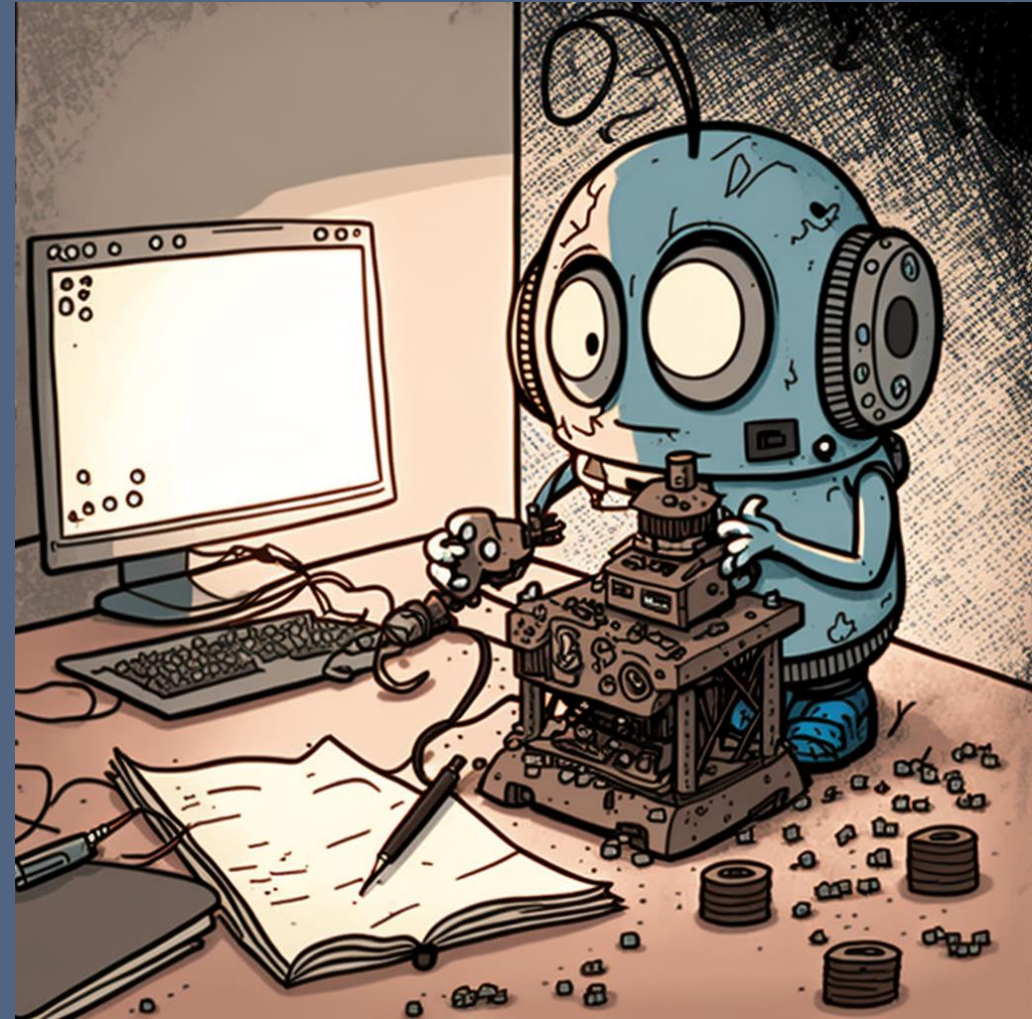


Beginn: 17:30

# KI und Audio

HFF

Lehrstuhl für KI in der Medienproduktion



# Termine (Mittwochs 17:30)

1. Mittwoch im Monat: KI – Sprechstunde (Q&A) - Zoom
  - 2. Mittwoch im Monat: Themenspezifische Veranstaltungen**
  3. Mittwoch im Monat: Open KI-Lab - [i.miller@hff-muc.de](mailto:i.miller@hff-muc.de)
  4. Mittwoch im Monat: Technische Tutorials
- Nicht in vorlesungsfreier Zeit



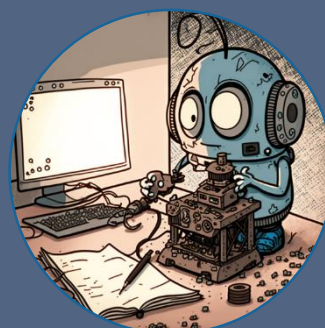
Text



Bild



Video



3D



Audio



Daten

# Termine, Tutorials, Aufzeichnungen

HOCHSCHULE  
FÜR FERNSEHEN UND  
FILM MÜNCHEN



## Lehrstuhl KI

Seit Mai 2022 gibt es an der HFF den Lehrstuhl für KI in der Medienproduktion. Dies wurde durch einen Wettbewerb der Hightech Agenda Bayern möglich, bei dem die HFF eine der begehrten KI-Professuren gewinnen konnte.

Wir möchten das Thema KI in der Lehre sehr praxisnah vermitteln und dabei u.a. deren Einsatz in der Filmproduktion aufzuzeigen. Beispiele sind das Durchsuchen großer Filmdatenmengen, der Einsatz im VFX-Bereich und das Generieren spezieller Filmsequenzen mittels KI-Technologien.

Das Thema KI ist fachübergreifend und betrifft viele Bereiche der Filmproduktion, in deren Lehrangebote die KI-Themen in der nächsten Zeit integriert werden. Dabei wollen wir das Potential dieser Technologie für die Filmproduktion aufzeigen, aber auch die rechtlichen und ethischen Aspekte einbeziehen. Eine große Herausforderung dabei ist die wahnsinnig schnelle Entwicklung dieses Gebietes.

Neben der Vermittlung von Wissen über den Einsatz von KI in der Filmproduktion unterstützen wir auch künstlerische Projekte, die KI im Produktionsprozess nutzen oder sich mit dem Thema KI auseinandersetzen.

---

Lehre: Wintersemester 2022/23



---

regelmäßige öffentliche Veranstaltungen



---

Tutorials / Aufzeichnungen



[https://www.hff-muenchen.de/de\\_DE/lehrstuhl-ki-allgemein](https://www.hff-muenchen.de/de_DE/lehrstuhl-ki-allgemein)

## Tutorials / Aufzeichnungen



### Maschine Learning für Medienschaffende Grundlagen

#### 1.0 KI-nema: KI und Film

##### 1.1 Einführung (Begriffe: KI, Maschine Learning, Deep Learning)

##### 1.2 Neuronale Netze (Was ist ein Neuronales Netz? Wie funktioniert es?)

##### 1.3 Aktuelle KI-Entwicklungen im Medienbereich

### Textgenerierung für Medienschaffende

#### 2.1 ChatGPT

#### 2.2 Überblick über Große Sprachmodelle

#### 2.3 KI als Co-Autor?

### TechTutorials

Diffusionzauber1111 - Techtutorial StableDiffusion für Fortgeschrittene

# Geplante Themen

- Januar: Textgenerierung
- Februar: KI in Hochschulen
- April: Bild- und Videogenerierung
- Mai: 3D
- **Juni: Audio**
- Juli: Rechtliche und Ethische Aspekte



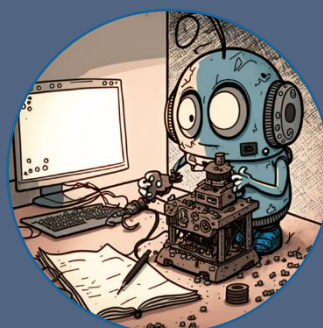
Text



Bild



Video



3D



Audio



Daten





# KI für Musik und Audio

A BRIEF OVERVIEW



- Transkription
  - Übersetzung, Untertitel
  - Text-based Editing
- Audiorestauration und -modifikation
  - Cleaning
  - Modifikation
- Generierung
  - Sprache
  - Musik
  - Soundeffekte

# Transkription

Transkription

Übersetzung, Untertitel

Text-based Editing

# Transkribierung, Untertitel

## Adobe Premiere

<https://helpx.adobe.com/de/premiere-pro/using/speech-to-text.html>

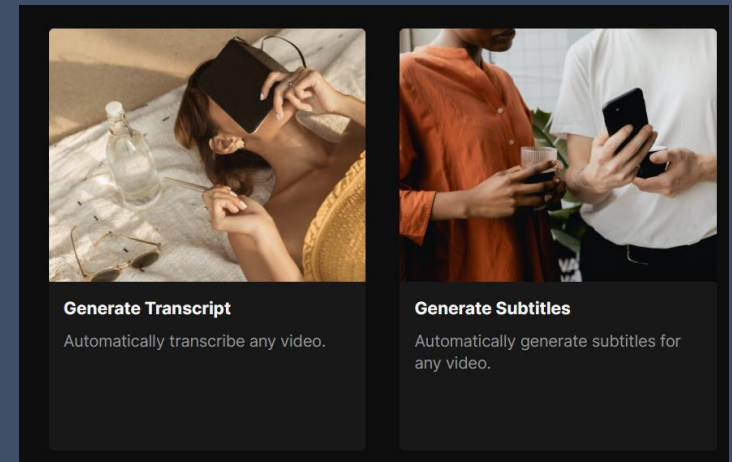
## DaVinci Resolve

<https://dvr.solve.com/tutorial/transcribing-audio-in-davinci-resolve/>

Runway <https://runwayml.com/>

## Whisper (OpenAI)

- Web: <https://replicate.com/openai/whisper>
- Lokal: [https://www.youtube.com/watch?v=ABFqbY\\_rmEk](https://www.youtube.com/watch?v=ABFqbY_rmEk)





# Weitere Tools Speech-to-Text (Transkribierung)

- <https://sonix.ai/de> (Transkribierung, Übersetzung)
- <https://www.descript.com/> (Transkribierung, Video/Audio Editing)
- <https://otter.ai/> (Transkribierung, Meetings, Zusammenfassungen)
- <https://cloud.google.com/speech-to-text?hl=de> (API)



Importieren

Bearbeiten

Exportieren

Quelle: (keine Clips)

Text

Effekteinstellungen

Transkript

Untertitel

Grafiken

Suchen

Sequenz transkribieren

Neue Untertitelspur erstellen

Untertiteldatei importieren

ABC

ABC

ABC

Projekt: TonTestTranskribierung

Media-Browser

TonTestTranskribierung.prproj

1 von 2 EL...

Diese Frau ist ni... 4:05:06

Diese Frau ist ni... 4:05:06

00:00:00:00

V3

V2

V1

A1

A2

A3

Mix

0,0

TonTestTranskribierung

Bearbeitet

Programm: Diese Frau ist nicht real! Erstes heute journal-Interview mit Künstlicher Intelligenz

2DF

00:04:05:06

1/2

00:02:30:00

00:03:00:00

00:03:30:00

00:04:00:00

00:04:30:00

00:05:00:00

0

-6

-12

-18

-24

-30

-36

-42

-48

-54

-60

dB

Transkript erstellen

Sequenzname: Diese Frau ist nicht real! Erstes heute journal-I...

Sequenzlänge: 00:04:06

Sprache: Deutsch

Audioanalyse: Audio-Clips mit dem Tag „Dialog“

Audio auf Spur: Mix

☐ Nur In-Point in Out-Point transkribieren

☐ Ausgabe mit vorhandener Transkription zusammenführen

☒ Erkennen, wann verschiedene Sprecher sprechen

Abbrechen

Transkribieren





HOCHSCHULE  
FÜR FERNSEHEN UND  
FILM MÜNCHEN

The screenshot displays the Adobe Premiere Pro 2023 interface. On the left, the **Transcript** panel shows a list of clips with their start and end times. A red line connects a text selection in the Transcript panel to a corresponding edit in the Timeline. The Program Monitor shows a video preview of a woman's face. The Timeline shows a sequence of clips with a red line indicating the current position.

Transcript Panel:

- Source: SOPHIEDALE\_INT03\_SONY.mp4
- Program: Interview Rough Cut
- Project: The science of color
- History: Effi
- 1 of 17 IL
- 01 Interview
  - SOPHIEDALE\_INT01\_A 13:20:35:14
  - SOPHIEDALE\_INT01\_C 13:25:18:08
  - SOPHIEDALE\_INT01\_S 13:22:56:23
  - SOPHIEDALE\_INT02\_A 13:28:40:20
  - SOPHIEDALE\_INT02\_C 13:33:54:08
  - SOPHIEDALE\_INT02\_S 13:31:17:14
  - SOPHIEDALE\_INT03\_A 13:38:40:08
  - SOPHIEDALE\_INT03\_C 13:47:45:12
  - SOPHIEDALE\_INT03\_S 13:43:12:22
  - SOPHIEDALE\_INT01\_IP 11:59:59:23
  - SOPHIEDALE\_INT03\_IP 11:59:59:23
- 02 B-roll
- 03 Audio
- 04 Graphics
- B-Roll 11:59:16:19
- Interview Rough Cut 11:59:16:19

Timeline:

- Interview Rough Cut
- B-Roll
- 00:02:44:11
- 00:02:29:23
- 00:05:59:22
- 00:01:29:21
- 00:01:59:21
- 00:02:29:20
- 00:02:59:19

[https://www.youtube.com/watch?v=Y\\_QxSET1dO4](https://www.youtube.com/watch?v=Y_QxSET1dO4)

# Audiorestauration

# Audiorestauration

## Beispiele

- Noise Remover
- Voice Separation
- Reverb Remover
- De-Esser
- Room Tone Match, Ambience Match
- Rekonstruktion, Modifikation

# Clean Audio

Goyo (z.B. Plugin in Premiere)

Runway

Adobe Podcast <https://podcast.adobe.com/>

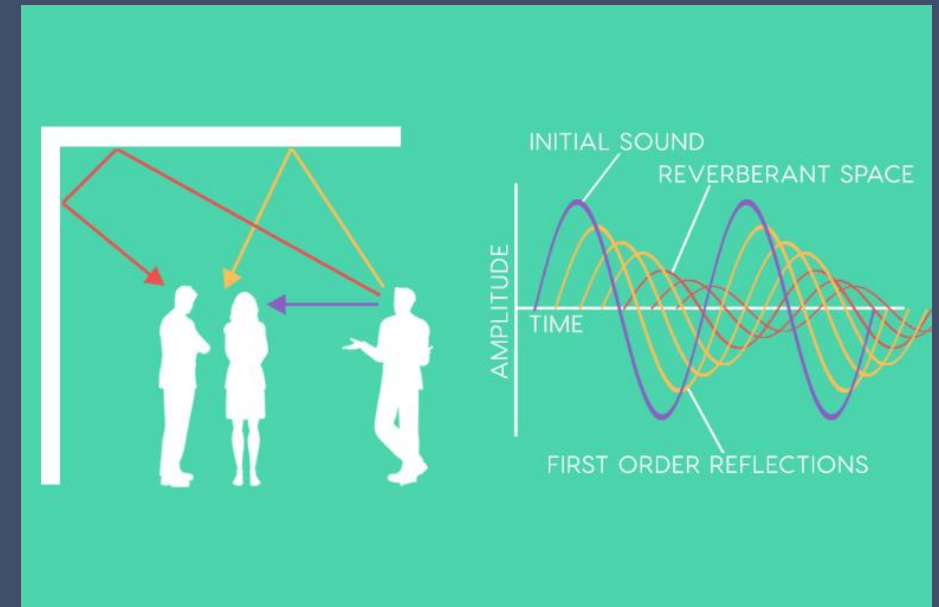




# Reverb

- engl. Hall oder Nachhall
- komplexes Echo, das Informationen über den physikalischen Raum enthält
- gibt einem Klang eine natürliche Räumlichkeit oder besondere Größe

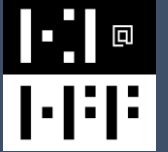
<https://www.waves.com/plugins/clarity-vx-dereverb-pro>  
<https://www.youtube.com/watch?v=m-ub4FaE9nc>



<https://flypaper.soundfly.com/produce/artificial-space-place-reverb-tech-primer/>

# Goyo - Voice Separator

HOCHSCHULE  
FÜR FERNSEHEN UND  
FILM MÜNCHEN

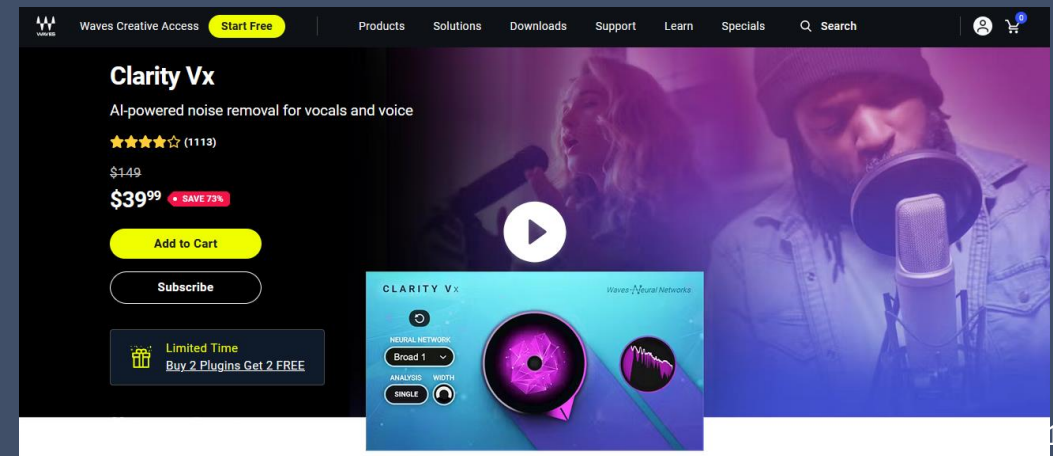


- Audio plugin trennt 3 Kanäle (Ambience, Voice, Voice Reverb)
- De-Noise
- De-Reverb
- De-Voice
- FREE OPEN BETA

# Audiorestauration – Clarity Vx

- Noise Remover
  - Reduktion von Rauschen, Brummen oder Hintergrundgeräuschen
  - Trennen von Stimmen
  - Reverb
- 
- Waves Clarity Vx (Waves Neural Networks) AI-powered noise removal for vocals and voice  
<https://www.waves.com/plugins/clarity-vx>  
<https://www.waves.com/plugins/clarity-vx-dereverb-pro>
  - <https://www.youtube.com/watch?v=m-ub4FaE9nc>

Unterstützt Pro Tools, Audition, Premiere Pro, Reaper u.v.a.



# Audiorestauration

Weitere Tools:

- Accusonus ERA Bundle
- iZotope RX 10



# Generierung

Sprache

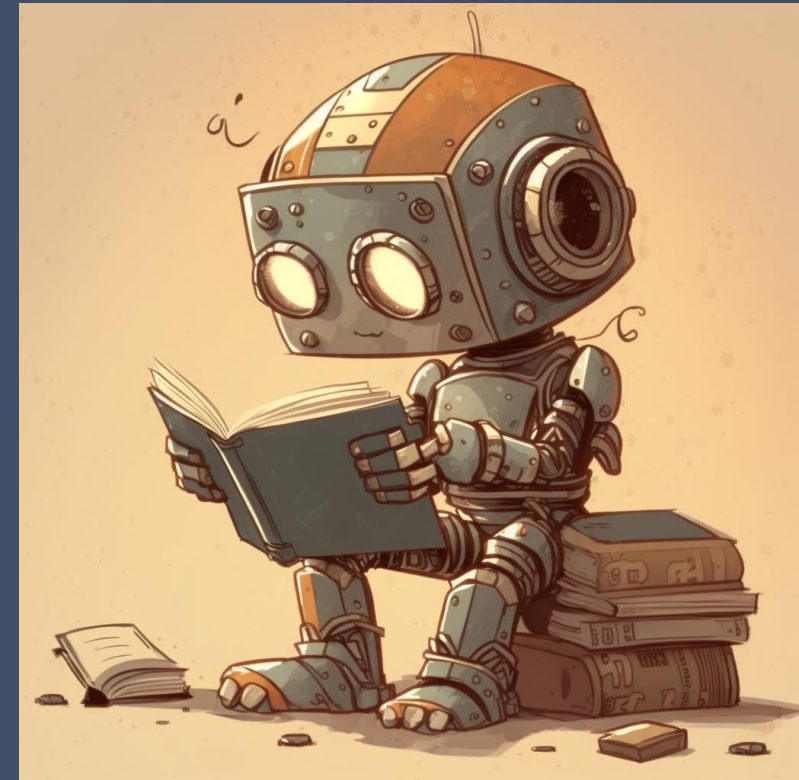
Musik

Geräusche

# Generierung - Sprache

# Text-to-Speech

- wandelt Text in eine akustische Sprachausgabe
- künstliche Erzeugung der menschlichen Sprechstimme
- Zwei Arten:
  - Neue synthetische Stimme
  - Clone einer existierenden Stimme



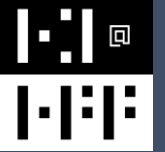
# Tools Text-to-Speech

- <https://speechify.com>
- <https://speechelo.com/>
- Deutsch:
- Murf AI: <https://murf.ai/text-to-speech/german>
- <https://ttsmp3.com/text-to-speech/German/>
- <https://www.narakeet.com/languages/text-vorlesen-lassen/>
- <https://www.readspeaker.com/de/>
- <https://cloud.google.com/text-to-speech?hl=de>
- <https://ttsfree.com/text-to-speech/german>



# Text-to-Speech: Speechify

HOCHSCHULE  
FÜR FERNSEHEN UND  
FILM MÜNCHEN



**Speechify** Text-to-speech ▾ AI Voice Over Audiobooks ▾ Groups ▾ Learn ▾ Company ▾ [Login](#) [Try for free](#)

## The #1 AI Text to Speech Reader

★★★★★ 20M+ downloads

Power through docs, articles, PDFs, email — anything you read — by listening with our leading AI text to speech reader.

**Gwyneth Paltrow**  
English Female Voice

**Snoop Dogg**  
English Male Voice

**John**  
English Male Voice

**Narrator**  
British Male Voice

[Continue](#)

## Artificial intelligence

Article Talk

Read View source View history Tools

From Wikipedia, the free encyclopedia

*"AI" redirects here. For other uses, see [AI \(disambiguation\)](#), [Artificial intelligence \(disambiguation\)](#), and [Intelligent agent](#).*

**Artificial intelligence (AI)** is *intelligence*—perceiving, synthesizing, and inferring information—demonstrated by machines, as opposed to *intelligence displayed by humans* or by other animals. Example tasks in which this is done include speech recognition, computer vision, translation between (natural) languages, as well as other mappings of inputs.<sup>[1]</sup>

AI applications include advanced web search engines (e.g., Google Search), recommendation systems (used by YouTube, Amazon, and Netflix), understanding human speech (such as Siri and Alexa), self-driving cars (e.g., Waymo), generative or creative tools (ChatGPT and AI art), automated decision-making, and competing at the highest level in strategic game systems (such as chess and Go).<sup>[2]</sup>

As machines become increasingly capable, tasks considered to require "intelligence" are often removed from the definition of AI, a phenomenon known as the *AI effect*.<sup>[3]</sup> For instance, optical character recognition is frequently excluded from things considered to be AI, having become a routine technology.<sup>[4]</sup>

Artificial intelligence was founded as an academic discipline in 1956, and in the years since it has experienced several waves of optimism,<sup>[5][6]</sup> followed by disappointment and the loss of funding (known as an "AI winter"),<sup>[7][8]</sup> followed by new approaches, success, and renewed funding.<sup>[6][9]</sup> AI research has tried and discarded many different approaches, including simulating the brain, modeling human problem solving, formal logic, large databases of knowledge, and imitating animal behavior. In the first decades of the 21st century, highly mathematical and statistical *machine learning* has dominated the field, and this technique has proved highly successful, helping to solve many challenging problems throughout industry and academia.<sup>[9][10]</sup>

The various sub-fields of AI research are centered around particular goals and the use of particular tools. The traditional goals of AI research include reasoning, knowledge representation, planning, learning, natural language processing, perception, and the ability to move and manipulate objects.<sup>[a]</sup>

Part of a series on  
**Artificial intelligence**

- Major goals [show]
- Approaches [show]
- Philosophy [show]
- History [show]
- Technology [show]
- Glossary [show]

**Select Voice**

Search language or voice name

Recommended Recent All

ENGLISH ▾

- Tanner Male • GB
- Jessica Female • US
- Gwyneth Paltrow Female • US **Beta**
- Snoop Dogg Male • US **Beta** ✓

0:02 1:27:46

⏮ ⏪ ⏩ ⏭ 1.0x

<https://speechify.com/> und Plugin für GoogleChrome

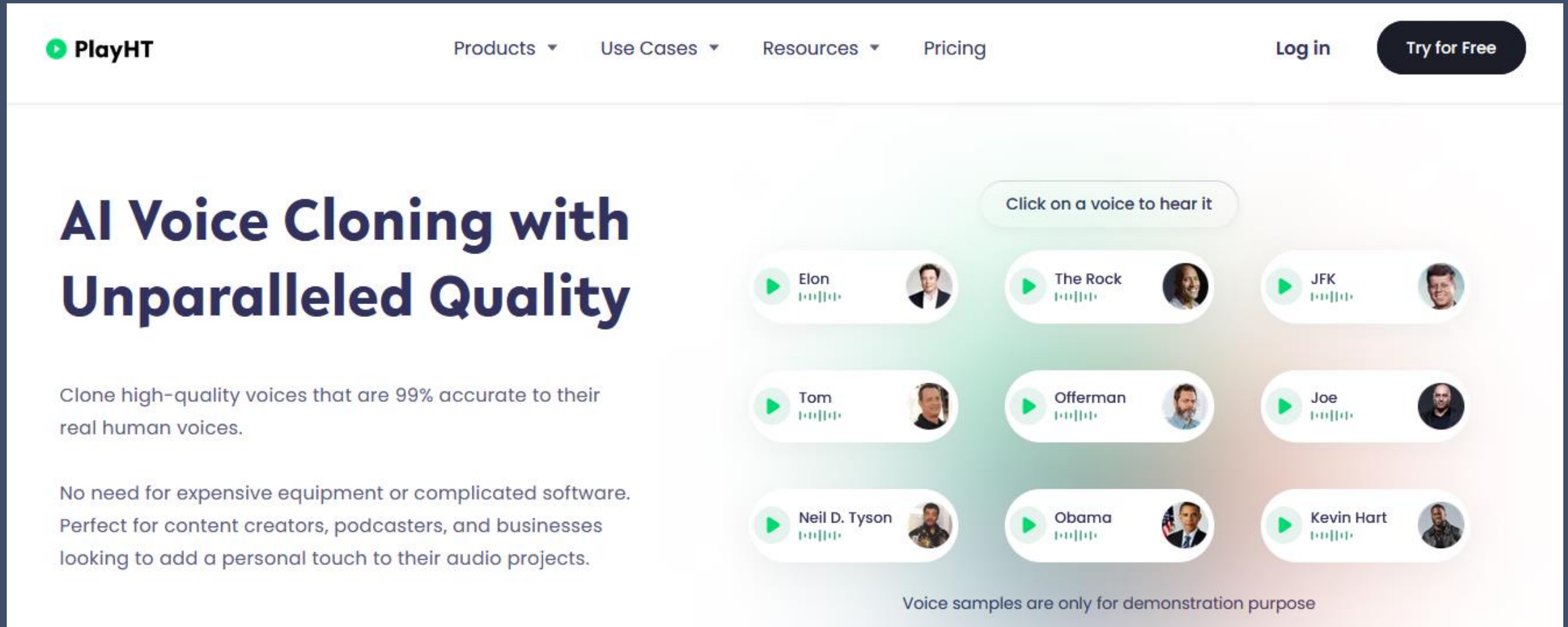
# Voice Cloning

- KI-basierte Deepfake-Technik, die verwendet wird, um eine menschliche Stimme zu analysieren und dann zu replizieren.
- Bisher meist Englisch

## Tools:

- Murf AI: <https://murf.ai/voice-cloning>
- Play HT: <https://play.ht/voice-cloning/>
- Resemble AI: <https://app.resemble.ai/playground>

# Voice Cloning



The screenshot shows the PlayHT website's homepage. At the top is a navigation bar with the PlayHT logo, links for Products, Use Cases, Resources, and Pricing, and buttons for Log in and Try for Free. The main heading reads "AI Voice Cloning with Unparalleled Quality". Below this, two paragraphs describe the service: "Clone high-quality voices that are 99% accurate to their real human voices." and "No need for expensive equipment or complicated software. Perfect for content creators, podcasters, and businesses looking to add a personal touch to their audio projects." To the right, a grid of nine voice samples is displayed, each with a play button, a name, a small audio waveform, and a profile picture. A callout box says "Click on a voice to hear it". The samples include Elon, The Rock, JFK, Tom, Offerman, Joe, Neil D. Tyson, Obama, and Kevin Hart. A disclaimer at the bottom states "Voice samples are only for demonstration purpose".

**AI Voice Cloning with Unparalleled Quality**

Clone high-quality voices that are 99% accurate to their real human voices.

No need for expensive equipment or complicated software. Perfect for content creators, podcasters, and businesses looking to add a personal touch to their audio projects.

Click on a voice to hear it

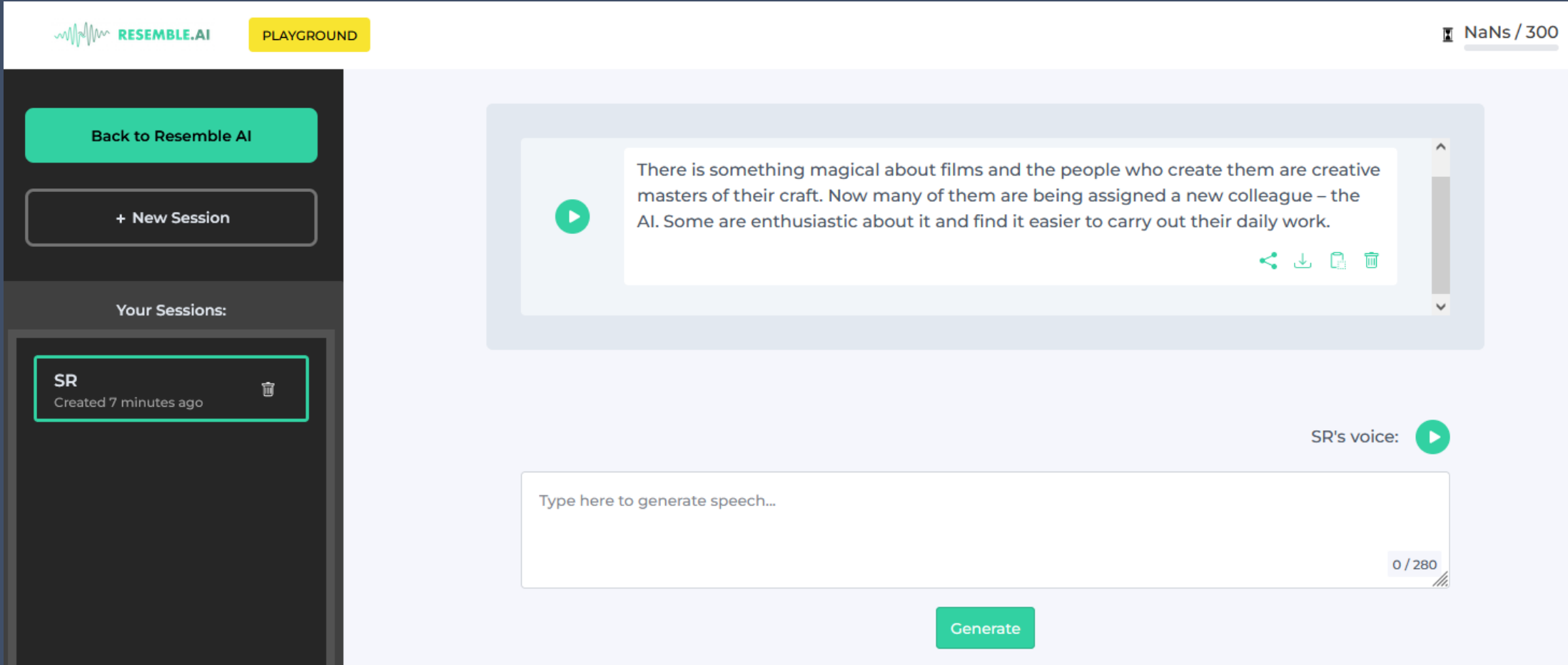
Name	Profile Picture
Elon	[Profile Picture]
The Rock	[Profile Picture]
JFK	[Profile Picture]
Tom	[Profile Picture]
Offerman	[Profile Picture]
Joe	[Profile Picture]
Neil D. Tyson	[Profile Picture]
Obama	[Profile Picture]
Kevin Hart	[Profile Picture]

Voice samples are only for demonstration purpose

text-to speech: <https://play.ht/app/voices>

<https://play.ht/voice-cloning/> (nur mit Kreditkarte)

# Voice Cloning – Resemble AI



The screenshot shows the Resemble AI Playground interface. On the left sidebar, there is a 'Back to Resemble AI' button, a '+ New Session' button, and a 'Your Sessions:' section with a session named 'SR' created 7 minutes ago. The main area displays a text input field with the text: 'There is something magical about films and the people who create them are creative masters of their craft. Now many of them are being assigned a new colleague – the AI. Some are enthusiastic about it and find it easier to carry out their daily work.' Below this, there is a 'Generate' button. To the right of the text input, there is a 'SR's voice:' label and a play button. The bottom right corner shows a character count '0 / 280'.

<https://app.resemble.ai/playground>

# Weitere Text-to-Speech Tools

## Web (auch deutsch)

- Beepbooply
- ElevenLabs
- Murf
- Revoicer

## Stimmen Clonen

- Coqui
- ElevenLabs
- Murf
- Uberduck

## Online-Dienste für Text-to-Speech

Dienst	Beepbooply	Coqui	ElevenLabs	Murf	Revoicer	Speecheasy	Überduck
URL	<a href="https://beepbooply.com">beepbooply.com</a>	<a href="https://coqui.ai">coqui.ai</a>	<a href="https://elevenlabs.com">elevenlabs.com</a>	<a href="https://murf.ai">murf.ai</a>	<a href="https://revoicer.com">revoicer.com</a>	<a href="https://speecheasyapp.com">speecheasyapp.com</a>	<a href="https://uberduck.ai">uberduck.ai</a>
Sprachen	81	1 (Englisch)	6	über 20	40	7	keine Angaben
Anzahl der Stimmen	über 900	33	9	über 120	80	92	über 5000
deutsche Stimmen	✓	–	✓	✓	✓	–	–
Stimmklone	–	✓	✓	Englisch auf Anfrage	–	–	✓
manuelle Bearbeitung	Tonhöhe, Geschwindigkeit, Pause	Dialog-Editor, Advanced Editor mit Phonem-Level-Anpassung	Stability, Clarity	Emphasis, Aussprache (IPA, alternative Schreibung)	Geschwindigkeit, Tonhöhe, Pause	SSML	–

## Wertung und Preise

Qualität deutsch / englisch	⊕ / ⊕	– / ⊕⊕	○ / ⊕⊕	⊕ / ⊕⊕	⊖ / ○	– / ○	– / ⊖⊖
Sprecherauswahl	⊕⊕	⊕	⊕	⊕⊕	⊕	○	⊖
Steuerung	⊕	⊕⊕	⊕	⊕⊕	○	○	⊖
kostenlose Version / Download erlaubt	10.000 Zeichen / –	30 Minuten / ✓	10.000 Zeichen / ✓	10 Minuten / –	– / –	je 20-Sekunden-Clips / ✓	5 Minuten / ✓
Abopreise pro Monat	ab 7 US-\$ für 100.000 Zeichen	ab 20 € für 4 Stunden (kein Abo)	ab 5 US-\$ für 30.000 Zeichen	ab 29 \$ für 2 Stunden Audio	ab 27 US-\$ für 60.000 Zeichen	ab 6 US-\$ für 15- Minuten-Clips	ab 10 US-\$ für 1 Stunde

⊕⊕ sehr gut ⊕ gut ○ zufriedenstellend ⊖ schlecht ⊖⊖ sehr schlecht ✓ vorhanden – nicht vorhanden

<https://www.heise.de/tests/KI-Stimmen-Sieben-Anbieter-fuer-Text-to-Speech-im-Vergleich-9156580.html?seite=all>



# Text-to-Audio - bark

erzeugt realistische Sprache und andere Audioinhalte (Musik, Hintergrundgeräusche, einfache Soundeffekte)

[laughter]

- [laughs]
- [sighs] seufzen
- [music]
- [gasps] keuchen
- [clears throat] räuspern
- — or ... for hesitations
- 🎵 for song lyrics
- capitalization for emphasis of a word
- MAN/WOMAN: for bias towards speaker

<https://github.com/suno-ai/bark>

<https://colab.research.google.com/drive/1eJfA2XUa-mXwdMy7DoYKVYHI1iTd9Vkt?usp=sharing#scrollTo=NyYQ--3YksJY>

# Text-to-Audio - bark

## Beispiele:

[clears throat] Hello, my name is Suno. And, uh — and I like pizza. [laughs] But I also have other interests such as... ♪ singing ♪.



WOMAN: [clears throat] Hallo, ich heiße Susi [laughter] und mag Pizza [laughs]. Ich bin gerade ein gerannt [gasps] und ein wenig k.o. [sighs].



Hello, my name is Tom. And, I like singing ... ♪ singing all the day and everywhere ♪.

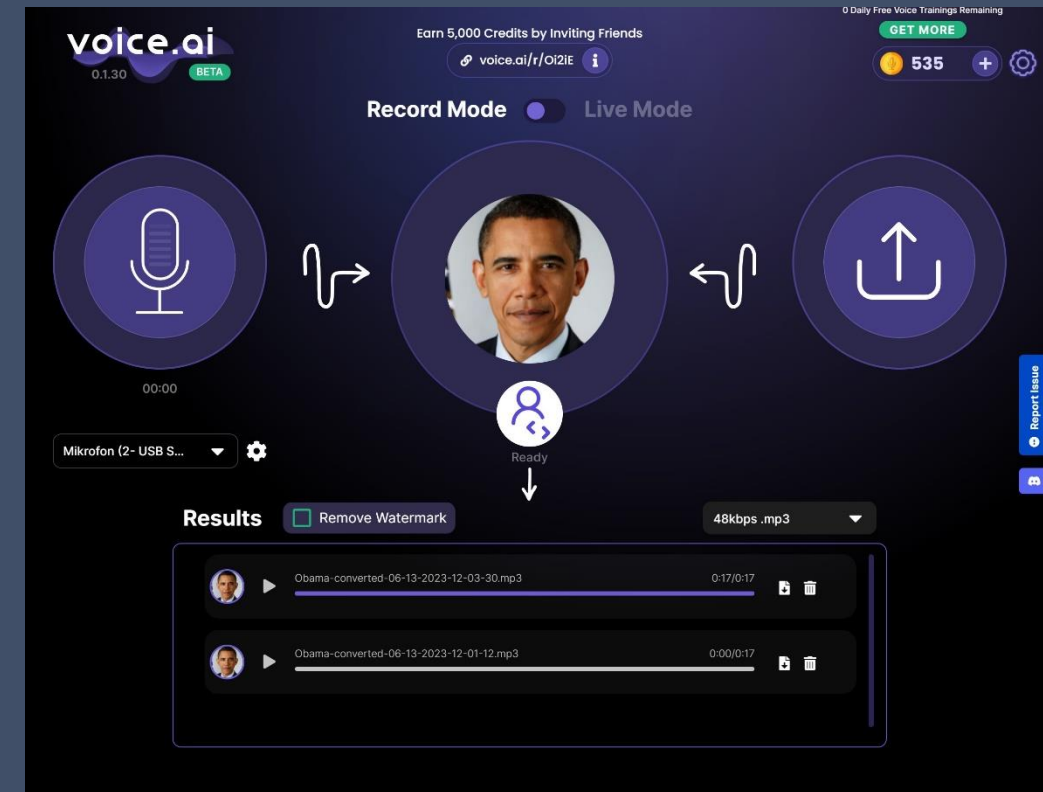
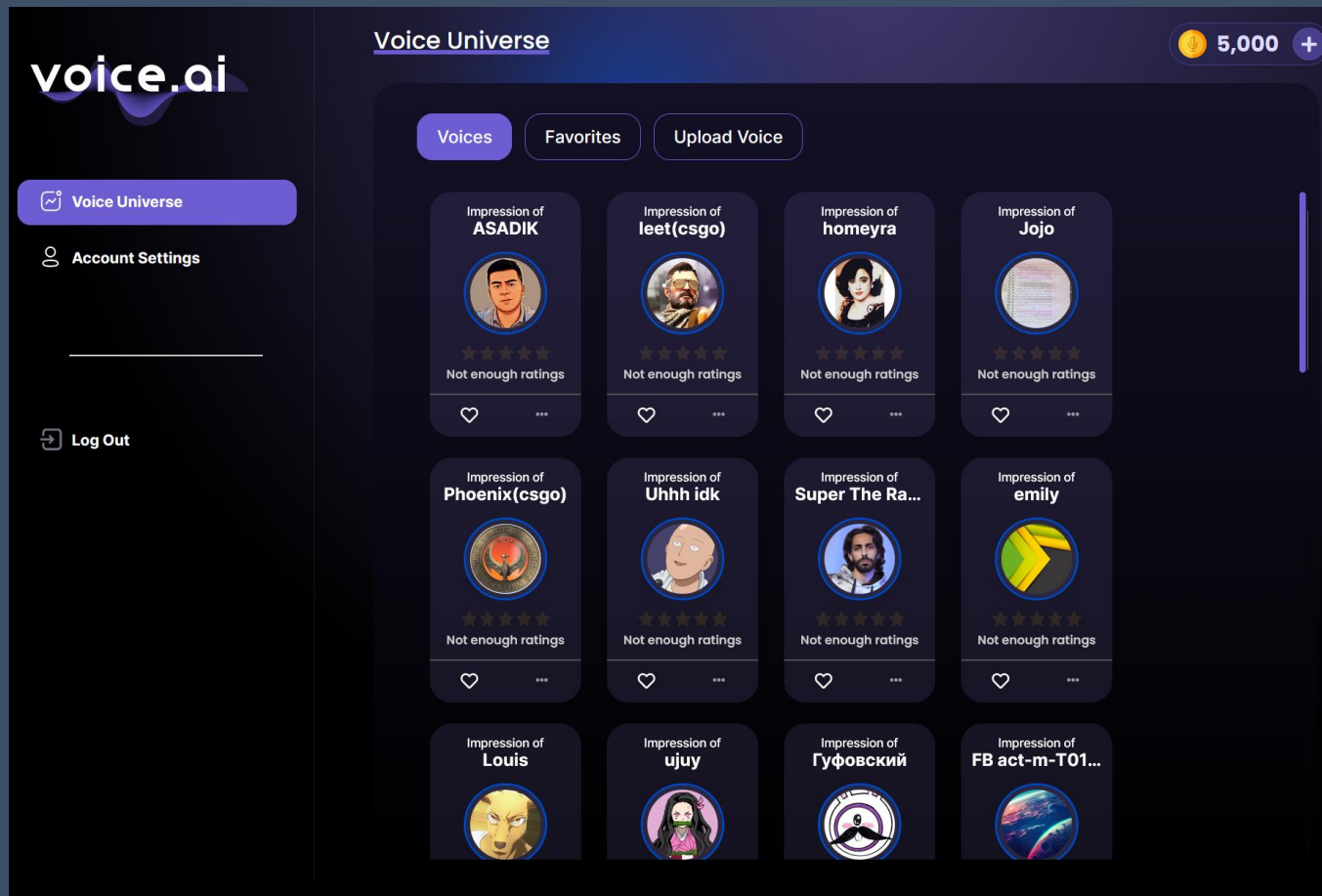


<https://github.com/suno-ai/bark>

<https://colab.research.google.com/drive/1eJfA2XUa-mXwdMy7DoYKVYH1iTd9Vkt?usp=sharing#scrollTo=NyYQ--3YksJY>

# Voice AI

HOCHSCHULE  
FÜR FERNSEHEN UND  
FILM MÜNCHEN



<https://voice.ai/>





# Text/Speech/Image-to-Video



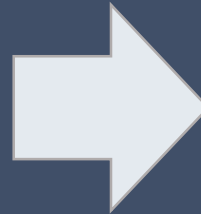
<https://studio.d-id.com/>

[iZotope RX 10](#)





# Text/Speech/Image-to-Video



<https://www.heygen.com/>

# Automatic1111: SadTalker

- <https://github.com/OpenTalker/SadTalker>
- <https://huggingface.co/spaces/vinthony/SadTalker>

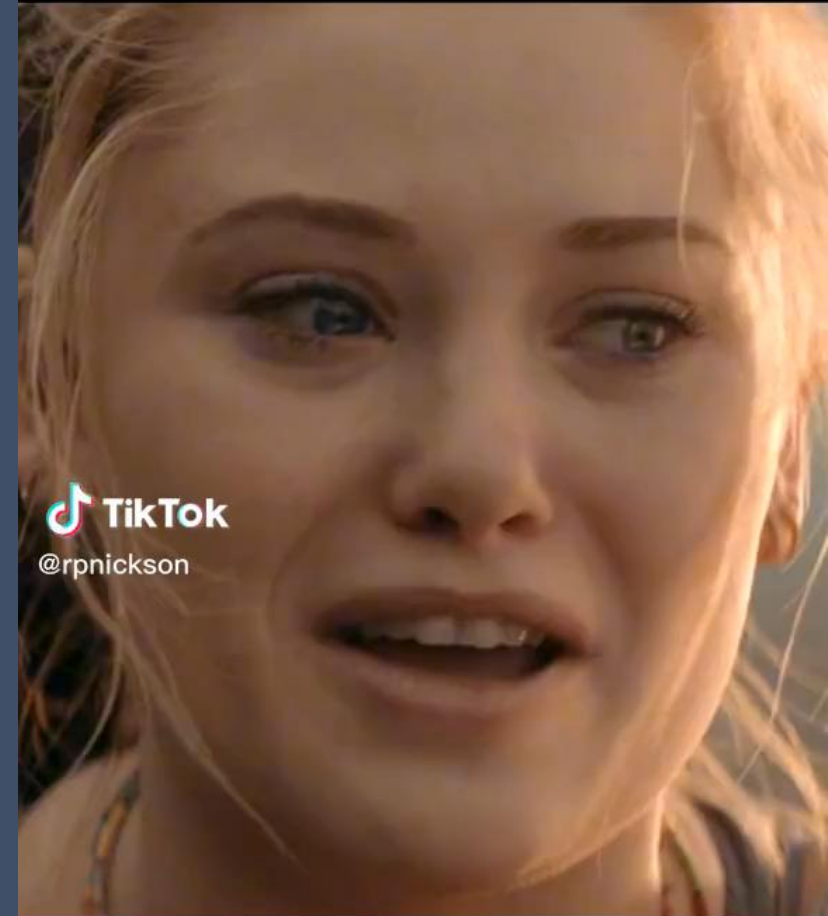




# Modifikation

## TrueSync (Flawless)

- Beispiel: Synchronisation
- Mundbewegungen der Schauspieler werden mit dem synchronisierten Dialog abgeglichen
- Service, kein Produkt
- <https://www.flawlessai.com/>



ORIGINAL SCENE - R RATED

# Generierung - Musik



# Musik

- Längen Anpassungen (z.B. Adobe Premiere)
- Musikgenerierung
- Styletransfer

# Tools Musikgenerierung

## Web-basiert (bisher nicht überzeugend):

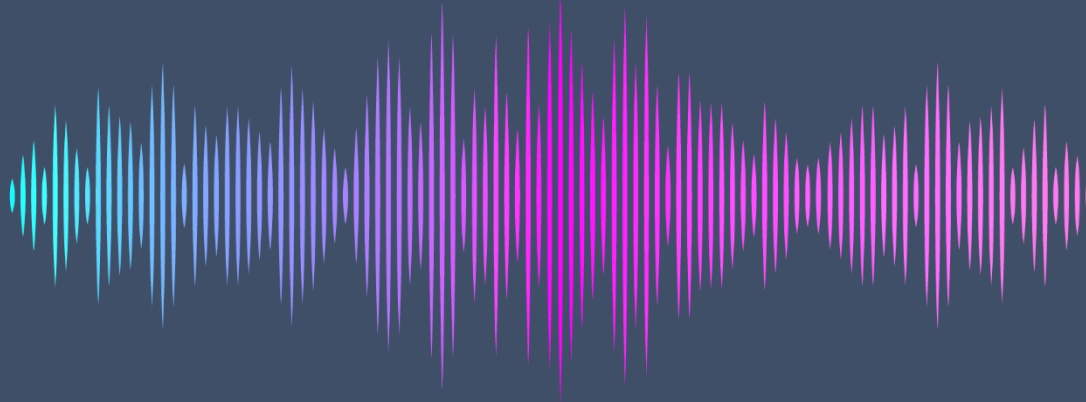
- <https://aiva.ai/>
- <https://melobytes.com/en/app/melobytes> (Text-to-Music)
- <https://melobytes.com/en/app/image2music> (Video-to-Music)
- <https://huggingface.co/spaces/Mubert/Text-to-Music> (Text-to-Music)
- <https://huggingface.co/spaces/fffiloni/img-to-music> (Image-to-Music)
- <https://openai.com/blog/musenet/> (Try, Advanced Settings)
- <https://soundraw.io/>

## Installieren:

- <https://openai.com/blog/jukebox/>
- <https://magenta.tensorflow.org/studio>

# Musik – Synthetisierung

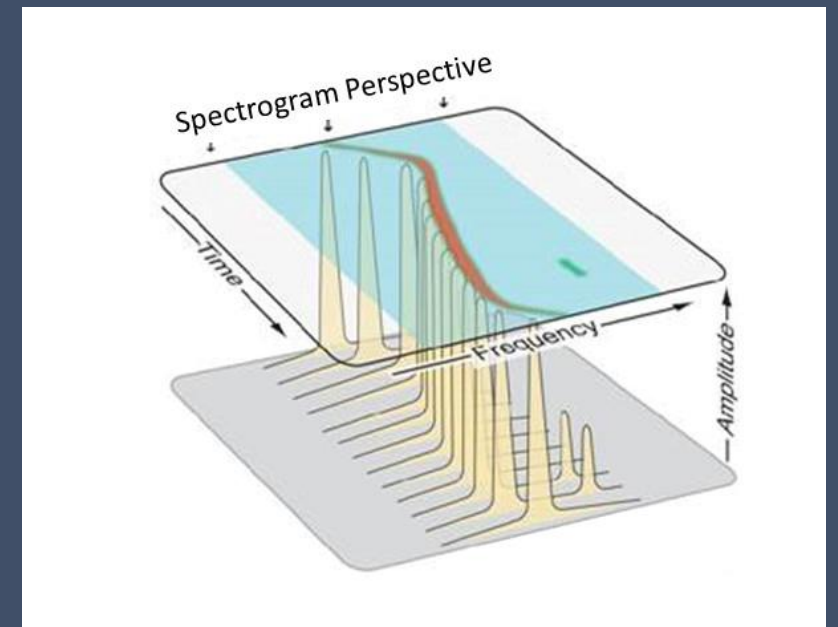
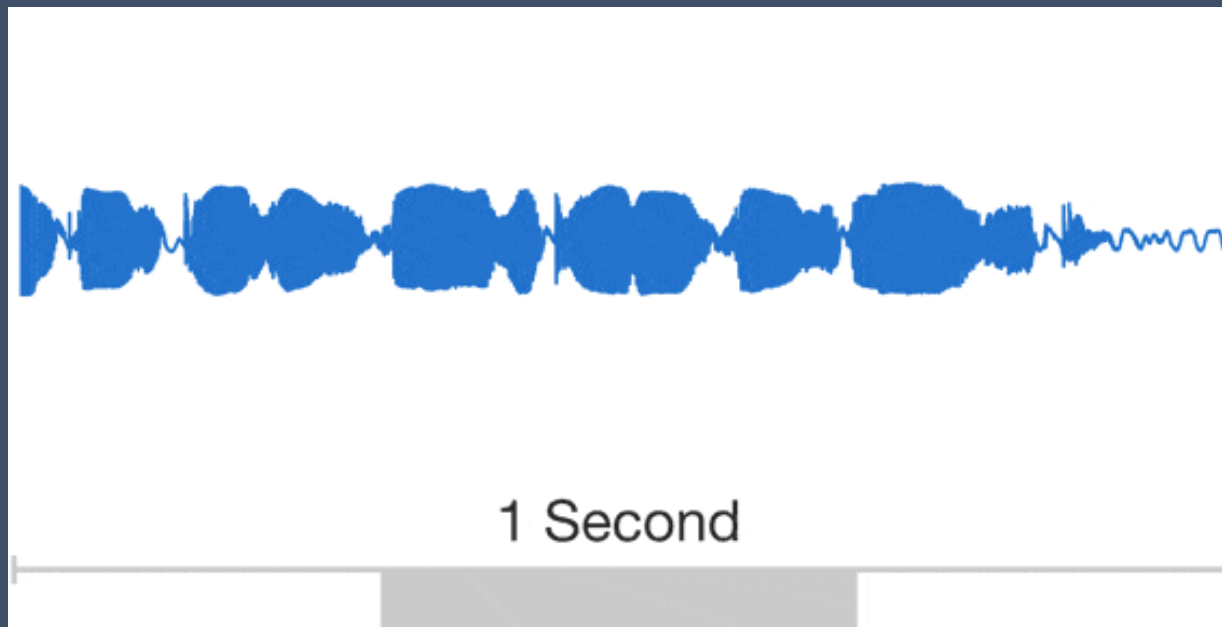
- WAV besteht aus Daten



- KI kann lernen diese Daten zu generieren
- Daten müssen zeitlich kohärent, melodisch, harmonisch und rhythmisch sein

# Komplexität von Tönen

Frequenz, Amplitude, Zeit





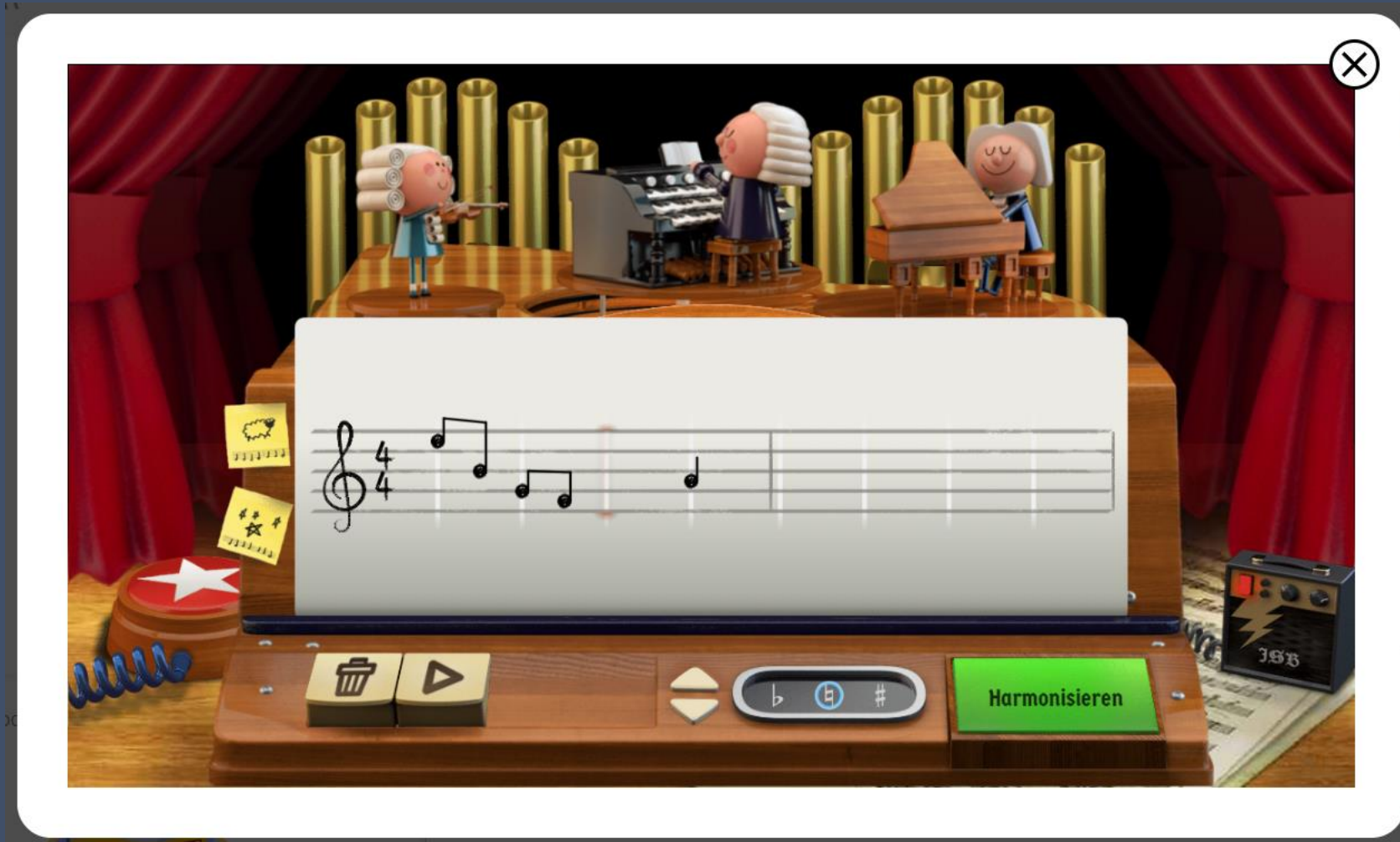
# Musik - Komposition

- KI komponiert MIDI Files / Noten
- Kann über MIDI Player oder auch von Menschen gespielt werden



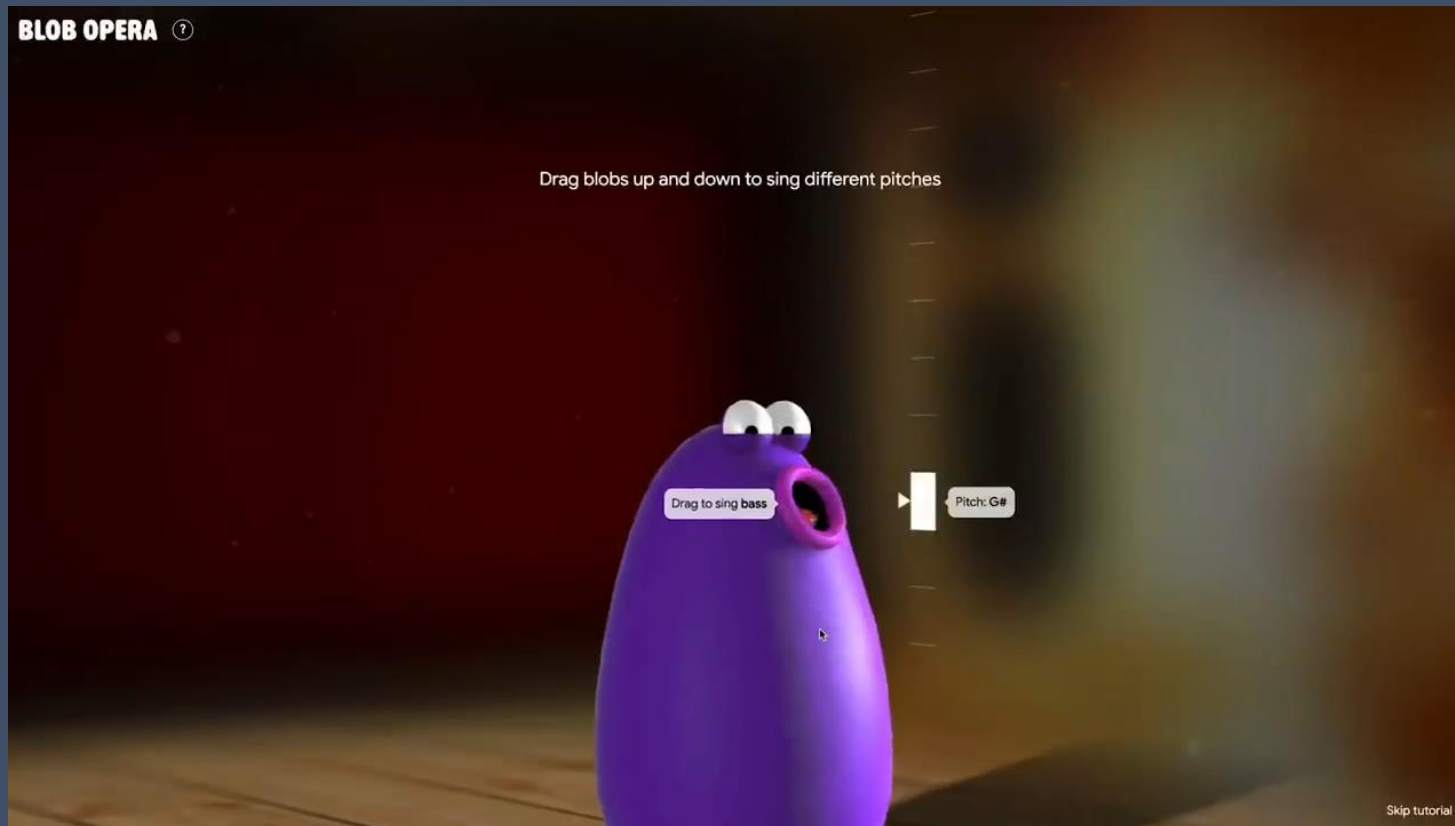
<https://www.aiva.ai/>

# Musikgenerierung (2019)



<https://www.google.com/doodles/celebrating-johann-sebastian-bach>

# Künstliche Intelligenz und Musik



2 KI Modelle:

- Harmonien
- Übergänge zwischen den Tönen

# MusicLM - Text-to-Music

Musik:



Prompt:

We can hear a choir, singing a Gregorian chant, and a drum machine, creating a rhythmic beat. The slow, stately sounds of strings provide a calming backdrop for the fast, complex sounds of futuristic electronic music.

Musik:

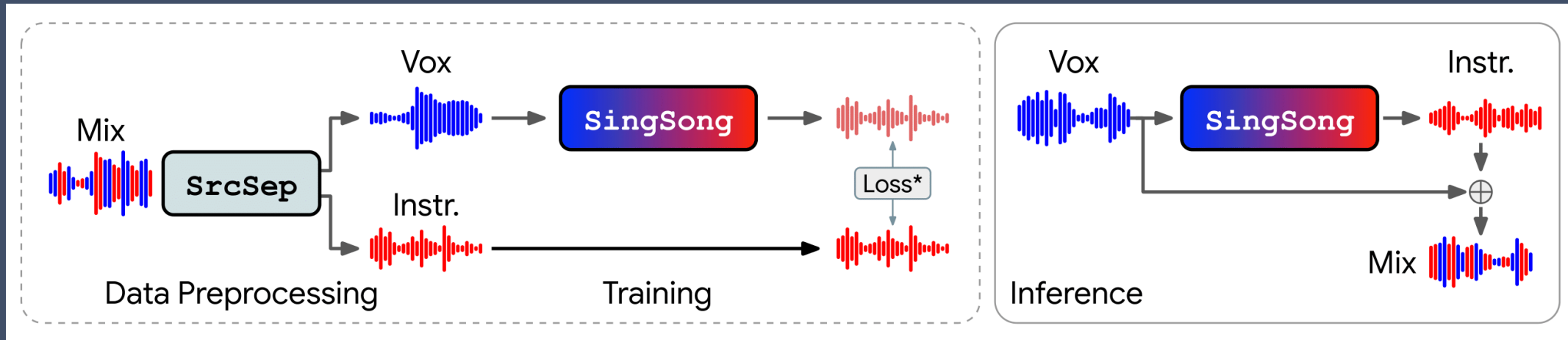


Prompt:

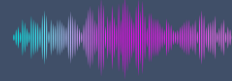
This is an r&b/hip-hop music piece. There is a male vocal rapping and a female vocal singing in a rap-like manner. The beat is comprised of a piano playing the chords of the tune with an electronic drum backing. The atmosphere of the piece is playful and energetic. This piece could be used in the soundtrack of a high school drama movie/TV show. It could also be played at birthday parties or beach parties.

<https://google-research.github.io/seanet/musiclm/examples/>

# SingSong - Musik Assistent



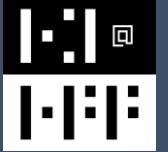
- SingSong
- Ground Truth:
- Voice Separated:
- SingSong:



<https://storage.googleapis.com/sing-song/index.html>

# Video-to-Music

HOCHSCHULE  
FÜR FERNSEHEN UND  
FILM MÜNCHEN



<https://melobytes.com/en/app/image2music>

14.06.2023

KI Bild- und Videogenerierung 2023

**Melobytes** Where creativity meets AI Subscribe Support Community

### Image / Video to Music

**How it works**

1. Upload your image or video
2. The app generates using special algorithms a unique result based on your content
3. Witness your creation and share it with the world, or keep it for yourself — it's yours ❤️

**Input data**

Upload your Image/video (R) infinite\_zoom\_1685446208.mp4

Allowed file types: aac, m4a, mp3, ogg, wav, aiff, mov, wmv, avi, mp4, gif, mkv, flv, mid, kar, jpeg, jpg, bmp, wmf, gif, png, ico, tiff, emf, rle  
Max size : 10 MB  
[Drag and drop files anywhere](#)

Generate video (R) No

**Convert**

Free user limits apply

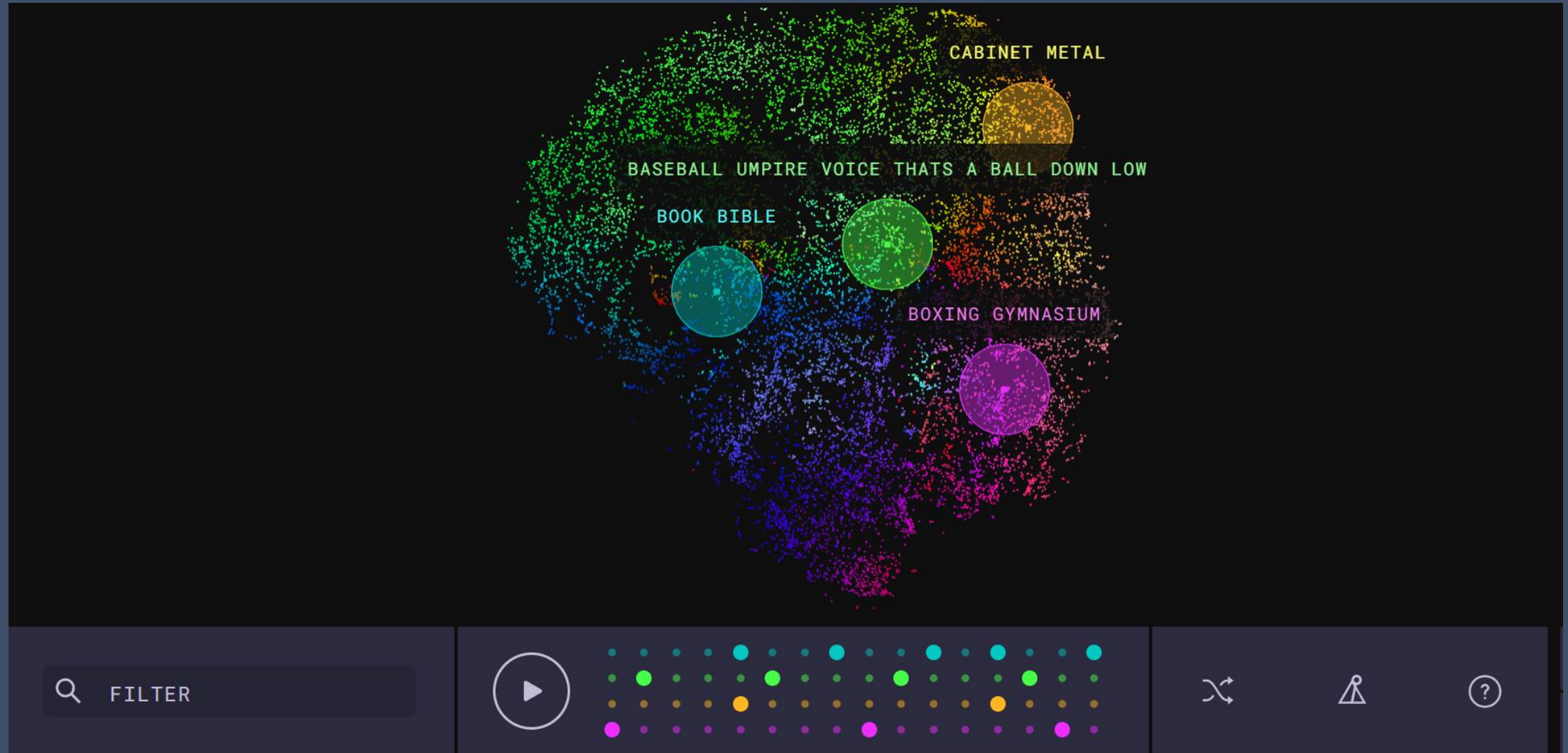
**Results**

File  
[Download MIDI](#)

Audio  
0:30 / 1:42 Share

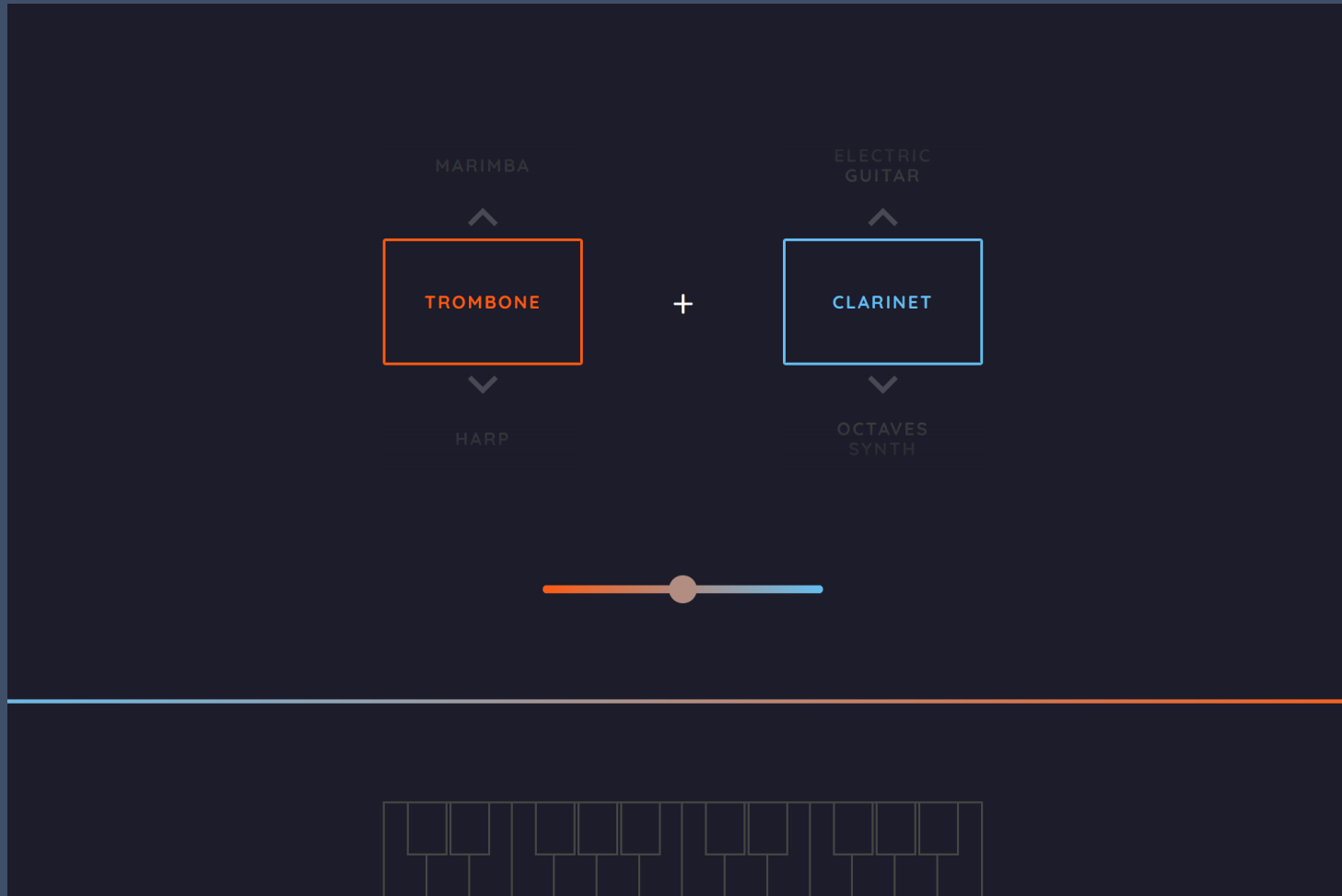


# Musikgenerierung



<https://experiments.withgoogle.com/ai/drum-machine/view/>

# Musikgenerierung



<https://experiments.withgoogle.com/ai/sound-maker/view/>

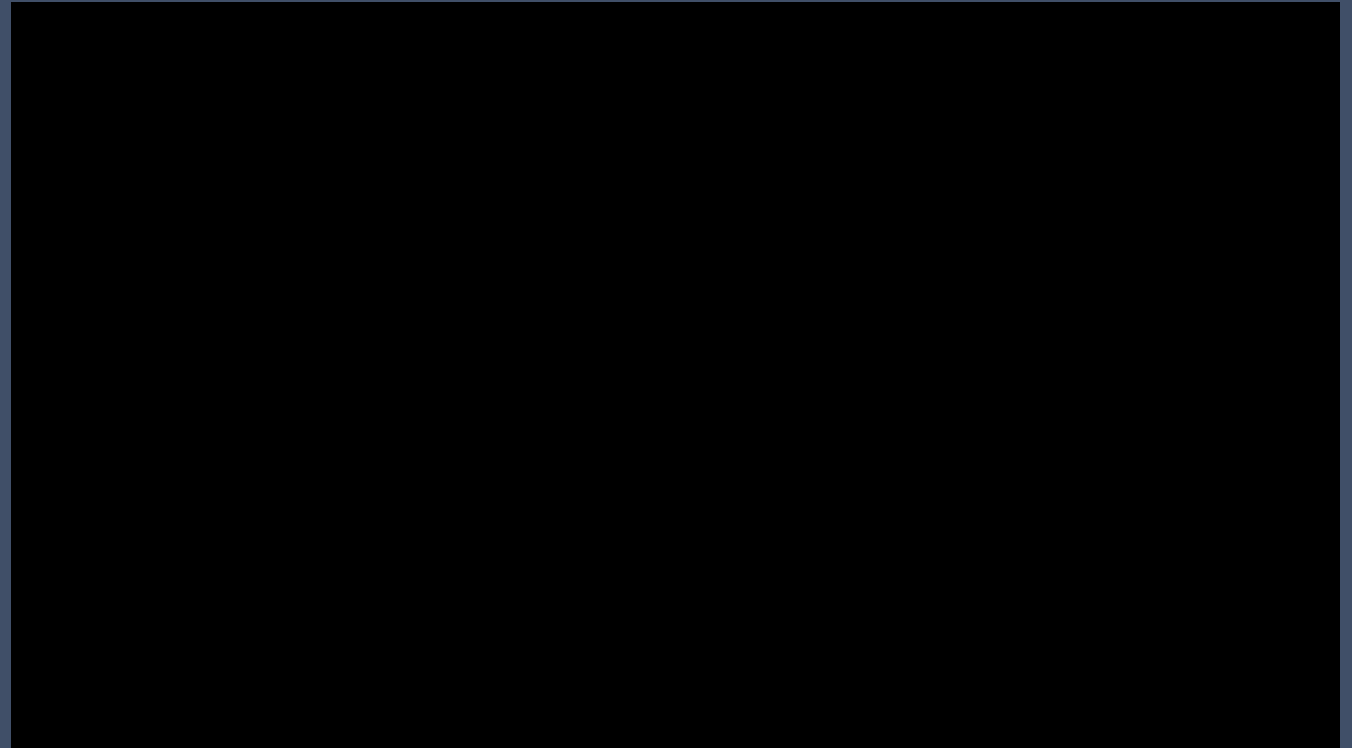
# Generierung – Geräusche (Forschung)

Soundeffekte & Foleys

# Soundeffekte & Foleys

## Text-to-Audio:

- Aukustische Umgebung
- Material Kontrolle



<https://audioldm.github.io/>

# Soundeffekte & Foleys

## Text-to-Audio:

- Aukustische Umgebung
- Material Kontrolle

## Style Transfer



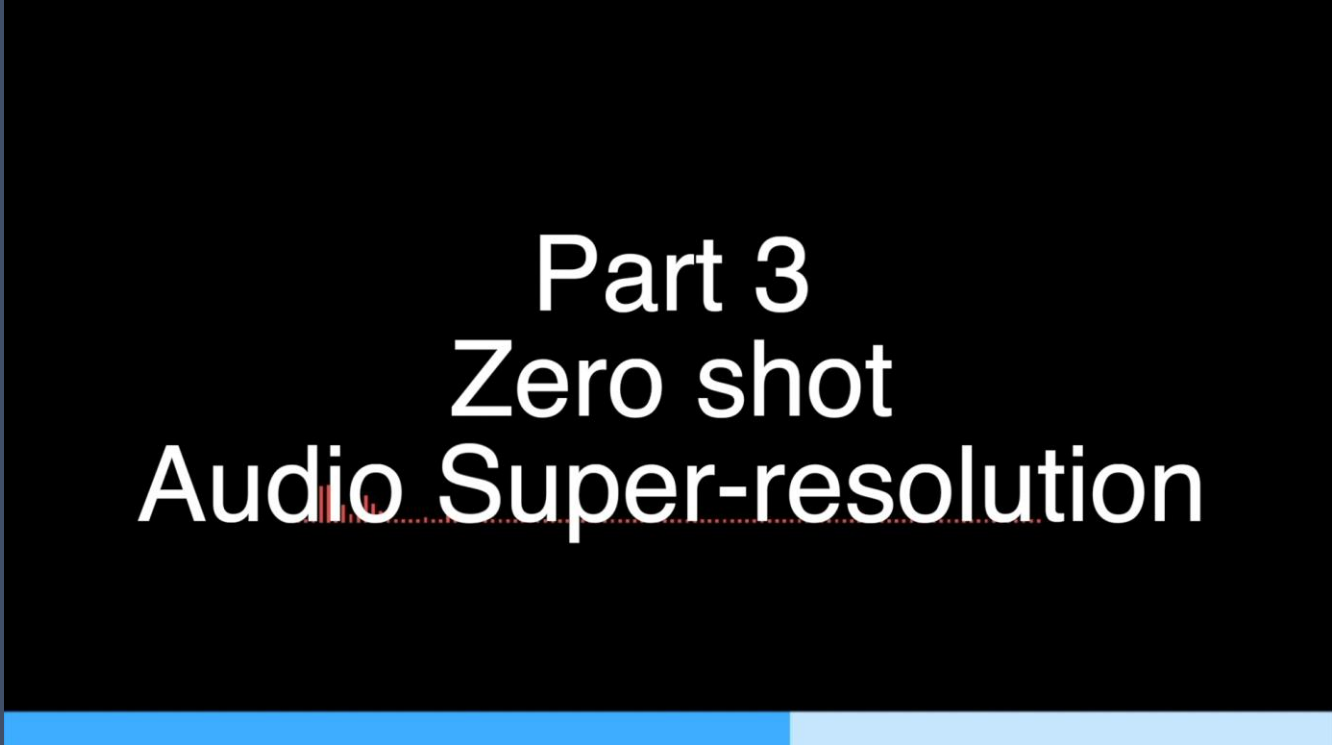
<https://audioldm.github.io/>

# Soundeffekte & Foleys

## Text-to-Audio:

- Aukustische Umgebung
- Material Kontrolle

Style Transfer  
Upsampling



Part 3  
Zero shot  
Audio Super-resolution

<https://audioldm.github.io/>



# Soundeffekte & Foleys

## Text-to-Audio:

- Akustische Umgebung
- Material Kontrolle

Style Transfer

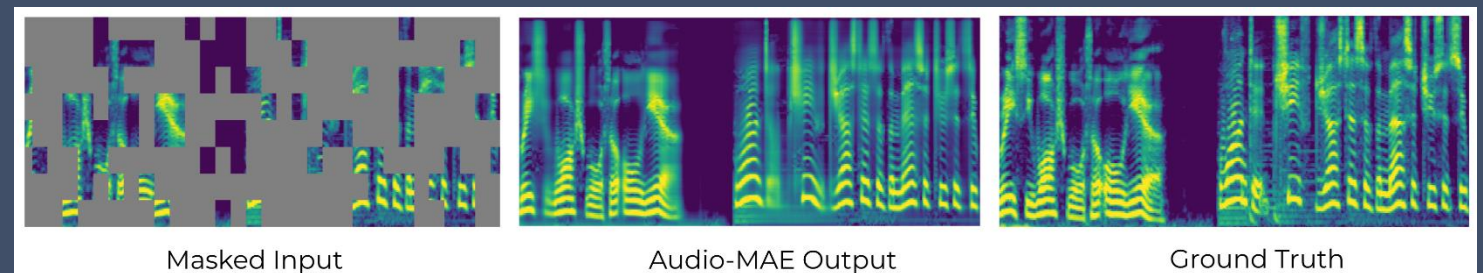
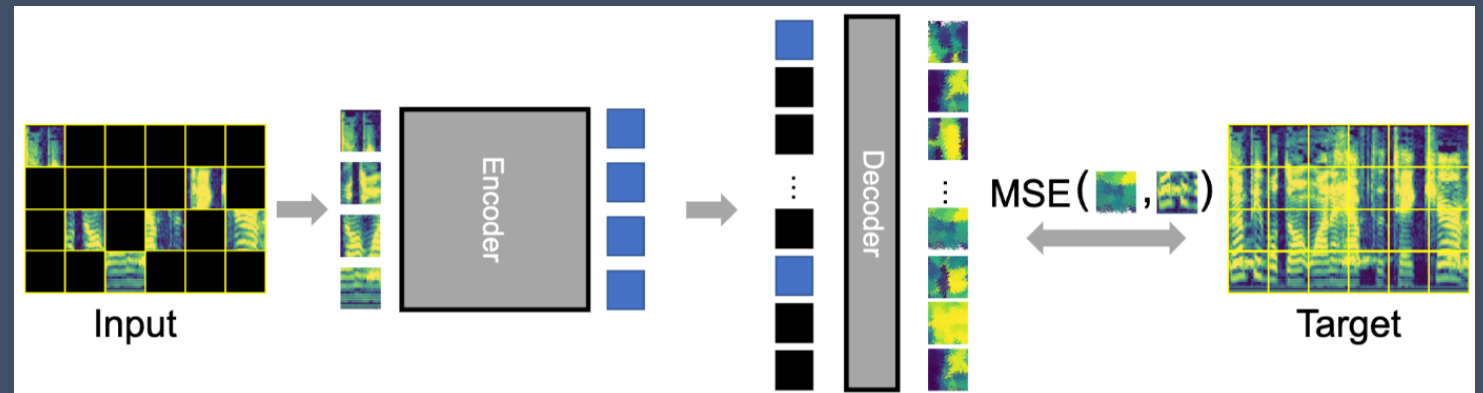
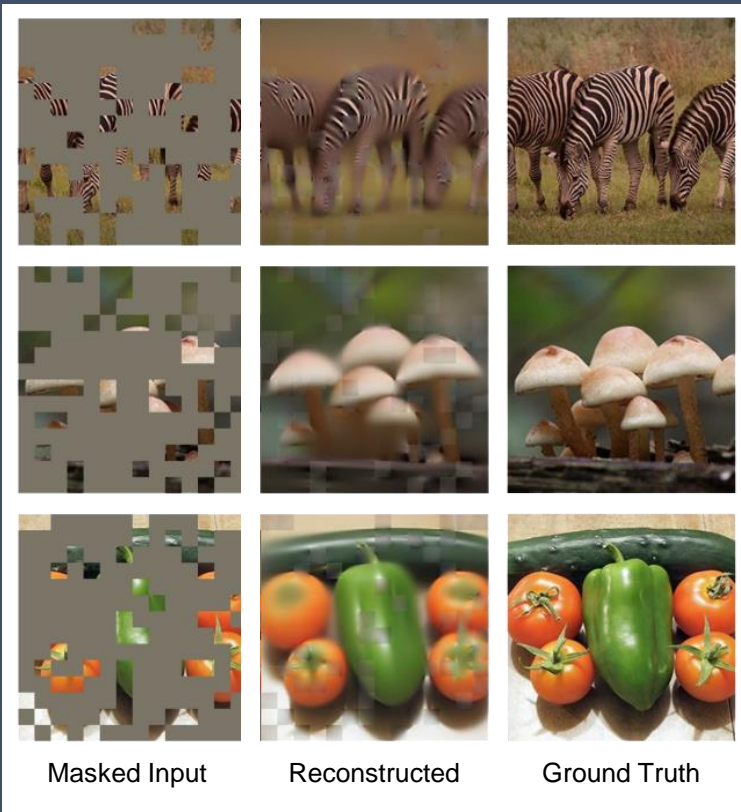
Upsampling

„Audio Inpainting“

<https://audioldm.github.io/>

Part 4  
Zero shot  
Audio inpainting

# Audio Rekonstruktion



<https://github.com/facebookresearch/AudioMAE>

# Masked Autoencoders that Listen

Masked Input

<https://github.com/facebookresearch/AudioMAE>