



Aikuma:

A mobile app for collaborative language documentation

Steven Bird^{1,2}, Florian Hanke¹, Oliver Adams¹, Haejoong Lee²

¹Department of Computing and Information Systems, University of Melbourne

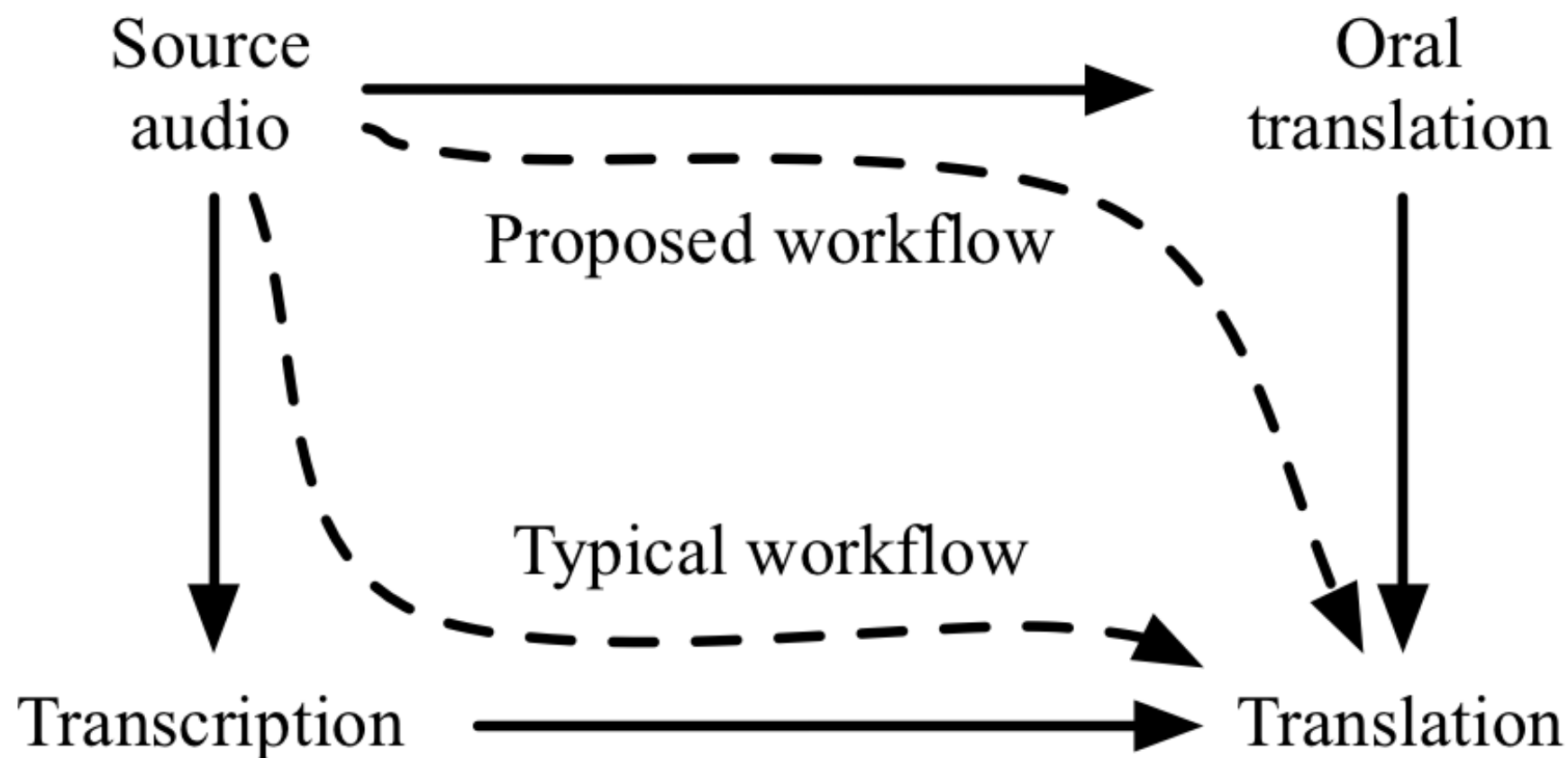
²Linguistic Data Consortium, University of Pennsylvania

Supported by NSF Grant to LDC, UPenn “Language Preservation 2.0: Crowdsourcing Oral Language Documentation using Mobile Devices”



Google Play: install Aikuma

Postponing the Transcription Task



Thought Experiment: The future philologist

- ▶ time is shorter than we think
- ▶ what should we archive today?
- ▶ Boasian trilogy: texts, grammar, lexicon
- ▶ the primary resource?

Careful Respeaking

- ▶ Listen to an original recording and repeat what was said, slowly and carefully
- ▶ recordings made “in the wild”
- ▶ phonetic analysis (Bettinson 2013)
- ▶ enables us to postpone the transcription task

Language Documentation

- ▶ what it consists of
- ▶ who it involves
- ▶ the challenge
- ▶ ensuring interpretability
 - ▶ what was said; what it meant

Design Principles

- ▶ text free
- ▶ synchronisation
- ▶ metadata

Transcription (mk 1)



Transcription (mk 2)



lp20.org/transcriber/test/ x

lp20.org/transcriber/test/web/transcriber.html

File Play Stop Create Segment Remove Segment Play Respeaking

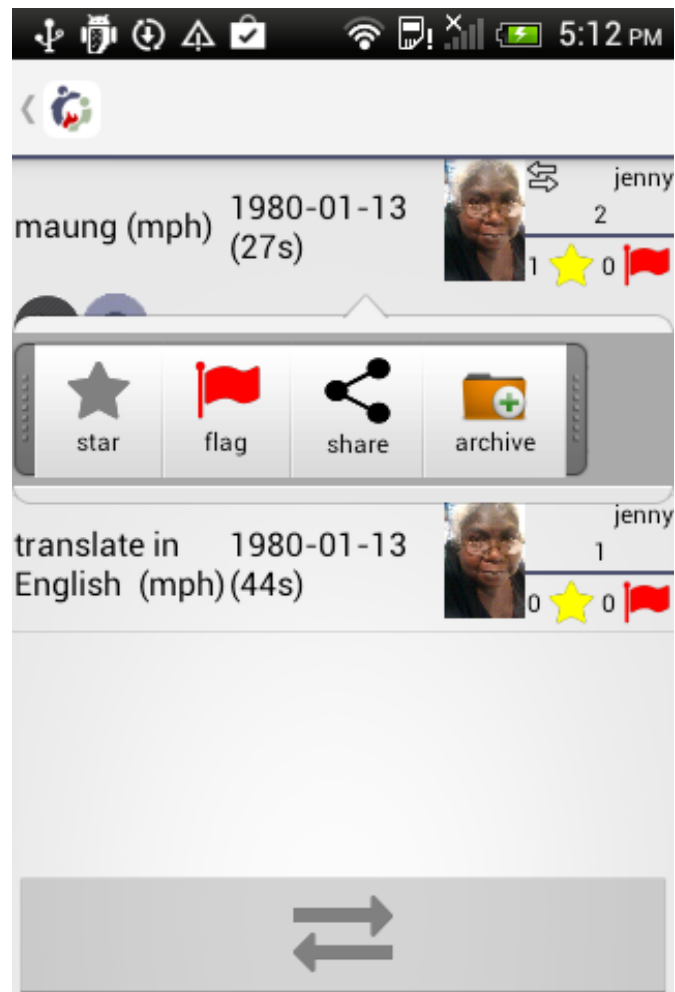
18.3078
9.305
8.353

historia do curupira

Open

Time	Transcription	Translation
18.3078	Certo dia um homem foi para o mato.	One particular day a man went into the forest
9.305	Ele gostava muito de trabalhar de tucumã, cipó e piassaba.	He very much liked to collect tucumã, cipó and piassaba
8.353	Ele passa durante o seu trabalho mais ou menos uma..duas ou três semanas.	He went out to the forest more or less every one, two or three weeks
	Trabalha muito longe, quase no meio do..do centro da mata.	He works very far away almost in the middle of the..In the center of the forest
	Era muito longe.	It was very far away

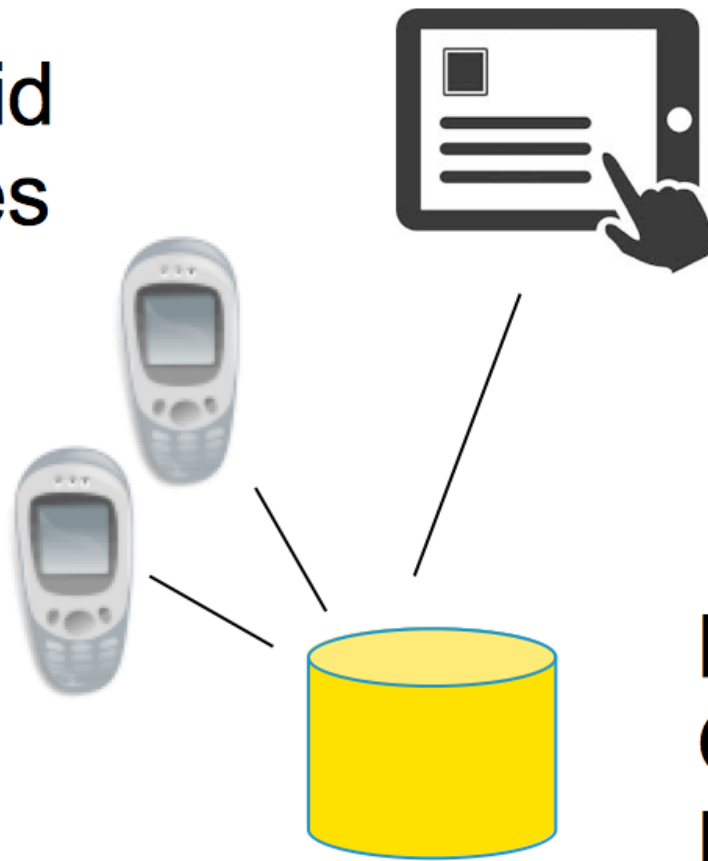
Sharing and Archiving



Architecture 1:

Backup and synchronization for several devices

**Android
devices**



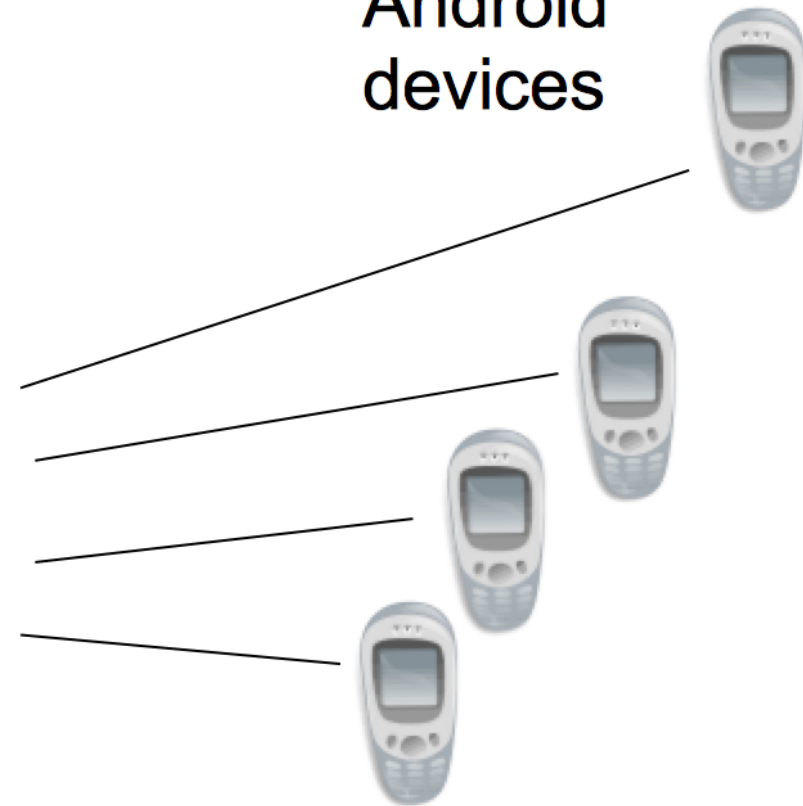
**Personal
Google
Drive**

Architecture 2: Village-based networking

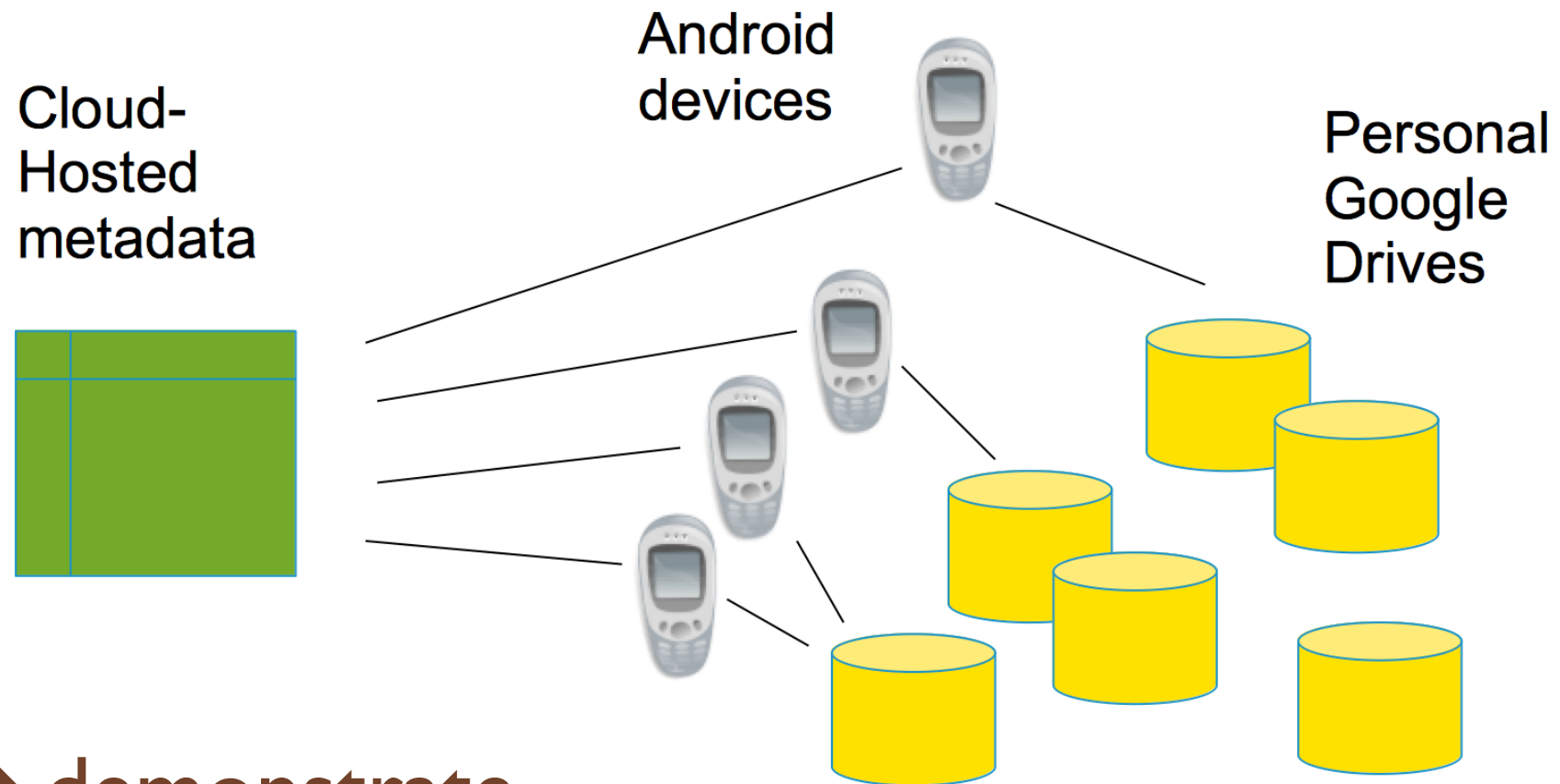
Wifi
router



Android
devices



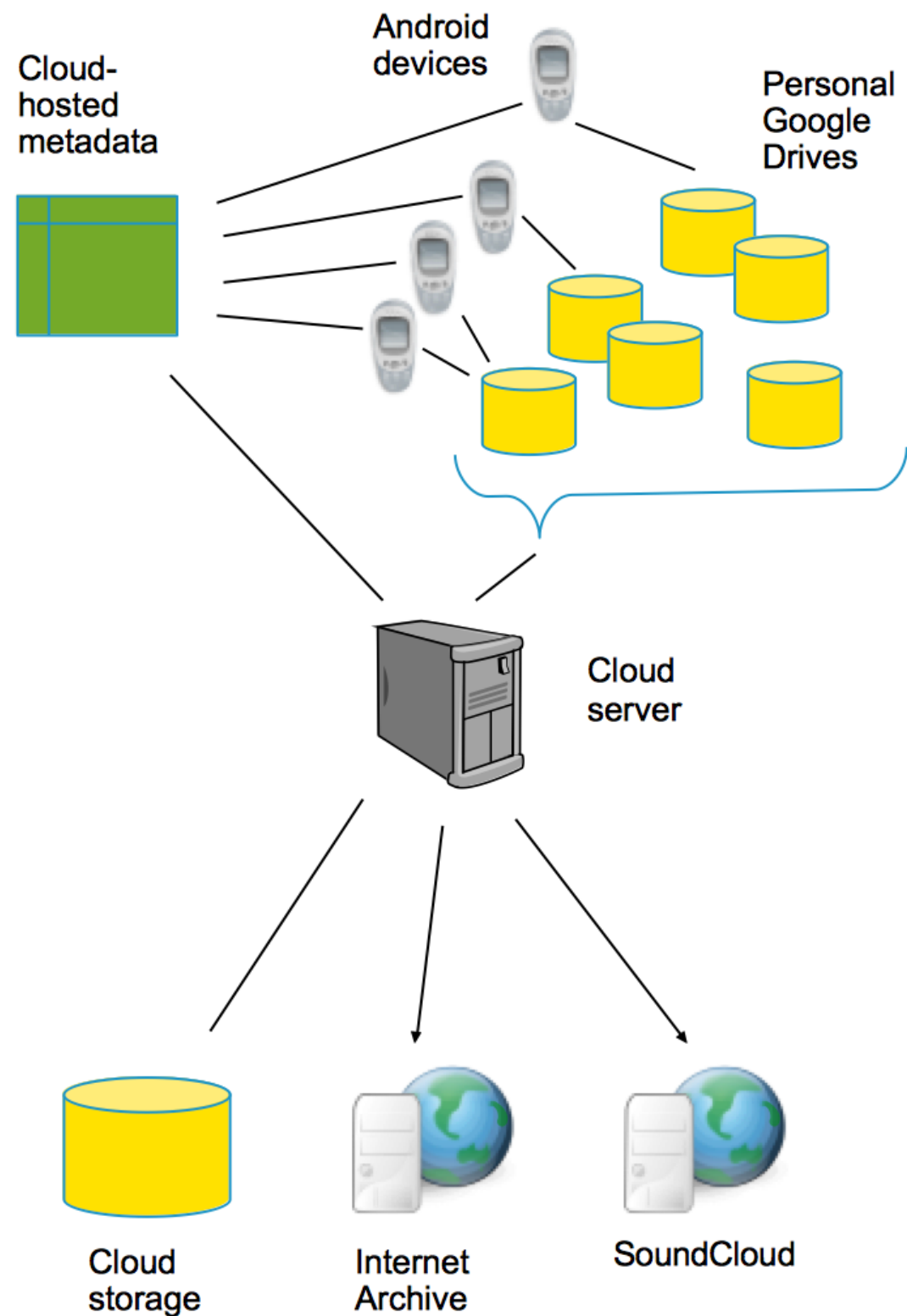
Architecture 3: Cloud-based sharing



► demonstrate
cloud server

Architecture 4: Preservation

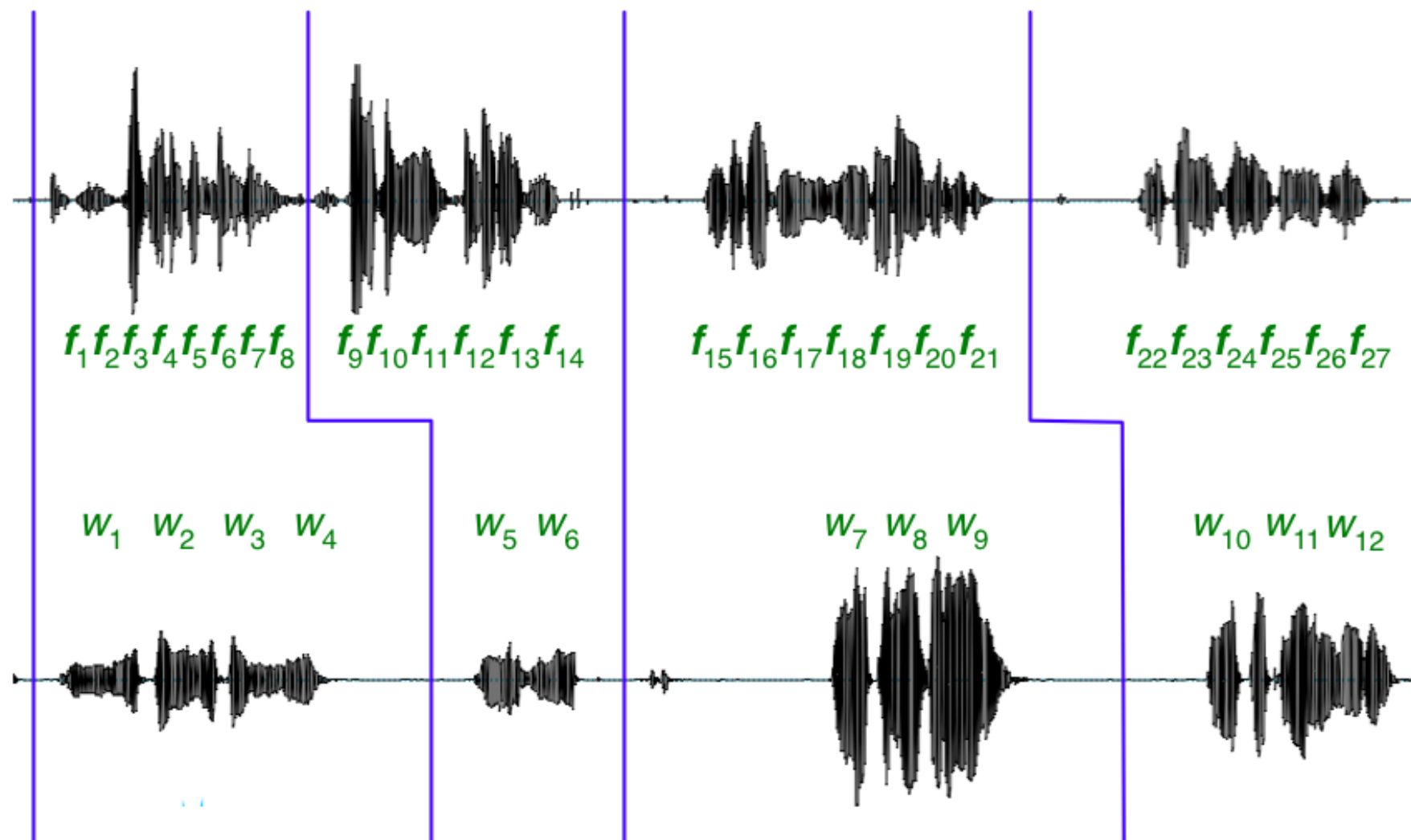
► not yet
implemented



Ongoing work: More features

- ▶ video recording and playback
 - ▶ commentary not yet supported
- ▶ Informed consent
 - ▶ storage and dissemination
 - ▶ documenting informed consent
- ▶ Lexical elicitation

Future work: applying MT algorithms



Conclusion

- ▶ what does it take to preserve a language?
- ▶ the allure of MT
- ▶ MT in a week
- ▶ contribute:
lp20.org

