Hannah Loukusa

Major

Saint Cloud State University

Saint Cloud, United States

hlloukusa@stcloudstate.edu

**Abstract – In this paper we describe a method for extracting vocal features from an audio recording such that speech may be reconstructed from these vocal measurements while maintaining emphasis and stress with minimal annotation. We also present a script for the Linguistic research tool PRAAT for performing this extraction and reconstruction via Klatt synthesis. [1] This technique contributes to the ongoing research of Dr. Ettien Koffi and Dr. Mark Petzold to preserve the Betine language, one of 10 critically endangered languages spoken in Côte d’Ivoire, West Africa.**

***Keywords–Speech Synthesis, Formant Synthesis, Formant Extraction, Formant Bandwidths, Endangered Languages, Betine***

**I. Introduction**

Many languages are rapidly going extinct, in that the number of native speakers has been or currently is rapidly decreasing. [2] Linguists are attempting to preserve these languages for their historic and cultural value through study ofspeech samples,identification of speech patterns, and of phonetic components. Currently, these preservation techniques require large amounts of data and researcher time to model the language, normalize text, and model these in software. [3] This paper details the implementation and use of PRAAT scripting to automate the extraction of this data, and to generate a KlattGrid object that captures these measurements. Given sufficient formant, bandwidth, intensity, and duration data for each phoneme, it should be possible to reconstruct an approximation of any possible combination of voiced phonemes within the language, including simple fricatives.

Jacob Haapoja

Computer Science

Saint Cloud State University

Saint Cloud, MN, United States

jlhaapoja@stcloudstate.edu

**II. Implementation**

The method consists of two parts which are performed simultaneously in the script. The script has two initial requirements. The first is a sample of speech audio, and the second is a PRAAT TextGrid with each phoneme to replicate marked with boundary intervals and labeled with a 1 or 0 to indicate frication. The script initializes with an information dialogue prompting for analysis parameters that will be used in the extraction and subsequent reconstruction. After this it will verify that preselection requirements are met, and initialize a new KlattGrid object with default settings.

**A. Automated Feature Extraction**

In this part, pitch, intensity, and formants are extracted using PRAAT’s built in algorithms. Then a count of intervals on the tier specified in the initial dialogue, along with the start and end times of each interval are taken from the TextGrid object. These are then used to average the pitch, intensity, and formants over those intervals. Measurements for frequency and bandwidth are taken in Hz while intensity is taken in dB. This is done with a for loop that counts through the number of intervals and extracts the relevant data into variables that are then passed to the next part.

# paste example code excerpt

**B. KlattGrid Construction**

This part is performed simultaneously to the data extraction as the script loops through the intervals. At each interval, the data extracted is used to add a new grid point in the KlattGrid object at the start time of that interval. If a particular interval is marked as fricative by having a 1 for the interval label, the intensity of that region is added as a frication amplitude point which is scaled down to 90%. This is done to soften the raw frication generated by the Klatt synthesizer. Conversely, if the interval is marked as non-fricative with a label of 0 the intensity is still added as a fricative amplitude point, but instead scaled down to 10%. This adds some slight variation to the produced sound which results in a smoother and less robotic sounding reconstruction.

# paste example code excerpt

**III. Results & Analysis**

**IV. Conclusion**

**V. References**

**[1]** Dennis H. Klatt & L C Klatt. 1990. Analysis, synthesis, and perception of voice quality variations among female and male talkers. The Journal of the Acoustical Society of America. Vol. 87, 2, 820-57, 1990

**[2]** Hale, Krauss. 1992. Language Endangerment and the Human Value of Linguistic Diversity. Language 68 (1) 35-42.

**[3]** Koffi, Ettien & Petzold, Mark. (2022). A Tutorial on Formant-Based Speech Synthesis for the Documentation of Critically Endangered Languages. Volume 11. 26-55.

**Appendix**

**A. Program Code**

**B. Design**

**C. Testing**

**D. User Manual**