Technology
Arts Sciences
TH Köln

TRANSCRIPTION OF SCIENTIFIC PODCASTS USING AUTOMATIC SPEECH RECOGNITION SYSTEMS

Bachelor Thesis

Robin Hilbrecht

23. Oktober 2023

Content

- Introduction
- Hypothesis and Goal
- Methods
- Results
- Limitations
- Future Research

Introduction

- *New Media* greatly shaped the digital information and communication landscape in the past two decades, also affecting science communication
- Recent Events such as the COVID-19 pandemic further accelerated this trend
- Podcasts have emerged simultaneously as a unique and diverse medium in modern media landscape (exponential growth since 2010 in science podcasts)
- However, auditory nature and the need for evidence-based communication pose challenges for science podcasts

Hypothesis and Goal

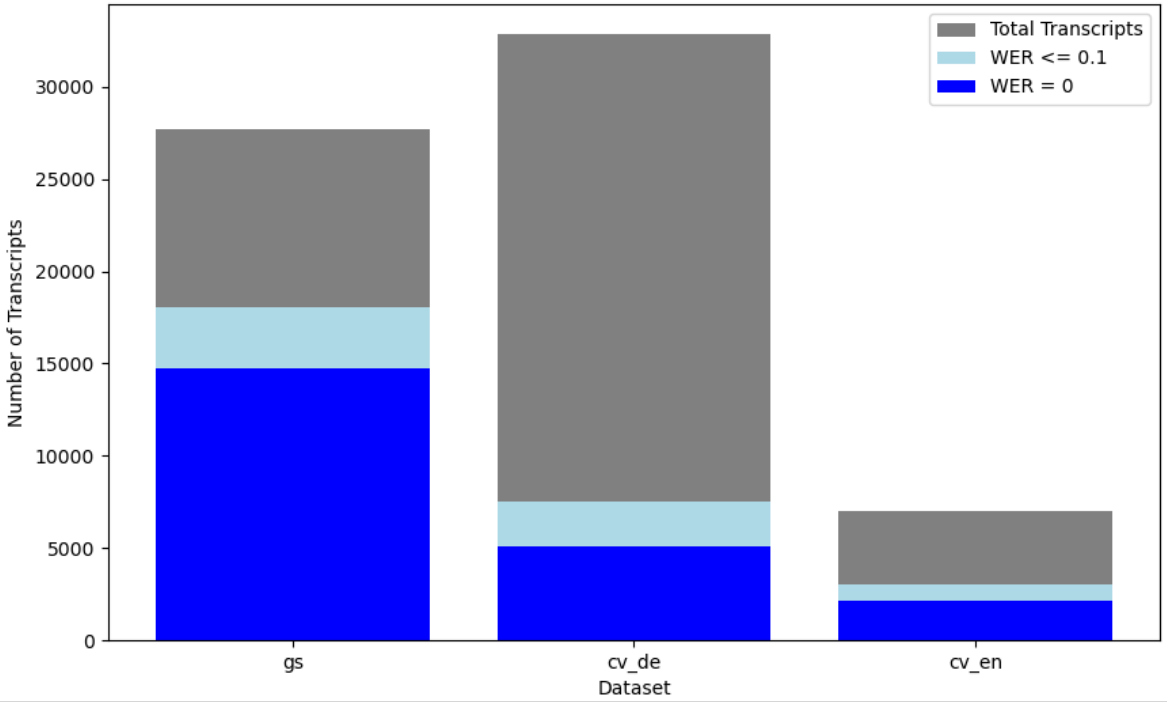
“The integration of AI-powered transcription methods into scientific podcasts not only enables efficient and high-quality transcriptions but also opens new possibilities to enhance discoverability, accessibility, and dissemination of scientific content.”

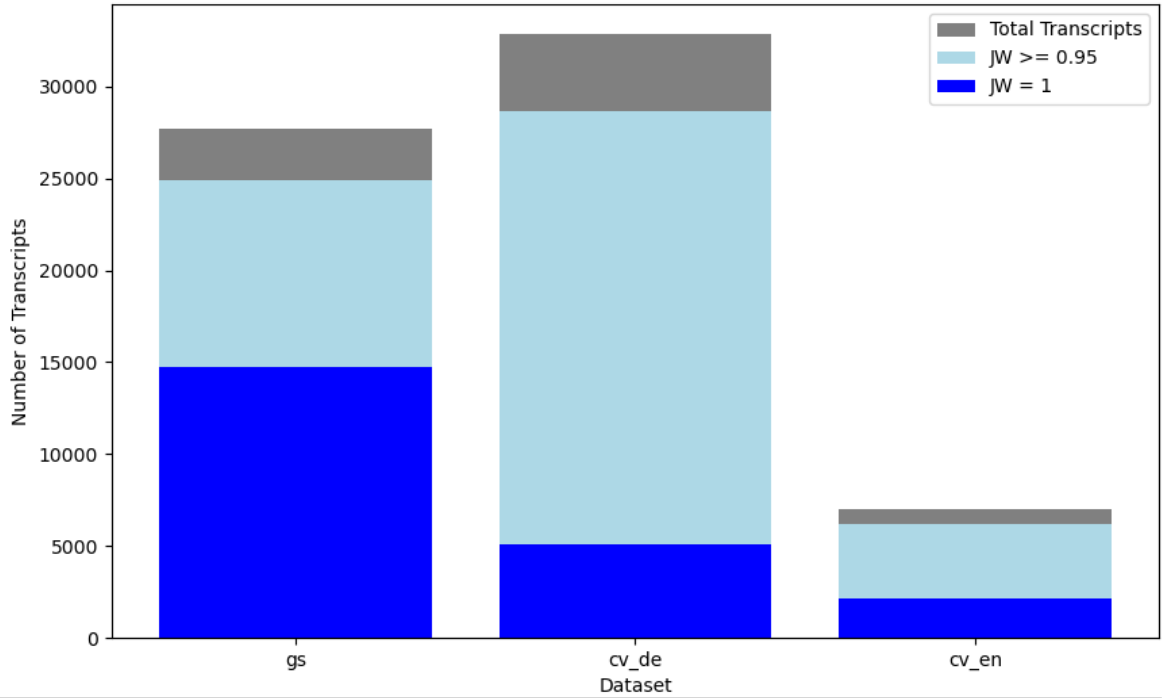
Goal: Propose an accessible and reuseable method to create high-quality transcripts for podcasts and elaborate on how transcripts enable content-based search, ultimately improving visibility and accessibility of podcast content

Methods

- Testing on Speech Transfromer *Whisper* from OpenAI
  - ASR system trained on 680,000 hours of multilingual and multitask supervised data collected from the web
  - open-source
  - Performance close to that of a professional human transcriber (average WER around 8.83%)
- Used parts of two Datasets for evaluation in german and english: Common Voice and GigaSpeech
- Metrics: WER and Jaro-Winkler-Similarity

Results

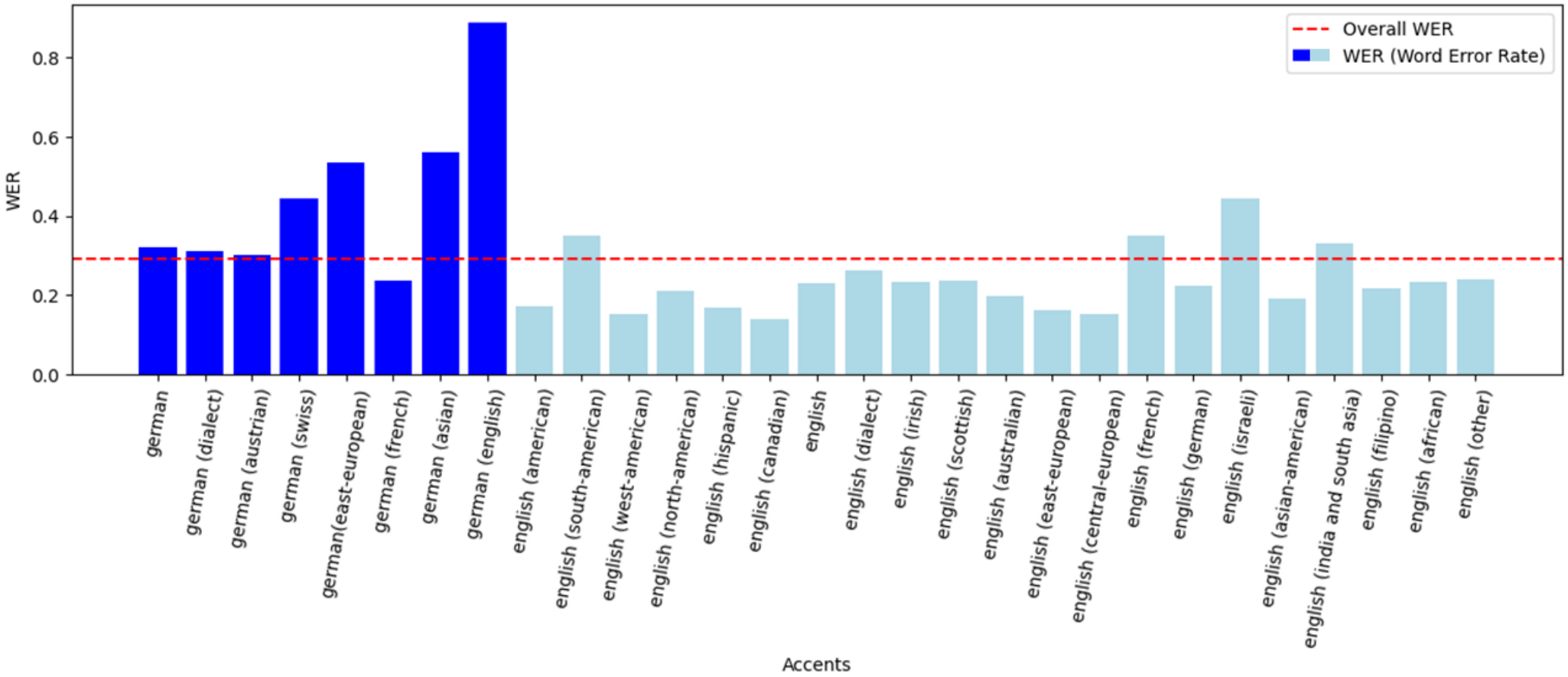




dataset	total amount	WER=0	WER<=0.1	sim_w=1	sim_w>=0.95	average WER
CV-DE	32852	15,5%	22,9%	15,5%	87,4%	31%
CV-EN	6983	30,8%	43,6%	30,8%	88,8%	19%
GS	27700	53,2%	65,1%	53,2%	89,8%	11%

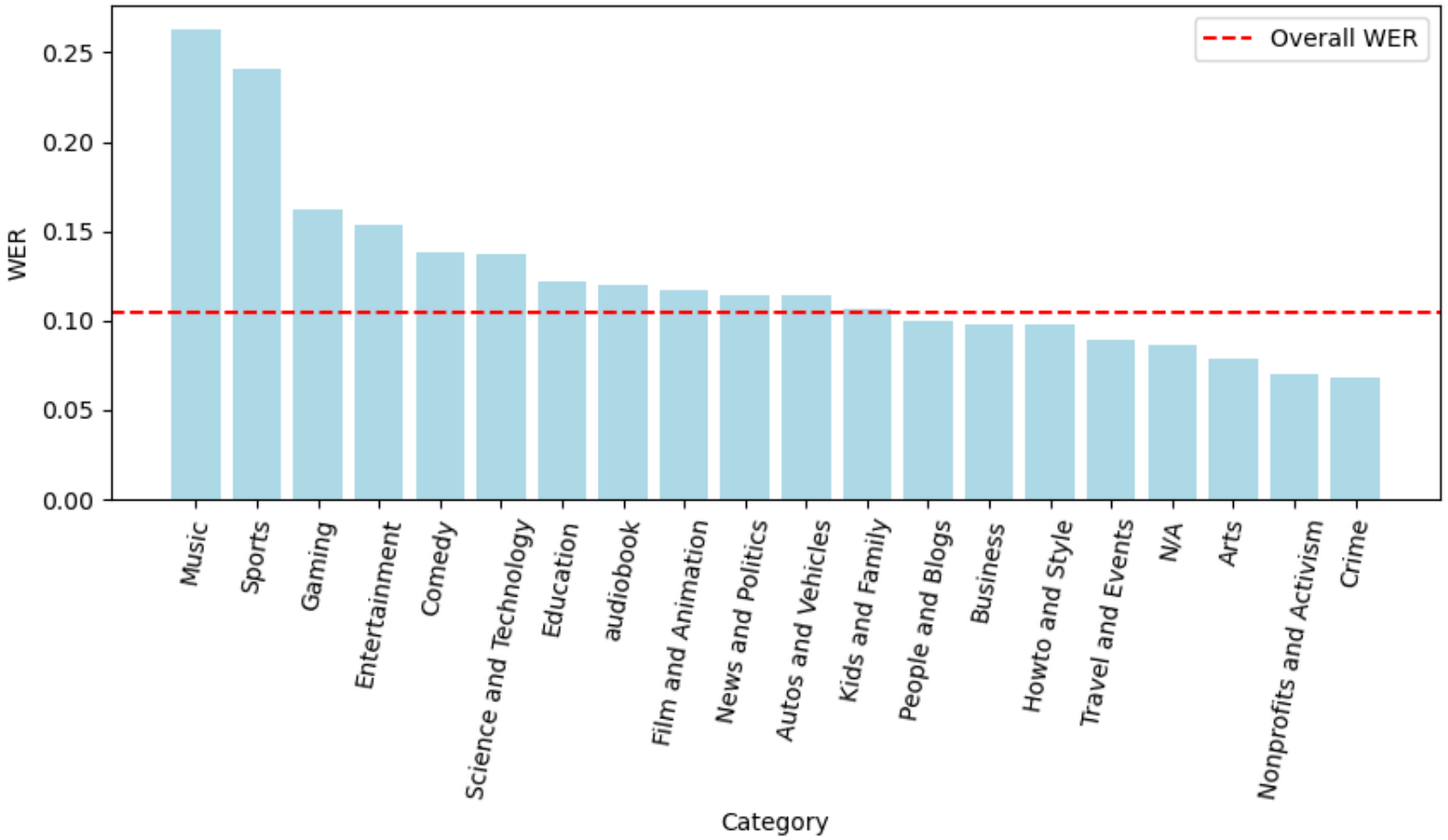
Results

Common Voice - Accents



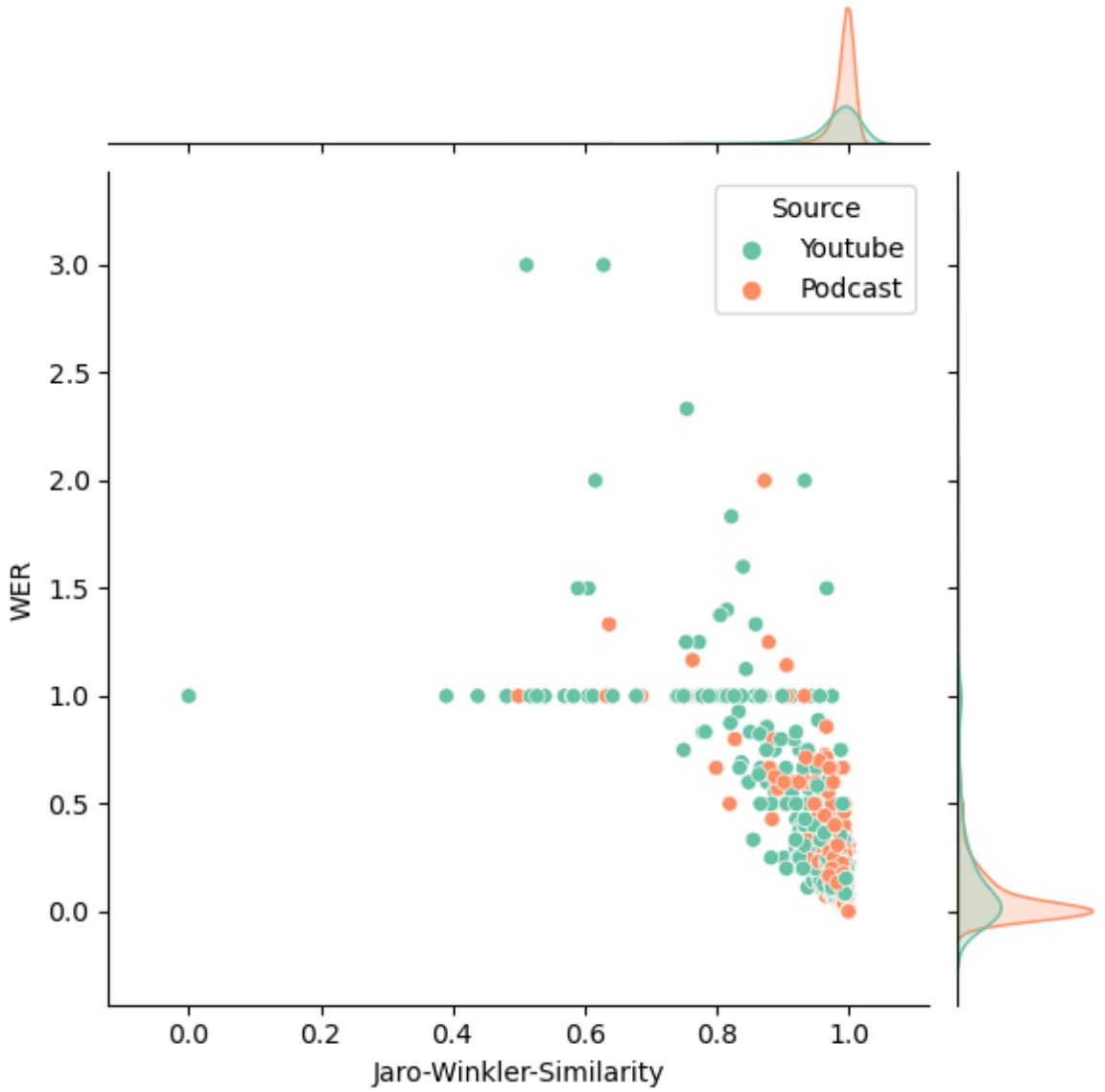
Results

GigaSpeech- Categories

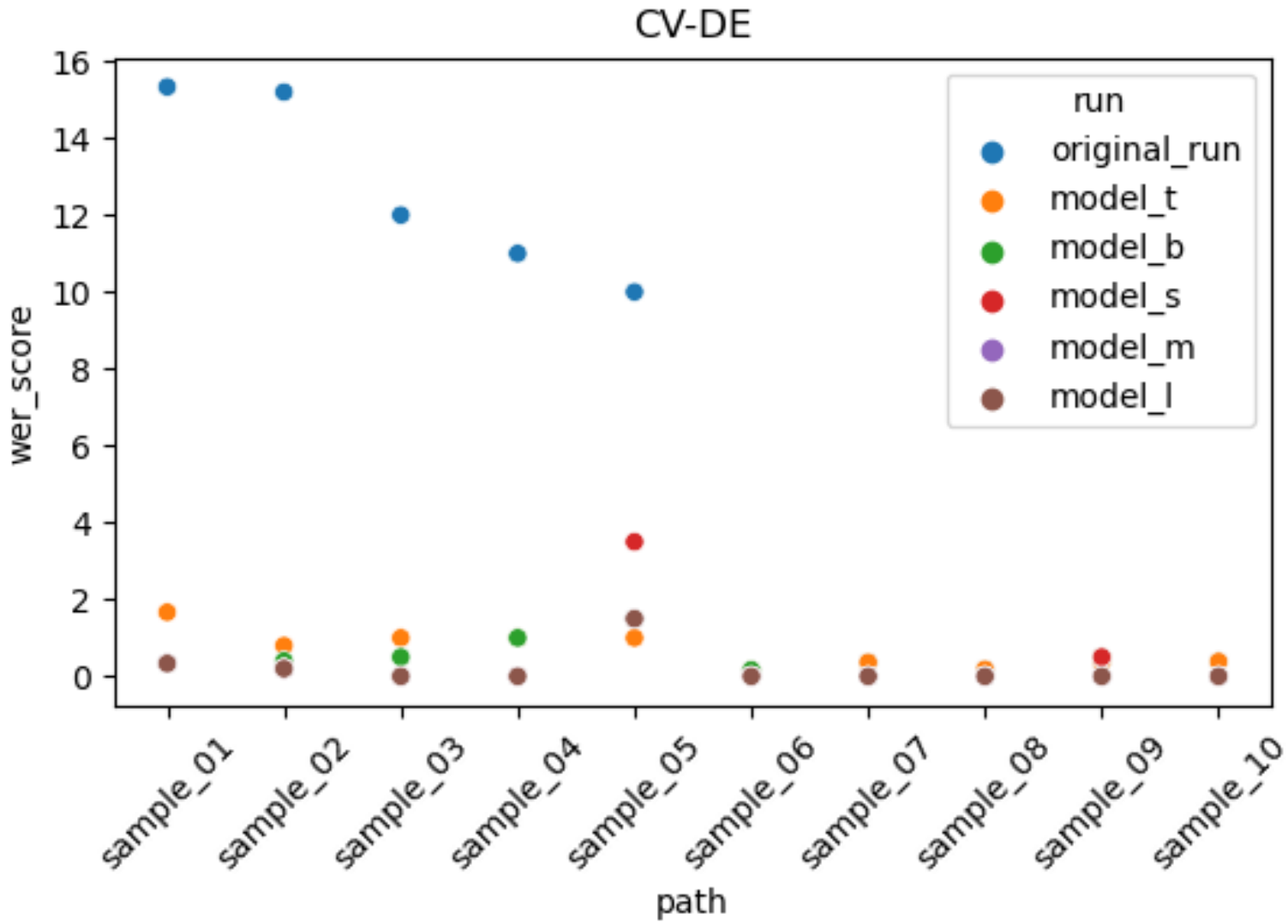


Results

Source	total	Avg. WER	WER =0	Sim_w =1	Sim_w >=0.95
Youtube	954	14%	48,2%	48,2%	83,5%
Podcast	1425	8%	55,5%	55,5%	96,5%



GigaSpeech – Category Science and Technology


CV-DE
wer_score
path
run
original_run
model_t
model_b
model_s
model_m
model_l
0
2
4
6
8
10
12
14
16
sample_01
sample_02
sample_03
sample_04
sample_05
sample_06
sample_07
sample_08
sample_09
sample_10

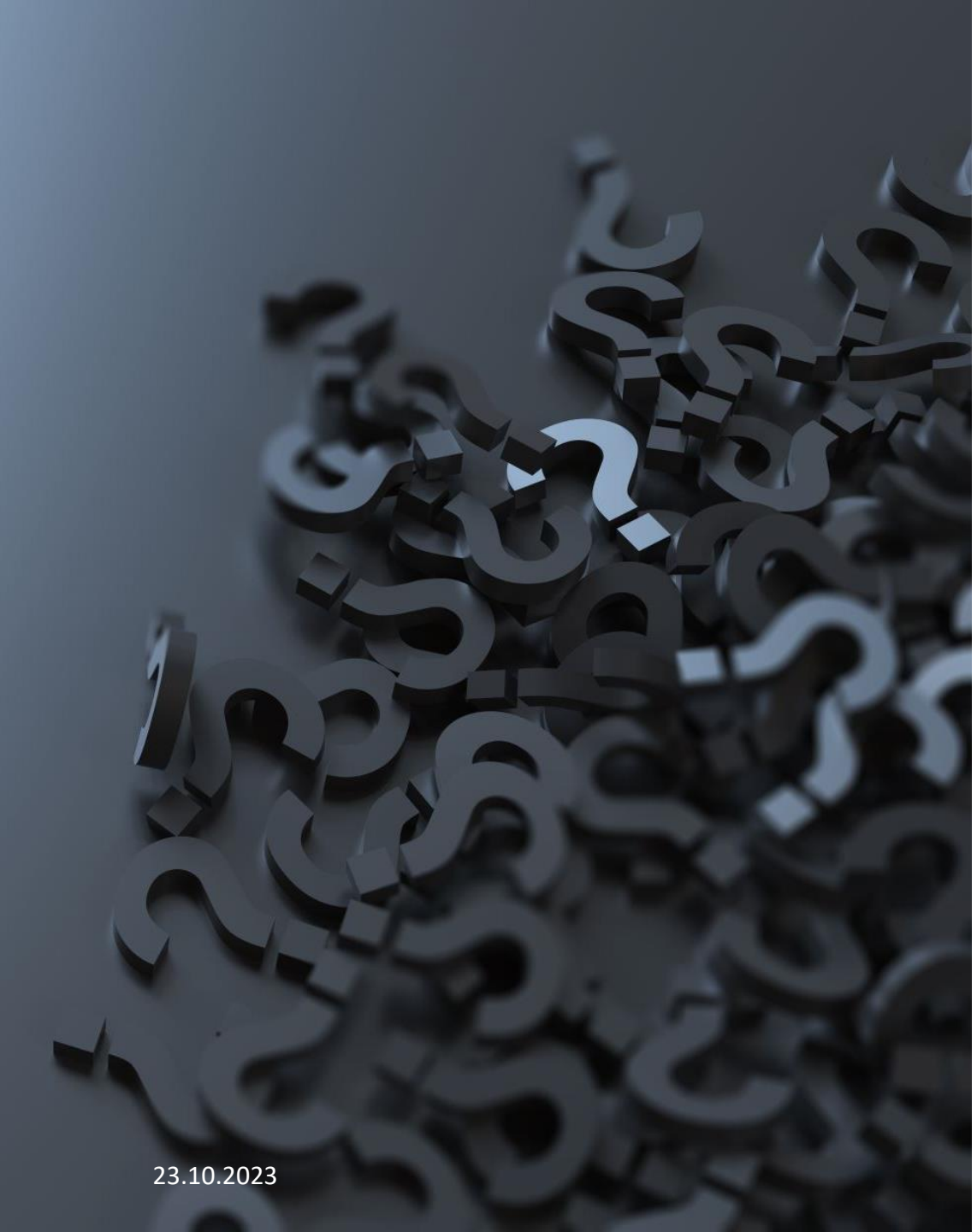
Results

Results

- WER varies depending on the language and dataset used. The results indicate that English data were transcribed more accurately than German data
- The study confirmed that accents and dialects tend to make transcription more challenging
- Despite a high WER, the transcripts displayed a high similarity to the ground truth, suggesting that most errors are minor
- Specifically, in the "Science and Technology" category, podcasts performed well with a high similarity to the ground truth and a low WER of 7.8%.
- A direct model comparison revealed that the WER decreases as the model size increases, indicating that the ASR model's performance can be enhanced with more resources and training data
- *Overall, Whisper seems to be well-suited for the transcription of science podcasts*

Limitations

- Limited Computing Resources
- Irregularly distributed data
- Limited Data Fields
- Need for Future Research



Future Research

- Transcripts serve as a foundation for various NLP methods, enabling the extraction of references, entities, keywords, text categorizations/summarization and segmentation
- Retrieved entities could then be used to enrich metadata an enhance podcast representation
- Entities can also be linked within a knowledge graph, a strategy that has been demonstrated to be effective for news recommendations in newsgraphs
- Fine-tuning a model could reduce WER for specific accents and specialized terminology
- While preserving the meaning of spoken words, it is also crucial to ensure that technical terms are correctly recognized and identified as such

You can access the data, code, and thesis, along with additional sources, on GitHub by following this link:

https://github.com/DrBilboArriba/ASR_and_science_podcasts

Thank you for your attention!

If you have any further questions, please feel free to ask them