VON DER STIMME ZUM TEXT

Spracherkennung in der Österreichischen Mediathek

technisches
museum wien
mit österreichischer mediathek

ERFAHRUNGEN 2008

- Erster Einsatz einer automatisierten Spracherkennung in der Österreichischen Mediathek 2008
- Fehlerrate – zu hoch, z.T. auch für grobe inhaltliche Erschließung ungeeignet
- Kosten – für eine Erschließung des Gesamtbestandes zu hoch
- Verarbeitungsgeschwindigkeit und Usability – für Regelworkflow ungeeignet

FAZIT 2008

- Grundsätzlich sinnvoller Ansatz
- Weitere Beobachtung der Entwicklung
- Es bedarf aber noch weiterer Forschung, verbesserter Programmierung und grundlegender Fehleroptimierung

ERFAHRUNGEN 2021

- Zweiter Einsatz einer automatisierten Spracherkennung in der Österreichischen Mediathek 2021
- Fehlerrate – akzeptabel, aber im Bereich Umgangssprache zu hoch
- Kosten – für eine Erschließung des Gesamtbestandes zu hoch
- Verarbeitungsgeschwindigkeit und Usability – akzeptabel

FAZIT 2021

- Für Teile des Bestandes sehr brauchbar
- Aufgrund der Kosten für einen umfassenden Einsatz nicht verwendbar

AB 2023 WHISPER(X)

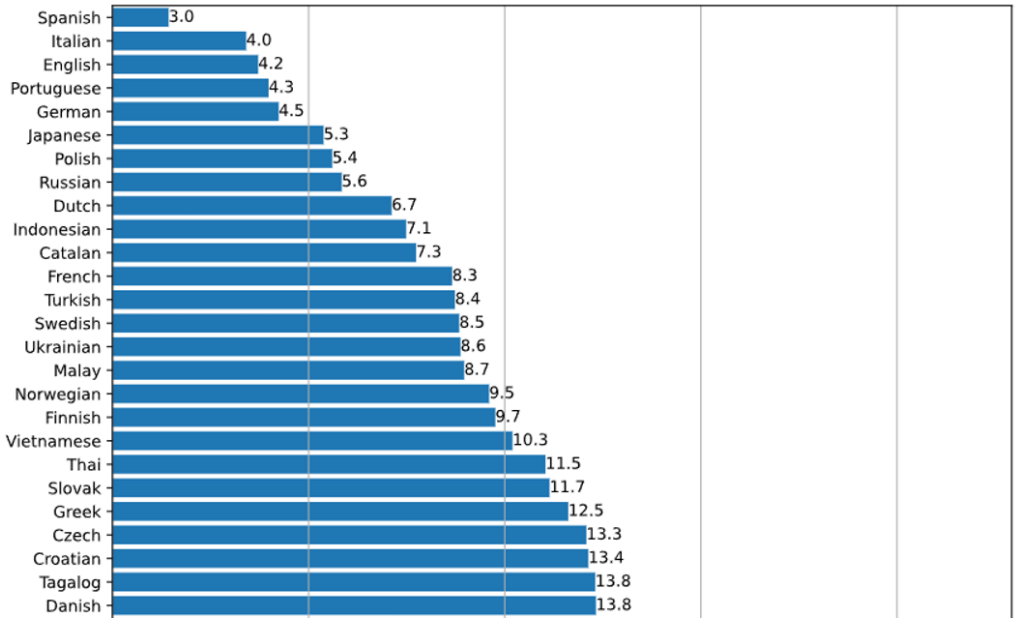
- Dritter Einsatz einer automatisierten Spracherkennung in der Österreichischen Mediathek ab 2023 mit Whisper und WhisperX
- Fehlerrate – niedrig
- Kosten – gering da *open weight* Model und *open source* Software
- Verarbeitungsgeschwindigkeit und Usability – gut

OPENAI UND WHISPER

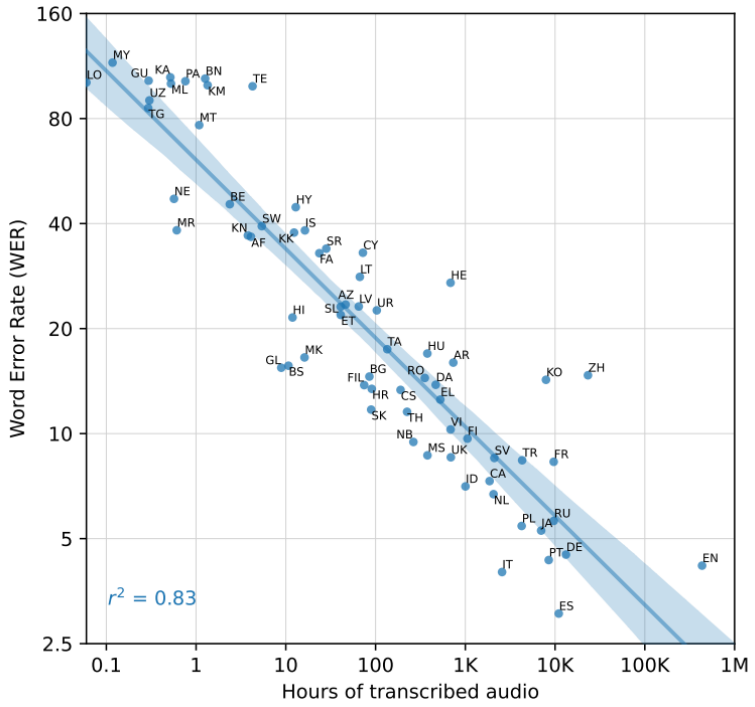
- OpenAI gegründet 2015
- Whisper: neuronales Netz, schwach überwachtes, tief lernendes Akustikmodel für Spracherkennung ("weakly-supervised deep learning acoustic model")
- Whisper wurde mit 680 tausend Stunden trainiert (nur 117 tausend Stunden für die restlichen 96 Sprachen!)
- Wichtig: recht freigiebige MIT Lizenz!
- Nach wie vor bestes *open weight* ASR-Modell (*automatic speech recognition* = automatische Spracherkennung)

ERGEBNISSE

Word Error Rate (WER)



WER und transkribierte Tonaufnahmen pro Sprache

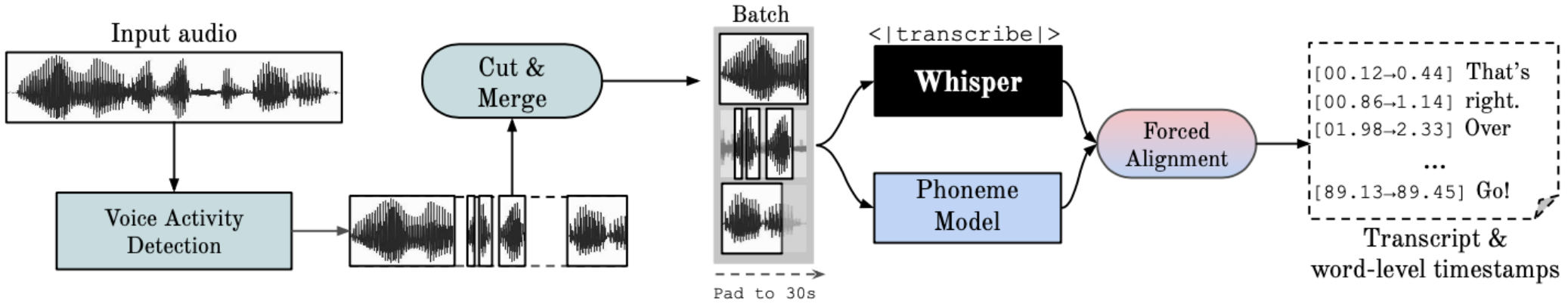


technisches museum wien
mit österreichischer mediathek

Quelle: github.com/openai/whisper

WAS IST WHISPERX?

- Projekt WhisperX baut auf Whisper auf
- WhisperX erstellt von der Visual Geometry Group an der Oxford University, ursprünglich um genauere Zeitstempel zu erhalten
- Umfangreiche Aufbereitung der Tonaufnahme vor und nach der Transkription:



Quelle: github.com/m-bain/whisperX

VORTEILE WHISPERX

- Akkurate Zeitstempel
- Kein Zeitdrift bei längeren Aufnahmen (Untertitel und Aufnahme gehen auseinander)
- Etwas bessere Ergebnisqualität als Whisper oder Faster Whisper
- Halluzinationen sind auf 30-Sekunden-Segmente beschränkt
- 25–30-fache Echtzeitgeschwindigkeit (mit 12 GB Nvidia Grafikkarte)

UNTERSCHIEDE WHISPER UND WHISPERX NR. 1

- Beispiel 1: Whisper "glättet" Sätze. Ein recht häufiges Phänomen ist das Auslassen von Füllwörtern bei Whisper:

  "Ich habe diese Angst, die er hatte, so dermaßen übernommen.

  Ich habe alles übernommen, was er durchlebt hat in der Zeit.

  Ich habe seine Nervosität auch geerbt.

  Das ist eines der prägendsten."

- WhisperX bleibt genauer beim Text:

  "Und ich habe diese Angst, die er hatte, ich habe das so dermaßen übernommen.

  Also ich habe alles übernommen, was er durchlebt hat in der Zeit.

  Ich habe seine Nervosität auch geerbt.

  Ja, das ist so eines der prägendsten."

UNTERSCHIEDE WHISPER UND WHISPERX NR. 2

- Bei Whisper lässt sich etwas häufiger das Phänomen der erfundenen Transkriptstellen (Halluzination) feststellen:

  Ja.

  Ich habe sie immer gut akzeptiert.

  Ich habe sie immer gut akzeptiert.

  Ich habe sie immer gut akzeptiert.

  Ich habe sie immer gut akzeptiert.

  Ja, schon. Die bis heute halten.

  Mit einer musiziere ich regelmäßig, mit der Sturmtradl.

- Bei WhisperX seltener:

  Ja, schon, die bis heute halten.

  Mit einem musiziere ich regelmäßig, mit der Sturmtradl.

ERFAHRUNG MIT WHISPERX

- Prinzipiell sehr gute Qualität
- Umgangsprache wird gut erkannt, wenn auch ins Hochdeutsche übertragen
- Starker Dialekt funktioniert nicht
- Elipsen und Wiederholungen können zu “Frankenstein-Sätzen” führen
- Spracherkennung schwierig:
  - wenn innerhalb der ersten 30 Sekunden nicht gesprochen wird
  - oder eine andere Sprache als die “Hauptsprache” gesprochen wird
- Eigennamen und Austriazismen tendenziell problematisch

LIVE DEMONSTRATION

technisches
museum wien
mit österreichischer mediathek

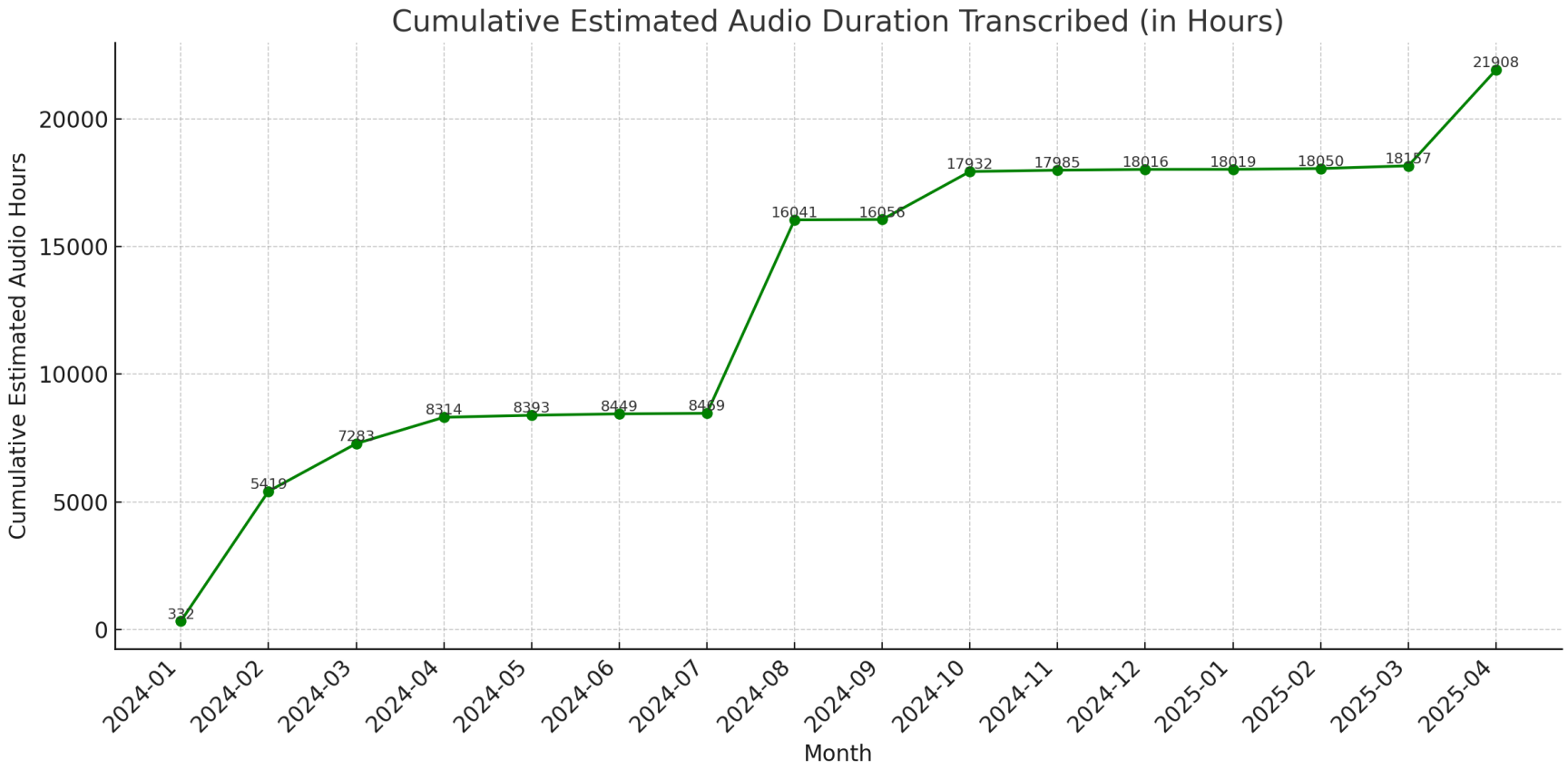
FAZIT ZU WHISPERX FUNKTIONIERT GUT, ABER NICHT FÜR ALLES

- Gut geeignet für:
  - Überblick über große Transkriptmengen
  - Volltextsuche & inhaltliche Erschließung
  - (Teilweise) als Grundlage für Beschlagwortung
  - Big-Data-Analysen *(mit Einschränkungen)*
- Derzeit ungeeignet für:
  - Phonetische, prosodische und linguistische Analyse
  - Emotions- und Affektforschung
  - Dialektforschung
- Vorsicht bei:
  - Analyse ohne Rückbezug zur Originalaufnahme
  - Interpretation einzelner Textstellen:
    - Könnten aus Modellartefakten statt Sprecherintention resultieren

FAZIT ZU SPRACHERKENNUNG ALLGEMEIN

- Transkripte sind keine Primärquellen:
  - Sie sind inhärent fehlerhafte Repräsentationen – keine neutralen Abbilder
  - Zugriff auf Originalaufnahme soll jederzeit niederschwellig möglich sein
- Notwendige Reflexion und Transparenz im Umgang mit automatisch generierten Texten
  - Forscher:innen sollten sich der Potenziale und Grenzen maschinell erzeugter Inhalte bewusst sein
  - Für Nutzer:innen muss klar erkennbar sein, wenn Texte maschinell erstellt wurden
- Manuelle Korrektur nur in gut argumentierten Ausnahmefällen (besseres Modell könnte Arbeit obsolet machen)

STATISTIKEN



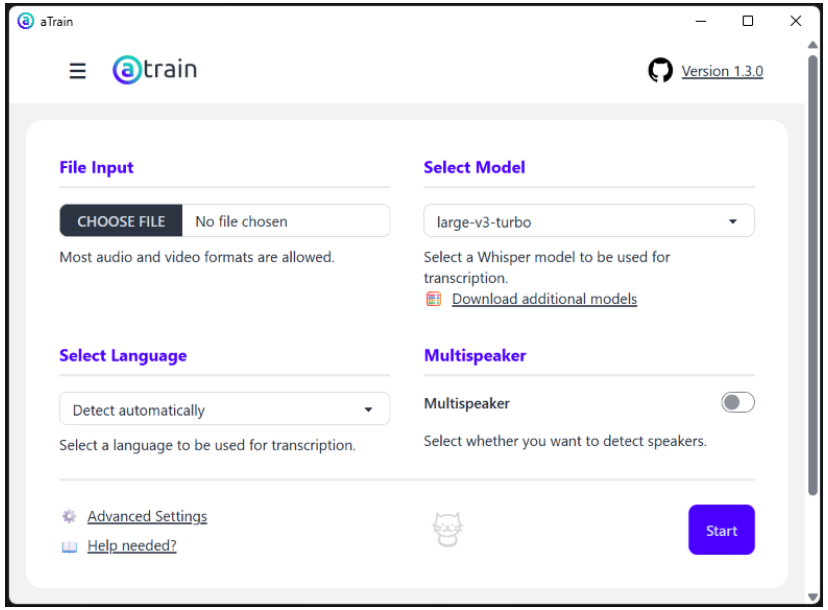
Total: **876,33** Stunden Transkriptionsdauer, ~ **21.908,22** Stunden Aufnahmen

Oder ca. 60 Jahre Vollzeitarbeit

technisches museum wien
mit österreichischer mediathek

WIE KANN ICH WHISPERX SELBST NUTZEN?

- WhisperX eher schwierig für IT-Laien (Kommandozeile muss bedient werden)
- Whisper ohne X ist jedoch recht einfach, ein paar Empfehlungen:
- „aTrain“ von der Uni Graz (Windows, Debian Linux, MacOS), mit Sprechererkennung



technisches museum wien
mit österreichischer mediathek

WEITERFÜHRENDE LINKS UND INFORMATIONEN

- Mittagsjournale mit Transkripten von WhisperX von 1967 bis 1999: https://www.mediathek.at
- Website Whisper: https://github.com/openai/whisper
- Website WhisperX: https://github.com/m-bain/whisperX
- Website aTrain: https://business-analytics.uni-graz.at/en/research/atrain
- Website Buzz: https://github.com/chidiwilliams/buzz
- Kontakt: martin.fellner@tmw.at
- Fragen?