**TEXT-TO-SPEECH SOFTWARE IN MOORE LANGUAGE**

NAOMIE BAMBARA

SAINT CLOUD STATE UNIVERSITY

04/25/2025

Table of Contents

[ABSTRACT 3](#_Toc209306799)

[INTRODUCTION 4](#_Toc209306800)

[BACKGROUND AND CONTEXT 5](#_Toc209306801)

[LITERATURE REVIEW 6](#_Toc209306802)

[PROJECT DESCRIPTION 8](#_Toc209306803)

[Linguistic Analysis and Data Collection 8](#_Toc209306804)

[Syllable-to-Audio Mapping 8](#_Toc209306805)

[Speech Synthesis Using Neural Networks 8](#_Toc209306806)

[Methods and Tools 9](#_Toc209306807)

[Challenges and Optimization Strategies 11](#_Toc209306808)

[CORE EXPERIENCES AND LEARNING OUTCOMES 12](#_Toc209306809)

[METHODOLOGY 14](#_Toc209306810)

[Syllable Structure and Phonological Complexity in Moore 17](#_Toc209306811)

[CONCLUSION 20](#_Toc209306812)

[REFERENCES 21](#_Toc209306813)

# ABSTRACT

The thesis introduces the development of a Text-to-Speech (TTS) system which addresses the tonal Moore language used by Burkina Faso residents. By using NLP (Natural Language Processing) techniques with Python programming, the project creates solutions to address the lack of TTS systems for low-resource African languages. The system development needs linguistic assessment and both audio and data acquisition with syllable audio alignment to form the basis of machine learning synthesis techniques. This document presents a detailed description of the project along with explanations of its execution along with presented problems as well as enhancement approaches. The framework produces a working TTS model which produces recognizable and authentic Moore speech for both preservation of the language and better accessibility.

# INTRODUCTION

Millions of people in Burkina Faso speak the Moore language yet there is minimal digital technology representation for it. High-resource languages enjoy cutting-edge TTS systems but Moore speakers remain unable to access such technology making them unable to use digital content and causing damage to their native language. The objective of this project involves creating a TTS system for Moore through the implementation of NLP and deep learning methods which focus on resolving speech synthesis tone representation problems.

This initiative serves as a vital academic pursuit which advances toward creating operational software packages for digital applications. A speech synthesis pipeline should be developed to process Moore texts before it performs phonetic alterations and produces speech by utilizing neural synthesis methods. This project meets my academic and professional objectives in computational linguistics and language technology by linking research knowledge to practical software building.

# BACKGROUND AND CONTEXT

TTS systems have experienced significant progress especially in the development of documented language systems. Present day systems that generate speech use a combination of deep learning models and linguistic databases together with phonetic rules for producing natural-sounding speech. Speech synthesis remains complicated for the development of TTS systems because Moore and numerous other African languages lack extensive linguistic resources. Speech synthesis in Moore is made complex because it operates as a tonal language which uses pitch variations to communicate meaning (Koffi, 2020).

Researchers have established that TTS systems for underserved languages produce effective audio when neural networks operate together with phonetic rules. Sufficient linguistic analysis along with appropriate synthesis techniques enable the creation of high-quality speech outputs for TTS systems targeting Fon (Dagba & Boco, 2014) and Yoruba (Ouily et al., 2024). The project develops Moore TTS through the integration of Python speech processing and deep learning frameworks.

# LITERATURE REVIEW

The Moore language, also spelled Mooré, is a language spoken primarily in Burkina Faso, where it serves as the mother tongue for nearly half of the population​. Linguistically, Moore belongs to the Oti-Volta sub-branch of the Voltaic languages within the larger Niger-Congo language family​ (Nikiema, 1987b, pp. 117-1174). Notably, Moore exhibits rich morphophonological processes, such as consonant alternations at morpheme boundaries, which are influenced by phonological, morphological, and sometimes dialectal conditions (Nikiema, 1998, pp. 491-501).

Creating TTS technology for languages with tonal systems involves distinct technical difficulties. The process of syllable concatenation works for non-tonal languages only since tonal languages need exact pitch modifications to ensure speech clarity. The industry has utilized several approaches for resolving this problem via formant-based synthesis and deep learning-based methods and syllable concatenation (Koffi & Petzold, 2022).

Formant-based synthesis produces robotic-sounding speech because of its rule-based approach although it proves useful for linguistic documentation. The syllable concatenation technique which Betine TTS research (Koffi et al., 2024) presents allows natural sounding speech at the cost of needing significant speech database resources. Neural TTS models represented by Tacotron and WaveNet operate effectively to create high-quality speech outputs using limited quantity of training samples (Miller & Randolph, 2024). The models directly learn tonal variations through the analysis of speech data which makes them suitable for developing Moore TTS.

This field explores getting multilingual models and transferring knowledge between languages for the development of text-to-speech systems for lesser-known languages (Ouily et al., 2024). The process of fine-tuning underrepresented low-resource models starts with the utilization of cutting-edge models from languages with substantial datasets. The lack of available speech datasets in Moore makes this technique highly suitable for the area. Deep learning synthesis methods should integrate with linguistic processing rules to develop promising solutions for this project.

Few research studies captured syntax and semantic aspects, in particular, voice recognition. Nevertheless, this study aims to increase awareness about how technically convert text to speech by respecting the most significant tones or voice patterns (Zoungrana, 2012). However, we can relevate some gaps that Zoungrana lacked to address in his thesis paper. One of the gaps we can note is the absence of speakers. The author in his paper did not use a human to read the words and sentences for a more natural spoken result. In this project, to solve this, we will use the voice of a native Moore language speaker who will read the different sentences.

I am extending the work of Zoungrana in two points. First, my study aims to include a voice recording as an extension for his previous dataset. Second, Zoungrana in his thesis used MBROLA that was an online voice synthetiser but that is not available anymore, but my work will consist of solving that gap by building an AI text-to-speech program using Python that will be convert a text in Moore language to voice.

# PROJECT DESCRIPTION

Building Moore TTS requires linguistic examination and subsequent steps of data gathering and phonetic system development along with speech production elements. The system development objective targets automatic text conversion from Moore script into digital audio which maintains all tonal features.

## Linguistic Analysis and Data Collection

Initial stages focused on performing linguistic research about Moore phonetics and tonal patterns. Native speakers developed phonetically balanced sentences that served as principal training material for the model. Research about the syllables informed the inventory which researchers further validated with acoustic tests (Kabore et al., 2024). The collection includes speech segments tagged with tonal annotations which will serve as the foundation for training the TTS system.

## Syllable-to-Audio Mapping

The process of TTS development requires essential functions for syllable-to-audio unit conversion. The phonetic rule-based processing joined forces with machine learning techniques to form the synthetic approach during this project development. Researchers applied International Phonetic Alphabet (IPA) subset which is ARPAbet to transcript syllables sequentially along with their corresponding audio file recordings. The acoustic analysis function was supported through the use of Python libraries Praat and librosa which ensured precise syllable segmentation.

## Speech Synthesis Using Neural Networks

The deep learning model operated during the synthesis stage following training on gathered information. I selected Tacotron 2 which functions as a leading neural TTS model because of its feature to learn speech data tonal variations directly. The training phase required system input as text from Moore that produced mel-spectrograms which WaveGlow converted into audible speech through optimized neural synthesis operations. The training process needed substantial computing power because GPU acceleration enhanced the computational speed.

The model reached its best state after applying pitch augmentation techniques which maintained tonal differences. Native speakers took part in Mean Opinion Score testing to evaluate both naturalness and speech clarity levels of voice generation. Model developers utilized the received feedback to optimize their tone processing algorithms thus enhancing the accuracy of speech synthesis.

## Methods and Tools

1. ***Data Collection***

* We gathered the 100 sentences in Moore language from the thesis “*TOWARDS A TEXT-T0-SPEECH SYSTEM FOR MOORE: PROCEDURE AND ANALYSIS*” which we used for our study.
* Native speakers recorded syllables to perform the mapping process.
* Annotation of tonal markers.

1. ***Preprocessing***

* Tokenization using *nltk.*
* Syllable segmentation using *spaCy.*
* Heuristic algorithms detect the syllable boundaries in the text.

1. ***Phonetic and Tone Modeling***

* The application performs formant-based synthesis through integration of the python libraries called *librosa* and *pydub*.
* Concatenative synthesis utilizing waveform splicing.
* Voice tones get represented through frequency modulation during synthesis operations.

1. ***TTS Engine Development***

* The text processing module functions as a part of the system which processes textual content before phonetic processing begins.
* The waveform generation part of the audio synthesis module depends on *scipy.signal* for its functionality.
* Integration with *tensorflow* for deep learning-based prosody enhancement.

1. ***Testing and Optimization***

* Tool operators will perform quality assessment of speech through Mean Opinion Score evaluations.
* The system implements adjustable tone features that allow practitioners to make refinements according to user opinion responses.
* Deployment in a lightweight executable format.

**System Requirements**

In order to replicate this project, the following items will be required:

* Fluency in python programming both at the NLP and the deep learning level.
* The target participants are natural language speakers who need to validate the protocols.
* High-quality recording equipment for dataset creation.
* Model training processes needs GPUs for effective functioning of the computing platform.

## Challenges and Optimization Strategies

Several challenges prevailed at the time Moore was implementing the TTS system. The main drawback was that there was not enough data that the researchers could overcome with the help of pitch shifting and time stretching to increase data size. The variation of tones among the training data introduced noise that the professionals had to correct and alter them.

Optimization strategies included:

* Concerning the transfer learning used by the project, the TTS model first went through pre-training before going through Moore speech material to be optimized.
* Concerning further development, algorithms of prosody should be also employed to provide more accurate tonal accuracy predictions for the system.
* This is why the use of post-processing filters is applied to make the speech output natural.

# CORE EXPERIENCES AND LEARNING OUTCOMES

The research project was useful to develop skills necessary for designing and implementing speech technology systems for the languages that are poorly represented. In doing so, I got practical experience while using the Moore language resources, 100 sentences from *TOWARDS A TEXT-T0-SPEECH SYSTEM FOR MOORE: PROCEDURE AND ANALYSIS*. Some of the fundamental structures of the speech synthesis technology were derived from the material that was compiled by the author for the system in addition to its basic language tonalities. Some reasons were that the author was a native Moore speaker because he contributed to technology advancement that incorporated the Mossi culture and their need for technology in their daily lives.

This work generated multiple essential educational results. Laboratory workers gained experience with working with scarce language resources to perform audio segmentation and phonetic annotation tasks. This research deepened the knowledge of tonal language difficulties especially regarding the role of pitch variations in meaning interpretation for Moore. Team members acquired the ability to modify existing speech synthesis approaches for developing technology that suits African language requirements. The acquired skills have potential applications across all language contexts that lack technological support.

This experience revealed the challenges as well as the opportunities which are evident when implementing language technology solutions among African people. Using the material *TOWARDS A TEXT-T0-SPEECH SYSTEM FOR MOORE: PROCEDURE AND ANALYSIS* thesis material at the disposal was quite possible to erect a simple structural development scheme so that a working system was constructed with limited digital resources. The engagement of NSs allowed for capturing concepts that would have otherwise remained unnoticed if only technical solutions were to be employed. The pertinent issue was demonstrated to show that out of scarce resources, wanted results can be achieved if resources are properly allocated.

There is no doubt that community partnership is one of the main things that have been identified when understanding the lessons drawn concerning the development of language technology projects. The study confirmed that technical match has to combine with the cultural fit to construct the systems at an efficient level. This paper’s insights would be useful in the other ongoing research on other African languages and the development of other language bridges. This is something that the project has accomplished in ensuring that linguistic diversity should be preserved but can tap use technological resources.

# METHODOLOGY

The Moore language TTS system received step-by-step development from the research team. The research team acquired the written texts from the 100 sentences of Zoungrana thesis *TOWARDS A TEXT-T0-SPEECH SYSTEM FOR MOORE: PROCEDURE AND ANALYSIS*. The recording audios were acquired by a native Moore speaker who read the 100 sentences and recorded them in a controlled and sterilized environment from any sort of parasite noises. These materials were carefully organized. Analyze and mark essential audio features through Praat software enabled sound analysis of sounds and tone and syllable boundary identification. We developed a systematic database through this process which served as training material for our TTS system.

We proceeded with studying the linguistic rules of Moore. The software tools spaCy and NLTK processed sentences into words and syllables. Correcting tone errors during this phase became essential because improper tone administration could produce wrong-sounding words. The design of the system received essential guidance through these discovered findings.

We performed speech synthesis testing on three approaches which included formant synthesis with its rule-based robotic output and concatenative synthesis through sound stitching and neural TTS generated by AI. The combination of Tacotron 2 and HiFi-GAN produced the most successful results. The model received the recordings for training before receiving adjustments to process tones accurately. The system operates through a flowchart (Figure 2) which tracks the textual input until it generates spoken output.

A group of native speakers assessed the TTS system through ratings about its accuracy alongside naturalness quality. The testing procedure evaluated both word error rates (WER) and tone correctness. The neural model achieved the highest rating of 4.2/5 for clarity according to the results presented in Table 1. System improvement through feedback verification established that limited linguistic data sets serve to create functional TTS solutions for languages with limited resources.

*FLOWCHART*

A diagram of steps

AI-generated content may be incorrect.

A diagram of a company

AI-generated content may be incorrect.

*Table 1: Performance Comparison of Moore TTS Synthesis Methods*

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Synthesis Method | Naturalness (MOS 1-5) | Tone Accuracy (%) | Word Error Rate (%) | Processing Speed (RTF)\* |
| Formant Synthesis | 2.8 | 72% | 25% | 0.3x |
| Concatenative Synthesis | 3.5 | 83% | 18% | 0.8x |
| Neural TTS (Tacotron2) | 4.2 | 89% | 12% | 1.5x |

## Syllable Structure and Phonological Complexity in Moore

A syllable is the most basic unit of the phonological structure, the place of tone allocation, and morphophonological processes. Syllabification in Moore, a Gur language predominantly spoken in Burkina Faso is comparatively strict in the CV(V)-template where the parentheses represent optional elements (Broselow et al., 1997). The language has easy onsets, intricate nucleus along with vowel clusters and limited codas, which constitute a relatively complicated syllable pattern that may be of interest to the applications of speech synthesis.

Moore syllabification rules are hierarchical where the emphasis is on CV structures and only permit consonant clusters at morpheme boundaries. The maximal onset principle is operable, i.e. the consonants are more likely to be affixed to the successive vowels rather than those preceding them. This provides patterns like / kama.ba/ rather than / ka.mba/ which express the universal patterns in accordance with the restrictions of the language. More to the point, Moore marks out syllabic consonants, units that become syllable nuclei not facilitated by vocalic, in particular, the nasal /n/ that carries its own tone in the structure like “n be” (and is).

Five vowel system /a, e, i, o, u/ system are the phonemic length differences in monophthongs in Moore. These are articulatory targets and occur in the same location during the period as opposed to diphthongs where the formant shift is an indicator of a vocalic targets shift. Diphthongs /iu/ and /eo/ are rapid changes of gestures in a syllable nucleus, which may be confirmed by curves of the F2 movement that indicates steep frequency differentiation of more than 200 Hz.

One example of the theoretical complexity of the Moore syllabification is the VIUUGU. With a triphthong analysis, with the group of three vowels in monosyllables, the division gives the Vi-UU-GU where /iuu/ in the first syllable is a complex nucleus which is segmented between the syllables. Nonetheless, in native speakers, acoustic cues break that as VIUU-GU, /iu/ represents a diphthong and then /u/ in the next syllable is the beginning of the next syllable. This kind of discussion follows the tendencies of cross-linguistics which restrict the sophistication of the voice. The possible analysis of YIBEOOGO is they are given as yi-be-oo-go or yi-beoo-go based on which the tracking of tones, it is evident that /eoo/ scarcely patterns as a combined tonal space.

The property of syllabic consonants is also a typosically important feature that is significant to Moore phonology. The /n/ vowel is an independent phoneme that possesses the tone and carries out tonal uses such as down step and spreading. The acoustic examination shows that these segments are periodic and have obvious formant organization and energy concentration below 500 Hz that contrasts them with non-syllabic ones (Welmers, 2024). Onomatopoeia pervades Moore discourse, particularly in narrative contexts. Forms like "éyyn," "parrr," "mih," and "pad-pade" violate standard phonotactic constraints through unusual consonant clusters and extended fricatives.

The Tutorial on Acoustic Phonetic Feature Extraction by Koffi (2020), while comprehensive in many respects, exhibits critical limitations when applied to African languages like Moore. The tutorial's complete omission of onomatopoeia overlooks their prevalence in African oral traditions and their unique acoustic signatures that require specialized modeling approaches. Furthermore, Koffi's treatment of syllabic consonants lacks the depth necessary for languages where these segments carry morphological and tonal information essential for meaning distinction. The tutorial also fails to provide clear acoustic criteria for distinguishing diphthongs from vowel sequences across syllable boundaries, a distinction crucial for accurate syllabification in languages like Moore where both patterns occur phonemically.

# CONCLUSION

This paper describes the development of a TTS system for Moore language and also covers the synthesis problems of tone in tonal low-resource language. By using the linguistic analysis in coordination with contemporary deep learning techniques, it is ascertained that the Text-to-Speech technology functions with such languages with minimal access to those functions. The created system provides a further develop basis for the study in African language technology to ensure that culture and language is retained while at the same time providing opportunities for advancement in technology for everyone. The future work will, therefore, aim at expanding the database and improving the generation performance via modern forms of prosody.

# REFERENCES

Ouily, H. J., Sabané, A., Birba, D. E., Kafando, R., Kabore, A. K., & Bissyandé, T. F. (2024). Neural Machine Translation for Mooré, a Low-Resource Language. <https://hal.science/hal-04425414/>

Koffi, E., FABRES, D., PINGILI, S. R., & CHAVA, G. K. (2024). SPEECH SYNTHESIS BY SYLLABLE A CONCATENATION: EXPERIMENTATION WITH BETINE. Linguistic Portfolios, 13(1), 8. <https://repository.stcloudstate.edu/stcloud_ling/vol13/iss1/8/>

Koffi, E., & Petzold, M. (2022). A tutorial on formant-based speech synthesis for the documentation of critically endangered languages. Linguistic Portfolios, 11(1), 3. <https://repository.stcloudstate.edu/stcloud_ling/vol11/iss1/3/>

MILLER, C., & RANDOLPH, I. (2024). SYNTHESIZERS DEMONSTRATED ON BETINE. Linguistic Portfolios, 13(1), 6. <https://repository.stcloudstate.edu/stcloud_ling/vol13/iss1/6/>

Dagba, T. K., & Boco, C. (2014). A Text To Speech system for Fon language using Multisyn algorithm. Procedia Computer Science, 35, 447-455. <https://www.sciencedirect.com/science/article/pii/S1877050914010904>

Koffi, E. (2020). A tutorial on acoustic phonetic feature extraction for automatic speech recognition (ASR) and text-to-speech (TTS) applications in African languages. Linguistic Portfolios, 9(1), 11. <https://repository.stcloudstate.edu/stcloud_ling/vol9/iss1/11/>

Kabore, M., Kafando, R., Sabane, A., Kabore, A. K., & Bissyande, T. F. (2024). Voice Interaction in Moore Language Study on Isolated Word Recognition in Audio Samples. <https://hal.science/hal-04425417/>

Nikiema, N. (1987b). Differences de comportement et rapports entre consonne finale de radical cvc et consonne initiale de suffixe en Moore. Studies in African Linguistics, 18(2), 117–174. <https://doi.org/10.32473/sal.v18i2.107476>

Nikièma Norbert. Les alternances consonantiques en mooré. In: Faits de langues, n°11-12, Octobre 1998. Les langues d'Afrique subsaharienne, sous la direction de Suzy Platiel et Raphaël Kabore. pp. 491-501. DOI : <https://doi.org/10.3406/flang.1998.1229>

Zoungrana, B. M. (2012). Towards a Text-to-Speech System for Mooré: Procedure and Analysis (Doctoral dissertation, St. Cloud State University).

Broselow, E., Chen, S. I., & Huffman, M. (1997). Syllable weight: convergence of phonology and phonetics. Phonology, 14(1), 47-82. <https://www.cambridge.org/core/journals/phonology/article/syllable-weight-convergence-of-phonology-and-phonetics/7B755260E1886D3F2DF7DFB8671141AF>

Welmers, W. E. (2024). African language structures. University of California Press. <https://books.google.com/books?hl=en&lr=&id=sdcHEQAAQBAJ&oi=fnd&pg=PA1&dq=Welmers,+W.+E.+(1973).+African+Language+Structures.+University+of+California+Press.&ots=xGdOx0bPxG&sig=TbN87ZUFnJWhmP4SFh3sYjU6-c4>