

Theoretisches Paper

Marcel Plaga (MT)

Inhaltliche Gliederung:

- 1 Unkonventionelle Ansätze der Klangsteuerung
 - 1.1 Tasten- und Berührungssensitive-Steuerungen
 - 1.2 Steuerungen per Hand- und Fingerbewegungen im Raum
 - 1.3 Steuerung mit dem gesamten Körper
 - 1.4 Weitere Ansätze
 - 1.5 Zusammenfassende Feststellungen
- 2 Klangparameter
- 3 Zuordnung von einem Gesichtsausdruck zu einem Klang
- 4 Echtzeitanforderungen

1 Unkonventionelle Ansätze der Klangsteuerung

Für die kreative und innovative Klangsteuerung gibt es mittlerweile zahlreiche unkonventionelle Ansätze, welche eine breite Spanne zwischen wenig produktiven Spielereien aber durchaus auch sehr brauchbaren und interessanten Entwicklungen abbilden. Die einzelnen Ansätze lassen sich dabei in Kategorien zusammenfassen.

1.1 Tasten- und Berührungssensitive-Steuerungen

Hier gibt es vor allem viele Tasten-Controller, welche unkonventionelle Tastenanordnungen aufweisen. Meist geht dies auch mit einer deutlichen Erhöhung der Tastenanzahl einher und soll so die Möglichkeiten schaffen, mit nur geringen Positionsänderungen der Finger auf eine größere Anzahl von Tönen und Funktionen und Effekten zugreifen zu können. Letztendlich kann hierdurch die Komplexität der resultierenden Komposition erhöht werden, jedoch sind keine erheblich neuen Steuerungsfunktionen geschaffen worden, da dies immer noch auf dem üblichen Tastendruck basiert.

Einen interessanteren Ansatz liefert hier das „Continuum Fingerboard“ der Firma Haken Audio. Auf einer Berührungssensitiven Fläche, mit zunächst typischem Pianoaufdruck, erfolgen die Eingaben in drei Dimensionen, also X- und Y-Position sowie Fingerdruck. Dadurch können die Töne beispielsweise Stufenlos gespielt bzw. in einander übergeführt werden. Der eingebaute programmierbare Synthesizer ermöglicht dann viele unterschiedliche Funktionen und Effekte zuzuordnen und einzustellen. Dadurch ergibt sich die Möglichkeit bis zu 3 Parameter eines Klangs gleichzeitig mit nur einem Finger stufenlos sehr genau regeln zu können. Durch beliebige X-Y-Drucksensitive-Slides und mit der Verwendung mehrerer Finger ergeben sich sehr besondere Möglichkeiten interessante Sounds zu erzeugen, was mit klassischen Synthesizern nicht oder nur sehr kompliziert realisierbar wäre. Hier gibt es bereits auch weitere Controller, welche vom Prinzip her den gleichen Ansatz nutzen, bei denen jedoch die Anordnung der Berührungsflächen, die Gehäuseformen und Funktionen angepasst sind.

Weiterhin gibt es in dieser Kategorie noch eine Vielzahl von Ansätzen, bei denen beliebige Objekte (wie z.B. Obst, Schalen, Flaschen etc.) in MIDI-Controller verwandelt werden. Meist kapazitiv oder vibrationssensitiv basierend, werden hierbei vereinfacht gesagt jedoch nur die Form der Tasten verändert und keine wirklich neue besondere Möglichkeit der Klangsteuerung geschaffen. Kann aber durchaus, für beispielsweise Klanginstallationen, einen interessanten Effekt erzielen, wenn Besucher auf eine neue Art und Weise mit Objekten interagieren können.

1.2 Steuerungen per Hand- und Fingerbewegungen im Raum

Die einfachsten Vertreter dieser Kategorie stellen beispielsweise das Airpiano oder der Crystal Ball Controller dar. Über Distanzsensoren, welche auf einem Board oder Halbkugel angeordnet sind, werden hier die vertikalen Abstände der Hände erkannt. Über diese können dann Noten, Effekte und Fader stufenlos angepasst und gesteuert werden. Einen ganz ähnlichen Ansatz und Funktionsweise bieten LaserHarp Controller, bei denen Laserstrahlen mit der Hand unterbrochen werden. Bei diesen Techniken sind die horizontalen Handbewegungen jedoch Stufenbehaftet, abhängig in der Anzahl der verbauten Distanzsensoren (oft zwischen 5 und 8).

Um nach dem ähnlichen Prinzip die vollen drei Dimensionen der Handbewegungen stufenlos zu erfassen und für die Klangsteuerung nutzbar zu machen, gibt es mehrere Softwarecontroller, wie z.B. der Geco MIDI Gestural Controller, die es ermöglichen die Handbewegungen, die über einen Leap Motion Controller erfasst werden, in Parameter für die Klangsteuerung umzuwandeln. Zusätzlich gibt es hier noch die Möglichkeiten zumindest Ansatzweise einzelne Finger zu erfassen und es können auch Gesten definiert und detektiert werden (wie z.B. das ballen einer Faust). Die resultierenden Sounds können so stufenlos sehr umfassend gesteuert werden.

In Verbindung mit dem Leap Motion Controller gibt es noch weitere Ansätze in Kombination mit VR-Brillen, bei denen Interfaces im 3-Dimensionalen raum die Klangsteuerung ermöglichen. Bei diesen kann man dann beispielsweise mit den Händen virtuelle Klangobjekte im Raum platzieren, bewegen oder verbinden. Beispielsweise wäre hier Lyra VR zu nennen.

Ein weiterer Ansatz sind Controller (MOTUS) sowie Smartphone-Apps (AC Sabre), welche Positions-, Gyroskop- und Beschleunigungs-Sensor-Daten Auswerten und für die Klanggestaltung verfügbar machen und so nicht mehr an einen festen Eingabebereich gebunden sind. Die dadurch ohnehin zahlreichen Parameter werden zusätzlich noch um z.B. Shake-, Shuffle- und Tilt-Gesten ergänzt. Eine Erweiterung davon sind die Imogen Heap Gloves oder der Remidi T8 Midi Glove, bei denen in Form von Handschuhen, zum einen auch Gyroskop- und Beschleunigungssensoren an den Händen zum Einsatz kommen, gleichzeitig aber auch noch die einzelnen Fingerbewegungen sowie das spreizen der Finger wie auch Fingertips miterfasst werden. Auch der Umfang der Gesten ist hier erheblich gesteigert, da diese von Gesten mit einem einzelnen Finger bis hin zu Kombinationen mit beiden Händen reichen können. Dadurch wird eine sehr freie und komplexe Möglichkeit der Klangformung geschaffen, da viele Tonparameter und Effekte gleichzeitig gesteuert werden können.

1.3 Steuerung mit dem gesamten Körper

Natürlich gibt es bereits auch Ansätze den Klang mit der Bewegung des gesamten Körpers beeinflussen zu können. Hier gibt es zum einen das Projekt Nagual Dance, welches die Kinect nutzt um bis zu zwei Personen bewegungstechnisch zu erfassen. Dabei werden den verschiedenen Körperteilen unterschiedliche Instrumente zugeordnet welche durch Tanzbewegungen dann beeinflusst werden. Ein weiteres Projekt ist Kagura, bei dem eine Person über eine Kamera erfasst wird. Auf dem Bildschirm können dann verschiedene Instrumente und Klangparameter positioniert werden, welche dann durch die Bewegung verschiedener Körperteile erreicht und ausgelöst bzw. gesteuert werden können. Die freie Konfigurierbarkeit und Klangformung ist bei diesen jedoch eher eingeschränkt und durch viele Vorgaben geprägt. Mehr Möglichkeiten bietet hierfür der Gypsy Exo-Suite. Dieser erfasst stufenlos die Bewegungen des Oberkörpers und ermöglicht das Mapping zu verschiedenen Synthesizer-Funktionen.

Drumpants bieten eine weitere Möglichkeit den Körper zur Klangsteuerung zu nutzen. Dabei gibt es verschiedenste Trigger, welche an beliebigen Körperpositionen auch unter der Kleidung angebracht werden können. Durch Berührung dieser können dann verschiedene Instrumente oder Samples ausgelöst werden. Diese weisen aber meist nur tastenähnliche Funktionalitäten auf und ermöglichen keine stufenlosen Steuerungen.

Weiterhin wäre hier noch die positions- und drehungsabhängige Klangsteuerung von Personen innerhalb eines Raumes bzw. Bereiches anzumerken. Auf einer Ausstellung oder Kunstinstallation können so beispielsweise die Positionen und Drehungen der einzelnen Besucher zu einem Gesamtklangerlebnis beitragen. Erste Ansätze dazu gibt es hier bereits im Rahmen der WFS-Anlage an der HAW am Standort Berliner Tor oder bei der Firma Usomo.

1.4 Weitere Ansätze

Hier wäre zunächst der BBC USB MIDI Breath and Bite Controller zu nennen, welcher speziell für die Steuerung von Klangparametern mit dem Mund gedacht ist. Auf Grund der begrenzten Parameter ist dieser auch vor allem als Ergänzung gedacht um neben den Fingern und Füßen eine zusätzliche Steuerungsmöglichkeit zu schaffen.

Auch gibt es bereits erste experimentelle Ansätze mit FaceOSC Synthesizer-Parameter per Gesichtserkennung zu steuern. Hierbei werden Augenbrauen, Augen, Mund, Gesichtposition, Gesichtgröße sowie Gesichtsdrehung erfasst und Parametern zugeordnet. Kommerzielle Produkte gibt es hier allerdings noch nicht.

Einen sehr unkonventionellen Ansatz bildet noch MIDI-Sprout, bei dem Pflanzen als MIDI-Steuerung genutzt werden. Hierbei werden kleinste Ströme auf der Oberfläche von Pflanzen ausgewertet und in nutzbare MIDI-Signale umgesetzt.

1.5 Zusammenfassende Feststellungen

Zunächst lässt sich feststellen, dass die insgesamt sehr gut funktionierenden Ansätze dem Benutzer die Möglichkeit geben das Mapping der Parameter frei zu beeinflussen. Dennoch lassen sich aber auch ein paar häufiger vorkommende Zuordnungen erkennen. So wird zum Beispiel mit horizontalen Handbewegungen oft das Panning gesteuert, also die Position eines Tones oder Samples im Stereopanorama definiert. Mit vertikalen Handbewegungen hingegen wird eher die Tonhöhe oder die Stärke eines Effektes gesteuert. Insgesamt werden auch im Zusammenhang von Tonparametern stufenlose Steuerungsmöglichkeiten bevorzugt genutzt.

Bei den Effekten fällt vor allem auf das ein Reverb gerne genutzt und gesteuert wird und daher auch bei fast allen Ansätzen implementiert ist. Weiterhin ist ein Vibrato auch etwas häufiger zu finden. Sonst sind alle anderen Effekte und Filter jedoch sehr individuell eingesetzt. Zudem sind aber oft auch noch Arpeggio- und Trill-Funktionalitäten vorzufinden. Neben Tönen wird auch gerne die Verwendung und Zuordnung von Samples genutzt.

2 Klangparameter

Um auch mit dem Gesicht interessante Sounds erzeugen zu können, sollten hier zumindest die Grundelemente vieler Synthesizer abgedeckt werden.

Oszillatoren

Hier sollte die Frequenz, also der Ton, gesteuert werden können und Standardsignalformen wie Sägezahn, Rechteck, Dreieck und Sinus ausgewählt werden können. Hier könnte vielleicht auch eine Stufenlose Überblendung der Signalformen interessant sein. Zudem sollte es vielleicht zwei Oszillatoren und einen Suboszillator, der ein oder zwei Oktaven unter den Hauptoszillatoren schwingt, geben. So könnte der Sound insgesamt angedickt werden und ein Fundament geschaffen werden können.

Rauschgenerator

Die Auswahl zwischen weißen und pinken Rauschen sollte vielleicht möglich sein und das Rauschen sollte Stufenlos dem Hauptsignal zugemischt werden können.

Filter

Oft ist zumindest ein Multimode-Filter implementiert, welches beispielsweise zwischen TP, HP und BP umgeschaltet werden kann. Hier werden dann Parameter wie Grenz- und Resonanz-Frequenz wie auch die Flankensteilheit (vielleicht 12/24/36 dB pro Oktave) des Filters gesteuert.

Hüllkurven

Um wirklich interessante Sounds erzeugen zu können, sollte es drei einstellbare Hüllkurven nach dem ADSR-Prinzip geben. Je eine für die Tonhöhe, die Lautstärke und die Grenzfrequenz des Filters. Für jede Hüllkurve können dann Attack, Decay, Sustain und Release eingestellt werden. Es könnte dabei auch bereits voreingestellte Hüllkurven geben, welche nur durchgeschaltet werden.

Modulation

Zusätzlich sollte es noch die Möglichkeit der Modulation geben. Hier wäre zunächst ein LFO (Low Frequency Oscillator) bei vielleicht 20 Hz als Grundlage denkbar. Darüber könnten dann wieder die Tonhöhe, die Lautstärke oder die Grenzfrequenz des Filters moduliert werden. Die Modulation der Tonhöhe würde dabei zu einem Vibrato-Effekt führen und die Modulation der Lautstärke zu einem Tremolo-Effekt.

Reverb

Zusätzlich wäre die Implementierung eines Reverb-Effektes auch sehr interessant. Wobei dieser von den Hallparametern jedoch voreingestellt sein könnte und letztendlich nur die Intensität steuerbar wäre.

3 Zuordnung von einem Gesichtsausdruck zu einem Klang

Wie gut die Zuordnung von einzelnen Gesichtsausdrücken zu einem Klang letztendlich funktioniert, muss sich in Versuchen erst zeigen. Jedoch lassen sich trotzdem schon mal ein paar theoretische Überlegungen anstellen.

Zunächst sollten dabei die möglichen Wertebereiche der Steuerparameter wie auch Eigenheiten, der einzelnen Gesichtselemente berücksichtigt werden.

Augen

Diese lassen sich eher Digital auswerten, geschlossen oder offen. Werte dazwischen wären auf Grund der verhältnismäßig geringen Abstände der Augenlider und den Wimpern, welche eine sehr genaue Bilderkennung sicherlich erschweren, kaum umsetzbar. Hierfür wäre also z.B. das Ein- und Ausschalten von Effekten denkbar oder auch das um- bzw. durchschalten von beispielsweise Signalformen oder Filtern (HP, TP, BP, Flankensteilheiten). Bei den Augen ergibt sich jedoch auch eine kleine Schwierigkeit, das unwillkürliche Schließen. Hier gäbe es drei Möglichkeiten damit umzugehen. Zum einen könnte man die Augen als Steuerparameter ausschließen, wodurch die Möglichkeiten insgesamt jedoch eingeschränkter würden. Eine weitere Möglichkeit bestände darin diese nur in einem bestimmten Modus auszuwerten. Oder man würde diese trotzdem zuordnen und damit erzielen das sich gewisse Parameter unwillkürlich verändern. Letzteres könnte zu einem interessanteren Erlebnis beitragen und man würde eher subtilere Parameter steuern.

Augenbrauen

Hier ist zumindest ein gewisser stufenloser Eingabebereich möglich wenn auch mit eher geringer Dynamik. Wenn man die tatsächliche bewusste Steuerbarkeit mit berücksichtigt, würde man jedoch eher nur auf drei Werte kommen (weit runtergezogen, normal, weit hochgezogen). Da für die Klanggestaltung jedoch eher stufenlose Steuerparameter gewünscht sind, könnte man hier die genaue Steuerbarkeit ggf. vernachlässigen.

Mund

Der Mund liefert gleich zwei Stufenlose Steuerungsmöglichkeiten mit höherer und mittlerer Dynamik. Die vertikale Öffnung sowie die horizontale Öffnung (Mundform). Zudem können die meisten Menschen diese Parameter relativ präzise steuern. Da der Mund an der Stimmenbildung beteiligt ist, wäre hier eine naheliegende Assoziation darüber auch die Tonhöhe zu steuern.

Gesichtsposition im Bild

Auch diese ergibt zwei Steuerparameter, die X- und Y-Position. Die Dynamik dieser hängt allerdings auch von der Gesichtsgröße im Bild ab. Daher wäre hier eher sinnvoll relative Wertänderungen zur Steuerung zu nutzen anstatt absolute Werte. Diese könnte man dann als Nick- und Schüttel-Gesten vielleicht Effektparametern zuordnen.

Gesichtsgröße im Bild

Diese liefert einen weiteren Stufenlosen Parameter, dessen Grenzwerte jedoch etwas schwieriger festzulegen sind. Dennoch aber sehr gut zur Steuerung von Klangparametern geeignet, wo keine hohe Präzision erforderlich ist. Hier wäre zum Beispiel die Stärke eines Reverbs vorstellbar.

Gesichtsdrehung

Die vertikale Gesichtsdrehung hingegen liefert wieder einen sehr präzise steuerbaren Parameter, welcher gleichzeitig auch einen relativ hochaufgelösten Wertebereich aufweist. Da die Drehung ja um eine Mittelstellung herum erfolgt, wären hier auch die Aufteilung in positive und negative Werte denkbar oder man steuert durch die verschiedenen Drehrichtungen zwei verschiedene Parameter.

Eine erste vorsichtige Zuordnung für Versuche könnte dabei dann folgendermaßen aussehen:

Auge Links	Signalformen Oszillator 2 durchschalten
Auge Rechts	Flankensteilheit des Filters durchschalten
Augenbraue Links	Resonanzfrequenz des Filters
Augenbraue Rechts	Grenzfrequenz des Filters
Mund Höhe	Tonhöhe / MIDI-Noten
Mund Breite	Hüllkurven der Tonhöhe
Gesichtsposition X	Signalformen Oszillator1 durchschalten
Gesichtsposition Y	Filter TP/HP/BP durchschalten
Gesichtsgröße	Stärke des Reverb
Gesichtsdrehung Links	Modulation der Tonhöhe
Gesichtsdrehung Rechts	Modulation der Lautstärke

4 Echtzeitanforderungen

Die tolerierbaren Latenzzeiten bei der Audioverarbeitung sind Gegenstand unzähliger Debatten und Diskussionen. Bis heute ist hier noch kein eindeutiger Wert festgelegt worden. Das Problem ist, dass das menschliche Ohr und Gehirn sehr tolerant mit dem Phänomen umgeht und jeder Mensch eine etwas andere Wahrnehmung davon hat. Auch die Beschaffenheit des bearbeiteten Klangs kann bedeutenden Einfluss auf die Akzeptanz von Latenzzeiten nehmen. So können Klänge mit langsamer Einschwingzeit durchaus auch mit größeren Latenzen noch gefühlvoll gespielt werden, während die Toleranz für hohe Latenzen bei Sounds mit kurzen Einschwingzeiten zunehmend sinkt. Schall breitet sich mit einer Geschwindigkeit von 340 m/s aus. Das bedeutet, dass ein Meter in einer Zeit von 2,9 Millisekunden zurückgelegt wird. Eine Schallquelle in einer Entfernung von einem Meter hören wir also mit 2,9 ms Verzögerung, was bedeutet, dass die Abhörposition bereits einen Beitrag zur Gesamtlatenz leistet. Insgesamt richtet man sich dann jedoch aber oft an einer von der Mehrheit akzeptierten Wahrnehmungsschwelle von 10 bis 11 Millisekunden.