SSRG International Journal of Electrical and Electronics Engineering
ISSN: 2348-8379 / https://doi.org/10.14445/23488379/IJEEE-V10I2P107

Volume 10 Issue 2, 75-83, February 2023

Original Article

Development of a Kiswahili Text-to-Speech System based on Tacotron 2 and Wave Net Vocoder

Kelvin Kiptoo Rono[1], Ciira Wa Maina[2], Elijah Mwangi[3]

[1]*Department of Electrical and Electronic Engineering, Dedan Kimathi University of Technology, Nyeri, Kenya.*
[2]*Center for Data Science and Artificial Intelligence, Dedan Kimathi University of Technology, Nyeri, Kenya.*
[3] *Faculty of Engineering, University of Nairobi, Nairobi, Kenya.*

[1]*Corresponding author: kiptookelvin96@gmail.com*

Received: 07 January 2023 Revised: 07 February 2023 Accepted: 16 February 2023 Published: 28 February 2023

Abstract *- Text-to-Speech (TTS) system converts an input text into a synthetic speech output. The paper provides a detailed description of developing a Kiswahili TTS system using Tacotron 2 architecture and WaveNet vocoder. Kiswahili TTS system will help visually impaired persons to learn, assist persons with communication disorders, and provide a source of input in Voice Alarm and Public Address equipments. Tacotron 2 is a sequence-to-sequence model for building TTS systems. The model consists of three steps: encoder, decoder, and vocoder. The encoder step converts the input text into a sequence of characters. The decoder predicts the sequence of Mel-spectrograms for each generated character sequence. Finally, the vocoder transforms the Mel-spectrograms into speech waveforms. The Kiswahili TTS system developed based on Tacotron 2 obtained a mean opinion score (MOS) of 4.05. The score shows that the speech generated by the system is comparable to human speech. The system implementation is available in the link below https://github.com/Rono8/kiswahili_tts*

Keywords *- Kiswahili text-to-speech system, Language modeling, Natural language processing, Speech processing, Text-to-speech system.*

1. Introduction

Kiswahili is a regional language in the Eastern African region. The countries speaking Kiswahili include Kenya, Burundi, Comoros, Tanzania, Uganda, Rwanda, and parts of Sudan and Ethiopia [1]. Millions of Kiswahili speakers provide a larger market for Kiswahili software products. Kiswahili Text-to-Speech (TTS) system will help in the following areas:

- Helping visually impaired persons to learn.
- Input to Public Address (PA) and Voice Alarm (VA) systems.
- Helping people with communication impairments.
- Application to telecommunication devices to read Kiswahili text messages.

The research contributes to advancements in speech processing applications by developing a TTS system based on state-of-art technologies.

2. Related Works

There are a number of approaches to developing TTS systems. Some of these include [2]:

- Parametric TTS
- Concatenative TTS
- Artificial Neural Network (ANN)-based TTS system

Mean Opinion Score (MOS) is used to evaluate speech quality. MOS is a number between 1 to 5, 1 being poor quality and 5, the excellent quality. The score checks the intelligibility and naturalness of the TTS system. Intelligibility evaluates how the system speech output matches the input text. Conversely, the system's naturalness analyzes the easiness of listening to the generated speech. Table 1 shows the MOS for the three approaches.

2.1. Concatenative Text-to-Speech System

Concatenation TTS generates speech output by concatenating parts of the pre-recorded speech. The pre-recorded speech units can be sentences, words, diphones, or syllables [5,6]. The method chooses one unit from the speech corpus best fits the required output [7,8]. The speech unit is selected from the target and concatenation cost. Concatenation features determine how two speech units match after being concatenated. Target cost is made up of duration, pitch, and energy variables. Target cost evaluates these variables to determine if the predicted speech waveform matches the required output.

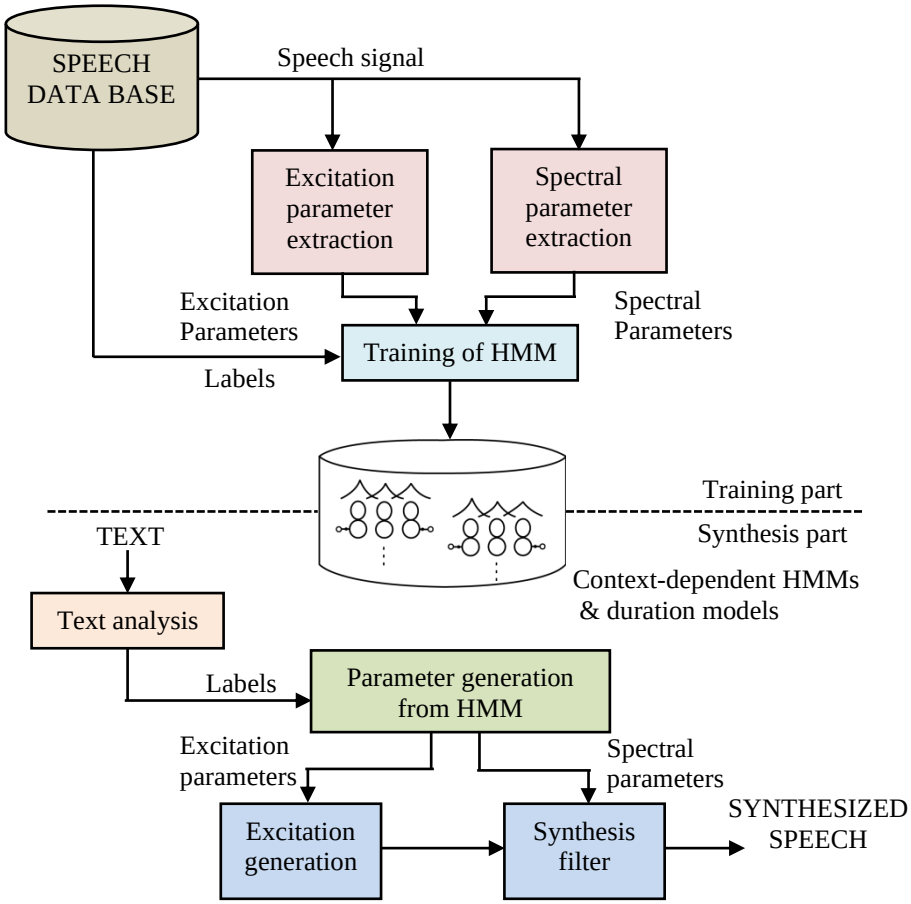

Fig. 1 HMM-based speech synthesis [9]

Table 1. Mean opinion scores for speech approaches [3,4]

Text-to-Speech System	MOS
Statistical-Parametric	3.49±0.096
Concatenative	4.17±0.091
Tacotron 2	4.52±0.051
Professional Speaker	4.58±0.053

The best Concatenation TTS has been made for North American English with a MOS of 4.17. This shows that the concatenation approach performs better in speech quality and intelligence. However, the challenge with the concatenation approach is that it requires a larger dataset to build the system.

2.2. Statistical Parametric Text-to-Speech System

There are several techniques for developing a Statistical Parametric TTS-based system. Hidden Markov Model (HMM), a statistical parametric speech synthesis technique, generates speech output effectively. HMM TTS has two parts: training and synthesis parts. The training part checks excitation and spectral parameters of speech. Excitation parameters include dynamic features and fundamental frequency. Speech spectral parameters are the MFCCs [9,11]. The synthesis part generates the parametric sequence for each text input. Figure 1 shows an HMM-based speech synthesis consisting of the training and the synthesis part. The system generates HMM characters for each input text, which the system uses to create the corresponding spectrum and excitation parameters. Finally, the synthesis converts the excitation and spectrum features to a time-domain speech waveform. The best work in statistical parametric TTS has been done for US English, which had a MOS of 3.47 [3,10], showing fair speech quality.

2.3. Artificial Neural Network Text-to-Speech Systems

Neural TTS first trains the Deep Neural Network (DNN) model on human recordings. The generated speech output voice sounds like the input data used in the recording. Training of the system associates the audio recordings with the corresponding text. Some ANN-based approaches for constructing a TTS system include WaveNet, Tacotron, Deep voice and Character-to-Wave.

2.3.1. Wave Net

WaveNet is autoregressive model. It predicts the next speech output based on the previous speech samples [3,26]. Therefore, the probability score of the next speech sample is based on the previous time steps, as shown by equation 1.

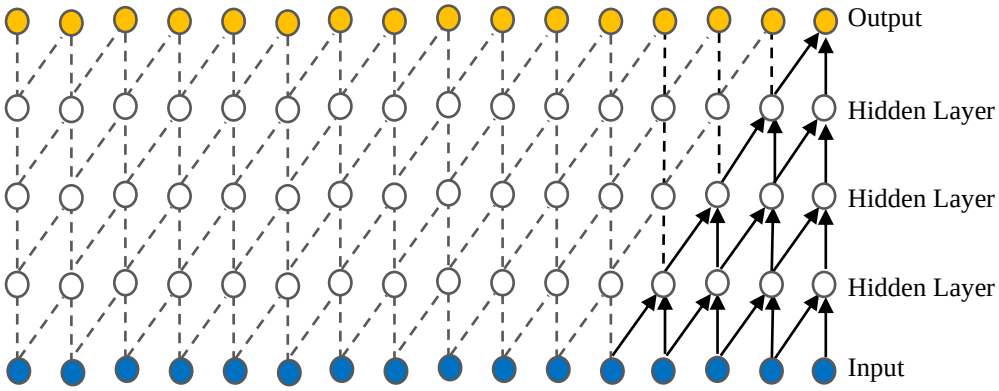

Fig. 2 Wave net stacked convolutional layers [3]

$$x_\lambda=\{x_1,x_2,x_3,\ldots,x_{t-1}\} \quad (1)$$

Thus, for each new input character, λ, the joint probability is given by equation 2.

$$p(x|\lambda)=\prod p(x_\lambda|x_1,x_2,x_3,\ldots,x_\lambda,\lambda) \quad (2)$$

Convolution Neural Networks (CNNs) are used to check the previous time step conditions and predict the next audio sample. Figure 2 shows the dilated convolution layers used to check the previous time steps [12-15]. Finally, WaveNet uses softmax distribution to output the probability score of the predicted audio sample, which then gives the best prediction for a given input text.

2.3.2. Character-to-Wave

The system implementation is similar to WaveNets. It is also an autoregressive model. However, unlike WaveNets, Character-to-Wave uses Recurrent Neural Networks (RNNs) instead of CNNs to check previous time steps [16, 17,27].

2.3.3. Tacotron

Tacotron 2 is a Sequence-to-Sequence model. It converts the character sequence to the Mel-spectrograms sequence. During training, the text files are converted to a character sequence. Next, the audio files are converted to Mel-spectrograms. Each character sequence is assigned to the corresponding Mel-spectrograms sequence. Inputting the text, the system generates character sequences to predict the Mel-spectrograms. Finally, the Mel-spectrograms is transformed into speech waveforms.

Tacotron 2 models used in developing the TTS system for US English provides the highest MOS of 4.52, which is close to the MOS of a professional speaker. Tacotron 2 and WaveNet vocoder was used to develop the Kiswahili TTS system described in this paper. The system achieved a MOS of 4.05, showing that the TTS system is both natural and intelligent. The paper differs from other Tacotron papers since it is used to develop a TTS system for a low-resource language. Other Tacotron papers have been used to develop TTS systems for high-resource languages.

3. Dataset

3.1. Creation of Dataset

Languages are grouped into low-resource and high-resource due to open-source and free accessibility. Despite the Kiswahili language having millions of speakers, it is still grouped as a low-resource language because it has little data for developing Language processing tasks.

This project required a larger dataset. The dataset was made using the Kiswahili audio Bible. Kiswahili Bible is free of copyright and open source. Also, the authors allow use for educational, non-profit, and public benefits [18,19]. The Bible has a larger text corpus, capturing different language properties. Also, the Bible provided an audio recording for each chapter.

The dataset consists of both text and audio files. For building Tacotron 2-based TTS, each audio file parameters were [20,28];

(1) 16-bit mono-channel and unsigned Pulse Code Modulated (PCM) WAVE file
(2) A sampling frequency of 22,050 Hz

Tacotron 2 requires setting the sampling frequency, frameshift, number of frequency bins, and frame length hyperparameters during training and testing the system. Tacotron 2 model attributes are encoder depths, decoder depths, and output for each step.

Equation 3 shows the maximum audio length determined from the Model parameters and audio hyperparameters.

$$\text{Audio length} = \text{iteration} \times \text{output per step} \times \text{frame shift} \quad (3)$$

Table 2 shows the model parameters for determining the maximum allowed audio length for training.

Therefore, the maximum audio file length was set to 12.5s.

Table 2. Model parameters for implementing TTS system based on tacotron 2 [22,23]

TTS Parameter	Value
Iterations	200
Output for each Step	5
Length of frameshift	12.5

Table 3. An example of a text file consisting of a unique ID, transcribed text and normalized text

Unique ID	Transcribed Text	Normalized Text
KISWA -00467	Mlango 1. Kitabu cha ukoo wa yesu kristo.	Mlango wa Kwanza. Kitabu cha ukoo wa yesu kristo.

The model parameters determine the maximum audio clip length, which was set to 12.5s. A maximum length of 12.5 s ensures each text is correctly assigned to the corresponding audio. If this length is exceeded, then there is a possibility that a text is wrongly assigned to the audio part.

Bible provides an audio file for each chapter, whose length was more than the maximum length of 12.5s. Therefore, it was required to cut the audio file length to the recommended size, which was done using Python. First, each downloaded audio file was read as a Numpy array. Silent and speech parts have different values of the Numpy array. Thus, a threshold value was set to determine the silent and speech parts. After detection, the audio file was cut and combined based on various lengths while ensuring the audio file length was between 1s to 12.5s. After combining it, the audio clips were saved as a mono-channel with a sampling frequency of 22.05 kHz and an unsigned 16 PCM WAVE file.

The cut audio file was manually mapped to the corresponding text section. Then, text files were created in CSV format, each having three parts: The unique ID, the Transcribed text, and the normalized text. The unique ID is the number that identifies the text file and the audio file. The transcribed text is the copied words from the text corpus, while the normalized words are the expansion of the transcribed text, which consists of expanding numbers and abbreviations. Table 3 shows an example of a text file.

4. Methodology

4.1. Model Architecture

Tacotron 2 architecture was used to build the TTS system. The architecture has encoder, decoder, and vocoder steps. The encoder generates a character sequence for each input text. The decoder converts the speech into spectrograms and Mel-spectrograms [28,24]. Spectrograms were generated by applying Fourier transform to the speech waveforms. Mel-spectrograms were generated by converting the amplitude of the spectrograms to the Mel scale. Once the Mel-spectrograms have been generated, the vocoder predicts the time speech waveform. Figure 3 shows the model architecture with the encoder, decoder, and vocoder sections.

The encoder converts the input text to a sequence of characters. Character sequence goes through convolution layers, where each layer checks the previous characters' time steps. The output of the final convolution layer is fed to the Long short-term memory (LSTM), which generates the encoded character embedding. The attention mechanism produces the encoded sequence and feeds it to the decoder section. The Decoder section has causal RNN, which predicts the Spectrograms and Mel-spectrograms sequence. The predicted Mel-spectrograms pass through the Rectified linear unit (ReLu) and PreNet to generate a linear projection. The output of the linear projection passes through convolutional layers to accurately predict the attributes of the speech sample. The attribute includes energy, pitch, and duration. Finally, the WaveNet vocoder converts the Mel-spectrograms to speech waveform using the Mixture of logistic (MOL) distributions.

5. Results

5.1. Kiswahili Dataset

The dataset created had 7,108 text and audio files from a single speaker. The total length of the audio file is more than 16 hours, divided into 7,108 files, each having a length of 1s to 12.5s. The text file was manually mapped to the audio file and saved in CSV format. The text file was divided into three sections: unique ID, transcribed, and normalized texts. The dataset can be accessed at the link below:

https://data.mendeley.com/datasets/vbvj6j6pm9/1

The dataset has approximately 106,000 Kiswahili words. The dataset captured Non-standard words, including numbers, acronyms, and abbreviations. Each text file had an average of 14.9128 words, with a minimum audio length of 1s and a maximum audio length of 12.5s. The created dataset was used successfully to build the Kiswahili TTS system. Furthermore, the dataset can be used for any NLP project.

5.2. Trained System

The system was trained using supervised learning. Each character sequence was matched to its correct Mel-spectrograms. Adam optimizer, a feedback signal, was used to update the training weights. Other model parameters were set: a batch of 8 samples and a learning rate 10^{-3}. The predicted Mel-spectrogram was assigned to the correct encoded sequence. This ensured that the predicted sequence was correctly aligned with the speech waveform. The dataset was grouped into training, validation, and test data. The model was trained and tested using 0.9 of the dataset's total size and 0.1 for testing and validation. The system randomly grouped the data into three sections. Accuracy was the metric used to assess the system's success. The accuracy of the system improved as the iteration increased.

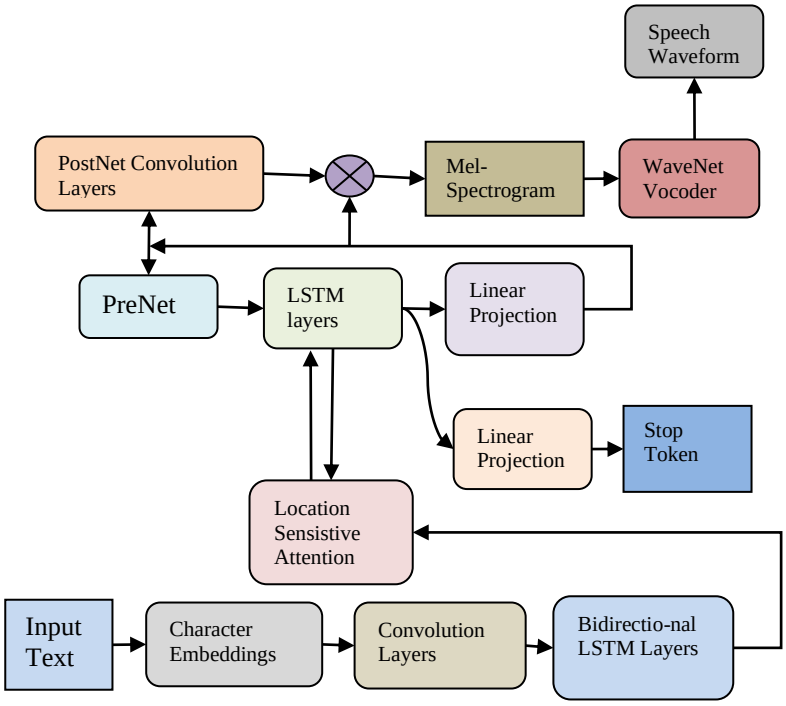

Fig. 3 Kiswahili text-to-speech system architectur

The system was trained to 151,000 iterations, and a model was saved after 1000 iterations. The model at 151,000 iterations had the highest accuracy of 92.1%. The model can be downloaded at the link below.

https://drive.google.com/file/d/1-LvxuDxfxJ8CyUurNkjeo_lXiQeXfuMY/view?usp=sharing

The model accuracy of 92.1% shows that the system correctly predicts the speech output for the test data. The system also correctly generates the speech waveform for new data. The system can be tested in the following link.

https://colab.research.google.com/drive/17ZZKB54T1cdPB6j47wBDkADG8UMW1jQd

5.3. Mel-Spectrograms and Alignment Steps

An alignment plot shows how the encoder and the decoder correctly match in predicting the correct Mel-spectrograms for the character sequence. A linear alignment plot shows that the system correctly predicts the Mel-spectrogram sequence for a given input text. The model at 151,000 iterations was used to test the system. Figure 4 shows a linear alignment between the encoder and the decoder. The alignment plot was generated from the input text, “Naweza ingia kesho. Wewe unaweza ingia Jumamosi.”

Mel-spectrogram plot shows the Mel-scale of Fourier Transform amplitude of a speech sample on the y-axis and frequency on the x-axis. Figure 5 shows the Mel-spectrograms of the input text “Naweza ingia kesho. Wewe unaweza ingia Jumamosi.” The speech is in the yellow sections, while the silent part is in the blue areas.

Mel-spectrograms of the correct speech waveform were compared to the predicted waveform. First, Mel-spectrograms of the correct speech waveform was generated by selecting one audio file from the dataset. Then, the system was used to generate the audio for the text section from the dataset, which was used to create the predicted Mel-spectrograms. Figure 6 and 7 shows the plot of the correct Mel-spectrogram and the Predicted Mel-spectrograms. The plots are similar, showing that the system accurately outputs the speech waveforms.

The model at 151,000 iterations was used to test the system with NSWs. Figure 8 shows the alignment plot for the text “Prof. anaingia Leo. Dkt. anaingia kesho”. Dkt. Requires expansion to “Daktari” and prof to “Profesa”’. The text was input in its non-standard form. The system correctly predicted the speech waveform for the NSWs. The system-generated audio sample can be accessed using the link below.

https://data.mendeley.com/datasets/x2p68n8dzz/1

The alignment shows how the encoder sequence matches the decoder sequence. A linear alignment plot implies that the character embeddings are correctly matched to the Mel-spectrogram, showing that for each character sequence, the system accurately predicts its Mel-spectrograms, which the vocoder consumes to generate the speech waveform.

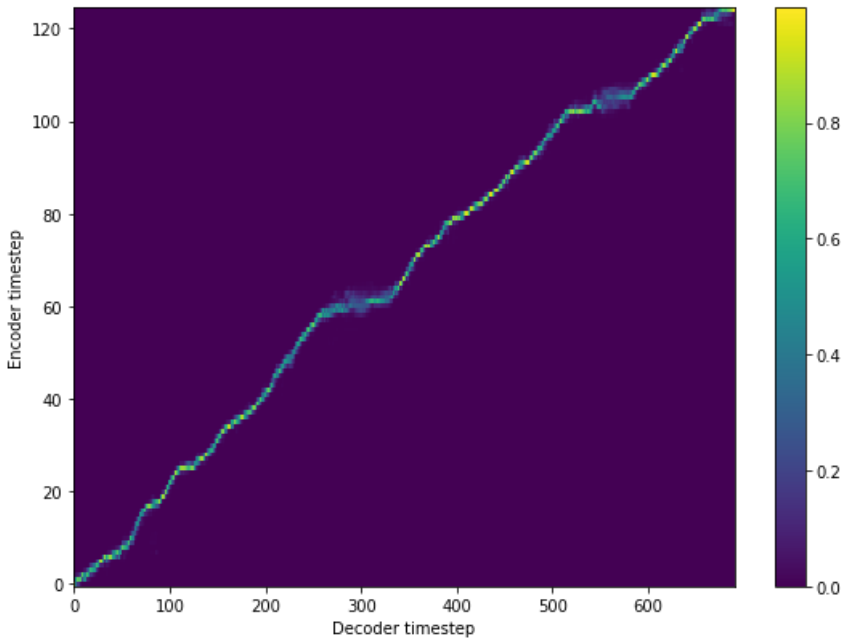

Fig. 4 Alignment plot for the encoder and decoder steps for the input text "Naweza ingia kesho. Wewe unaweza ingia jumamosi" the english translation is "I can come tomorrow. You can come on Saturday."

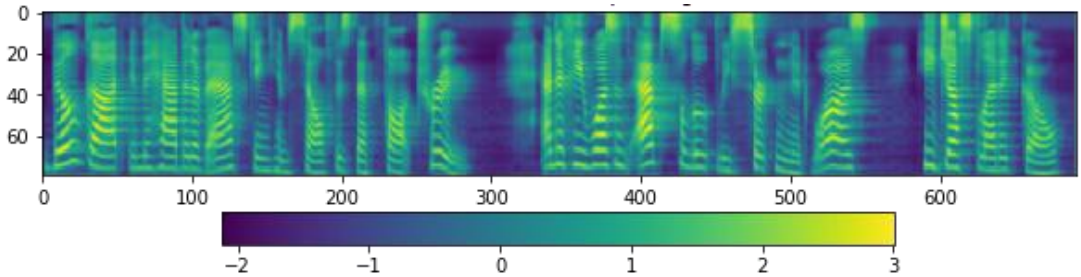

Fig. 5 Alignment plot for the encoder and decoder steps for the input text "Naweza ingia kesho. Wewe unaweza ingia Jumamosi" the english translation is "I can come tomorrow. You can come on saturday."

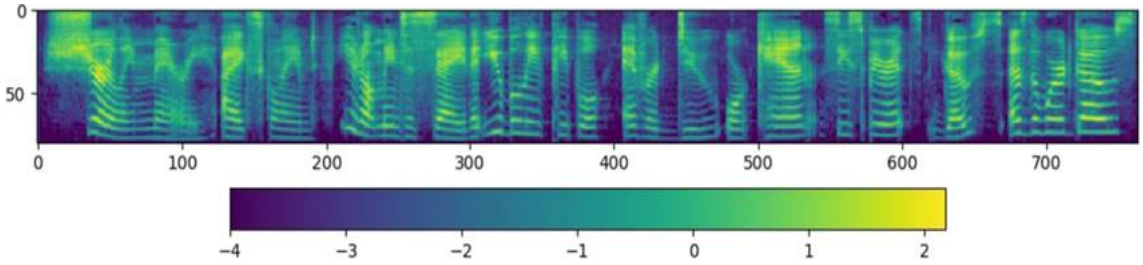

Fig. 6 Target mel-spectrograms

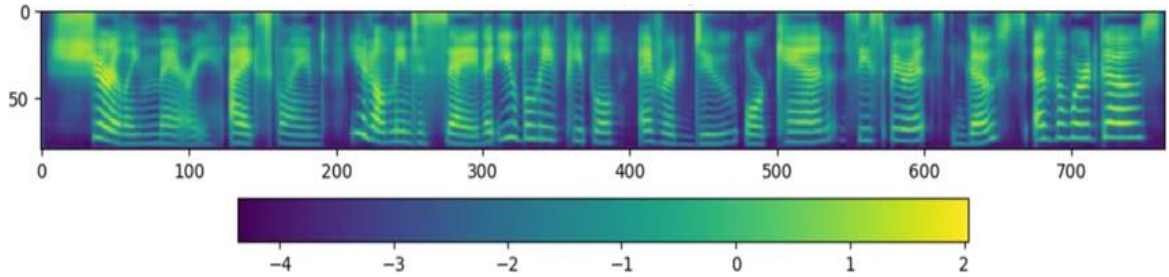

Fig. 7 predicted mel-spectrograms

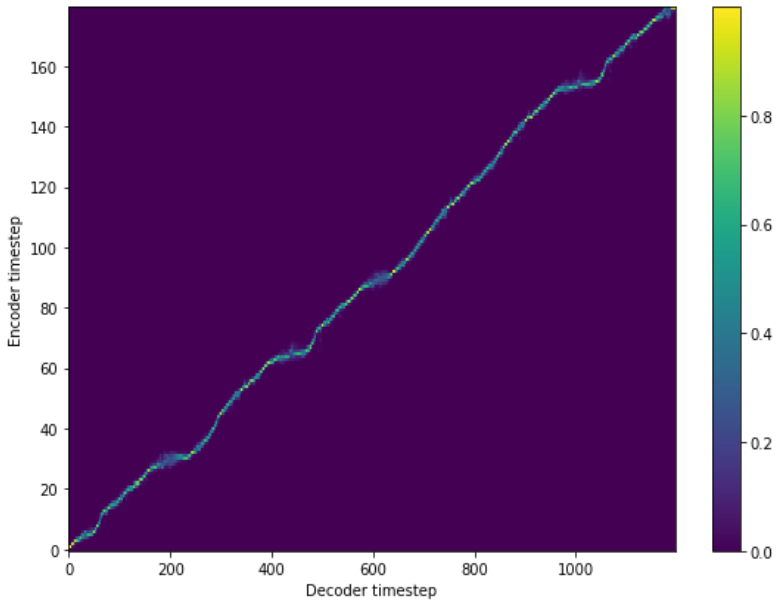

Fig. 8 Alignment plot for the input text "Prof. anaingia leo. Dkt. Anaingia kesho" the english translation is "Prof. is coming today. Dr. is coming tomorrow."

Table 4. PESQ MOS for tacotron 2-based TTS System

TTS System	MOS
Speech_sample1.wav	3.960
Speech_sample2wav	4.055
Speech_sample3.wav	4.135
Mean PESQ MOS.wav	4.050

Table 5. PESQ MOS for concatenation-based TTS system

TTS System	MOS
Sample1.wav	3.605
Sample2.wav	3.540
Sample3.wav	3.655
Mean PESQ MOS	3.600

Figure 4 and Figure 8 show that the system has a linear and stable alignment in generating Mel-spectrograms from the character sequence. Furthermore, the speech output was natural sounding and intelligent.

5.4. Evaluation of the System

The saved model at 151,000 iterations had the highest accuracy for the test data. The model was used to evaluate the system. MOS was used to assess the system. Perceptual Evaluation of Speech Quality (PESQ) is perceived to assess the system in terms of intelligibility and naturalness. PESQ MOS was computed for the Tacotron model sample audio, achieving a MOS of 4.05. Three audio samples were generated and downloaded for use in computing the PESQ MOS.

The concatenation approach had a PESQ MOS of 3.60. The speech samples downloaded from festival voices were used to compute the PESQ MOS of the concatenation-based TTS [25]. LLSTI site provides festival voices for the Kiswahili language developed based on Concatenation. The audio clips were downloaded, and the PESQ MOS was computed. Table 4 and Table 5 display the MOS for the audio samples generated from Tacotron 2 and Concatenation approaches.

PESQ MOS shows that the Kiswahili TTS developed in this paper performs better in naturalness and intelligibility.

6. Conclusion

The Kiswahili TTS system was developed using the Tacotron model. The system is a sequence-to-sequence model. Each text file was converted to a sequence of characters. Also, the audio file was converted to a Mel-spectrograms sequence. During the training, character sequence was assigned to Mel-spectrograms by applying the supervised learning approach. For testing the system, the input text is first converted to a character sequence, which the system predicts the Mel-spectrograms. Finally, the WaveNet Vocoder generates the speech waveform based on the Mel-spectrograms.

The saved model at 151,000 iterations produced the best speech quality for a given input text. The speech generated was also intelligent. The model saved after 151,000 iterations generated audio output, which was used to compute the PESQ MOS. Kiswahili TTS system developed based on Tacotron performs better over Concatenation approach-based TTS, with a MOS of 4.05 over a MOS of 3.60 for the concatenation-based TTS. Therefore, Kiswahili TTS built on Tacotron 2 is more natural sounding and intelligible.

Conflicts of Interest

The authors declare that they have no known competing financial interests or personal relationships which have or could be perceived to have influenced the work reported in this article.

Funding Statement

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

References

[1] The University of Arizona, “Critical Languages Program,” 2016.

[2] Suman K. Saksamudre, P.P. Shrishrimal, and R.R. Deshmukh, "A Review on Different Approaches for Speech," *International Journal of Computer Applications*, vol. 115, no. 22, pp. 23-28, 2015. [CrossRef]

[3] K. Simonyan et al., "*WaveNet: Generative Model for Raw Audio*," pp. 1–15, 2016. [CrossRef]

[4] Sercan O. Arik et al., "Deep Voice: Real-Time Neural Text-to-Speech," *axXiv:1702.07825*, 2017. [CrossRef]

[5] Mucemi Gakuru et al., “Development of Kiswahili Text to Speech,” *2005 9th European Conference on Speech Communication and Technology*, Lisbon, Portugal, pp. 1481-1484, 2005. [CrossRef]

[6] K. Ngugi, P. W. Wagacha, and P.W Wagacha, "Swahili Text-to-Speech System," *African Journal of Science and Technology*, vol. 6, no. 1, pp. 80–89, 2005. [CrossRef]

[7] R. A. Khan, and J. S. Chitode, "Concatenative Speech Synthesis : A Review," *International Journal of Computer Applications*, New York, USA, vol. 136, no. 3, pp. 1–6, 2016. [CrossRef]

[8] Vinayak K. Bairagi, Sarang L. Joshi, and Vastav Bharambe, "Implementing Concatenative Text-To-Speech Synthesis System for Marathi Language using Python," *SSRG International Journal of Electrical and Electronics Engineering*, vol. 9, no. 9, pp. 1-11, 2022. [CrossRef]

[9] Sneha Lukose, and Savitha S. Upadhya, "Text to Speech Synthesizer-Formant Synthesis," *2017 International Conference on Nascent Technologies in Engineering*, Vashi, India, pp. 1-4, 2017. [CrossRef]

[10] Lauri Juvela et al., “GlotNet—A Raw Waveform Model for the Glottal Excitation in Statistical Parametric Speech Synthesis," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 27, no. 6, pp. 1019-1030, 2019. [CrossRef]

[11] Abitha A, and Lincy K, "A Faster RCNN Based Image Text Detection and Text to Speech Conversion," *SSRG International Journal of Electronics and Communication Engineering*, vol. 5, no. 5, pp. 11-14, 2018. [CrossRef]

[12] S. R. Hertz, "Integration of Rule-Based Formant Synthesis and Waveform Concatenation: A Hybrid Approach to Text-to-Speech Synthesis," *Proceedings of 2002 IEEE Workshop on Speech Synthesis*, pp. 87-90, 2002. [CrossRef]

[13] Heiga Zen, Keiichi Tokuda, and Alan W. Black, “Statistical Parametric Synthesis,” *Journal of Speech Communication*, vol. 51, no. 11, pp. 1039-1064, 2009. [CrossRef]

[14] Ryan Prenger, Rafael Valle, and Bryan Catanzaro, “WaveGlow: A Flow-Based Generative Network for Speech Synthesis," *IEEE International Conference on Acoustics, Speech and Signal Processing*, Brighton, United Kingdom, pp. 3617-3621, 2019. [CrossRef]

[15] Fisher Yu, and Vladlen Koltun, "Multi-Scale Context Aggregation by Dilated Convolutions," *arXiv:1511.07122*, 2015. [CrossRef]

[16] Anamika Baradiya, and Vinay Jain, "Speech and Speaker Recognition Technology using MFCC and SVM," *SSRG International Journal of Electronics and Communication Engineering*, vol. 2, no. 5, pp. 6-9, 2015. [CrossRef]

[17] Martin Strauss, and Bernd Edler, "A Flow-Based Neural Network for Time Domain Speech Enhancement," *ICASSP 2021 - 2021 IEEE International Conference on Acoustics, Speech, and Signal Processing*, pp. 5754-5758, 2021. [CrossRef]

[18] WordProject, Swahili Audio Bible, 2022.

[19] WordProject, Swahili Audio Bible, 2022.

[20] Tom Le Paine et al., "Fast Wavenet Generation Algorithm," *Proceedings* of the 26th International *Conference* on World Wide Web, Geneva, Switzerland, no. 98, pp. 381-389, 2016.

[21] Prashanth Kannadaguli and Vidya Bhat, "Phoneme Modeling for Speech Recognition in Kannada using Multivariate Bayesian Classifier," *SSRG International Journal of Electronics and Communication Engineering*, vol. 1, no. 9, pp. 1-4, 2014. [CrossRef]

[22] Soroush Mehri et al., “SampleRNN: An Unconditional End-to-End Neural Audio Generation Model,” *arXiv:1612.07837*, 2017. [CrossRef]

[23] GitHub, An Implementation of Tacotron Speech Synthesis in Tensorflow, 2017.

[24] K.Sureshkumar, and P.Thatchinamoorthy, “Speech and Spectral Landscapes Using Mel-Frequency Cepstral Coefficients Signal Processing,” *SSRG International Journal of VLSI & Signal Processing*, vol. 3, no. 1, pp. 5-8, 2016. [CrossRef]

[25] The Local Language Speech Technology Initiative, 2022.

[26] Yuxuan Wang et al., “Tacotron: Towards End-to-End Speech Synthesis,” *Interspeech*, 2017. [CrossRef]

[27] Ryuichi Yamamoto, Eunwoo Song, and Jae-Min Kim, “Probability Density Distillation with Generative Adversarial Networks for High-Quality Parallel Waveform Generation,” *arXiv: 1904.04472*, 2019. [CrossRef]

[28] Jonathan Shen et al., "Natural TTS Synthesis by Conditioning WaveNet on Mel Spectrogram Predictions," *2018 IEEE International Conference on Acoustics, Speech, and Signal Processing*, Calgary, Alberta, Canada, pp. 4779-4783, 2018. [CrossRef]