SDS-200: A Swiss German Speech to Standard German Text Corpus

Michel Plüss, University of Applied Sciences and Arts Northwestern Switzerland, Windisch

LREC 2022

Full author list:
Michel Plüss, Manuela Hürlimann, Marc Cuny, Alla Stöckli, Nikolaos Kapotis, Julia Hartmann, Malgorzata Anna Ulasik, Christian Scheller, Yanick Schraner, Amit Jain, Jan Deriu, Mark Cieliebak, Manfred Vogel

Swiss German

- Family of German dialects
- 5 million speakers
- Differs from Standard German in phonetics, vocabulary, morphology, syntax
- Mostly spoken
- No standardized writing system

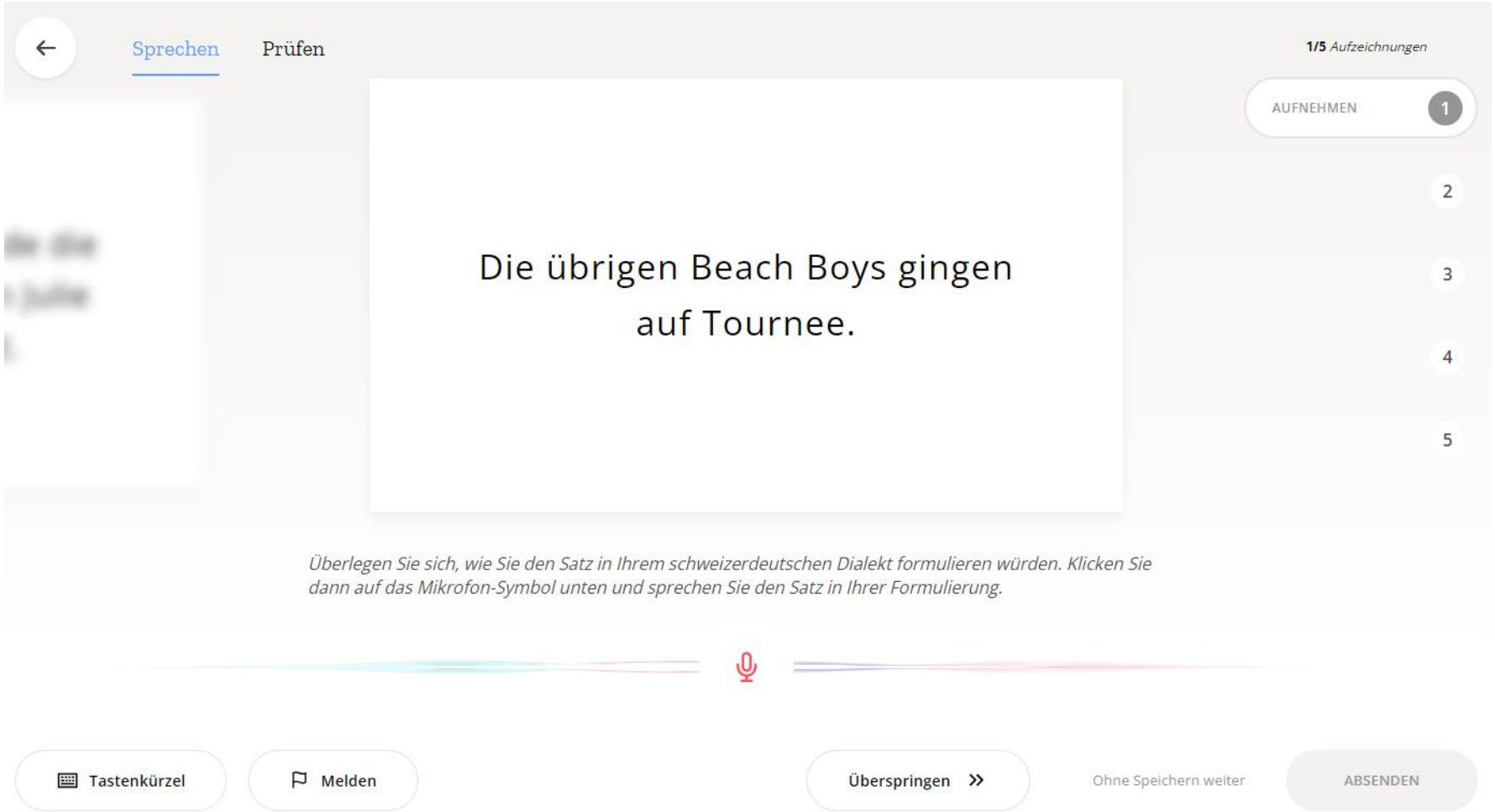
Challenges Swiss German STT

- Many written variants for the same word
- Huge vocabulary size
- Desired language of output: Standard German → Translation
- Diversity of dialects, especially vocabulary and phonetics
- Limited amount of publicly available training data:
  - SwissDial [1], 26 h
  - Radio Rottu [2], 2 h
  - Swiss Parliaments Corpus [3], 293 h (noisy)

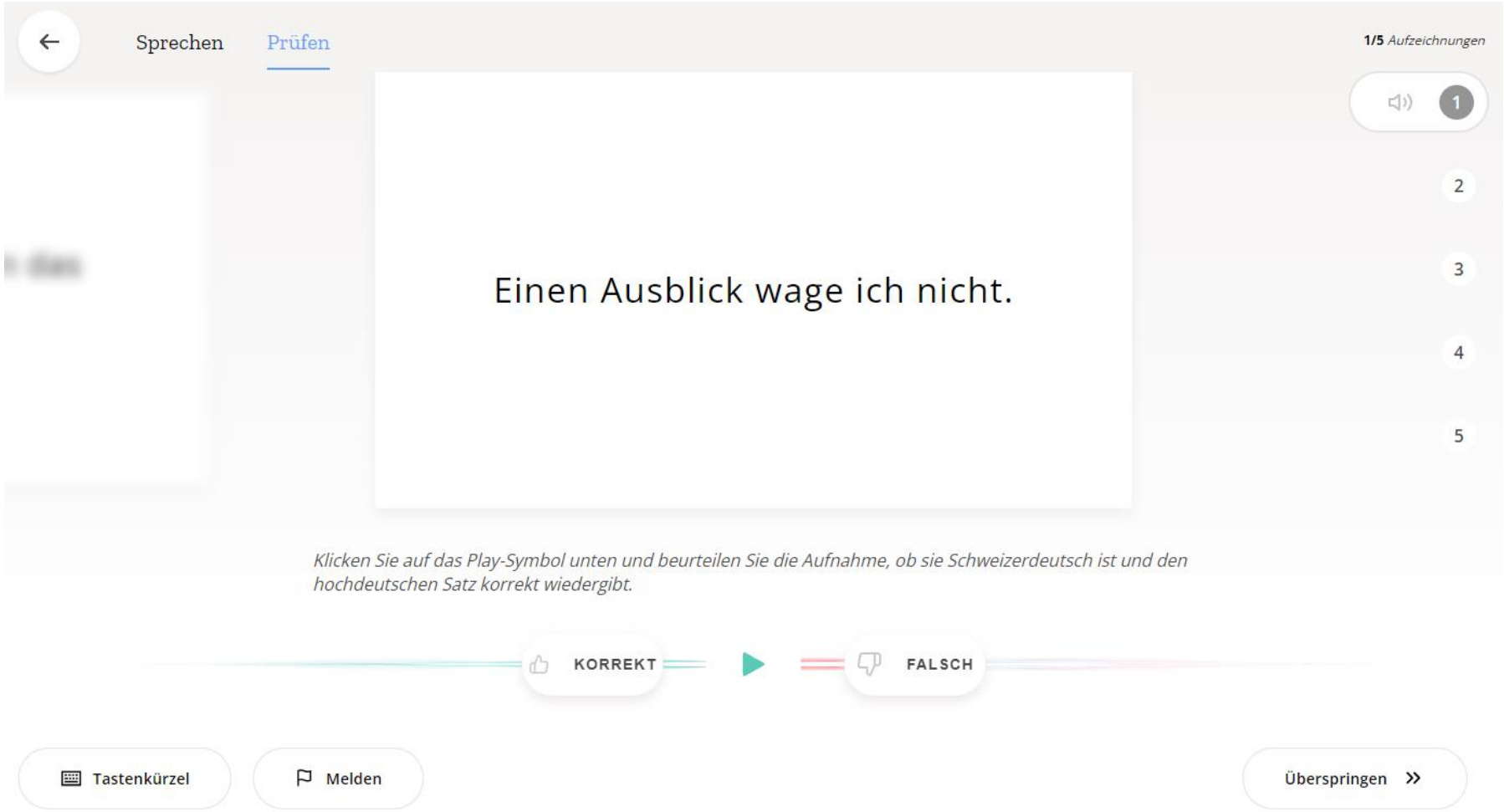
Data collection

- Crowdsourcing
- Web recording tool https://dialektsammlung.ch/
- Based on Common Voice [4]
- Recording:
  - Standard German sentence
  - How would I say this in Swiss German?
  - Record in Swiss German
- Validation:
  - Is the recording in Swiss German?
  - Is it a faithful Swiss German translation of the Standard German sentence?

Data collection - recording



Data collection - validation



Data collection - metadata

- Dialect:
  - Zip code of origin of the dialect
  - E.g. where the participant went to school
  - Allows to group dialects in dialect regions, cantons, municipalities
- Age
- Gender

Data collection - sentences

- Source:
  - 80 % from Swiss newspapers
  - 20 % from the German Common Voice [5]
- 5-12 tokens

Data collection process

- Interviews and reports in Swiss television, radio, newspapers
- Support videos by 4 well-known Swiss comedians, shared on their social media accounts
- Leaderboard contest:
  - Individual contest
  - Score based on number of recordings, validations, quality of recordings
  - Attractive Switzerland-themed prices for the top 10 participants
- Clash of Cantons:
  - Canton contest
  - Spark competition between cantons
  - Score based on number of recordings, weighted by average quality, normalized by population

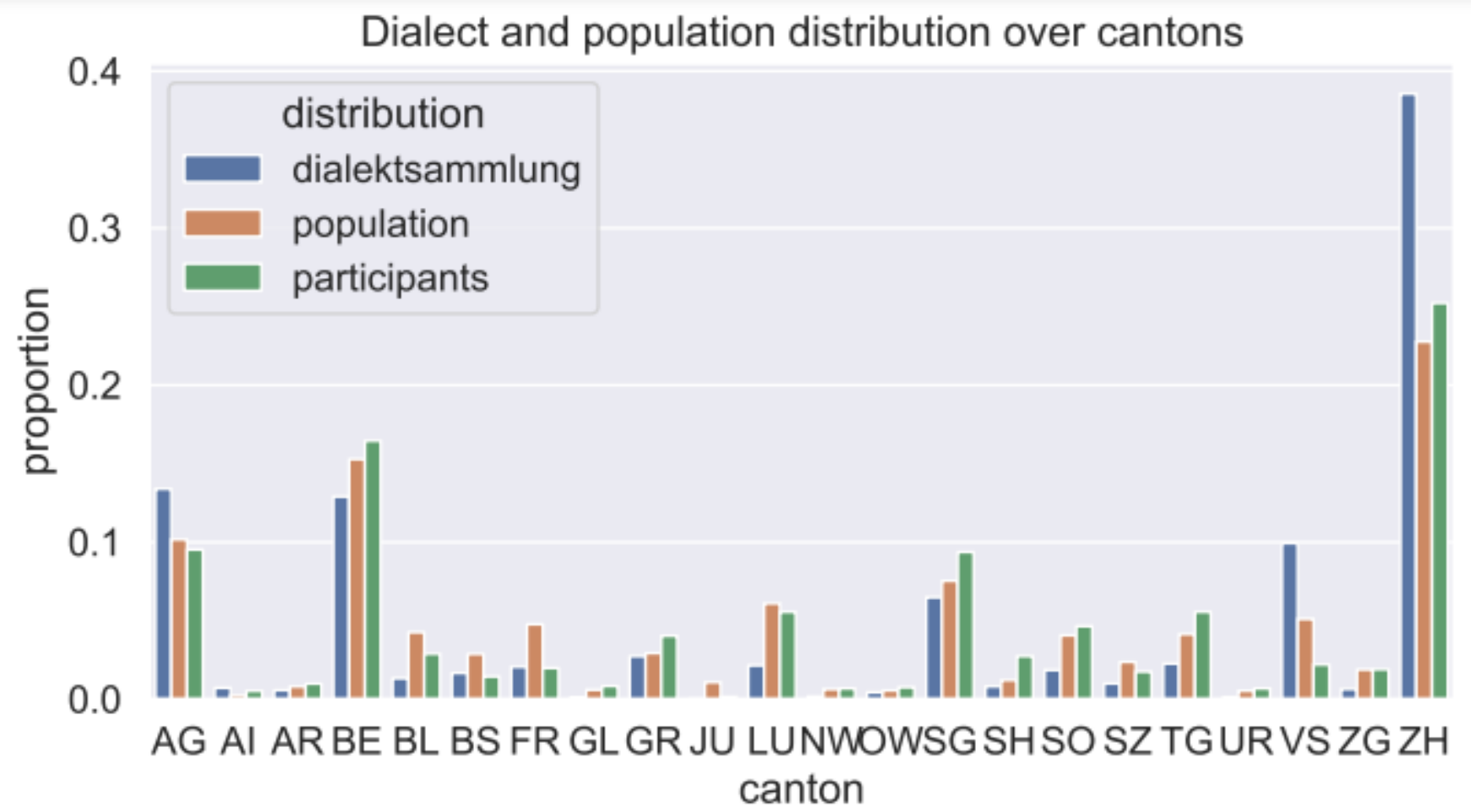
Corpus

Split	Hours
train (raw)	188.9
train (filtered)	178.3
validation	5.2
test	5.4

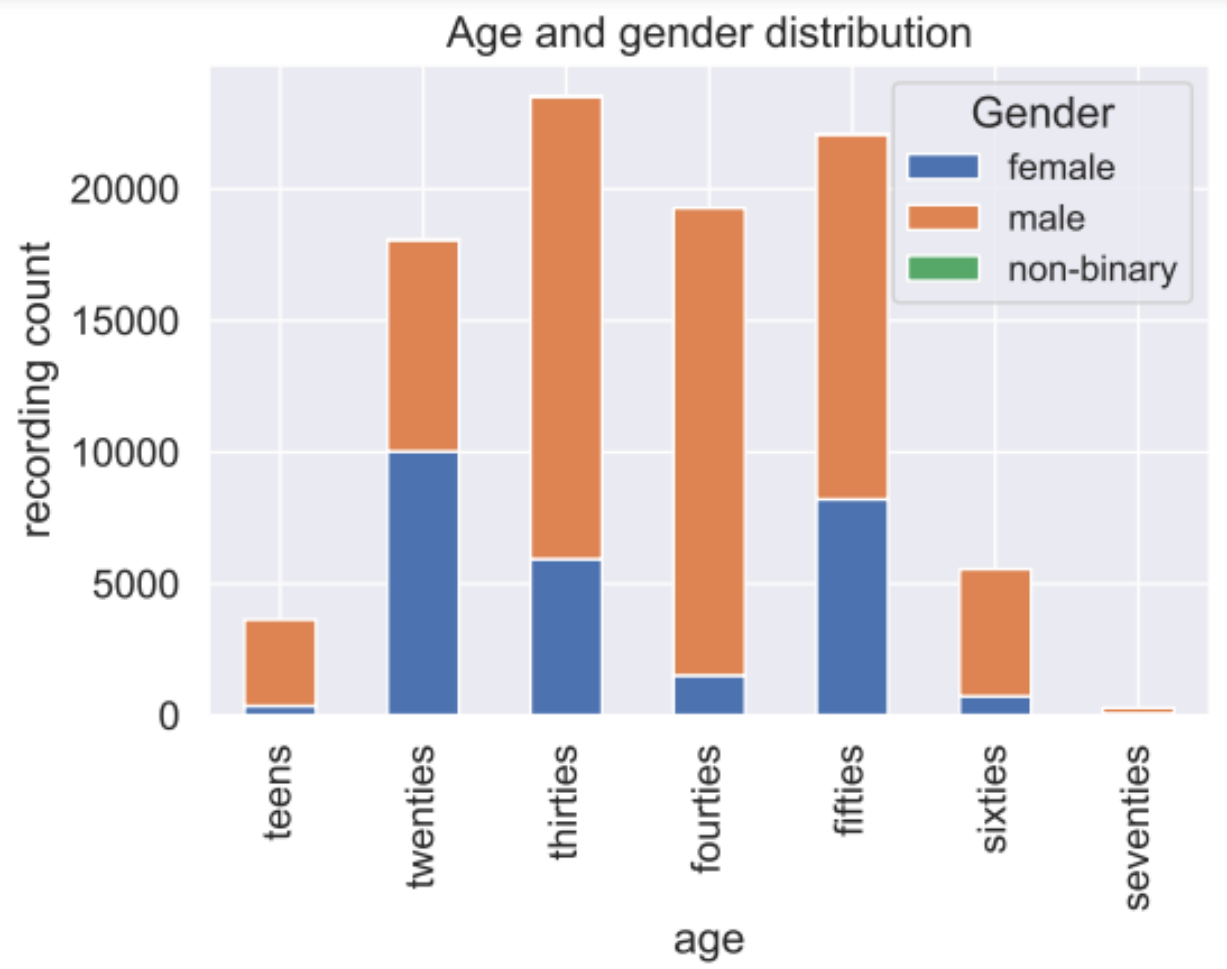
Corpus

- train (filtered), validation, test:
  - Filtered using validations
  - 142’545 recordings
  - 3’816 speakers
  - Recording duration: 4.8 s +/- 1.3 s

Corpus - dialects



Corpus – age and gender



- Speakers:
  - 8 % male
  - 6 % female
  - 86 % unknown
- Recordings:
  - 46 % male
  - 19 % female
  - 35 % unknown

Models

- All models are trained in an end-to-end fashion (Swiss German speech to Standard German text)
- *SDS-200 train (filtered)* is the only labeled corpus used for these models
- WER and BLEU reported on *SDS-200 test*

Models

Model	Params	Train method	WER	BLEU
Transformer	72M	from scratch	30.3	53.1
XLS-R 1B	965M	finetuning	21.6	64.0
XLS-R 1B + LM	965M	finetuning + LM	**17.9**	**70.3**

- XLS-R [6]:
  - wav2vec 2.0 [7] models
  - Self-supervised pre-training of speech representations
  - 436’000 h unlabeled speech data in 128 languages (no Swiss German)
- LM:
  - KenLM [8]
  - 68M Standard German sentences

References

[1] Dogan-Schönberger, P., Mäder, J., and Hofmann, T. (2021). SwissDial: Parallel Multidialectal Corpus of Spoken Swiss German.

[2] Garner, P. N., Imseng, D., and Meyer, T. (2014). Automatic Speech Recognition and Translation of a Swiss German Dialect: Walliserdeutsch. In Proceedings of Interspeech 2014.

[3] Plüss, M., Neukom, L., Scheller, C., and Vogel, M. (2021). Swiss Parliaments Corpus, an Automatically Aligned Swiss German Speech to Standard German Text Corpus. In Proceedings of the Swiss Text Analytics Conference 2021.

[4] Ardila, R., Branson, M., Davis, K., Henretty, M., Kohler, M., Meyer, J., Morais, R., Saunders, L., Tyers, F. M., and Weber, G. (2020). Common Voice: A Massively-Multilingual Speech Corpus. In Proceedings of the 12th Conference on Language Resources and Evaluation (LREC 2020), pages 4211–4215.

[5] https://github.com/common-voice/common-voice/tree/main/server/data/de

[6] Babu, A., Wang, C., Tjandra, A., Lakhotia, K., Xu, Q., Goyal, N., Singh, K., von Platen, P., Saraf, Y., Pino, J., Baevski, A., Conneau, A., and Auli, M. (2021). Xls-r: Self-supervised cross-lingual speech representation learning at scale. arXiv, abs/2111.09296.

[7] Baevski, A., Zhou, Y., Mohamed A., and Auli, M. (2020). wav2vec 2.0: A framework for self-supervised learning of speech representations. In Advances in Neural Information Processing Systems, 2020.

[8] Heafield, K. (2011). KenLM: Faster and Smaller Language Model Queries. In Proceedings of the Sixth Workshop on Statistical Machine Translation.