Speech Synthesis for the Chuvash Language

Dante Razo

Department of Linguistics at

Indiana University Bloomington

Background

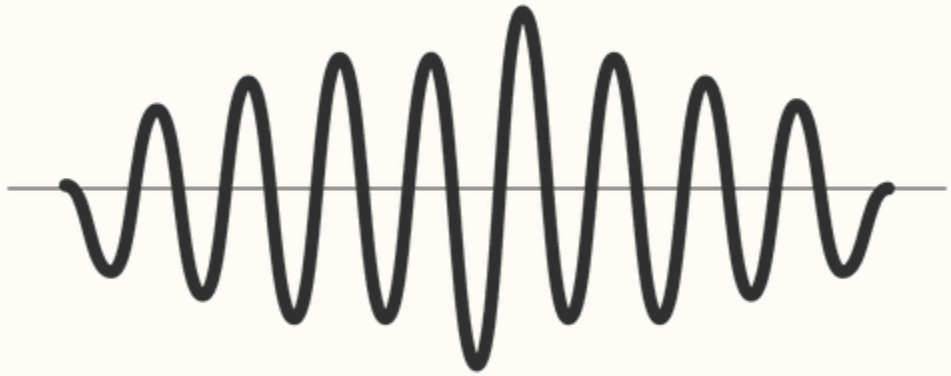
- Chuvash (Чăвашла) is a minority language spoken by roughly one million people in European Russia
- Turkic language that utilizes the Cyrillic alphabet
- This project aimed to train popular speech-synthesis systems and compare them

Introduction to Speech Synthesis

- Speech synthesis is the production of artificial human speech
  - e.g. Bloomington Transit & IU buses, Google Assistant, Amazon Alexa
- Text-to-speech (TTS) is a subset of speech synthesis
  - Self-explanatory name
  - Take text, parse it, conduct linguistic analysis on it, then produce audio waveforms
  - e.g. Microsoft Sam, NOAA Severe Weather Alerts, Stephen Hawking

Speech Synthesis Systems

- Used:
  - Ossian & Merlin
- Attempted:
  - eSpeakNG
  - Mozilla TTS
  - Mozilla LPCNet
- Considered:
  - Festival



Corpora & Repositories

Corpora:

- ***Turkic_TTS*** by Francis M. Tyers (ftyers)
- ***Apertium-chv*** (GPL-3.0) by Apertium (apertium)

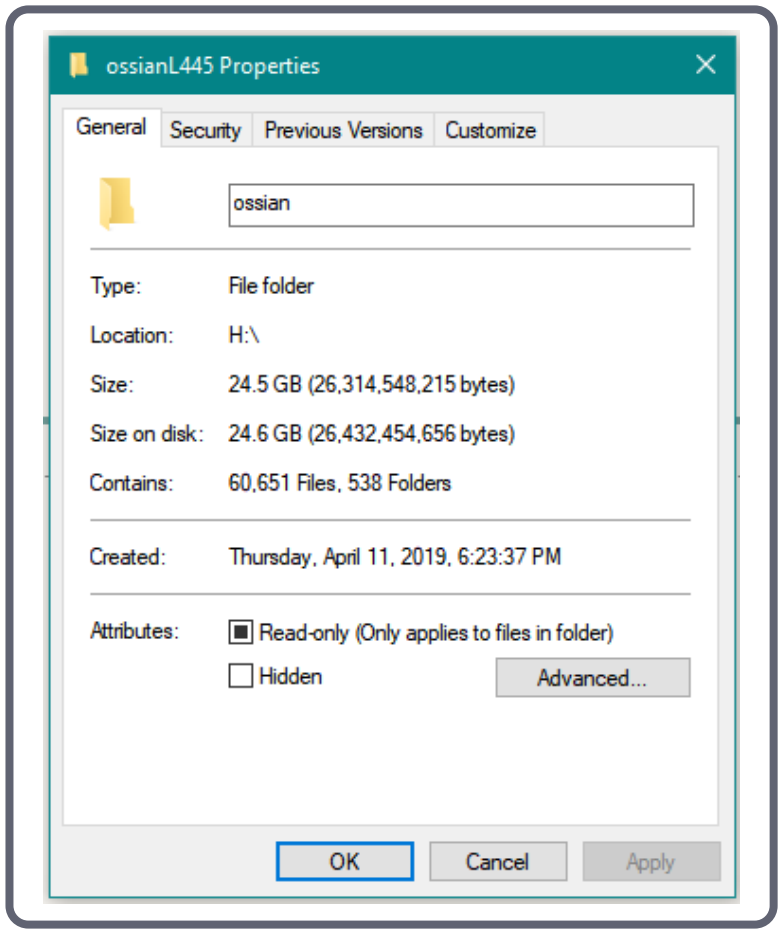
Repositories:

- ***eSpeakNG (-cv)*** by Harry Zhang (contextualist)
- ***Mozilla TTS*** (MPL-2.0) by Mozilla (mozilla)
- ***Ossian*** (Apache-2.0) by CSTR Edinburgh (cstr-edinburgh)
- ***Mozilla LPCNet*** (BSD-3) by Mozilla (mozilla)

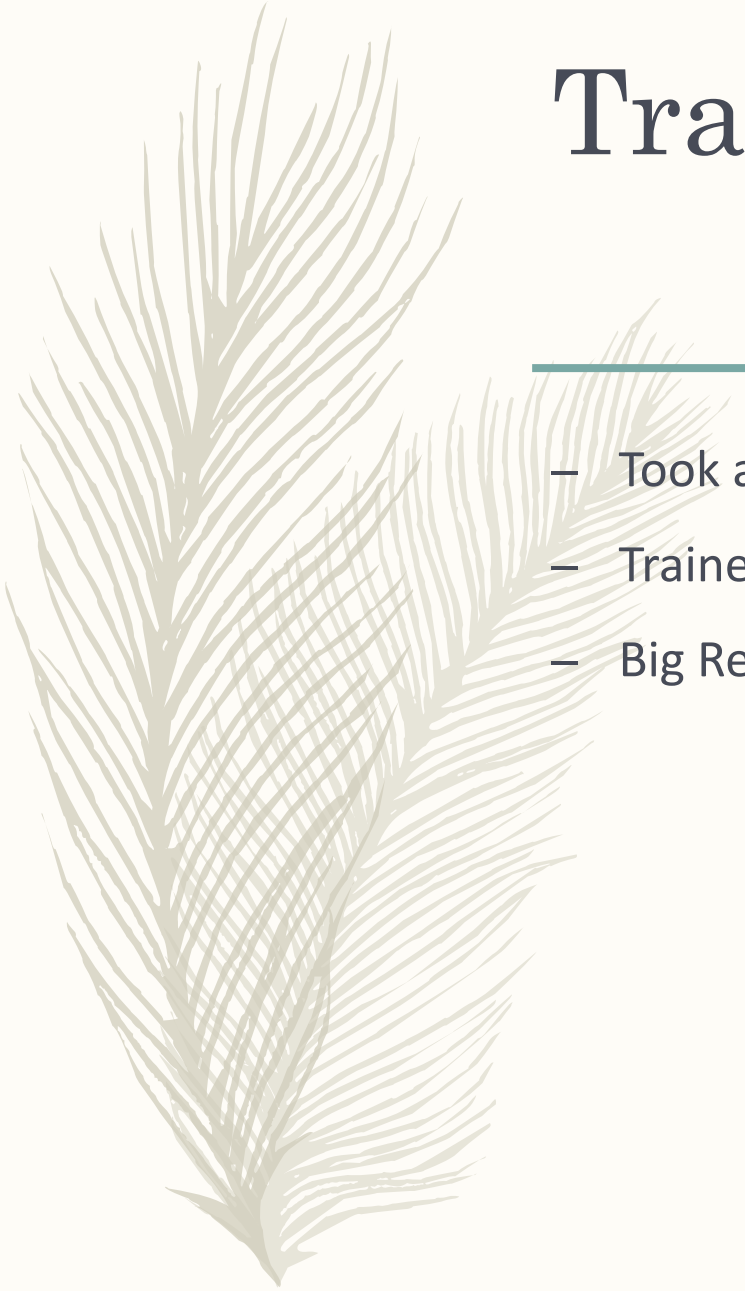
Ossian & Merlin

- Ossian is a front-end for speech synthesis development
  - Developed by the Centre for Speech Technology Research (CSTR) at The University of Edinburgh
- Merlin library is for neural-net based speech synthesis
  - Uses a Deep Neural Network (DNN)
  - Developed by the same team (CSTR)

Data Preprocessing



- Extracting *Turkic_TTS*' data increased its size by a factor of 8
  - 3GB -> 24GB
  - Both text and audio from Chuvash-language news clips
- Steps:
  1. Remove trailing ends from files (where silence was most common)
  2. Segment audio files and pair with transcriptions
  3. Match audio and text by renaming files



Training Ossian

- Took a few hours
- Trained on Ubuntu 18.04 in an 8-core virtual machine with 8GB of RAM
- Big Red II ambitions

eSpeakNG

- Formant synthesis
- English & Spanish TTS work perfectly
- Unable to test Chuvash due to corrupted installation of custom repo
  - Tested with ***sudo apt-get install espeak-ng*** (package manager)

Model Evaluation & Results

- Ossian produced good-sounding Chuvash from *Apertium* text corpora samples
- LPCNet, given the correct type of data, would've likely worked just as well
- Unable to test eSpeak due to compilation and installation issues

Examples

- “Ытти чĕлхесемпе пĕрлех ку хатĕрте чăваш чĕлхи валли те вырăн тупăннă”
  - From *apertium_chv* corpus **(/texts/cvorg-commonvoice.txt**, line 2)
- “Dante was here”
  - English (United States)
- “Dante estaba aquí”
  - Spanish (Latin American)

Future

- Get more than one neural-net based system working
- Implement Chuvash in eSpeak-NG to have a non-NN point of comparison

Special Thanks

- Harry Zhang

Special Thanks

- Harry Zhang
- And viewers like you. Thank you.

References

- Russian Bureau of Statistics: Владение Языками Населением Российской Федерации [*Population of the Russian Federation by Languages*]

http://www.gks.ru/free_doc/new_site/perepis2010/croc/Documents/Vol4/pub-04-05.pdf

- Wikipedia: Speech Synthesis

https://en.wikipedia.org/wiki/Speech_synthesis

- Wikipedia: Chuvash Language

https://en.wikipedia.org/wiki/Chuvash_language

Fin