A. Weisbecker, M. Burmester & A. Schmidt (Hrsg.): Mensch und Computer 2015 Workshopband, Stuttgart: Oldenbourg Wissenschaftsverlag, 2015, S. 325-330.

Gedanken zur Nutzung von Handgesten in der Musikproduktion

Axel Berndt, Simon Waloschek, Aristotelis Hadjakos

Zentrum für Musik- und Filminformatik, HfM Detmold / HS Ostwestfalen-Lippe

Zusammenfassung

Die digitale Musikproduktion ist ein Gebiet, welches mit seinen komplexen Anwendungen die Entwicklung von Benutzungsschnittstellen regelmäßig vor große Herausforderungen stellt. Hier herrschen meist graphische Bedienoberflächen vor, die dem WIMP-Paradigma verhaftet sind. Eine darüber hinaus gehende Auseinandersetzung zur Einbindung von Post-WIMP-Interfaces in diesem Anwendungsszenario ist bislang weder in der Wissenschaft noch in der Praxis nennenswert. Am Beispiel der Freihandgesteninteraktion will der vorliegende Text sinnvolle Anknüpfungspunkte aufzeigen und motivieren. Ihre Einbindung in die etablierten Setups der Digital Audio Workstations wird ausgearbeitet und diskutiert.

1 Anknüpfungspunkte an bestehende Szenarien

Tonstudios sind von je her Multi-Device-Umgebungen. Spezialisierte Geräte kümmern sich um die Klangerzeugung, Generierung von Audioeffekten, Tonaufnahme, Abmischung usw. Sie sind zumeist über Audio- und MIDI-Kabel miteinander vernetzt. Im Zentrum dieser Umgebung steht heute die *Digital Audio Workstation* (kurz DAW), ein Computer, welcher alle Komponenten integriert und um Software-Komponenten ergänzt. Auch die Software-Architektur, nicht nur des alle Signalströme zusammen führenden Sequencer-Programms, das gleichermaßen als Host fungiert, sondern der Gesamtheit aller Komponenten, ist mittlerweile hochkomplex. Software-Module (Plugins) erweitern die Architektur und werden über Schnittstellen wie VST, VST3, AAX, AU und RE eingebunden. Wer nicht auf die speziellen Klangeigenschaften der analogen Geräte angewiesen ist, kann für alle Funktionen auch entsprechende und oft kostengünstigere Software-Alternativen finden, sodass eine moderne Musikproduktionsumgebung sogar als reine Software-Lösung möglich und praktikabel ist.

Trotz dieser hochgradigen Verteilung von Funktionen auf eine Vielzahl miteinander vernetzter, spezialisierter Komponenten sind deren Benutzungsschnittstellen einer aus der klassischen Hardware kommenden, relativ einheitlichen Tradition verhaftet. Es dominieren Fader, Drehregler, Buttons und sogar digitale Kabel in überladenen, graphischen Benutzeroberflächen, wo sie mittels Maus und Tastatur bedient werden. Dass dies bei weitem keine optimale

Lösung darstellt, ist hinlänglich bekannt. So ist etwa die Maus für radiale Eingaben denkbar ungeeignet, weshalb Drehregler meist über eindimensional horizontale oder vertikale Bewegungen bedient werden. Während Multitouch- und Stifteingabe einen allmählichen Einzug in diese Umgebung halten, findet der Reichtum an Post-WIMP-Paradigmen, welche die Mensch-Computer-Interaktion darüber hinaus kennt, kaum Widerhall in diesem Anwendungsfeld.

Neue Interface-Technologien kommen vor allem im Zusammenhang mit digitalen Musikinstrumenten zum Einsatz, was die Tagungsbände der alljährlich stattfindenden Konferenz NIME anschaulich widerspiegeln. Eine vergleichbar umfangreiche und tiefgründige Auseinandersetzung und Befruchtung für die Musikproduktion fehlt aber fast gänzlich, also gerade dort, wo die Vielzahl von spezialisierten Geräten eine entsprechende Spezialisierung und Optimierung von Benutzungsschnittstellen nahelegen, ja geradezu einfordern würde. Die wenigen Ansätze zur in diesem Text beispielhaft besprochenen Freihandgesteninteraktion kommen in ihrer Wahrnehmung selten über den Status von Spielereien hinaus. Neuartige Eingabemodalitäten werden häufig als Ersatz für Bestehendes eingeführt. Der Eingriff in etablierte Workflows ist dabei oft so gravierend, dass sie in der Praxis nicht angenommen werden. Wie aber kann diese Hürde überwunden und Freihandgesteneingabe als praktikable Bereicherung in diese Szenarien eingeführt werden? Mit dieser Frage mag sich jeder Entwickler in diesem Feld konfrontiert sehen. Deren Gedanken dazu werden jedoch kaum festgehalten. Deshalb ist es das Anliegen dieses Textes, sie einmal systematisch nachzuzeichnen und zur Diskussion zu stellen.

Bevor die technische Einbindung der Freihandgestensteuerung in klassische DAW-Setups diskutiert wird, sollen die verschiedenen Anwendungsszenarien vorgestellt werden. Dabei wird das Feld der digitalen Musikinstrumente, das neuartigen Eingabemodalitäten sehr wohl offen gegenübersteht, bewusst ausgespart. Im Fokus dieser Ausführungen steht die Musikproduktion, im Besonderen die Arbeit in der DAW. Hier kann und soll diese Freihandgesteneingabe kein Ersatz von etablierten Modalitäten wie Maus- und Tastatureingabe sein, die sich an vielen Stellen im besten Sinne bewährt haben. Vielmehr ist sie dort als Ergänzung gedacht, wo noch Optimierungspotential besteht und ein Mehrwert zu erwarten ist. Diese Anknüpfungspunkte werden bei den weiteren Betrachtungen identifiziert.

Die **professionelle Musikproduktion** findet im Tonstudio unter entsprechend eingerichteten Abhörbedingungen statt. Hier ist die Nutzung unterschiedlichster, spezialisierter Geräte am deutlichsten ausgeprägt. Ansätze zum Mixing via Freihandgesten existieren für dieses Szenario zwar bereits (Balin & Loviscach 2011, Ratcliffe 2014), haben sich bislang aber nicht etabliert – wohl auch deshalb, weil der Hardware-Fader am Mischpult bereits als eine optimale Lösung angesehen werden kann. Das größte Potential für die Freihandgestensteuerung liegt vielmehr in der höherdimensionalen Parametersteuerung, etwa von Effekten und der Klangsynthese. Hier werden, zumeist über Drehregler, mehrere zueinander in Bezug stehende Parameter eingestellt. Mit zwei Händen kann ein Nutzer lediglich zwei Drehregler gleichzeitig bedienen, mit der Maus sogar nur einen. Diese Begrenzung kann die deutlich höherdimensionale Freihandgesteneingabe überwinden. Für die Eingabe statischer Werte mögen Maus und Tastatur dennoch die bessere Lösung sein, denn die sich frei in der Luft bewegende Hand korrespondiert eher mit einer kontinuierlichen als mit einer diskreten Werteingabe. Das größte

Potential ergibt sich deshalb dann, wenn die Effektsteuerung kontinuierlich und live zur laufenden Musik geschieht. Die aus den Gesten abgeleiteten Steuerdaten können als Sequenz von MIDI-Befehlen in der DAW aufgezeichnet werden. Bei einer rein am Computer produzierten Musik (ohne Einspielungen durch menschliche Musiker) kann dieses Verfahren auch zur Steuerung expressiver Parameter (Tempo, Dynamik, Artikulation, Klangfarbe; siehe Berndt 2011) verwendet werden. So kann vergleichsweise schnell eine bereits recht facettenreiche „Rohinterpretation" eines vorgegebenen musikalischen Materials mittels Handgesten erstellt werden. Deren Details können dann über etablierte Methoden (Maus, Tastatur, Stift) ergänzt und bearbeitet werden.

Auch die **Musikwissenschaft**, speziell die **Interpretationsforschung** und **Musikedition**, sind potentielle Anwendungsfelder. Im Zentrum der Editionsarbeit mag zwar der musikwissenschaftlich aufbereitete und (meist noch) gedruckte Notentext stehen. Die im Editionsprozess gefällten Entscheidungen etwa über Vortragsanweisungen werden aber die mit dem Text entstehenden Darbietungen prägen und damit auch wieder das klingende Resultat. Dabei steht der Editor oft vor der Entscheidung zwischen mehreren Varianten, die sich je nach Quelle unterscheiden können. Typische Fragestellungen können sein: Wo beginnt ein Crescendo-Pfeil und bis wohin geht er? Wie schnell ist das geforderte Aufführungstempo? Welche Artikulation ist am plausibelsten? Will der Bearbeiter verschiedene Lösungen am klingenden Musikstück erproben, so ist er entweder auf die gegebenen Funktionalitäten von Notensatzprogrammen angewiesen oder mit einem unzumutbaren Produktionsaufwand im Sequencer konfrontiert. Auch die Ergänzung des kritischen Berichtes um Klangbeispiele stößt auf zunehmendes Interesse. Mit Techniken zur ausdrucksvollen Bearbeitung einer rein notengetreuen Wiedergabe in eine „Rohinterpretation", wie sie bereits im vorherigen Abschnitt beschrieben wurden, kann deshalb auch in der musikwissenschaftlichen Editionsarbeit ein Mehrwert geschaffen werden.

In den Bereichen der **hobbymäßigen Musikproduktion** und des **Home Recordings** können Handgesten als Eingabemodalität einen leichteren und im besten Falle auch intuitiveren Zugang zu allgemein als eher abschreckend komplex empfundenen Themenfeldern wie der Klangsynthese und Effektsteuerung geben. Das kann positiv zu deren Bedien- und Erlernbarkeit beitragen. Sensoren wie der Leap Motion Controller (Leap Motion, Inc 2013) sind zudem unkompliziert anzuschließen und in Betrieb zu nehmen. Damit ist diese Eingabemodalität direkt verfügbar, was in diesem Anwendungsfeld von ganz entscheidender Bedeutung ist. Ein Nutzer, der nur gelegentlich für ein oder zwei Stunden in der DAW arbeitet, nimmt keine langwierigen Aufbau- und Startphasen in Kauf. Noch drastischer ist das im Bereich des **Jammings** der Fall. Entsprechende Programme, wie etwa der *Music Maker Jam* (MAGIX Software GmbH 2015) müssen ad hoc ausführbar sein und die kreative Arbeit mit dem musikalischen Material sofort, ohne jede Vorbereitung ermöglichen. Eine Reduktion der interaktiven Parameter zu Gunsten der leichteren Bedienbarkeit wird von fortgeschrittenen Nutzern häufig als künstlerische Einschränkung wahrgenommen und abgelehnt. Hier können Freihandgesten einen schnellen Zugang zu einer größeren Anzahl bedienbarer Parameter geben. Vor allem aber die Kombination mehrerer Modalitäten (z.B. Freihand und Touch) scheint in diesem Szenario vielversprechend.

2 Der Griff in die DAW

Im Musikproduktionsprozess bieten sich Handgesten vor allem dort an, wo Effekte, Instrumentenklänge und expressive Parameter wie Tempo und Dynamik kontinuierlich über die Zeit gesteuert werden sollen. Die Einbindung in eine typische DAW-Architektur wird in Abbildung 1 veranschaulicht und einzelne Gesichtspunkte im Weiteren diskutiert. Als Klangquellen können (akustisches) Audiomaterial und MIDI-Daten dienen. Letztere gestatten tiefergehende Eingriffe in die Musik, da zusätzlich zu klangspezifischen Eigenschaften auch Tonhöhe und -dauer deutlich differenzierter beeinflusst werden können.

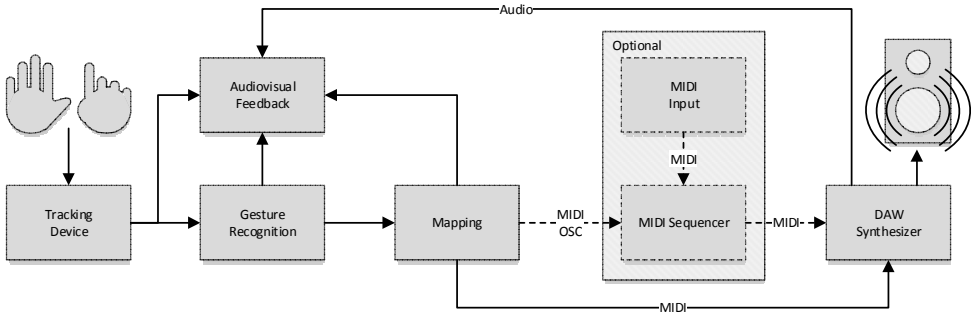

Abbildung 1: Integration der Freihandgesteninteraktion in die DAW.

2.1 Handerfassung und Gestenerkennung

Für die Erfassung der Hände gibt es heute bereits eine Vielzahl hoch entwickelter, kommerzieller Lösungen, die in unterschiedlichem Maße für bestimmte Arbeitsumgebungen geeignet sind. Populär ist die Microsoft Kinect 2 (Microsoft 2013), eine auf Infrarotlicht basierende Tiefenbildkamera. Ihre Genauigkeit wird allerdings durch die vergleichsweise geringe Auflösung von etwa 3 Millimetern bei 2 Metern Abstand deutlich begrenzt. Zugleich ist die Bildwiederholrate mit 30 Bildern pro Sekunde für die Erfassung hochfrequenter Bewegungen ungeeignet. Die Kinect eignet sich also vornehmlich für größere, weit ausholende (Körper-)Gesten und damit auch eher für die Grobsteuerung von musikalischen Parametern, bspw. für langsame und kontinuierliche Dynamik- und Tempoänderungen. Für feingranulare und hochfrequente Parameter, wie etwa notenweise variable Artikulationen ist die Kinect auch aufgrund ihrer geringen zeitlichen Auflösung nicht geeignet.

Vergleichsweise jung ist der Leap Motion Controller, der speziell für das Tracking von Händen optimiert ist. Sein Arbeitsraum ist mit rund 1m³ deutlich geringer, bietet aber mit maximal 0,01 Millimetern eine höhere Genauigkeit. Dank der 300 Bilder pro Sekunde können auch schnelle Bewegungen zuverlässig erkannt werden. Der Controller ist vorwiegend für die Nutzung am Schreibtisch optimiert und ist dann das Mittel der Wahl, wenn fein differenzierte Handgesten und -posen mit geringer Latenz erkannt werden sollen. Der Leap Motion Controller wird sich vorwiegend für professionelle Arbeitsumgebungen am Schreibtisch oder Mischpult eignen. Der Nutzer kann aus seiner Arbeitshaltung heraus schnell mit einer oder beiden

Händen in den Tracking-Bereich greifen, Eingaben tätigen und zurück zu Maus und Tastatur kehren. Die Kinect wird sich hingegen eher für performative und spielerische Anwendungszwecke eignen.

2.2 Mapping und Anbindung an die DAW

Die Gestenerkennung ermittelt aus den Rohdaten der Tracking Hardware im nächsten Schritt die Gesten, aus denen das Mapping musikalisch sinnvolle Steuersignale ableitet. Im einfachsten Fall sind das MIDI-Signale[1], welche direkt an die DAW geschickt und dort aufgezeichnet werden. Hier steuern sie die Parameter von Klangsynthese- und Effekt-Plugins sowie die Abmischung. Je intuitiver dieses Mapping definiert ist, desto leichter lassen sich die Gesten vom Nutzer erlernen und anwenden. Frameworks wie das kommerzielle GECO (Uwyn bvba/sprl 2013) gestatten dem Nutzer sogar die Definition eigener Mappings. Zur Steuerung ausschließlich klangbezogener Parameter einer schon vorhandenen Musik ist das bereits ausreichend. Komplexere Eingriffe in das Notenmaterial und seine Zeitstruktur (Tempo, Micro-Timing) sind so allerdings noch nicht möglich.

Für komplexere Bearbeitungen bis hin zur Handgesten-gesteuerten, algorithmischen Musikgenerierung in Echtzeit muss die MIDI-Sequencer-Funktionalität auf ein eigenständiges Modul ausgelagert werden. Die DAW fungiert nun nur noch als Tongeber (Klangsynthese, Effekte) und als Aufnahmegerät. Der ausgelagerte MIDI Sequencer kann nun selbst MIDI-Daten einlesen oder generieren[2]. Sämtliche Funktionalität kann interaktiv gestaltet und durch Steuersignale aus dem Mapping-Modul angesteuert werden[3]. Eingriffe in die Zeitstruktur des Musikstücks sind nun ebenso möglich wie die Veränderung und Erweiterung des Notenmaterials[4]. Tatsächlich kann an dieser Stelle auch ein digitales Musikinstrument zum Einsatz kommen, das eine eigene Audioausgabe generiert, was freilich weg führt vom hier betrachteten Anwendungsbereich der Musikproduktion. Nimmt man also den Umweg über einen externen MIDI Sequencer, sind weitreichende Bearbeitungen möglich. In beiden Fällen werden die MIDI-Datenströme abschließend von der DAW reproduzier- und editierbar aufgezeichnet.

2.3 Audiovisuelles Feedback

Direktes Feedback während der Eingabe trägt zur Erlernbarkeit der Eingabemodalität und der Gesten bei und verringert die mentale Belastung. Der Grund für eine nicht oder falsch erkannte Geste wird aus dem Feedback oft direkt klar, bleibt ohne dieses hingegen meist verschlossen

---

1 Zumeist werden es Controller-Befehle sein.

2 Z.B. aus dem Arbeitskontext einer Musikedition heraus (Wiedergabe der aktuell bearbeiteten Partiturstelle).

3 Hier sind dem Entwickler keine Grenzen gesetzt, welche Signale das Mapping generiert und verarbeitet werden. Es könnten MIDI- ebenso wie OSC-Nachrichten sein.

4 Z.B. die Änderung von Tonhöhen, die Generierung von Ornamenten aus Handgesten oder Sprünge in der MIDI-Datei.

und frustriert. Es bietet sich daher an, sowohl die erfassten Daten (Körper, Hände, Tiefenbild, Gesten) visuell als auch das Audioresultat in Echtzeit darzubieten, sodass der Nutzer sofort hört und sieht, was er durch seine Eingaben bewirkt.

Um präzises Feedback zu bieten, sollte die Latenz zwischen der Gesteneingabe und dem erfahrbaren Output möglichst gering sein. In Bezug auf die Eingabe expressiver Parameter kann diese Restriktion etwas sanfter aufgefasst werden, da hier keine diskreten, zeitgenauen Klangereignisse eingegeben werden, sondern vordefiniertes Klangmaterial um zusätzliche Informationen erweitert wird. Der Benutzer hat daher in der Regel genügend Zeit, sich vorzubereiten und seine Gesten entsprechend vorausschauend zu koordinieren.

3 Fazit

Obwohl im Bereich der Musikproduktion einige der komplexesten und traditionellsten Benutzungsschnittstellen anzutreffen sind und von neuartigen Interaktionsmodalitäten wie der Freihandgesteninteraktion profitieren können, konnten sich diese bislang nicht etablieren. Als Ersatz für etablierte Modalitäten sind die Auswirkungen auf bestehende Workflows zu gravierend. Wir sehen ihr größtes Potential vielmehr als Ergänzung in all jenen Situationen, in denen eine große Zahl von Parametern direkt und intuitiv zu steuern ist, in der Effektsteuerung, Klangsynthese und Ausarbeitung von ausdrucksvollen Interpretationen. Die entsprechende technische Integration der Handgesteneingabe in bestehende DAW-Setups wurde diskutiert.

Literaturverzeichnis

W. Balin and J. Loviscach (2011). Gestures to Operate DAW Software. In *130th Audio Engineering Society Convention*. London, UK: Audio Engineering Society.

A. Berndt (2011). *Musik für interaktive Medien: Arrangement- und Interpretationstechniken*. München: Verlag Dr. Hut.

Leap Motion, Inc. (2013). *Leap Motion controller*. www.leapmotion.com.

MAGIX Software GmbH (2015). *Music Maker Jam*. app on iTunes, Google play and Windows Store, version 1.2.3.

Microsoft (2013). *Kinect for Xbox One*. www.microsoft.com/en-us/kinectforwindows/.

J. Ratcliffe (2014). Hand Motion-Controlled Audio Mixing Interface. In *Proc. of New Interfaces for Musical Expression (NIME) 2014*, London, UK: Goldsmiths, University of London, S. 136-139.

Uwyn bvba/sprl (2013). *GECO: Multi-Dimensional MIDI/OSC/CopperLan Expression Through Hand Gestures*. app on Airspace store. version 1.3.0.

Kontaktinformationen

Axel Berndt, Simon Waloschek, Aristotelis Hadjakos
Zentrum für Musik- und Filminformatik, HfM Detmold, HS Ostwestfalen-Lippe
Email: {berndt; waloschek; hadjakos}@hfm-detmold.de