



**Qualitative und freie Sprachsynthese -
darf's ein bisschen weniger Cloud sein.**

06.08.2023



@ThorstenVoice




Thorsten-Voice

Vielen Dank an die
Organisatoren der
FrOSCon



Guude

- **Thorsten Müller** | Hessebub
- IT'ler aus Leidenschaft | Kein professioneller Sprecher
- Enthusiast für offenen Sprachtechnologien seit 2019
- Gründer und **Stimmenspender** vom “**Thorsten-Voice**” Projekt
-  **Thorsten-Voice** | Tutorials zu freien Sprachtechnologien

Thorsten-Voice

Technische **Tutorials** rund um offene
Sprachtechnologien (/)

Mein Wunsch:

Ein Youtube “Play” Button
(100.000 Abos)

Es fehlen “nur” noch 97.500

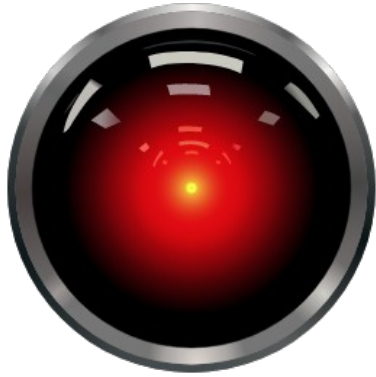


https://commons.wikimedia.org/wiki/File:Youtube_Creator_Award.jpg

Was ist Sprachsynthese | Text-to-Speech | TTS?

Künstliches Erzeugen von
menschlicher Stimme auf Basis
geschriebenen Textes.

Sprachsynthese war früher Science Fiction



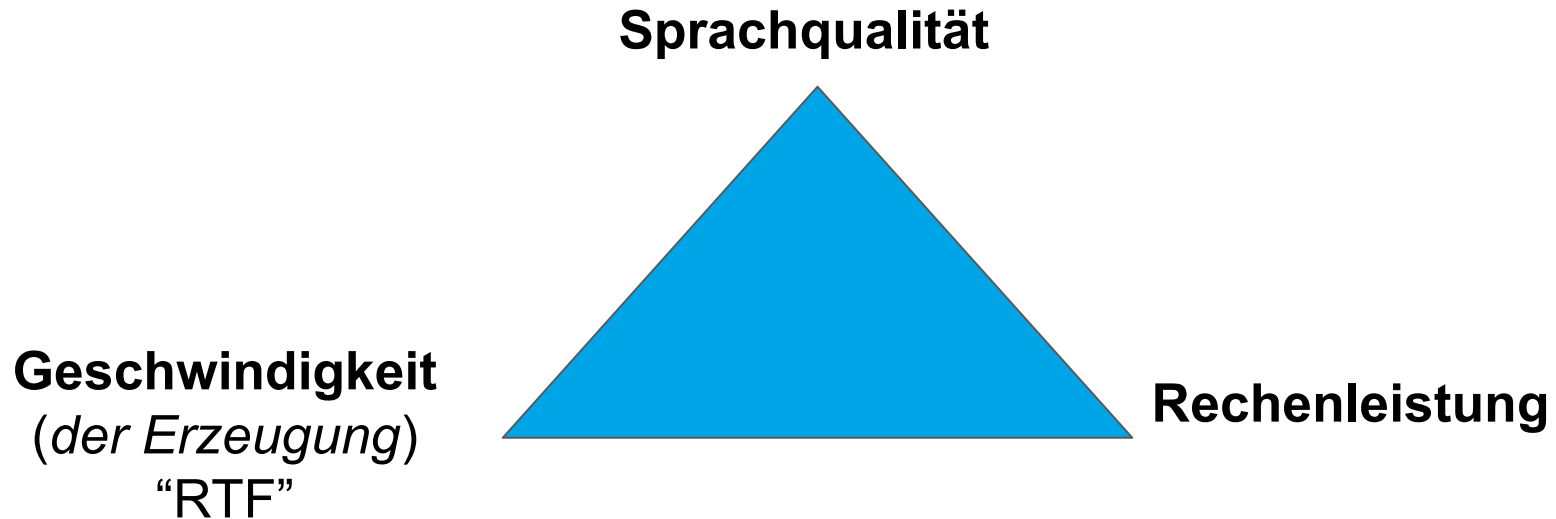
KNIGHT RIDER

Sprachsynthese ist im Alltag angekommen

- Sprachassistenten (*Alexa, Siri, Google Home*)
- Smartphone
- Smartwatch/Fitness-Tracker
- TV Steuerung
- Navigationssysteme
- ...
- **Barrierefreiheit** (Sehen/Sprechen)

**Also nicht nur “Schmuck am Nachthemd”,
sondern auch “Lebensqualität”**

Qualität hat einen “Preis” - Cloud oder Lokal / im eigenen Netz



Wie künstliche Stimme erzeugt werden kann

Cloud basiert / Kommerziell

- Google TTS
- Amazon Polly
- Microsoft Azure
- ...

Lokal / Frei (*Synthetisch*)

- eSpeak(-ng)
- Festival
- MBROLA
- SVOX/Pico
- ...

Lokal / Frei (*Deep Learning*)

- Coqui TTS
- Mimic 3
- Piper TTS
- ...

- Einfache API Einbindung
- Gute Geschwindigkeit / Qualität
- Techn./Finanzielle Abhängigkeiten
- Benötigt Internetzugang
- Datenschutzaspekte

- Sehr schnell
- Roboterhaft /
Metallisch
- Wenig natürlicher
Sprachfluss

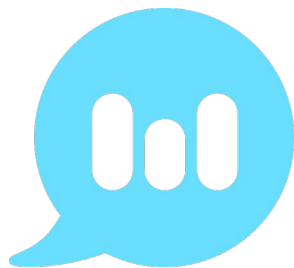
- Natürlicher Sprachfluss
- Gute bis sehr gute Qualität
- Gewisse Rechenleistung
nötig
- Basieren auf Stimmspenden

Beispiele künstlicher Stimmen




“Hallo liebe FrOSCon, ich freue mich heute als Sprecher dabei sein zu dürfen.”



Pico



Coqui TTS, Mimic 3 und Piper TTS - lokal lauffähig, kostenfrei, KI-basiert

	 coqui TTS	MYCROFT AI 	 P PER
Projekt gestartet	März 2021 (aus Mozilla TTS)	Juni 2022	Januar 2023 von Michael Hansen (Synthesiam)
Anzahl Sprachen	20+	24+	23+
Raspberry Pi kompatibel	Bedingt, nur sehr langsam	Ja	Ja
Emotionen / SSML	Nein	Ja	Ja
Bedienung per	Lokalem Webfront, Python API, Kommandozeile	Kommandozeile, Webfront	Kommandozeile, Python
Unterstützt "Thorsten-Voice"	Ja	Ja <i>(inkl. Emotionen)</i>	Ja



www.Thorsten-Voice.de

“Thorsten-Voice” | Vision

“Eine qualitativ hochwertige, deutsche, künstliche Stimme, die offline erzeugt werden kann, sollte jedem Projekt kostenlos und ohne lizenzrechtliche Einschränkungen zur Verfügung stehen.”

Warum habe ich meine Stimme gespendet?

Weil ich's kann :) und um die "Lücke" im deutschsprachigen Raum zu schließen. Bei allen Chancen und Risiken glaube ich fest daran:

“Die Menschheit ist gut und kann Großes erreichen, wenn wir uns gegenseitig vertrauen.”

Persönliches Statement:

Für mich sind alle Menschen gleich, unabhängig von Geschlecht, sexueller Orientierung, Religion, Hautfarbe oder Geokoordinaten der Geburt. Ich glaube an eine globale Welt, wo jeder überall willkommen ist und freies Wissen und Bildung kostenfrei für jeden zur Verfügung steht. Ich habe meine Stimme der Allgemeinheit gespendet, in der Hoffnung darauf, dass sie in diesem Sinne genutzt wird.

Thorsten-Voice - Die Fakten zur “Stimmspende”

- Drei CC0 Sprachdatensätze (*2x neutral, 1x emotional*) mit über **33.000 Aufnahmen** und einer reinen Audiodauer von ca. **34 Stunden!**
- Zwei TTS Modelle (mit Coqui TTS Open-Source Software) trainiert
- NVIDIA NeMo, Mycroft Mimic 3 (*inkl. Emotionen*), Piper TTS, TensorflowTTS, ...
- Mehrere Universitäten/Hochschulen/Schulen nutzen die Sprachdatensätze
- Verfügbar in der Smarthome Software “Home Assistant”

- <https://www.thorsten-voice.de/einfach-loslegen/>
- <https://huggingface.co/spaces/Thorsten-Voice/demo>

Mehr langweilige Fakten unter

<https://github.com/thorstenMueller/Thorsten-Voice>

Bei Nutzung von Thorsten-Voice gerne einen “Showcase-Issue” öffnen.

Kann ich auch
meine Stimme klonen?

Ja . . . aber es ist kein “Quick Win”

&

kein “Weiter -> Weiter -> Fertigstellen” Prozess

Deine eigene TTS-Stimme erzeugen (+ *meine "Fails"*)

- Tausende einzelne Sätze anhand Textkorpus aufnehmen (*bspw. mit Mimic-Recording-Studio*) und die Aufnahmen "Probehören"
- Neutral aber natürlich sprechen
- Die Aufnahmen als "Voice Dataset" packen - bspw. LJSpeech Struktur
- Mit Coqui oder Piper TTS trainieren (*GPU statt CPU*)
- Regelmäßige Kontrolle des Trainingsfortschrittes (*Tensorboard*)
- Stunden/Tage/Wochen trainieren lassen

Verwendet ein gutes Mikrofon und Aufnahmesetup!

Dankeschön Dominik



<https://github.com/domcross>

Nicht das "Abo" meines **Youtube Kanals** vergessen



Die eigene Stimme klonen
Text-to-Speech

Folge 1: Einleitung

Die eigene Stimme klonen (Text-to-Speech)

Thorsten-Voice

Öffentlich

7 Videos 3.498 Aufrufe Heute aktualisiert

Alle ansehen Zufallsmix

Sortieren

- Eigene Stimme klonen (1) - Einleitung**
Thorsten-Voice • 3337 Aufrufe • vor 1 Jahr
Video 1: Einleitung 5:23
- Eigene Stimme klonen (2) - Audioaufnahmen der eigenen Stimme**
Thorsten-Voice • 1997 Aufrufe • vor 1 Jahr
Video 2: Audioaufnahmen der Originalstimme 10:45
- Eigene Stimme klonen (3) - Tipps zur Textauswahl**
Thorsten-Voice • 1059 Aufrufe • vor 1 Jahr
Video 3: Tipps zur Textauswahl 8:58
- Eigene Stimme klonen (4) - Voice Dataset erstellen**
Thorsten-Voice • 2028 Aufrufe • vor 1 Jahr
Video 4: Voice Dataset erstellen 7:47

Youtube Playlist mit
Schritt-für-Schritt
Anleitungen zum **freien**
und **lokalen Klonen**
eurer **eigenen Stimme**.



Vielen Dank für eure Aufmerksamkeit.
Gibt's Fragen?



Mail: tm@thorsten-voice.de
Twitter: @ThorstenVoice
Youtube: Thorsten-Voice
Github: github.com/thorstenMueller/Thorsten-Voice

