

SALAN: A Massive ASR Dataset for the Languages of Niger

Mamadou K. Keita¹, Christopher M. Homan¹, Emily Prud'hommeaux², Abdoulaye Sako³, Seydou Diallo⁴

¹Rochester Institute of Technology, ²Boston College, ³ESEO, ⁴DAUST

Abstract

We introduce SALAN, a large-scale speech dataset covering eight of the major indigenous languages of Niger: Zarma, Hausa, Buduma, Gourmantchema, Tubu, Tamasheq, Fulfulde, and Kanuri. The final dataset exceeds 2,000 hours of audio, largely sourced from radio broadcasts and community recordings. We transcribed portions of the audio using the MMS model and conducted manual verification for 110 hours across Zarma and Hausa. We then used active learning to expand annotation to an additional 5 hours of high-uncertainty Zarma segments. To evaluate SALAN's utility for ASR, we fine-tuned both Wav2vec2 XLS-R and Whisper on Zarma subsets and carried out additional pre-training with multilingual unlabeled data. Our best model achieved a word error rate of 25.3% and a character error rate of 6.2%. SALAN and the trained models will be made publicly available for use by researchers and speakers, with the potential to impact over 20 million individuals in Niger and neighboring countries.

1. Introduction

Despite recent advances in automatic speech recognition (ASR), many African languages remain underserved in part due to limited availability of transcribed speech datasets (Olatunji et al., 2023). The languages of Niger are no exception to this problem. Over 20 million people speak languages such as Zarma, Hausa, or Fulfulde daily, yet the scarcity of labeled data slows the development of accurate ASR systems for these and other languages of Niger. As a result, speech technologies that support what many English speakers might consider basic functions of a typical smartphone, such as the ability to dictate text messages, are unavailable to the speakers of these languages.

In this paper, we present SALAN (derived from the Zarma word for *speak*), a new multilingual speech dataset that covers eight of Niger's widely-spoken languages. SALAN contains over 2,000 hours of curated audio, gathered from radio archives, community recordings, and other freely available sources. We perform denoising, segmentation, and transcription on this dataset to ensure a practical resource for the community. We manually verified 110 hours of transcripts for Zarma and Hausa, providing high-quality labeled data for ASR works.

Our main contributions are:

- 1. First massive dataset for Niger languages:** SALAN has over 2,000 hours of audio for eight distinct Niger languages, with 110 hours of manually verified transcripts for Zarma and Hausa.
- 2. Semi-automatic transcription pipeline + active learning:** We use the MMS model (Pratap et al., 2023) for initial recognition and apply human verification to refine 110 hours of transcripts for Zarma and Hausa. We then use an

active learning strategy to identify and annotate an additional 5 hours of high-uncertainty segments.

- 3. State-of-the-art ASR model for Zarma:** We fine-tuned Wav2vec2 XLS-R (Conneau et al., 2020) and Whisper (Radford et al., 2022) on a Zarma subset of SALAN. Our best model achieves Word Error Rate (WER) and Character Error Rate (CER) scores of 25.3% and 6.2%.
- 4. Experiments with leveraging unlabeled data:** We use unlabeled data to improve ASR performance of the wav2vec2 XLS-R model.
- 5. Public release:** We provide open access to our dataset and alignments with specific guidelines for ethical uses.

2. Related Work

Several initiatives address the lack of speech datasets for African languages. One early effort, the ALFFA (African Languages in the Field: speech Fundamentals and Automation) Project, collected several dozen hours of carefully transcribed audio for training ASR models for Swahili, Amharic, and Wolof (Abate et al., 2005; Gelas et al., 2012; Gauthier et al., 2016). (Zanon Boito et al., 2022) introduced two corpora, including one focused on Tamasheq and another exceeding 600 hours of unlabeled audio in Fulfulde, Hausa, Tamasheq, and Zarma. Their work gave particular attention to Tamasheq but did not cover other Niger languages. Likewise, (Diarra et al., 2022) presented the Griot dataset, which has over 31 hours of Bambara speech. This project is one of the few available ASR resources for Manding languages. A subset of this dataset has been used to train robust

ASR systems using both the architectures we explore here (Tapo et al., 2024). Mozilla’s Common Voice project actively gathers speech data for numerous underrepresented languages by leveraging crowdsourcing (Ardila et al., 2020).

Adelani et al. (2021) investigated multilingual modeling and self-supervised learning to develop ASR for 15 African languages, achieving substantial improvements over traditional methods. In the work by Dossou and Emezue (2021), end-to-end ASR systems for Fon and Igbo were developed, providing insights into handling the linguistic complexities of African languages such as tone and diacritics. Mohamud et al. (2021) demonstrated how self-supervised learning can be employed to develop ASR for low-resource African languages like Wolof, Ga, and Somali, even when using as little as one hour of transcribed data. ASR datasets for the languages of the Democratic Republic of the Congo, similar to our dataset but on a smaller scale with fewer languages, have been developed and released (Kimanuka et al., 2024). While not directly addressing African languages, (Olatunji et al., 2023) developed AfriSpeech-200, a pan-African accented speech dataset aimed at clinical and general-domain ASR tasks. Its coverage of diverse accents and dialects across Africa makes it especially relevant for cross-lingual research.

3. Languages of Niger

Our dataset includes data from the following eight languages spoken in Niger, which are members of three distinct language families. While a number of other languages are spoken in Niger – notably French and Arabic – we focus on these eight as particularly representative of the languages and cultures indigenous to Niger.

Zarma (dje) is spoken in parts of West Africa, particularly around the Niger River. It has over 6 million speakers, and it is one of the primary languages of Niamey—Niger’s capital city (Eberhard et al., 2023). Zarma belongs to the Nilo-Saharan language family.

Tubu (dzg) (or Dazaga) is used mainly in desert regions, including areas of Chad and northern Niger. It has approximately 1 million speakers, with the highest concentration in Chad and Niger (tou, 2022). Tubu falls under the Saharan branch of the Nilo-Saharan language family and has several regional variants.

Kanuri (krt) is widespread around the Lake Chad Basin, covering parts of Niger, Nigeria, Chad, and Cameroon. It has over 5 million speakers, including around 450,000 in Niger. Like Tubu, Kanuri also belongs to the Nilo-Saharan language family.

Gourmantche (gux) (or Gourmanché) is spoken in eastern Burkina Faso and neighboring regions

of Niger, Togo, and Benin. It is a Niger-Congo language and has more than 1.2 million speakers (Rédaction, 2019).

Buduma (bdm), sometimes called *Yedina*, is primarily spoken around Lake Chad. Although it has fewer than 200,000 speakers, we include it to broaden the dataset’s coverage of underrepresented languages. It is a language of the Afro-Asiatic family.

Hausa (hau) is one of the most spoken languages in West and Central Africa, with an estimated 50 million native speakers and an additional 20 million second-language speakers. More than half of Niger’s population speaks Hausa, making it a major language of daily life. As part of the Afro-Asiatic language family and its prevalence through the region, it has benefited more linguistic attention compared to other languages discussed here.

Tamasheq (taq), sometimes spelled *Tamashek*, is a Berber language of the Afro-Asiatic family. It is used by the Tuareg people and has around 1.2 million speakers. Tamasheq is noteworthy for its Tifinagh script, an ancient system that has been revived in recent years.

Fulfulde (fuh)—also known as *Fula* or *Fulani*—is spoken across several West African nations and has an estimated 25 million speakers. A Niger-Congo language, it shows marked dialectal diversity, reflecting the geographic spread of its communities.

4. Data Collection

4.1. Data Sources

Niger has a rich oral tradition, where most information is transmitted verbally rather than in written form. Radio stations play a major role in distributing news, making them the largest source of audio in the country. To build our dataset, we requested permission to acquire audio data from several major radio and television agencies: RTN¹, Kalangou Studio², Tambara RTVT³, and RTL-Niger⁴. They granted us access to their content, which consists mostly of local-language **podcasts** and **news programs**.

We followed a multi-phase approach to collect and legally access the audio resources. Initially, we reached out to several broadcasting agencies by email and phone, describing the project’s aims, scope, and ethical considerations. During these exchanges, we emphasized that the data would be released as an open-source resource. Each

¹<https://www.ortn.ne/>

²<https://www.studiokalangou.org/>

³<https://www.youtube.com/@tambarartvt106>

⁴<https://www.youtube.com/@GroupeLabariTV>

agency granted permission, stipulating conditions such as acknowledgment requirements.

After receiving approvals, we gathered publicly available audio from their respective repositories. We also obtained smaller datasets from community organizations – who possessed unarchived recordings in local languages—mostly religious recordings. Furthermore, we promised all the data contributors access to tools and models that will be trained on the SALAN dataset.

In addition to the radio broadcasts, we incorporated audio from other public sources. Some languages, for example, included readings from public-domain religious texts.

4.2. Data Processing and Cleaning

Many of the audio files contained unwanted elements—such as musical intros, background noise, and inconsistent frequency responses. Some of these issues arose from older archives or from equipment with limited recording capabilities. To address these problems, we first located the standard intros for each broadcast—usually the first 7 to 16 seconds—and removed them. We then applied a denoising and enhancement procedure using the Resemble Enhance model⁵. This helped remove transient noise while preserving the speaker’s voice clarity.

Next, we segmented the audio into smaller chunks. We adopted the breath-pause strategy presented by (Shu et al., 2023) but made slight modifications to accommodate language-specific phonetic structures. For instance, in tone languages like Zarma and Hausa, the duration and placement of pauses can carry information that affects meaning interpretation. In those cases, we retained mild pauses to safeguard this content.

We retained the original, unmodified and unsegmented audio files as an additional resource. Note that we did not perform speaker diarization. Our aim was to reduce the risk of errors from automatic speaker labeling, and we wanted to avoid injecting potential bias in the dataset.

4.3. Transcription and Alignment

We adopted a semi-automatic transcription pipeline that combines the MMS model — which offers at least some coverage of 363 African languages and dialects with a mean CER of approximately 2.6% overall (though not on African languages specifically) — with human annotation and alignment checks. Our alignment and transcription workflow used a mix of automated and human adjustments to produce the final transcriptions. Figure 1 shows

⁵<https://www.resemble.ai/introducing-resemble-enhance/>

an overview of this pipeline. and the annotator recruitment and training process are described in Section 5.1.

Using the MMS model, we generated automatic transcriptions at the utterance level for the available audio in each language. For languages with more than 50 hours of available audio, we transcribed 100 hours. For languages with less than 50 hours of audio, we transcribed all available hours. After the automated pass, our annotators then reviewed and corrected the MMS outputs across all languages. However, due to limited availability of annotators for languages other than Zarma and Hausa, we have so far completed manual verification for only these two languages. The transcripts for the other six languages currently remain at the stage of raw MMS outputs. For all transcripts, we used a forced alignment tool⁶ to create word-level time-alignments between transcripts and their corresponding audio segments. After human correction, we estimate the manually verified Zarma and Hausa transcripts has approximately 83-94% accuracy at the word level. This is due to the fact that some words do not have a single acceptable spelling; for instance in zarma: “Kwara” and “Kwaara” are both accepted. We, therefore, leave room for improvements.

While our approach accelerates the transcription process, it also carries a risk of transcription inaccuracies. Section 9 addresses the implications of relying on raw transcriptions for some languages and outlines our plans to include more manual verification in future dataset releases.

After cleaning and processing, the entire dataset spans *eight languages* and totals **110 hours** of manually corrected audio—50 hours for Hausa, and 60 for Zarma. Table 1 details the distribution of cleaned audio for each language.

5. Data Annotation

5.1. Annotators

In total, we enlisted 17 annotators for this project. Each annotator was between 19 and 28 years old and had some technical background, often in computer science or related fields. Below, we detail their recruitment process, profiles, annotation guidelines, and compensation arrangements.

We reached out to a local educational institution in Niger (**CODELOCCOL**) to find prospective annotators. Our recruitment email explained the project’s objectives, the expected time commitment, and the basic requirements for language knowledge. Once we received their confirmations, we invited the selected candidates to an orientation session. During this session, they reviewed our tran-

⁶<https://huggingface.co/MahmoudAshraf/mms-300m-1130-forced-aligner>

Languages	Total Hours	Distribution		
		Hrs Transcribed	Hrs Adjusted	Ad-
Zarma	783	460	60	
Tubu	12	12	—	
Kanuri	17.66	17.66	—	
Gourmantche	13.30	13.30	—	
Buduma	13.27	13.27	—	
Hausa	808.9	400	50	
Tamasheq	192.32	100	—	
Fulfulde	290.94	100	—	

Table 1: Data Statistics across languages. Below we show the total number of hours of audio recordings available for each language (Total Hours) followed by the number of hours transcribed with MMS (Hrs Transcribed). For the two languages for which we have completed manually-verified alignment and transcription so far, we also indicate the number of actual hours of speech that have been reviewed (Hrs Adjusted).

scription guidelines, performed short practice tasks, and learned how to operate our Google Sheets, the interface we used to carry out annotation.

All annotators had strong computer skills. They were either software developers or computer science students. While most had limited linguistics experience, they spoke at least one of the target languages — Hausa and Zarma — natively, which allowed them to understand the basic semantics, and phonetic variations.

We provided a small compensation to each annotator, acknowledging that it fell below the market rate. In addition, we granted them early access to the final dataset, along with priority use of any derived language models or software.

5.2. Annotation Guidelines

Because many covered languages—apart from Hausa and Fulfulde—lack a standardized writing system, we developed concise, language-specific guidelines. These guidelines addressed:

Orthographic Conventions: We adopted spellings for common words and speech sounds. For that, we refer to the Feriji dataset dictionary (Keita et al., 2024) as the primary guide for common words in Zarma.

Speech Overlap Annotation: In case of speech from multiple speakers talking at once, we asked them to focus on adjusting the transcription obtained from the automatic process regardless of the speaker.

Ambiguous Segments: When one of them came across a speech utterance or segment that was difficult to interpret, that segment was shared with the other annotators for disambiguation. We supported these guidelines with illustrative examples so that annotators had a reference for

ambiguous cases.

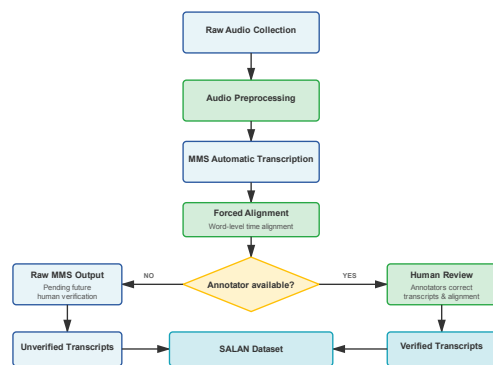


Figure 1: Overview of the alignment and transcription workflow.

6. Experiments

6.1. MMS Baseline

Before conducting our experiments, we evaluated the out-of-the-box performance of the MMS model on our manually verified transcripts. We tested it on the 10 hours of Zarma data and the 10 hours of Hausa data that had been all reviewed and corrected by humans. For Zarma, MMS achieved a **WER of 43.8%** and a **CER of 29.6%**. For Hausa, the model achieved a **WER of 35.3%** and a **CER of 22.7%**.

These results suggest that while MMS provides a descent starting point for initial automated transcription, its accuracy remains poor for accurate use in these languages. The high error rates justify our decision to develop fine-tuned ASR models for these Niger’s languages.

Additionally, we deliberately chose not to fine-tune MMS itself in our experiments, as this would introduce potential bias.

6.2. Trainings

To assess the practical utility of SALAN for ASR model development, we conducted experiments using Zarma, one of the two languages for which we have manually verified transcriptions. From the 60 hours of manually verified Zarma data, we selected a 10-hour subset for our training experiments and kept 2 hours for validation. As shown in Section 5, the baseline MMS model achieved 43.8% WER and 29.6% CER on our verified Zarma data, establishing the need for improved models. We conducted three experimental configurations: (1) Using wav2vec2 XLS-R and Whisper, we carry out **direct fine-tuning on the 10 hours of manually labeled Zarma data**, as is commonly done

Model	Setup	Size	Unlabeled	Device	Train/Val	WER(%)	CER(%)
<i>Whisper</i>	Baseline	769M	0 hr	A100	10h/2h	27.4	7.2
<i>wav2vec2</i>	Baseline	300M	0 hr	A100	10h/2h	32.6	11.3
	+ Unlabeled	300M	60 hr	A100	10h/2h	30.7	8.4
<i>wav2vec2</i>	+Active Learning	300M	450 hr	A100	15h/2h	25.3	6.2

Table 2: Settings and results of all experiments. "Baseline" experiments are those using only direct fine-tuning, with 10 hours of labeled Zarma data. The "+ Unlabeled" experiments adds 60 hours of unlabeled data from Tubu, Kanuri, Gourmantche, Buduma, Tamasheq, and Fulfulde during an initial round of continued pre-training, followed by fine-tuning to the 10 hours of labeled Zarma data. The "+Active Learning" experiment uses 450 hours of unlabeled Zarma audio from which we selected high-uncertainty segments for targeted annotation, expanding the training set to 15 hours.

for under-resourced ASR. (2) Within the *wav2vec2* XLS-R framework only, we carry out **continued pre-training** using unlabeled audio from six of the seven languages that lacked manually-verified transcriptions: Tubu, Kanuri, Gourmantche, Buduma, Tamasheq, Fulfulde. We then fine-tune that newly pre-trained model to the 10 hours subset of manually verified Zarma data. (3) We explore an **active learning** strategy to maximize annotation efficiency.

The direct fine-tuning experiments were conducted on a A100 GPU with 80GB of memory, and the pretraining experiments were conducted on same setting. We selected *Whisper-medium* and *Wav2vec2-XLS-R-53* models for our experiments. We used direct fine-tuning on *Whisper* because it was pre-trained on several African languages and already demonstrates good accuracy, especially for Hausa. We applied both approaches—direct fine-tuning and pre-training followed by fine-tuning—for the *Wav2vec2-XLS-R* architecture.

In the third configuration we explored an active-learning strategy to maximize the value of our limited annotation budget. We first fine-tuned *Wav2vec2-XLS-R* on the original 10-hour Zarma subset. We then scored the remaining 450 h of unlabeled Zarma audio with this model and selected the 1,000 segments with the highest frame-level CTC loss (≈ 5 h) for human re-annotation. After correction, the expanded 15-hour set was used for a second fine-tuning run. Table 2 shows this "+Active Learning" condition.

7. Results and Discussion

Table 2 presents the outcomes of the three training configurations. As mentioned in Section 5, the baseline MMS model achieved 43.8% WER and 29.6% CER on our manually verified Zarma data. The *Baseline* results in the table show the two direct fine-tuning experiments, where only the 10 hours of manually verified Zarma data is used. Both architectures outperform the MMS baseline, with

Whisper achieving a WER of 27.4% and a CER of 7.2%, and *Wav2vec2* yielding a WER of 32.6% and a CER of 11.3%. These results represent WER improvements of $\approx 37\%$ and 26% over MMS for *Whisper* and *Wav2vec2* respectively.

The "+*Unlabeled*" entry incorporates 60 hours of unlabeled data from the other languages during an initial continued pre-training step. We see that continued pre-training with unlabeled data from the other languages further improved performance, reducing the WER from 32.6% to 30.7% and the CER from 11.3% to 8.4%. Although not significant, this demonstrates that our large unlabeled set can enhance fine-tuning performance even with a limited labeled dataset for a specific language.

Despite these experiments used a small labeled dataset and limited training epochs, the obtained results represent a benchmark for Zarma. The performance can potentially be further enhanced by increasing the size of the training data, adjusting the number of training epochs, and implementing more elaborated training strategies.

Our active-learning experiment achieved the best results, reducing WER from 30.7% to 25.3% and CER from 8.4% to 6.2% compared to the continued pre-training baseline. This represents a 42% WER reduction from the original MMS baseline (43.8% to 25.3%) and a 79% CER reduction (29.6% to 6.2%). These results show that targeted annotation of high-uncertainty samples yields larger gains than uniform expansion of the labeled set.

These improvements confirm the utility of our large unlabeled set in improving the performance of fine-tuning using a limited set of labeled data for a specific language.

8. Conclusion & Future Work

In this paper, We presented SALAN, a large-scale speech dataset that focuses on eight of Niger's main indigenous languages. We combined a semi-automatic transcription pipeline with partial human verification, enabling efficient data curation under

resource constraints. Our experiments highlight how unlabeled data can further improve ASR performance in low-resource contexts, even with a small labeled subset.

Niger is a multilingual country where local languages hold an important role in everyday communication ([WorldAtlas, 2017](#)). Its literacy rate remains among the lowest in the world, with approximately 68.9% of youth unable to read or write ⁷. Given these realities, SALAN has the potential to benefit over **20 million people** across Niger by making speech-driven technologies available in their native languages. For instance, the dataset can be used to develop more accurate ASR systems, which in turn could help automate transcription for media broadcasters, enhance access to information with less efforts, or enable voice-based user interfaces for low-literacy communities. In addition, speech technologies can be used to support literacy education and to enhance literacy interventions in both children and adults.

One obvious future avenue for research is to expand manual verification and correction to unlabeled portions of the SALAN dataset. Our first goal will be to manually verify transcriptions for 10 hours of each of the 8 languages. Our work here demonstrates the potential of our pipeline, particularly through the participation of students who are eager to contribute to the development of new AI technologies. As the accuracy of our ASR models improves, the efficiency of our pipeline for producing manually verified transcripts will increase. We note that a number of the languages in our dataset have significant dialectal variation. In our future work, we will also investigate accent variations and dialects, which may reduce errors for localized usage. We also hope to carry out semi-supervised diarization. Finally, we intend to explore domain-adaptation strategies, larger training sets, and extended trainings to create state-of-the-art ASR models for the languages of Niger.

9. Limitations

Even though this project has a wide range of impact, we acknowledge some limitations.

Only a small portion of the data for only two of the languages – Hausa and Zarma – was transcribed and manually corrected. We relied on raw MMS outputs for the other languages, which almost certainly contain errors, particularly at the word level. Our limited annotation capacity makes full human verification infeasible at this time.

We collected audio from heterogeneous sources and time periods, causing differences in microphone setups, acoustic environments, and signal-

to-noise ratios. Although we applied cleaning and enhancement, some noise may remain.

We decided not to implement any automated speaker identification to avoid potential errors and biases. This choice restricts the dataset’s utility for tasks like speaker-specific analysis or gender balance research.

10. Acknowledgment

We are grateful to the Research Computing Department of RIT ([RIT, 2025](#)) for providing compute resources for this project.

References

2022. [Toubou people](#). Accessed: 2024-12-23.
- Solomon Teferra Abate, Wolfgang Menzel, and Bairu Tafila. 2005. An amharic speech corpus for large vocabulary continuous speech recognition. In *INTERSPEECH-2005*.
- David Ifeoluwa Adelani, Sandy Ritchie, You-Chi Cheng, Mingqing Chen, Rajiv Mathews, Daan van Esch, Bo Li, and Khe Chai Sim. 2021. [Large vocabulary speech recognition for languages of africa: multilingual modeling and self-supervised learning](#).
- Alexis Conneau, Alexei Baevski, Ronan Collobert, Abdelrahman Mohamed, and Michael Auli. 2020. [Unsupervised cross-lingual representation learning for speech recognition](#).
- Bonaventure F. P. Dossou and Chris Chinenye Emezue. 2021. [Okwugbé: End-to-end speech recognition for fon and igbo](#).
- M. Paul Eberhard, Gary F. Simons, and Charles D. Fennig. 2023. *Ethnologue: Languages of the World*. SIL International.
- Elodie Gauthier, Laurent Besacier, Sylvie Voisin, Michael Melese, and Uriel Pascal Elingui. 2016. Collecting resources in sub-saharan african languages for automatic speech recognition: a case study of wolof. In *LREC*.
- Hadrien Gelas, Laurent Besacier, and Francois Pellegrino. 2012. [Developments of Swahili resources for an automatic speech recognition system](#). In *SLTU - Workshop on Spoken Language Technologies for Under-Resourced Languages*, Cape-Town, Afrique Du Sud.
- Mamadou Keita, Elysabhete Ibrahim, Habibatou Alfari, and Christopher Homan. 2024. [Feriji: A](#)

⁷<https://www.anp.ne/>

French-Zarma parallel corpus, glossary & translator. In *Proceedings of the 62nd Annual Meeting of the Association for Computational Linguistics (Volume 4: Student Research Workshop)*, pages 1–9, Bangkok, Thailand. Association for Computational Linguistics.

Jama Hussein Mohamud, Laurent Besacier, Francis Muthumbi, and Justin T. Chibueze. 2021. [Fast development of asr in african languages using self supervised speech representation learning](#).

Tobi Olatunji, Tejumade Afonja, Aditya Yadavalli, Chris Chinenye Emezue, Sahib Singh, Bonaventure F. P. Dossou, Joanne Osuchukwu, Salomey Osei, Atnafu Lambebo Tonja, Naome Etori, and Clinton Mbataku. 2023. [Afrispeech-200: Pan-african accented speech dataset for clinical and general domain asr](#). *Transactions of the Association for Computational Linguistics*, 11:1669–1685.

Vineel Pratap, Andros Tjandra, Bowen Shi, Paden Tomasello, Arun Babu, Sayani Kundu, Ali Elkahky, Zhaoheng Ni, Apoorv Vyas, Maryam Fazel-Zarandi, Alexei Baevski, Yossi Adi, Xiaohui Zhang, Wei-Ning Hsu, Alexis Conneau, and Michael Auli. 2023. [Scaling speech technology to 1,000+ languages](#).

Alec Radford, Jong Wook Kim, Tao Xu, Greg Brockman, Christine McLeavey, and Ilya Sutskever. 2022. [Robust speech recognition via large-scale weak supervision](#).

RIT. 2025. [Research computing services](#).

La Rédaction. 2019. A la découverte du Gulmu - Gulmu Info — gulmu.info. <https://www.gulmu.info/a-la-decouverte-du-gulmu/>. [Accessed 23-12-2024].

Yuchun Shu, Haoneng Luo, Shiliang Zhang, Longbiao Wang, and Jianwu Dang. 2023. [A cif-based speech segmentation method for streaming e2e asr](#). *IEEE Signal Processing Letters*, 30:344–348.

Allahsera Tapo, Éric Le Ferrand, Zoey Liu, Christopher Homan, and Emily Prud'hommeaux. 2024. Leveraging speech data diversity to document indigenous heritage and culture. In *Proceedings of the Annual Conference of the International Speech Communication Association (Interspeech)*, pages 5088–5092.

WorldAtlas. 2017. What Languages Are Spoken In Niger? — worldatlas.com. <https://www.worldatlas.com/articles/what-languages-are-spoken-in-niger.html>. [Accessed 20-12-2024].

Language Resource References

Ardila, Rosana and Branson, Megan and Davis, Kelly and Kohler, Michael and Meyer, Josh and Henretty, Michael and Morais, Reuben and Saunders, Lindsay and Tyers, Francis and Weber, Gregor. 2020. [Common Voice: A Massively-Multilingual Speech Corpus](#). European Language Resources Association.

Sebastien Diarra and Michael Leventhal and Allahsera Auguste Tapo. 2022. [RobotsMali Griots Speech Dataset, and ASR](#).

Ussen Kimanuka and Ciira wa Maina and Osman Büyüç. 2024. [Speech recognition datasets for low-resource Congolese languages](#).

Olatunji, Tobi and Afonja, Tejumade and Yadavalli, Aditya and Emezue, Chris Chinenye and Singh, Sahib and Dossou, Bonaventure F. P. and Osuchukwu, Joanne and Osei, Salomey and Tonja, Atnafu Lambebo and Etori, Naome and Mbataku, Clinton. 2023. [AfrISpeech-200: Pan-African Accented Speech Dataset for Clinical and General Domain ASR](#).

Zanon Boito, Marcelly and Bougares, Fethi and Barbier, Florentin and Gahbiche, Souhir and Barrault, Loïc and Rouvier, Mickael and Estève, Yannick. 2022. [Speech Resources in the Tamasheq Language](#). European Language Resources Association.