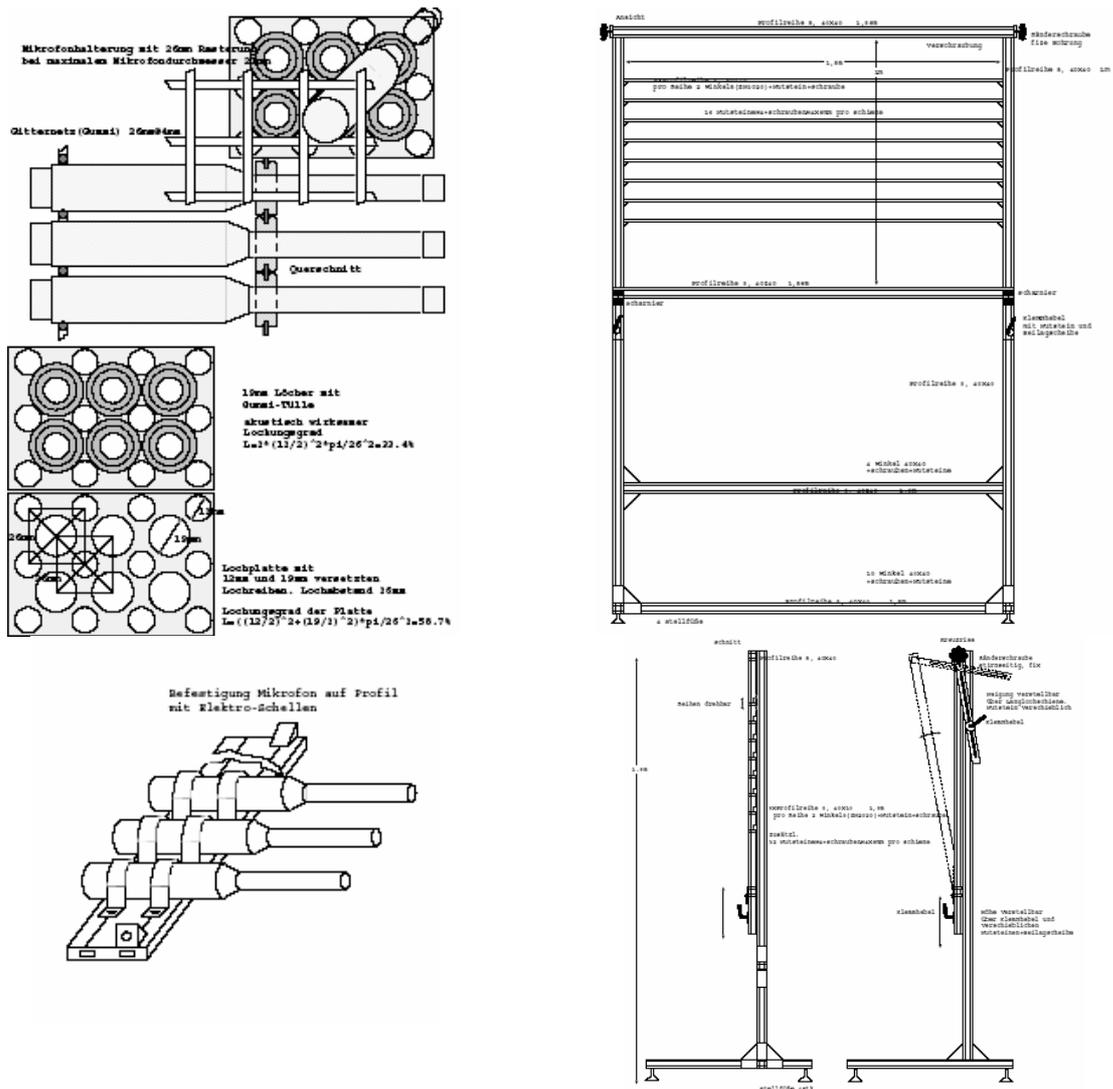
können diese Funktionsverläufe unterabgetastet einer FM-AM-Oszillatorbank zugeführt werden.

### **3.3 Ergebnisse Abstrahlverhalten – Abstrahlcharakteristik**

Für die exakte und reproduzierbare Vermessung der Klangabstrahlverhalten ausgewählter Klangkörper wird derzeit am IEM ein spezieller in-situ Messplatz eingereicht (siehe Abb. 6). Dieser ermöglicht eine Vermessung einer Schallquelle, durch die zusätzlich räumlich aufgebrachte hohe Absorption, unter nahezu Freifeldbedingungen in einem definierten Frequenzbereich. Des Weiteren wurde ein spezielles 64-Kanal Mikrofonarray (siehe Abb. 7) für die Erfassung von Abstrahlcharakteristika, als auch für die Schalldruckkartierung geplant und in Auftrag gegeben.



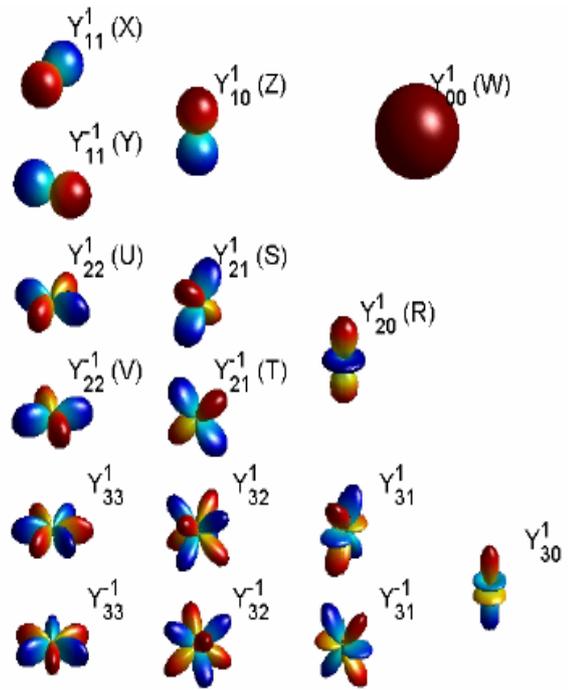
**Abbildung 6:** In-situ Messplatz am IEM im 1.Stock.



**Abbildung 7.:** Konstruktionspläne f. das 64-Kanal Mikrofonarray

Links: oben und unten 2 alternative Befestigungsmöglichkeiten f. unterschiedliche Mikrofonabstände.  
Rechts: Schwenkbarer Arrayhalterahmen in Frontal- und Seitenansicht.

Bei der Umsetzung der diversen Abstrahlcharakteristika kommt eine Lautsprecheranordnung in der Form eines Ikosaeders zum Einsatz (siehe Abb. 8). Für die breitbandige Wiedergabe kommen kombinierte Tief- und Hochtonlautsprecher (ähnlich einem Koaxialsystem) zum Einsatz. Die Reproduktion der diversen Richtcharakteristik erfolgt durch Zusammensetzung von definierten Kugelflächenfunktionen (*Multipole*, siehe Abb.8), die mit der Lautsprecheranordnung jeweils bis zu einer bestimmten Ordnung (Genauigkeit) entsprechend nachgebildet werden können. Somit erfolgt die Schallfeldzusammensetzung ähnlich wie im IEM-CUBE basierend auf dem erweiterten Ambisonic Ansatz nur jetzt nicht von außen nach innen sondern von innen nach außen.



**Abbildung 8:** Rekonstruktion beliebiger Richtcharakteristika.

Links: Lautsprecherbox für Iksaederarray.

Rechts: Kugelflächenfunktionen (Multipole) bis zur 3.Ordnung.