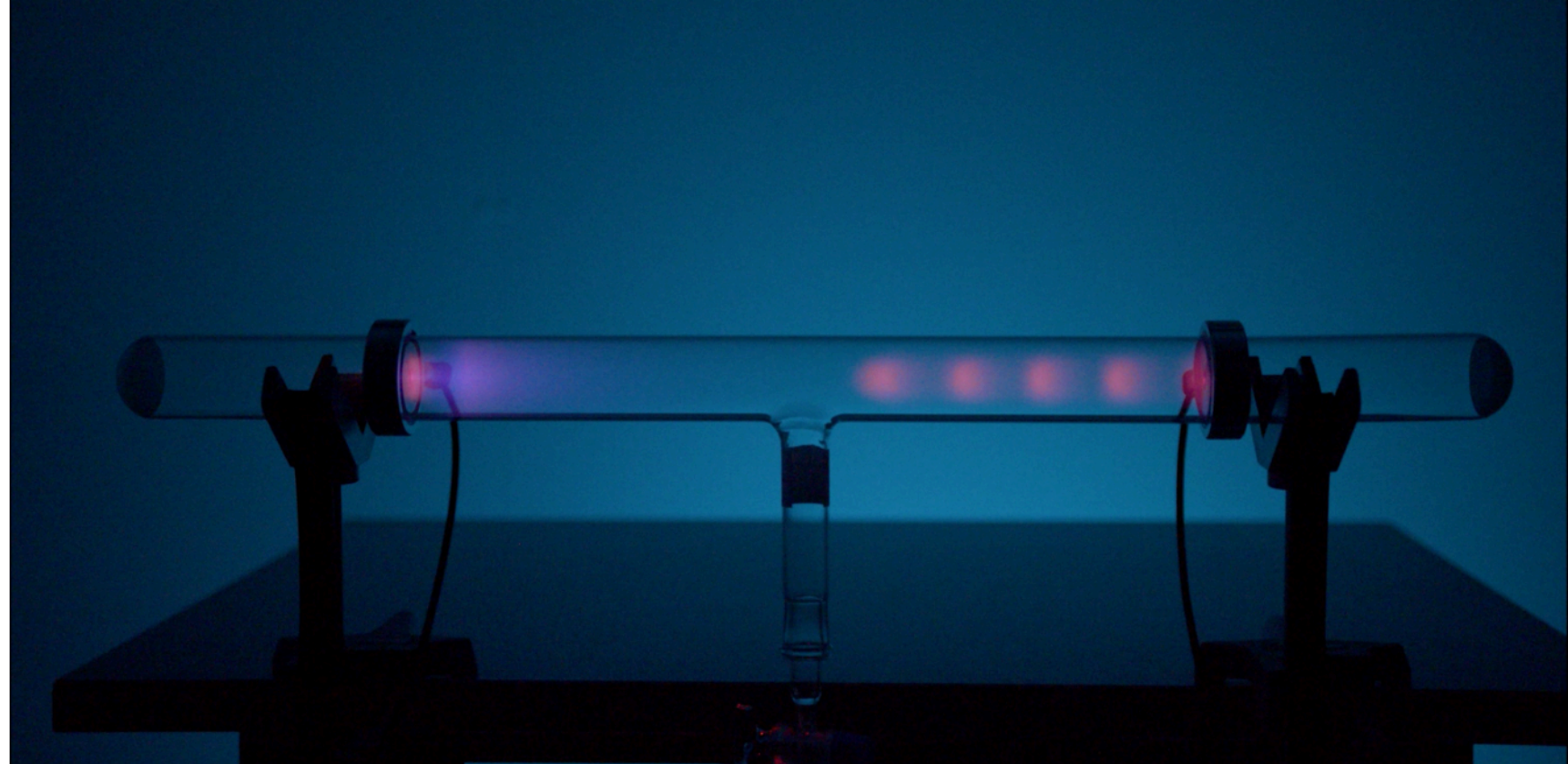


Text-to-Speech: effiziente Sprachsynthese

Dr. Gillian Kiliani

Fortbildungsreihe FO>KI, 07.03.2025



Struktur des Workshops

Teil 1

- Einführung
- Töne und Klänge
- erste Demonstrationen von TTS

Pause

Teil 2

- TTS im Webinterface
- TTS in der API
- TTS in Spektrogrammen

Was ist Text-to-Speech (TTS)?

Teil 1

Einführung

Töne und Klänge

Demonstration von TTS

Pause

Teil 2

TTS im Webinterface

TTS in der API

TTS in Spektrogrammen

Sprache wird durch einen Computer erzeugt

- Eingabe:
„Diese Audiospur wurde nicht von einer Person
eingesprochen, sondern mit Hilfe einer KI generiert.“
- Ausgabe:



- starke Weiterentwicklung durch KI

Vorlesungsexperimente in der Physik

„Ich mache niemals Videos.“

Ein Versprechen an mich selbst von 2013.

Teil 1

Einführung

Töne und Klänge

Demonstration von TTS

Pause

Teil 2

TTS im Webinterface

TTS in der API

TTS in Spektrogrammen



Dann kam Corona...

Und Corona änderte alles, auch die stärksten Versprechen.
Frühjahr 2020.

Vorlesungsexperimente in der Physik

Teil 1

Einführung

Töne und Klänge

Demonstration von TTS

Pause

Teil 2

TTS im Webinterface

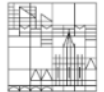
TTS in der API

TTS in Spektrogrammen

Während der Pandemie

- rund 200 Videos
- CC-BY-SA 4.0 Lizenz
- Repositorien: Cloud der Universität Konstanz, ZOERR

Universität
Konstanz



Creative Common Lizenz

Dr. Gillian Kiliani
Fachbereich Physik

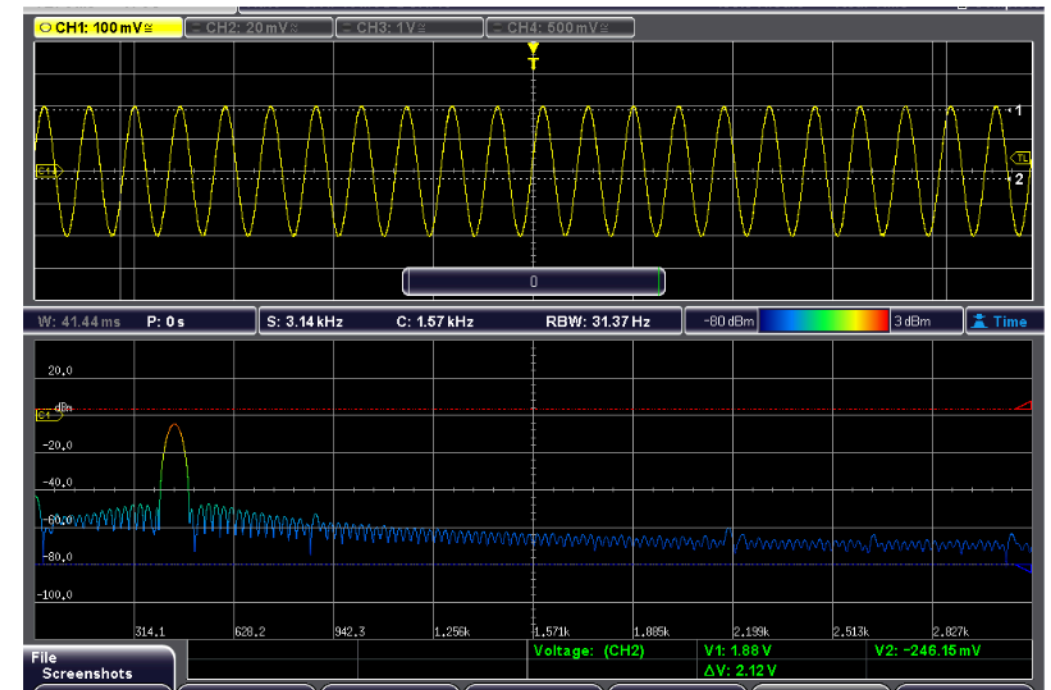
Dieses Video ist unter der Lizenz CC BY SA 4.0 verfügbar.

Das gezeigte Experiment erfordert spezielle Fachkenntnisse und darf nur von dafür ausgebildeten fachkundigen Personen durchgeführt werden. Von eigenmächtigem Nachbau und eigenmächtiger Durchführung wird abgeraten.

Was ist ein Ton?

physikalisch / mathematisch

- Sinus-förmiger Verlauf des Luftdrucks als Funktion der Zeit
- eine einzige Frequenz, z.B. Stimmgabel oder Metallophon



Ton

oben: zeitlicher Verlauf
unten: einzelne Frequenz

Teil 1

Einführung
Töne und Klänge
Demonstration von TTS

Pause

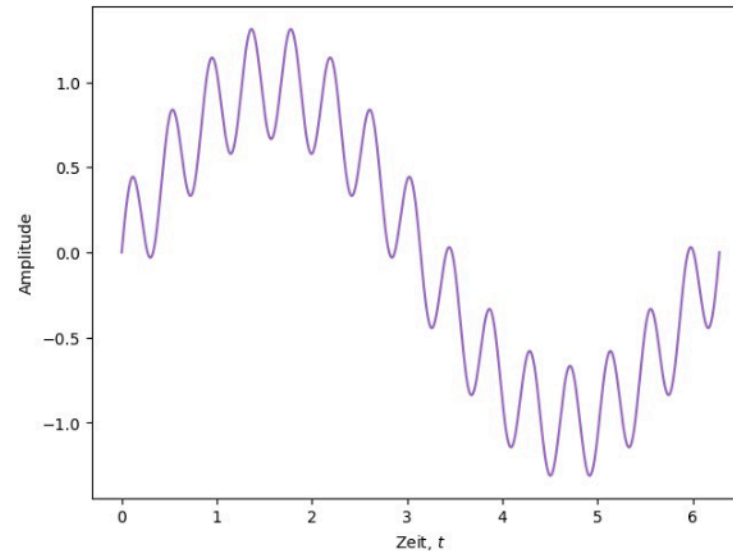
Teil 2

TTS im Webinterface
TTS in der API
TTS in Spektrogrammen

Was ist ein Klang?

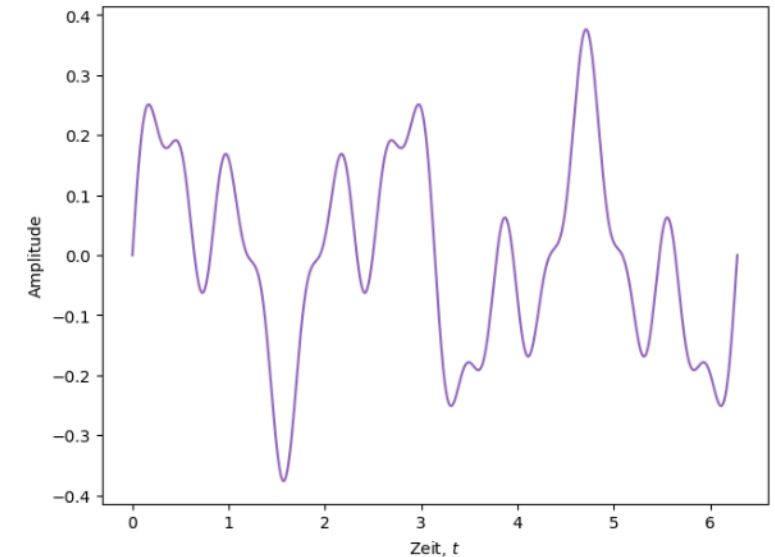
physikalisch / mathematisch

- mehr als eine Frequenz



einfacher Klang

Beitrag von zwei Frequenzen



komplexerer Klang

Beitrag von drei Frequenzen

Teil 1

Einführung

Töne und Klänge

Demonstration von TTS

Pause

Teil 2

TTS im Webinterface

TTS in der API

TTS in Spektrogrammen

TTS mit generativer KI

Teil 1

Einführung
Töne und Klänge
Demonstration von TTS

Pause

Teil 2

TTS im Webinterface
TTS in der API
TTS in Spektrogrammen

ohne generative KI

- roboterhafte Stimme
- regelbasiert

mit generativer KI

- Erzeugung von Wellenformen abhängig vom Kontext (Emotionen, Betonung)
- Echtzeit-TTS (Chatbot)
- Klonen von Stimmen

Was ist eine API?

Application Programming Interface

- Programmierschnittstelle / Anwendungsschnittstelle
- Programmierung von Anwendungen
- Zugriff auf (KI-)Dienste direkt vom eigenen Rechner

```
from pathlib import Path
from google.cloud import texttospeech
import os
import logging

logging.basicConfig(level=logging.INFO, format='%(asctime)s - %(levelname)s - %(message)s')

def text_to_speech(text: str, outfile: str) -> str:
    """
    Args:

    text: text to synthesize
    outfile: filename to use to store synthetic audio

    Returns:
    String of synthesized audio
    """

    # Instantiates a client
    client = texttospeech.TextToSpeechClient(
        client_options={"api_key": os.getenv("GOOGLE_API_KEY")}
    )
```

Teil 1

Einführung
Töne und Klänge
Demonstration von TTS

Pause

Teil 2

TTS im Webinterface
TTS in der API
TTS in Spektrogrammen

Teil 1

Einführung
Töne und Klänge
Demonstration von TTS

Pause

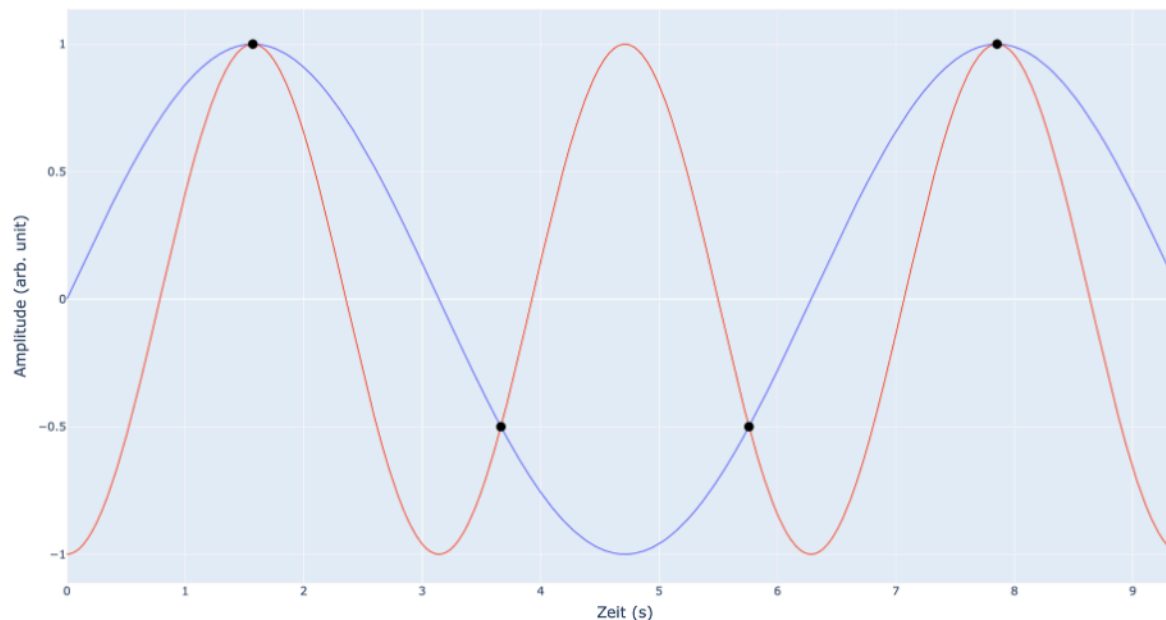
Teil 2

TTS im Webinterface
TTS in der API
TTS in Spektrogrammen

Samplingraten

höher = besser (bis zu einem gewissen Maß)

- maximale Tonhöhe = Hälfte der Samplingrate
- menschliches Gehör max: 20 kHz
typisch: 15 kHz
- CD-Qualität Sampling: 44,1 kHz



Software

Frei/kostenlos

- Audacity (Win, MacOS, Linus)
<https://www.audacityteam.org/download/>
- phyphox (Android, iOS, iPadOS)
<https://phyphox.org/de/home-de/>
- DaVinci Resolve (Win, MacOS, Linus)
<https://www.blackmagicdesign.com/de/products/davinciresolve>
- Python
<https://www.python.org>

Günstig

- SpectrumView (iOS, iPadOS)
<https://oxfordwaveresearch.com/products/spectrumviewapp/>

Professionell

- Adobe Audition
- Adobe Premiere Pro
<https://www.adobe.com/de/>

Herzlichen Dank!

- Open Science Team, Universität Konstanz
- ZOERR Support Universität Tübingen, Michael Menzel
- Stifterverband, bwDigiFellows II
- Kultusministerium Baden-Württemberg
- HND-BW, Matthias Bandtel