

Speech Resources in the Tamasheq Language

Marcely Zanon Boito¹, Fethi Bougares², Florentin Barbier³, Souhir Gahbiche³,
Loïc Barrault², Mickael Rouvier¹, Yannick Estève¹

¹LIA - Avignon University, France ²LIUM - Le Mans University, France ³Airbus - France

Contact: {marcely.zanon-boito, yannick.esteve} @univ-avignon.fr

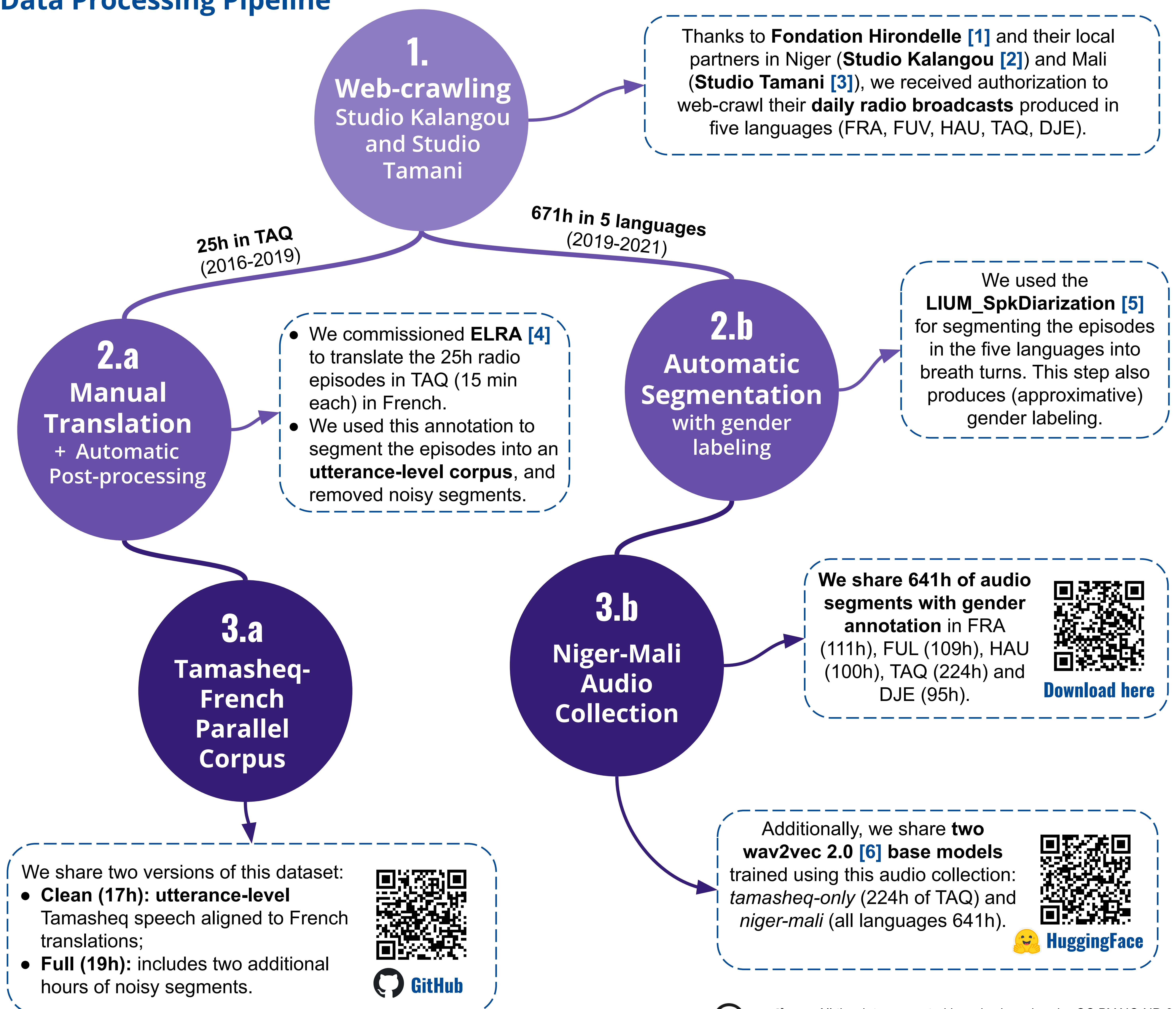
In this paper we...

1. Produce a Tamasheq to French **speech translation corpus** (17h).
→ **The Tamasheq language:** ~500 thousand speakers, Tuareg variety, Berber macro-language spoken by nomadic tribes across North Africa in Algeria, Mali, Niger and Burkina Faso.
2. Distribute a 641h **audio collection** in five languages from Niger:
→ French (FRA), Fulfulde (FUV), Hausa (HAU), Tamasheq (TAQ) and Zarma (DJE).



PAPER PDF

Data Processing Pipeline



All the data presented here is shared under CC BY-NC-ND 3.0: Attribution-NonCommercial-NoDerivs 3.0 Unported

[1] <https://www.hirondelle.org/en/>
[2] <https://www.studiokalangou.org/>
[3] <https://www.studiotamani.org/>
[4] <http://www.elra.info/en/about/elda/>

[5] Meignier&Merlin (2010). Lium spkdiazation: an open source toolkit for diarization. In CMU SPUD Workshop.
[6] Baevski et al. (2020). wav2vec 2.0: A framework for self-supervised learning of speech representations. arXiv:2006.11477