

MODULE 1

- [How to Use this Playbook](#)
- [PART I — BEFORE YOU COLLECT ANYTHING](#)
- [PART II — DESIGN THE DATASET](#)

How to Use this Playbook

This playbook is for anyone who wants to collect voice data in an African language, whether the project is:

- Small — 50–100 speakers
- Medium — 500–1,000 speakers
- Large — thousands of contributors
- Academic
- Community-led
- Government-led
- NGO-led
- Commercial
- Open-source
- Designed for ASR/TTS
- Intended for language documentation
- Intended for AI research

It is designed to answer a simple question:

“If I want to collect voice data in my language, what exactly do I need to do?”

The playbook takes you through the entire lifecycle:

“

Plan ? Design ? Prepare ? Mobilize ? Recruit ? Consent ?
Train ? Record ? Validate ? Transcribe ? QA ? Pay ?
Document ? Release ? Sustain

Voice collection is not simply a recording exercise. It is an operational system involving people, language expertise, community relationships, technology, quality control, ethics and data management. The AfriVoices-KE project organized these functions through language leads, resource persons, mobilizers, contributors, validators, transcribers, super-reviewers and administrators.

PART I — BEFORE YOU COLLECT ANYTHING

1. Define Why You Are Collecting Voice Data

Before recruiting anyone, answer:

1.1 What will the data be used for?

Examples:

- Automatic Speech Recognition
- Speech-to-text
- Text-to-Speech
- Speech translation
- Voice assistants
- Language documentation
- Linguistic research
- Benchmarking
- Conversational AI
- Digital government
- Healthcare
- Agriculture
- Education

Your intended use determines what type of recordings you need.

1.2 What kind of speech do you need?

There are two basic approaches used in the reference project.

a) Scripted speech

A participant reads a prepared sentence.

b) Unscripted speech

A participant responds naturally to a question or prompt.

The reference project deliberately combined the two, targeting approximately 75% unscripted and 25% scripted speech.

Practical recommendation

If your goal is a general-purpose speech dataset, consider collecting both.

Type	What it gives you
Scripted	Controlled sentences, predictable content
Unscripted	Natural speech, dialects, spontaneous phrasing
Both	Broader coverage

Do not assume that one approach automatically replaces the other.

2. Decide How Much Data You Need

Start with the target rather than starting with recruitment.

Define:

a)Target number of speakers

Example: 1,000 speakers

b)Target hours

Example: 500 hours

c)Target recordings per person

Example: 30 minutes validated speech per contributor

d)Target speech type

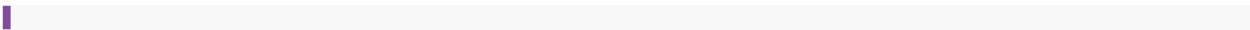
Example: 75% unscripted / 25% scripted

e)Target geographic coverage

Example: 10 counties

f) Target demographic representation

Example: Balanced across gender and five age groups.



The Africa Next Voices project targeted approximately 3,000 hours across five languages, with different targets per language.

g) Create a simple target sheet

Metric	Target
Speakers	1,000
Total hours	500
Scripted	125 hrs
Unscripted	375 hrs
Regions	10
Dialects	4
Female	50%
Male	50%
Age groups	5

Your numbers will depend on your project. The table is a planning tool, not a universal standard.

3. Define Your Language

Do not simply write:

“We are collecting Language X.”

Ask:

- Which variety?
- Which dialects?
- Which geographic areas?
- Which communities?
- Is the language internally diverse?
- Are some varieties more widely used?
- Are some varieties endangered?
- Which varieties are appropriate for the intended use?

“ Example : The AFN project explicitly considered dialect variation in languages including Kalenjin, Maasai, Somali, Dholuo and Kikuyu.

Create a language profile

Record:

1. Language:
2. Language family:
3. Country/countries:
4. Main regions:
5. Dialect groups:
6. Estimated speaker distribution:
7. Existing digital resources:
8. Existing speech datasets:
9. Native-speaker experts:
10. Relevant institutions:

4. Define Your Dialect Strategy

This is one of the most important steps.

A dataset can contain thousands of hours and still be poorly representative if most speakers come from one location or dialect.

Create a dialect matrix:

Dialect	Geography	Target speakers	Target hours
Dialect A	Region 1	200	100

Dialect B	Region 2	200	100
Dialect C	Region 3	200	100
Dialect D	Region 4	200	100

“ The Africa Next Voices project used explicit dialect targets and geographic recruitment rather than treating each language as homogeneous.

PART II — DESIGN THE DATASET

5. Choose Scripted vs Unscripted Collection

5.1 Scripted Collection

Participants read prepared sentences.

a) Use scripted data when you need:

- Consistent text/audio pairs
- Controlled vocabulary
- Sentence-level alignment
- Clear transcription
- Specific words or phrases
- Named entities
- Domain-specific terminology

b) Advantages

- Fast to collect
- Easy to validate
- Predictable content
- Easy to align with text

c) Limitations

The reference report notes that scripted speech can omit natural speech phenomena such as:

- Fillers
- False starts
- Slang
- Code-switching
- Natural prosody
- Informal speech

It can also introduce bias toward formal language.

6. Unscripted Collection

Unscripted collection asks contributors to speak naturally in response to a prompt.

For example:

“Describe what you normally do when you wake up in the morning.”

or:

“Tell us about a memorable journey you have taken.”

or:

Show an image and ask the participant to describe what is happening.

“ The reference project used both text and image prompts to elicit spontaneous speech.

Advantages

Captures:

- Natural phrasing
- Accent
- Dialect
- Spontaneous speech
- Informal vocabulary
- Prosody
- Natural discourse

Challenges

Unscripted recordings require substantially more:

- Validation
- Transcription
- QA
- Prompt design
- Human review

The reference project therefore used a multi-level QA and transcription process for this data.

7. Decide Your Domains

Do not collect speech around one topic only.

The reference project used multiple domains, including:

- Agriculture and Food
- Everyday Scenarios
- Financial Transactions
- Digital Government Services
- Named Entity Recognition
- Role Play
- Extempore Stories
- Healthcare
- News and Media
- Education and Technology
- Customer Care

Create your own domain matrix

Domain	Scripted	Unscripted	Target hours
Everyday life	✓	✓	50
Agriculture	✓	✓	100
Healthcare	✓	✓	50
Education	✓	✓	50
Government	✓	✓	50
Stories		✓	50
Customer care	✓	✓	50
Other	✓	✓	100

The purpose is to expose the eventual model to varied vocabulary and contexts.

8. Develop Your Scripted Sentences

Your scripted sentences should not simply be random translations.

Build them from:

- Existing corpora
- Existing language resources
- Domain-specific material
- Locally generated sentences
- Carefully reviewed translations

The reference project combined existing corpora with translated and newly generated material.

Every sentence should be checked for:

- Naturalness
- Correct grammar
- Correct spelling
- Correct terminology
- Cultural appropriateness
- Dialect
- Names
- Places
- Numbers
- Dates
- Abbreviations

Important rule

Do not assume an English sentence can simply be translated word-for-word.

The reference project specifically encountered problems with polysemous words, inconsistent translations and concepts that lacked direct equivalents in some languages.

9. Develop Unscripted Prompts

Good prompts produce speech.

Bad prompt:

“Do you like farming?”

Possible response:

“Yes.”

Good prompt:

“Tell us about the farming activities that people in your community usually do during the rainy season.”

This encourages a longer response.

Use several prompt types

a) Text prompts

A written question.

b) Image prompts

An image plus a question.

c) Story prompts

Ask participants to tell a story.

d) Scenario prompts

Ask participants to explain what they would do.

e) Role-play prompts

Example:

“Imagine you are calling a customer-care agent because your mobile money transaction failed.”