UnMute

Speech Interactions Designed With Minoritised Language Speakers

Emily Nielsen, Electra Wallington, Ondřej Klejch, Dani Kalarikalayil Raju, Nina Markl, Gavin Bailey,
Thomas Reitmaier, Jennifer Pearson, Matt Jones, Peter Bell, Simon Robinson

Motivation

- Speech-based systems are not currently available for the vast majority of languages
- This leaves many communities digitally ‘unheard’
  - Especially for languages that are not commonly written or entirely unwritten [1]
- Existing datasets do not represent the day-to-day languages used (e.g., code-switching) [2]
- This further embeds the digital and linguistic inequality and excludes millions of people worldwide

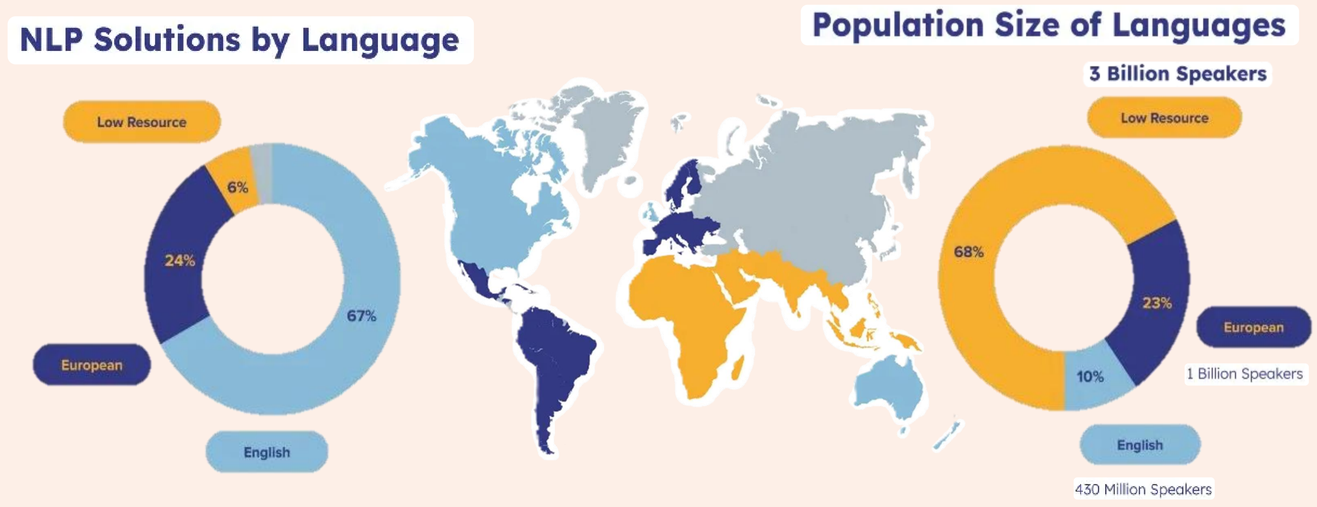


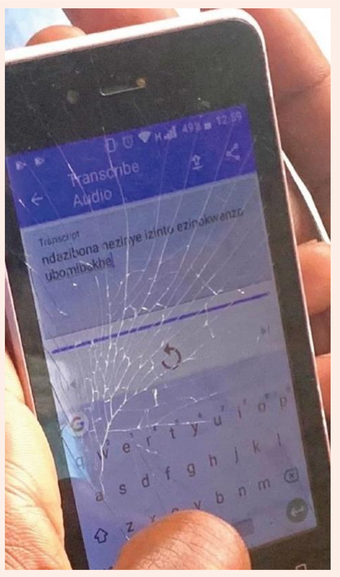
Image Source: https://medium.com/neuralspace/low-resource-language-what-does-it-mean-d067ec85dea5

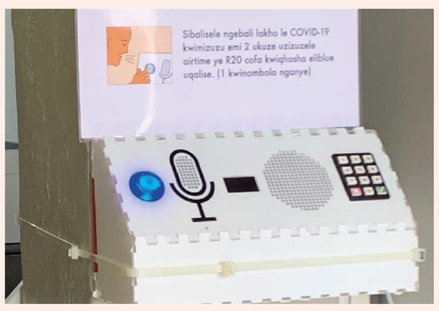
The UnMute Project

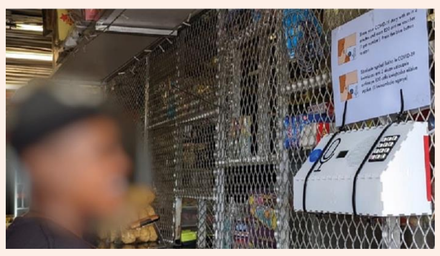
- Interdisciplinary project: HCI, NLP and linguistics researchers
- Aims to empower under-served voices from ‘low-resource’ language communities
- To do this, we co-create spoken-language technologies with communities in South Africa and India
  - Diverse, resource-constrained, and under-heard communities
- These co-created prototypes:
  - Demonstrate the benefits of human-centred methodologies, co-creating datasets and use-cases directly with communities
  - Illustrate how language technologies can broaden digital participation in communities of minoritised language speakers

Community-Based Data Collection and Engagement

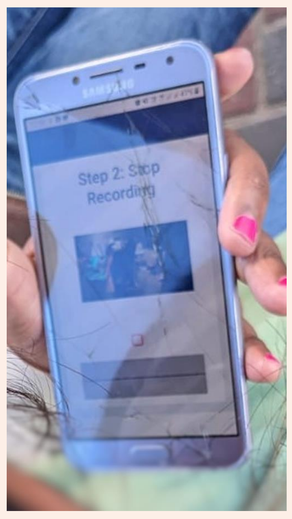
- Voices from Langa (left): a speech probe to engage minoritised language speakers in voice data-collection
- TranscriptTool (middle): an interface to involve community members in the transcription of collected speech data
- SpeechBox (right): a speech probe to gather community responses on their experience of COVID-19 lockdowns
- These probes and prototypes facilitate the collection of speech data from communities in their day-to-day context
  - Rich, every-day language
  - Code-switching
  - Reliable transcriptions

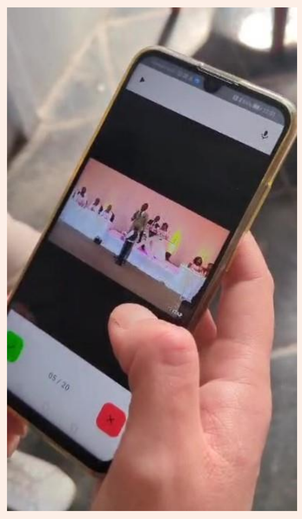






Speech Based Retrieval Application





- Two data collection platforms to gather audio annotations from communities
  - Verbally annotate images and videos
- Ranker trained on phone sounds of high resource languages and fine-tuned on small amounts of low-resource data
- Facilitates media retrieval without requiring written input
- Image retrieval on < 4 hours data
- 74% queries retrieve the correct image in the top 5 results

[1] S. Bird and D. Yibarbuk, “Centering the Speech Community,” in EACL’24, Mar. 2024, pp. 826–839.
[2] T. Reitmaier, E. Wallington, D. Kalarikalayil Raju, O. Klejch, J. Pearson, M. Jones, P. Bell, and S. Robinson, “Opportunities and Challenges of Automatic Speech Recognition Systems for Low-Resource Language Speakers,” in CHI ’22. Apr. 2022, pp. 1–17.

RAi UKRI The University of Edinburgh Swansea University Prifysgol Abertawe CSTR lab University of Essex