NaijaVoices — A Community-Built 1,838-Hour Speech Dataset So Voice Technology Can Understand Igbo, Hausa and Yoruba

Digital Inclusion · Good practice · Nigeria

Quality score: 76/100

Evidence strength: Strong

Summary

Community initiative that collected 1,838 hours of Igbo, Hausa and Yoruba speech from 5,455 speakers across all Nigerian zones; peer-reviewed at Interspeech 2025, it markedly improved speech recognition models.

Key dimensions

Dimension	Score
Innovation level	4/5
Sustainability (ongoing)	0/1
Evaluation evidence	1/1
Demonstrated reach	1/1
Documented learning	1/1

Evaluated by C-NAPSE

Data sources

C-NAPSE web research — accessed 2026-11-17

NaijaVoices — accessed 2026-11-17

Emezue et al. (2025), The NaijaVoices Dataset — arXiv / Interspeech 2025 — accessed 2026-11-17

ISCA Archive — Interspeech 2025 paper — accessed 2026-11-17

Technext (2025) — 'The machine now speaks our language' — accessed 2026-11-17

Cite this — Evidoria. *NaijaVoices — A Community-Built 1,838-Hour Speech Dataset So Voice Technology Can Understand Igbo, Hausa and Yoruba*. Persistent ID: 60d873cf-1ffb-48ff-beca-bea86427cc45.