

Challenges in creating speech recognition for endangered language CALL: A Chickasaw case study

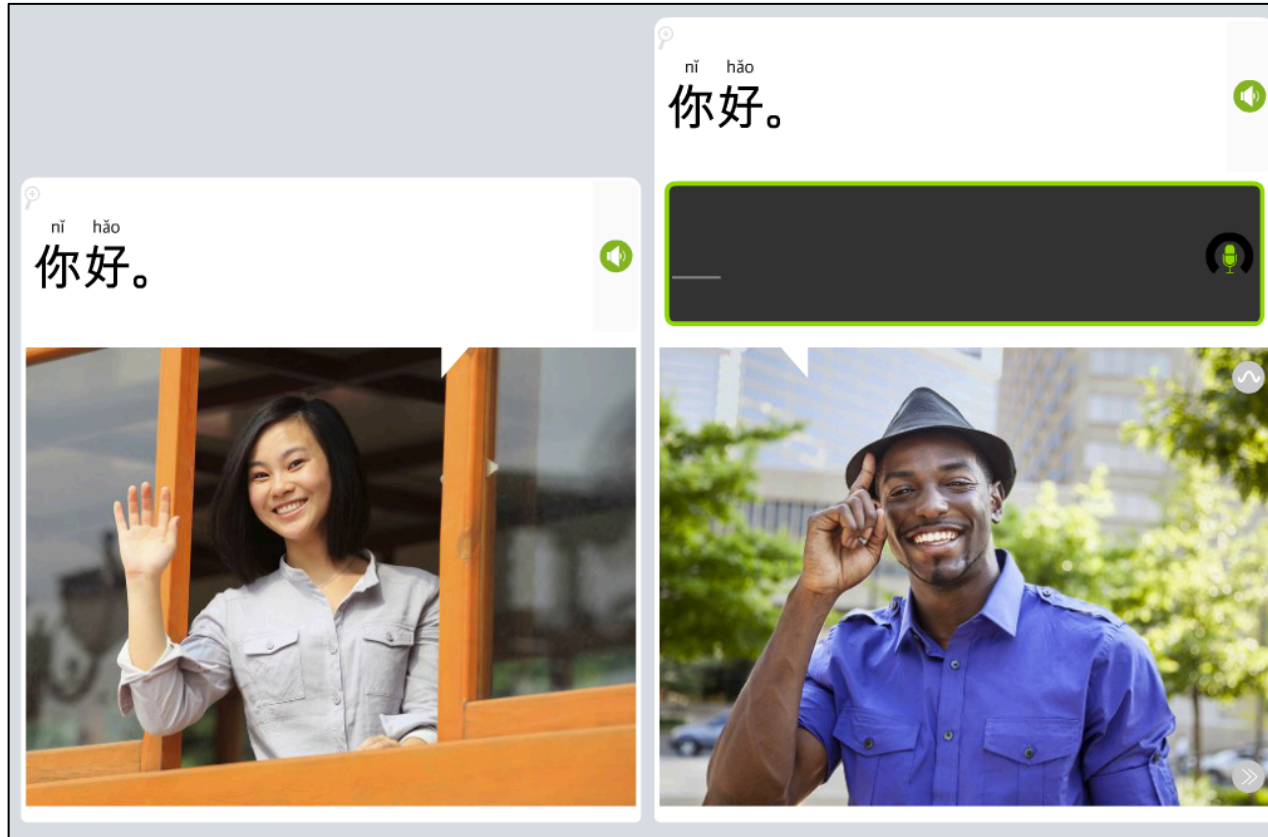


Danielle Lovaas
Yitagessu Gebremedhin
Emily Soward

Overview

1. Rosetta Stone and Chickasaw
2. Automatic Speech Recognition (ASR)
3. Resource collection overview and challenges
 - Speaker availability
 - Text resources
 - Speaker and network access
 - Technological familiarity
 - Target language literacy
 - Inter-speaker variation
4. Chickasaw Speech Recognition Engine (SRE)
5. Conclusions and future work

Rosetta Stone®



Chickasaw

Do you speak Chickasaw?

Chikashshanompa'
ishanompolita?

Hello.

Hallito. / Chokma.

Hello, how are you?

Hallito. / Chokma, chinchoma?

I am Chickasaw.

Chikashsha saya.

Have a nice day.

Chinittakat chinchokma'shki.

Native American language of the Muskogean family

- Closely related to Choctaw
- Now spoken mainly in Oklahoma

Endangered and under-resourced

- Around 50 native speakers remaining
- Active revitalization efforts since 2007
- Partnered with RS in 2015

<https://anompa.chickasaw.net/anompa/index.html>

Automatic Speech Recognition (ASR)

1. Language independent core engine
 - Software that processes input speech signal from user
2. Language dependent resources
 - Acoustic model set, lexicon, language model

Speech corpora

- High quality digital audio recordings
- Accurate text annotations
- High speaker variety

High performing ASR assumes abundant language resources are available: text, audio, and high speaker variety.

Endangered languages generally cannot offer the resources assumed by typical ASR: a different approach is needed and certain challenges must be met.

Audio Resource Collection


- 5 days on site; 5 speech team members to facilitate recording
- 4 recording stations
 - Laptop computer
 - Good quality headset
 - Internet access for ACT
- 1 hour sessions
 - Self-select levels from novice to native
 - 200 – 400 prompts per hour


In total, we collected 3,000+ recordings from about 60 participants of all ages and proficiency levels.


Audio Resource Collection


Audio Collection Tool (ACT)


Sachofata


 Listen to prompt

 Skip

 Record my voice

 Play my voice

 Note a problem

 Save

 Help

You have completed 0 out of 400 prompts

Challenges in Resource Collection

Speaker
availability

Text resources

Access:
physical and
network

Technological
familiarity

Target
language
literacy

Inter-speaker
variation

Speaker Availability

Considerations

How many native speakers, language learners, or children are available for recording?

Many endangered/under-resourced languages simply do not have enough speakers to gather the typical amounts of data

Strategies:

- Record native speakers multiple times
 - Gather more data while avoiding voice fatigue
- Use eligible native speaker archive data
 - UCLA audio corpus, material from the Chickasaw Nation

Speaker Availability

Considerations

How many native speakers, language learners, or children are available for recording?

Many endangered / under-resourced languages simply do not have enough speakers to gather the typical amounts of data

Strategies:

- Additional non-native and kids' data
 - Classify fluent or near-fluent speakers as “native”
 - Record as many novice speakers and kids as possible
- Any additional data manipulation as needed
 - Approximate kids' speech from female models

Text Resources

Considerations

Does the language have a standardized orthography?

How much text is available?

Endangered languages may not have the volume of text resources that exist for more widely-spoken languages.

Strategies:

- Use all appropriate and available text
 - UCLA corpus, Chickasaw children's books, Chickasaw dictionaries, grammars, and textbooks, various online sources
- Use product content, folded into the SRE
 - Free and simple way to generate more text
 - Level of the prompts will be appropriate

Physical and Network Access

Considerations

Where does the community exist: centralized location, or dispersed?

Is there internet connection?

Remote or dispersed communities will be more difficult to record;
may not have internet connectivity.

Strategies:

- Go on-site for data collection whenever possible
 - Most cost effective and efficient
 - More control over recording quality
- Internet connectivity
 - Bring a local server if connection may be slow
 - ACT needs reliable, fast connection

Technological Familiarity

Considerations

Is the speaking community relatively familiar with technology?

Is long-term data collection possible?

Speakers not comfortable or familiar with tech will need additional recording assistance throughout the session

Generally elders; always children

Strategies:

- A larger team may be necessary, or more time for recording
- Train speakers on the ACT
 - May be able to help facilitate
 - Allows for remote, ongoing data collection

Target Language Literacy

Considerations

Are speakers literate in the target language?

Speakers from an oral tradition and languages with multiple / unstandardized orthographies may not be able to rely on the ACT text prompts alone

Strategies:

- Always include native-spoken audio reference prompts
 - Also benefits beginning or children speakers – simply listen and repeat
 - Shorter text prompts so speakers can repeat from memory
- Pre-record native speakers whenever possible
 - Also allows for valuable input on prompt text before recording

Inter-Speaker Variation

Considerations

Is there a high level of inter-speaker variation within the language?

Ideally, the SRE recognizes all allowable variations in speech across speakers - but the greater the variation, the more difficult to capture within speech models

Strategies:

- Collect as much and as varied data as possible
- Collapse certain rare pronunciations, or
- Add all possible pronunciation variants

Challenges exist in resource collection and ASR building for endangered languages – but solutions are always available

Next Steps: Chickasaw SRE

HMM/GMM framework

Resources:

- Phone set
 - Defines phonemes using IPA
- Text normalization rules
 - How text is to be processed
- Lexicon
 - 3,500 + words and IPA transcriptions

Chickasaw	IPA
'aysha	ʔ aɪ ʃ a
a'hi	a ʔ h i
a'ma	a ʔ m a
a'shna	a ʔ ʃ n a
aabi	a b i

Training:

- Prepared corpus is used to generate acoustic models
 - Lexicon and language rules model the target language

Conclusions and Future Work

- Typical methods of corpus creation and resource gathering cannot be taken for granted with endangered languages
- Quality ASR is nevertheless possible with creative solutions to augment data
- Ongoing data collection to continuously improve recognition
- 6 considerations help frame the scope of the work, define challenges, and present solutions
 - Will also guide steps for future ASR work for endangered / under-resourced languages



Questions?



Danielle Lovaas: dlovaas@rosettastone.com

Yita Gebremedhin: ygebremedhin@rosettastone.com

Emily Soward: esoward@rosettastone.com