

Künstliche Intelligenz im (räumlichen) Sounddesign

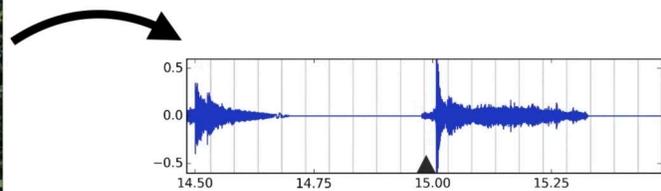
Ein Einblick in eine Vielzahl an Möglichkeiten

Lukas Münter

Masterstudiengang Audiovisuelle Medien, Schwerpunkt Ton
Hochschule der Medien Stuttgart



Silent video



Predicted soundtrack

Abbildung 1: Vertonung eines stummen Videos durch Vorhersage eines Sounds mittels KI [1]

ZUSAMMENFASSUNG

Im (räumlichen) Sounddesign bietet Künstliche Intelligenz bereits viele Vorteile. Tools und Programme auf KI-Basis helfen den Ton-Schaffenden, bestimmte Aufgaben zu erledigen oder zumindest zu vereinfachen, ersetzen den Menschen jedoch nicht. Sie machen Arbeitsabläufe möglich, die sonst nur schwer realisierbar wären. Sonst stumme Inhalte werden zum Leben erweckt und auch bereits bestehende Soundkulissen können umgestaltet werden. Hierbei findet die Künstliche Intelligenz sowohl bei der Vertonung von (Bewegt-)Bild als auch beim räumlichen und interaktiven Sound wie bei Video-spielen Anwendung. Oft liegt die Anwendung der Künstlichen Intelligenz jedoch nicht in der eigentlichen Erstellung der Sounds, sondern sie verbessert durch andere Aspekte das auditive Erlebnis.

So bietet Künstliche Intelligenz in dem Bereich bereits viele Vorteile, sowohl auf der Seite der Tonschaffenden als auch auf Seite der Hörer:innen; sie kann aber in einigen Aspekten noch verbessert werden. Im Folgenden werden einige ausgewählte Tools und Programme vorgestellt, um einen Einblick in die Vielfalt an Möglichkeiten zu bekommen - denn es gibt bereits zu viele Anwendungen in diesem Bereich, um sie alle zu betrachten.

1 EINLEITUNG

Künstliche Intelligenz (KI) wird für den Menschen und die Gesellschaft immer wichtiger. In fast allen Bereichen, in denen gedacht oder etwas geschaffen werden muss, kommt KI vermehrt zum Einsatz. Auch in der Produktion von Medien gibt es verschiedenste Tools und Programme auf KI-Basis, die den Medienschaffenden helfen sollen, bestimmte Arbeitsabläufe zu erledigen oder zu vereinfachen; und solche Tools werden ständig neu- und weiterentwickelt.

Im (räumlichen) Sounddesign kann KI verschiedenste Aufgaben übernehmen. KI kann z.B. verwendet werden, um (stumme) Videos zu vertonen. Sie kann durch Machine Learning¹ Bilder in musikalische Werke verwandeln. Oder aber ein neuronales Netz² hilft dabei, anhand eines Fotos die räumliche Verteilung von Sounds besser zu simulieren und zu individualisieren, für eine bessere Immersion und ein besseres auditives Erlebnis. Was das bedeutet und wie KI das macht wird im Folgenden nach und nach anhand von Beispielen besprochen und evaluiert.

2 ANWENDUNG IM AUDIOVISUELLEN BEREICH

Bei der Betrachtung von KI im Sounddesign kann (Bewegt-)Bild dennoch nicht ausgelassen werden. Schließlich kommt Sounddesign auch in Filmen, Serien und anderen visuellen Medien zum Einsatz. Hierbei sind jedoch häufig lange Suchen in Sound-Bibliotheken o.ä. nötig, um den richtigen Sound zu finden. Erst danach findet das eigentliche Designen und Bearbeiten des Sounds statt.

¹*Machine Learning* (engl. für *maschinelles Lernen*) beschreibt bei KI das Trainieren des künstlichen neuronalen Netzes und die dadurch von diesem Netz erworbene Fähigkeit, in ihm gegebenen Daten Regelmäßigkeiten oder sonstige neue Informationen zu erkennen [2].

²Das *neuronale Netz* bzw. im Bereich von KI eigentlich das *künstliche neuronale Netz* basiert auf dem menschlichen neuronalen Netz und ist die Grundlage für die Lernfähigkeit einer KI [3].

Hier schafft KI abhilfe. Wie das funktioniert, wird in diesem Kapitel anhand von zwei Beispielen betrachtet.

2.1 VISUALLY INDICATED SOUNDS

In [4] wird eine Anwendung entwickelt und betrachtet, die stumme Videos (realistisch) vertont. Hierbei wurde sich auf das Schlagen, Kratzen, Streifen, etc. von verschiedenen Objekten und Untergründen mit einem Schlägel konzentriert. Dahinter steckt ein Algorithmus mit einem neuronalen Netz, der aus den Aktionen in den zugespilten stummen Videos vorhersagt, wie diese Aktion klingt, und daraus dann einen Sound synthetisiert [4]. Abbildung 2 zeigt dies beispielhaft; die oberen Bilder sind Frames aus jeweils einem stummen Video, darunter ist die vorhergesagte Tonspur zu sehen. Die gepunktete Linie zeigt jeweils die zugehörige Position des Frames in der Tonspur [4]. Die Synthese der Tonspur liegt nicht in der Synthese der Sounds selbst. Was synthetisiert wird ist ein Chochleagramm³, zumindest in Teilen des Experiments [4]. Die Grundlage dazu entsteht durch das Training der KI:

Das neuronale Netz wurde mit 977 Videos, die insgesamt 46.577 Aktionen enthalten, trainiert. Diese Videos sind genau die Art von Videos, die der Algorithmus hinterher auch vertonen soll, enthalten jedoch Sound. Diese Trainingsvideos fungieren gleichzeitig als Datenbank für die Synthese [4]. Ein Einblick in diese Datenbank gibt es in [5]. Durch das Training lernt das Netz die verschiedenen Untergründe, die Bewegungen und die Reaktionen der Untergründe auf den Schlag oder das Streifen kennen und lernt, das Geräusch der gesamten Aktion vorherzusagen. Es lernt, die Eigenschaften des Videos auf die Eigenschaften des Sounds zu übertragen [4].

³Ein *Cochleagramm* ist eine zeit-frequenz-basierte Darstellung eines Audio-Signals. Es bezieht sich auf die Wahrnehmung in unserem Ohr, insbesondere auf die Wahrnehmung durch die Hörschnecke, die *Cochlea* [6].

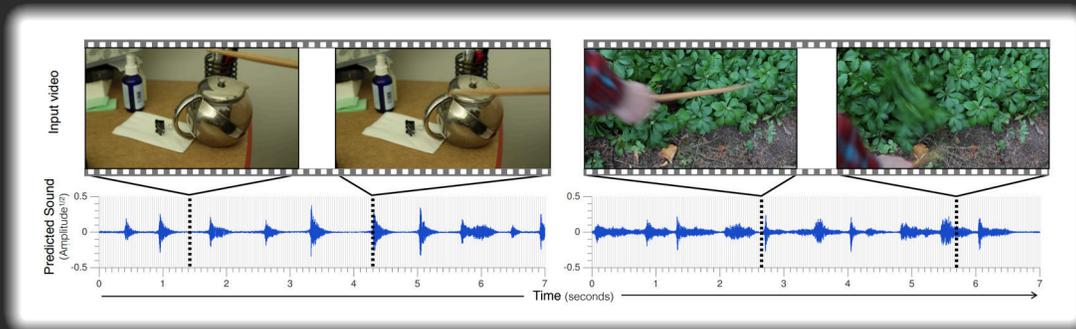


Abbildung 2: Beispielhafte Vertonung stummer Videos mit der zugehörigen synthetisierten Ton-Spur [4]

Zurück zur Synthese. Wie zuvor erwähnt, generiert der Algorithmus nicht den eigentlichen Sound, sondern die Eigenschaften des Sounds in Form eines Chochleagramms. Die synthetisierte Chochleagramme werden dann mit denen der Sounds aus den Trainingsvideos bzw. den Videos in der Datenbank verglichen. Die Sounds aus der Datenbank, die dem vorhergesagten Sound am ähnlichsten sind, werden für die Vertonung verwendet [1, 4]. Abbildung 3 zeigt dieses Prinzip beispielhaft und vergleicht die realen Chochleagramme des Videos mit den synthetisierten Chochleagrammen, die der Algorithmus für dieses Video hergestellt hat. Die Tonspur für das stumme Video entsteht also nicht aus künstlichen Sounds, sondern aus realen Sounds, die neu zusammengefügt wurden.

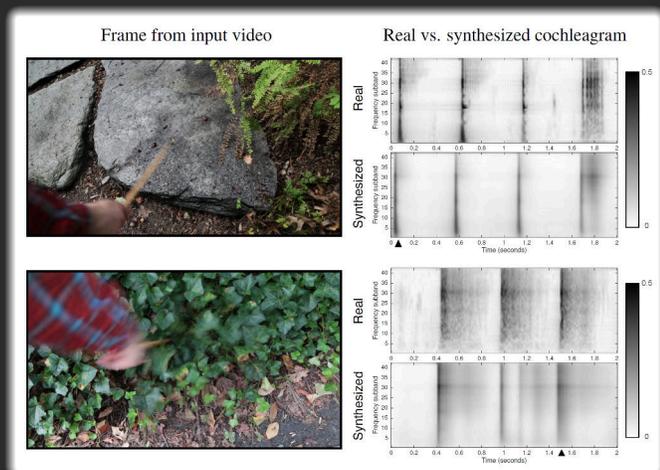


Abbildung 3: Beispielhafter Vergleich von synthetisierten Chochleagrammen mit den realen Chochleagrammen des jeweiligen Videos [4]

Bei genauerer Betrachtung der Chochleagramme in Abbildung 3 wird erkennbar, dass die synthetisierten Chochleagramme nicht exakt so aussehen, wie die realen, doch sie kommen sich in der grundsätzlichen Struktur schon sehr nah.

Wenn man sich das Video zu dem Experiment, verfügbar in [1, 5], anschaut und sich die Beispiele zu künstlich vertonten Videos anhört (ab Minute 1:27), wird klar, dass das grundsätzlich bei einigen Aktionen ganz gut funktioniert. Es gibt aber auch Beispiele, vor allem bei Vertonung von Schlägen auf Wasser in Minute 3:12, bei denen der Sound mehr oder weniger eindeutig nicht zu dem Video passt.

Das Prinzip der Vertonung von Videos durch KI, wie es hier dargestellt wird, hat also noch Schwächen. Darüber hinaus funktioniert dieser Algorithmus nur für einen sehr spezifischen Anwendungsfall und ist dadurch nicht für den Einsatz im Sounddesign durch z.B. Foley⁴ geeignet. Der Ansatz ist jedoch für möglicherweise zukünftig entstehende KIs sehr gut, indem dadurch, wie zuvor erwähnt, das ewige Suchen in Sound-Bibliotheken erspart wird. Das anschließende Weiterbearbeiten und Verfeinern, also das eigentliche Designen des Sounds, sollte

⁴Foley beschreibt die (realistische) Vertonung von Geräuschen wie Schritten, Kleider-Rascheln oder dem Abstellen einer Tasse. Diese Art der Nachvertonung findet sich in Filmen, lässt sich aber auch in interaktiven und räumlichen Formaten für die Vertonung von medialen Einspielern oder (interaktiven) Objekten anwenden.

dennoch von einem:r Sounddesigner:in übernommen werden. Das menschliche Gehör ist dann doch zu empfindlich, als dass die reinen synthetisierten Sounds für ein plausibles Ergebnis ausreichen. Vor allem dann, wenn der Sound eine bestimmte Stimmung oder Emotion transportieren soll.

2.2 PLAY A KANDINSKY

Play a Kandinsky ist ein Experiment vom *Centre Pompidou* in Zusammenarbeit mit *Google Arts & Culture*, bei dem ein Gemälde von Wassily Kandinsky, genauer gesagt „Gelb-Rot-Blau“, vertont wurde [7].

Das Experiment nimmt mittels Machine Learning und KI die Synästhesie⁵ von Wassily Kandinsky als Grundlage, um aus dem Visuellen des Gemäldes Sound zu kreieren. Dadurch wird ausgedrückt, „was Wassily Kandinsky gehört haben könnte, als er „Gelb-Rot-Blau“ 1925 gemalt hat“⁶ [7, 8].

Das Gemälde wurde nicht einfach nur vertont, man kann es auch spielen. Über eine Schaltfläche in [7] kommt man direkt zu der interaktiven Online-Anwendung. Man klickt sich zunächst durch ein paar erklärende Sequenzen, in denen man die Synästhesie von Kandinsky kennenlernt; Abbildung 4 zeigt Ausschnitte aus diesen Sequenzen.

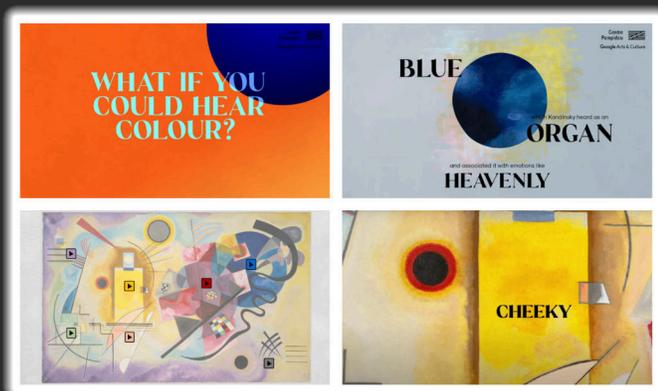


Abbildung 4: Ausschnitte der Sequenzen von *Play a Kandinsky* [7]

⁵Synästhesie beschreibt die Verbindung bzw. simultane Wahrnehmung verschiedener Sinneswahrnehmungen beim Menschen. So verbinden beispielsweise manche Menschen Farben mit einem bestimmten Geschmack. Kandinsky hat Farben gehört [8].

⁶Eigene Übersetzung.

Zum Schluss hat man komplette Freiheit dabei, das Gemälde mit Hilfe der Farben und Formen auf dem Bild zu spielen, wobei ein Soundtrack entsteht, den man anschließend auch teilen kann [7].

Hier wird KI verwendet, um visuelle und auditive Kunst zu verbinden. Es wird uns möglich, Einblicke in die mögliche Gedankenwelt des Künstlers zu bekommen, der dieses Gemälde erschaffen hat. Das Gemälde wird interaktiv spielbar und dadurch zum Leben erweckt. KI erschafft hier Möglichkeiten, die wir sonst u.U. nicht hätten.

3 ANWENDUNG IM RÄUMLICHEN UND INTERAKTIVEN BEREICH

Im folgenden Kapitel wird die Betrachtung von KI im Sounddesign auf den räumlichen und interaktiven Bereich erweitert. Hierbei werden zwei KI-Tools betrachtet. Die beiden Tools verfolgen vom Einsatz der KI her den gleichen Ansatz, unterscheiden sich jedoch in ihrem Anwendungsfall. Bei dem einen Tool liegt die Anwendung auf der Creator-Seite, d.h. es wird in erster Linie von z.B. Sounddesigner:innen verwendet. Das andere Tool zielt auf die Nutzung von Videospiele:innen ab, d.h. dessen Anwendung liegt auf der Seite der Nutzer:innen. Damit verfolgen die beiden Tools jeweils andere Ziele, die im Folgenden betrachtet werden.

3.1 IMMERSE VIRTUAL STUDIO

Das *Immerse Virtual Studio* von *Embod* (hier wird die *Alan Meyerson*⁷ *Signature Edition* [10] betrachtet) ist, wie der Name schon sagt, ein virtuelles Studio für die Audio-Produktion, d.h. man kann von zu Hause aus über Kopfhörer virtuell in einem anderen Studio-Setup mischen, in diesem Fall im Studio von Alan Meyerson [11].

⁷Alan Meyerson ist für seine Mischung von Filmmusik in verschiedenen Filmen bekannt [9].

Der Einsatz der KI liegt bei dieser Software nicht in der eigentlichen Erstellung oder Veränderung der Sounds, sondern in der Erstellung eines personalisierten HRTF⁸-Satzes des:der Nutzer:in. Dazu wird ein Foto vom rechten Ohr aufgenommen, wodurch die KI den HRTF-Satz erstellt (siehe Abbildung 5). Dieser HRTF-Satz wird dann auf das virtuelle Lautsprecher-Setup angewendet, sodass der:die Nutzer:in über Kopfhörer virtuell im Studio von Alan Meyerson sitzt [12].

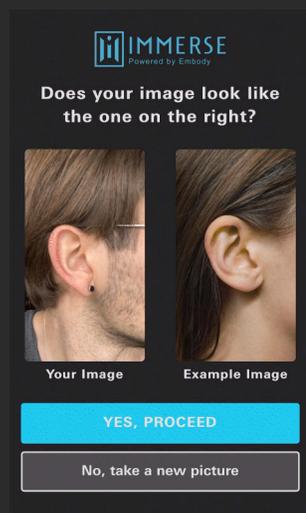


Abbildung 5: Screenshot eines Abschnitts der Erstellung der personalisierten HRTF in [10]

Um die Personalisierung zu ermöglichen, wurde ein Kunstkopf - ein künstlicher Kopf bestehend aus dem Kopf selbst, Gesicht, Ohren und in diesem Fall Rumpf mit Mikrofonen in den Ohren für die möglichst "menschenähnliche" Aufnahme von räumlichem Sound - an die Abhörposition gesetzt, dessen jeweiliger Abstand zu allen Lautsprechern ausgemessen und anschließend sein HRTF-Satz erstellt. Diese Messungen und HRTFs dienen als Referenz. Die HRTFs werden später durch die des:der Nutzer:in ersetzt, sodass das Lautsprecher-Setup auf diese Person abgestimmt ist [12].

⁸HRTF steht für *Head Related Transfer Function* (dt.: kopfbezogene Übertragungsfunktion). Sie beschreibt, wie sich unser Kopf, unsere Ohren und unser Rumpf auf unser räumliches Hören, also auf unsere Wahrnehmung von Sounds im dreidimensionalen Raum um uns herum, auswirkt [13].

Weitere Personalisierung des Hörerlebnisses in dem virtuellen Studio kann durch verschiedenste Einstellungen wie Anteil des hörbaren Raumes oder der hinzugemischten HRTF vorgenommen werden. Zusätzliche Optimierung liefert die Möglichkeit, im Programm einzustellen, welchen Kopfhörer(-Typ) man benutzt [11].

Die Personalisierung mittels KI in diesem virtuellen Studio ist insofern sinnvoll, als das räumliche Hören über Kopfhörer für die jeweilige Person durch Simulation ihres individuellen räumlichen Hörens optimiert wird. Bei simpler Stereo-Wiedergabe über Kopfhörer ist das natürliche räumliche Hören nicht getreu abbildbar. Eine Mischung von Surround-Sound ist hierbei nur bedingt möglich. Eine Auralisation - also eine "Verräumlichung" des Audiosignals - verbessert diese räumliche Wahrnehmung über Kopfhörer. Wenn die Auralisation, wie bei diesem virtuellen Studio, mittels HRTF an den:die Hörer:in angepasst wird, funktioniert das räumliche Hören noch besser [13].

Beim *Immerse Virtual Studio* wird die HRTF allerdings über ein Foto des Ohrs erstellt - andere Methoden, eine HRTF zu messen [13], sind i.d.R. genauer. Dennoch wird das räumliche Mischen über Kopfhörer mit dieser Software sicherlich verbessert.

Ein persönlicher Test des virtuellen Studios hat gezeigt, dass dediziertes Sounddesign einzelner Sounds hier nicht so gut funktioniert. Die Sounds klingen teils verfärbt, d.h. ihre Klangfarbe verändert sich, sie klingen etwas anders. Für die anschließende räumliche Positionierung und Abmischung der Sounds wiederum ist das Tool sehr gut geeignet.

3.2 IMMERSE GAMING

Immerse Gaming ist eine weitere Sparte von *Embodiment* mit verschiedenen Anwendungen für verschiedene Systeme. Diese Anwendungen beruhen ebenfalls auf KI, indem sie auf die gleiche Weise, wie beim *Immerse Virtual Studio* die HRTF des:der Nutzer:in erstellen [14].

Immerse Gaming ist auf die Optimierung von Spiele-Sounds ausgelegt. Sounds in Spielen sind grundsätzlich häufig räumlich produziert; und diese Räumlichkeit soll hier durch die Anwendung der persönlichen HRTFs auf den Spiele-Sound verbessert werden. Die Anwendung verspricht verschiedene Verbesserungen der Audio-Umgebung im Spiel, je nachdem, welches der drei Presets, die für verschiedene Spielegenres optimiert wurden, von dem:der Spieler:in aktuell ausgewählt ist [15].

Die Presets sind eine Möglichkeit zur Personalisierung des Sound-Erlebnisses mit *Immerse Gaming*. Eine Visualisierung im Programm (hier am Beispiel von *Immerse Gaming | Hive* [16]) zeigt, wie die Sounds verteilt werden (siehe Abbildung 6). Sie können sehr nah sein, was u.a. für erhöhte Intensität der Sounds sorgt (*Close Combat Mode*), sie können für verbesserte Richtungserkennung klar verteilt sein (*Immerse Mode*) oder sie können gleichmäßig und stark räumlich verteilt sein für höheres (auditives) Bewusstsein der Spielumgebung (*Awaken Mode*) [15].

Ein hörbarer Vergleich der drei Modi am Beispiel von *Batman: Arkham Knight* [17] befindet sich in [18].

Anders als beim *Immerse Virtual Studio*, bei dem die Anwendung auf Seite der Creator liegt, liegt die Anwendung von *Immerse Gaming* auf Seite der Konsument:innen. KI im Sounddesign kann also nicht nur z.B. Sounddesigner:innen helfen und ihre Arbeit erleichtern, sondern auch das Erlebnis von Videospiele:innen verbessern.

Ein wichtiger Aspekt bei Sound in interaktiven und räumlichen Formaten ist die Immersion. Auch Videospiele schaffen durch ihre dreidimensionalen Soundkulissen ein immersives Erlebnis. Sowohl auditive Einhüllung als auch erkennbar positionierte Sounds verbessern die Immersion [19]. Auf beides zielt *Immerse Gaming* ab und versucht, das (auditive) Spielerlebnis zu verbessern und immersiver zu machen.

Ein persönlicher Test der Anwendung zeigt, dass das (bedingt) funktioniert. Wie bei dem virtuellen Studio werden die Sounds durch die verschiedenen Presets unterschiedlich verfärbt, wie man auch im Videobeispiel [18] hören kann. Im *Close Combat Mode* sind die Sounds in der Tat sehr nah, was für den gewünschten Zweck dieses Modus sicherlich gut ist. Der *Immerse Mode* war beim Testen vom Klang her nicht großartig anders zu dem originalen



Abbildung 6: Screenshots der verfügbaren Presets in [16]. Von links nach rechts: *Close Combat Mode*, *Immerse Mode* und *Awaken Mode*.

Spiele-Sound. Der *Awaken Mode* schafft eine breite Soundkulisse, was tatsächlich für erhöhte Immersion sorgt. Beim Testen war hier der größte Unterschied zum originalen Sound zu hören. Hierbei werden die Sounds jedoch am meisten verfärbt und Sounds, die eigentlich nah dran sein sollten, werden auch nach außen bzw. "weiter weg gedrückt". Für Menschen mit "geschulten Ohren", weil sie z.B. im Ton-Bereich tätig sind, und auch sonst tonempfindliche Menschen ist das Tool also u.U. auf Dauer nicht geeignet, weil die Verfärbung als zu störend empfunden werden könnte. Für Menschen, die darauf nicht so empfindlich reagieren ist die Anwendung aber eine gute Möglichkeit, ihr Spielerlebnis zu verbessern. Ein weiterer Vorteil ist, dass 3D-Sound, wenn auch über Kopfhörer, dadurch zumindest teilweise verstärkt in den Fokus der Öffentlichkeit außerhalb der Medienbranche rückt, wodurch er sich womöglich in Zukunft noch mehr etabliert.

4 FAZIT & AUSBLICK

Zu Guter Letzt noch einmal die Erkenntnisse aus den oben besprochenen KI-Anwendungen für Sounddesign als Fazit zusammengefasst.

Tools wie *Visually Indicated Sounds* helfen bei der Vertonung von Videos und ersparen im Idealfall lange Suchen nach passenden Sounds in Sound-Bibliotheken. Es wird eine Grundlage geschaffen, mit der Sounddesigner:innen weiterarbeiten können. Ersetzen kann KI den Menschen hierbei aber noch nicht. Hierzu fehlt es noch an Präzision und auch Emotionen können den Sounds (bisher) nur durch Menschen verliehen werden.

Durch die Vertonung von Bildern oder Gemälden wie bei *Play a Kandinsky* erhalten wir dank KI Einblicke in die Gefühls- und Gedankenwelt anderer und können Visuelles interaktiv und auditiv erleben.

Tools wie *Immerse Virtual Studio* können für Tonschaffende hilfreich sein, die räumlich Abmischen möchten, aber nicht die entsprechenden Voraussetzungen dafür zu haben. KI hilft hier, ein virtuelles Studio zu erstellen, das dies von zu Hause möglich macht.

Immerse Gaming individualisiert mittels KI das Spielerlebnis, indem die Räumlichkeit an das Spiel und den:die Spieler:in angepasst wird; die Soundkulisse wird immersiver.

Wie zu Beginn erwähnt, sind diese Anwendungen nur ein Auszug aus vielen weiteren, die es in diesem Bereich gibt. KI bietet im (räumlichen) Sounddesign viele Vorteile, auch wenn sie noch nicht bei allem ihr volles Potential ausschöpft. Wir dürfen sicherlich mit Spannung erwarten, was sich in Zukunft in diesem Bereich tut und wie KI zukünftig Workflows erleichtert und verschnellert und die Erfahrung von Sound noch immersiver macht.

REFERENZEN

- [1] A. Owens, *Visually indicated sounds: results*, Apr 2016. Accessed on: Mar 16, 2022. [Video file]. Available: <https://www.youtube.com/watch?v=JpZUZ9ZDECE>.
- [2] K. P. Murphy, *Machine Learning: A Probabilistic Perspective*. Cambridge, London: The MIT Press, 2012.
- [3] W. Ertel, *Grundkurs Künstliche Intelligenz: Eine praxisorientierte Einführung*, 4., überarbeitete Auflage. Wiesbaden: Springer Vieweg, 2016, doi: <https://doi.org/10.1007/978-3-658-13549-2>.
- [4] A. Owens, P. Isola, J. McDermott, A. Torralba, E. H. Adelson and W. T. Freeman, "Visually Indicated Sounds", in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Las Vegas, NV, USA, 2016, pp. 2405-2413, doi: 10.1109/CVPR.2016.264.
- [5] A. Owens, P. Isola, J. McDermott, A. Torralba, E. H. Adelson and W. T. Freeman, *Visually Indicated Sounds*, andrewowens.com. Accessed on: Mar 16, 2022. [Online]. Available: <https://andrewowens.com/vis/>.
- [6] D. L. Wang and G. J. Brown, *Computational Auditory Scene Analysis: Principles, Algorithms, and Applications*, Hoboken, NJ: Wiley-IEEE Press, 2006.
- [7] Centre Pompidou and Google Arts & Culture Lab, *Play a Kandinsky*, Experiments with Google, Feb 2021. Accessed on: Mar 16, 2022. [Online]. Available: <https://experiments.withgoogle.com/play-a-kandinsky>.

- [8] Google Arts & Culture, *How to Play a Kandinsky: behind the scenes*, Feb 2021. Accessed on: Mar 16, 2022. [Video file]. Available: <https://www.youtube.com/watch?v=L-h6wQ2rP4M&t>.
- [9] IMDb, *Alan Meyerson*, IMDb. Accessed on: Mar 16, 2022. [Online]. Available: <https://www.imdb.com/name/nm0583657/>.
- [10] *Immerse Virtual Studio | Alan Meyerson Signature Edition*. Embody. [Software]. Available: <https://embody.co/pages/ivs-alan-meyerson>.
- [11] EMBODY personalized spatial audio, *Immerse Virtual Studio | Alan Meyerson*, Embody. Accessed on: Mar 16, 2022. [Online]. Available: <https://embody.co/pages/ivs-alan-meyerson>.
- [12] EMBODY personalized spatial audio, *Behind The Scenes: Studio M Room Virtualization*, Embody. Accessed on: Mar 16, 2022. [Online]. Available: <https://embody.co/blogs/proaudio/studiom-virtualization>.
- [13] S. Weinzerl, Ed. *Handbuch der Audiotechnik*. Berlin, Heidelberg: Springer, 2008, doi: <https://doi.org/10.1007/978-3-540-34301-1>.
- [14] EMBODY personalized spatial audio, *Immerse Gaming*, Embody. Accessed on: Mar 16, 2022. [Online]. Available: <https://embody.co/pages/gaming>.
- [15] EMBODY personalized spatial audio, *Immerse Gaming Software*, Embody. Accessed on: Mar 16, 2022. [Online]. Available: <https://embody.co/pages/immerse-gaming-software>.
- [16] *Immerse Gaming | Hive*. Embody. [Software]. Available: <https://embody.co/pages/gaming-hive>.
- [17] *Batman: Arkham Knight*. (2015), Warner Bros. Interactive. [Software].
- [18] L. Münter, *Immerse Gaming | Hive - Vergleich*, Jan 2022. Accessed on: Mar 16, 2022. [Video file]. Available: <https://www.youtube.com/watch?v=4r2Ww-0lLwU>.
- [19] L. Münter, "Diskussion und Vergleich von Erzählstrukturen mit Fokus auf Medien im Raum sowie auditive Narration und tontechnische Umsetzung," B.Eng. thesis, Electronic Media, Hochschule der Medien, Stuttgart, 2021.