SEED JOURNAL

# Efforts to Apply Natural Language Processing Technologies in Africa

**Kweku Andoh Yamoah**[1]
Computer Science & Information Systems Department, Ashesi University
[1]Computer Science; kweku.yamoah@ashesi.edu.gh

***Abstract*** **-** The primary objective of this paper is to investigate the developments and advancements made in Automatic Speech Recognition (ASR), Natural Language Processing (NLP), and Text-to-Speech (TTS) technologies for African languages and problems. The paper addresses the challenges the African continent faces in implementing these technologies. Additionally, the paper proposes a novel concept that combines computer vision, NLP, and TTS to aid visually impaired individuals in Ghana.

***Keywords*** *- Automatic Speech Recognition(ASR), Natural Language Processing(NLP), Text-to-Speech(TTS), Hidden Markov Model and Gaussian Mixture Model(CD-HMM/GMM), Deep Neural Networks(DNN), Convolutional Neural Networks(CNN).*

## 1. Introduction

The world has undergone a significant transformation, shifting from manual tasks to an era where machines can learn and perform these tasks with remarkable accuracy. This era is known as Machine Learning (ML). Within ML, Automatic Speech Recognition (ASR), Natural Language Processing (NLP), and Text-to-Speech (TTS) are subsets that enhance human interactions with computer systems. The impact of these technologies is evident in voice-based tools such as Alexa, Siri, and Google's Home Assistant. These tools can be used for various purposes, including playing music, setting reminders, asking questions, shopping online, and more. However, these technologies are mostly available in popular Western languages, notably English, and some Eastern languages, leaving African languages and problems underrepresented.

This paper aims to showcase the efforts being made across Africa to apply and use NLP tools and technologies on the continent, with a specific focus on Automatic Speech Recognition (ASR), Natural Language Processing (NLP), and Text To Speech (TTS) and their applications to native African Languages. Furthermore, this paper presents a creative idea on how these technologies can be used to address African problems and languages.

## 2. Challenges Encountered in Africa

Before we get into the current efforts to introduce these innovations in Africa, we need to discuss the obstacles preventing them from being implemented. There are currently 6900 languages spoken worldwide [1], and out of them, 2000 are spoken in Africa alone [2]. Unfortunately, African languages lack documentation and datasets [3], and their resources are dispersed and often challenging to access, as pointed out by Abbot and Martinus [3]. Additionally, the scarcity of data for African languages that can be used to train models discourages most researchers from attempting to do so [4]. For instance, ASR models require 100,000 hours of recorded speech to create an accurate model, which is not available for most native African languages [4]. Finally, code-switching presents a challenge for machine learning engineers in building a multilingual model for African languages [4]. The challenges faced by African languages hinder the adoption and implementation of ASR, NLP, and TTS technologies. In the next section, we will discuss the efforts being made to address these challenges, as it is crucial to improve the accessibility and effectiveness of these technologies in Africa.

## 3. Existing Efforts in Africa

Despite the challenges, startups and researchers have been working to apply these technologies to African problems. In this section, we will discuss the various strides that have been made across the continent. We will explore these efforts in three subsections: ASR, NLP and TTS.

### *A. Automatic Speech Recognition*

ASR is the task of mapping waveforms to the appropriate string of words [5]. In the African ASR ecosystem, Gauthier, Besacier, and Voisin have developed a system with a vowel length constraint for Hausa and Wolof languages [6]. Their systems for both languages made use of the hidden Markov Model and Gaussian mixture model (CD-HMM/GMM) and deep neural networks (DNN). They modeled four ASR systems, two for each language, while ensuring they considered vowel duration for the languages [6]. The Wolof ASR system obtained a word error rate (WER) of 31.9% and a character error rate (CER) of 12.9% for the CD-HMM/GMM acoustic model [6]. Similarly, on the

CD-DNN-HMM acoustic model, they obtained a WER of 27.7% and a CER of 10.5% [6]. The Hausa ASR system performed better on both acoustic models. The system achieved a WER of 12.9% and a CER of 3.7% for the CD-HMM/GMM acoustic model. On the CD-DNN-HMM acoustic model, it achieved a WER of 7.9% and a CER of 2.1% [6]. However, the ASR systems do not seem to be fine-tuned for longer duration models [6], and hence the performance may decline on such duration models. Still, Gauthier, Besacier, and Voisin's work laid the groundwork for other developers to use their ASR systems as baseline acoustic models to build upon.

Their work also resulted in a speech corpus for Wolof and the first-ever large vocabulary continuous speech recognition system [6]. In the future, the Wolof speech corpus may be used for NLP and TTS subtasks. An ASR system was also applied to Nigerian Pidgin English. Pidgin English, a variant of the structured English language, is commonly spoken across West Africa [7]. The Nigerian pidgin system was developed using the neural architecture called the NeMo toolkit [7]. In the implementation, the developers utilized two ASR architectures called Jasper and QuartzNet. Their model takes a speech recording and predicts the corresponding text, making it a speech-to-text system [7]. The team obtained a WER of 0.997% for the Jasper model with no data augmentation and a WER of 0.987% with data augmentation [7]. Similarly, they reported a WER of 0.772% for the QuartzNet with data augmentation and a 0.777% WER with no data augmentation [7].

According to [7], this ASR system was developed with the hope of building the first speech-to-text benchmark for Nigerian pidgin. Hence the team open-sourced their code and data to motivate further research on the language. Thus, their ASR model can be used as a benchmark due to the excellent results achieved for the WER.

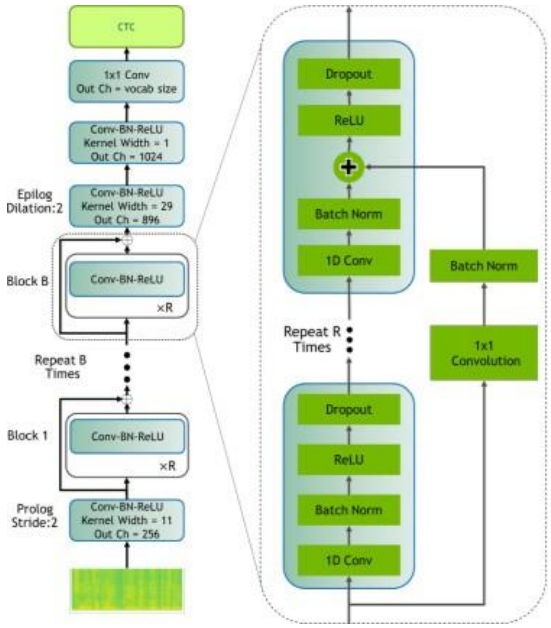


Fig. 1. Jasper Architecture [8]

### *B. Natural Language Processing*

NLP is an area that has seen a lot of applications with regards to African languages. NLP involves the application of computational techniques to the synthesis and analysis of natural language [5]. In Ghana, an NLP team has released a model called ABENA, which can perform some tasks using the Twi language [10]. Twi is a language spoken by a specific group of Ghanaians called the Akans. The ABENA model provides contextual word embedding for Twi [10]. Learned embeddings from the model show fairly accurate word embeddings when visualized. ABENA was then applied to sentiment classification, and it was discovered that the model's accuracy always lay in a range between 83% to 100% [10]. Although the results of ABENA look promising for the Twi language, it was discovered that the model had varying religious bias. The bias occurs due to data that emanates from religious context used for the model's training [10]. However, this current work presents an avenue for NLP research for Twi and other Ghanaian Languages.

Another NLP-focused team in Ghana is a promising startup called Nokwary Technologies. Nokwary's mission is to use NLP to build conversational WhatsApp bots for a variety of Ghanaian businesses [11]. Conversational bots are a subset of NLP that deals with dialogue systems. Nokwary employs NLP to make advanced technology accessible to Ghana's impoverished communities [12]. The company has created a WhatsApp banking app for Twi speakers that allows them to purchase airtime using natural language interaction in Twi [12]. Through this, Nokwary's product is helping many Ghanaians have access to financial services. Due to the fact that Nokwary has been able to apply NLP to Twi, it is possible they could do so for other African languages. The obstacle hindering them from doing so is most likely the lack of documented data mentioned earlier. Should the needed data required be obtained, it is possible to envision them evolving into a prominent provider of financial solutions across Africa. Given the lack of curated data discussed above, Nokwary's secrecy with plans to make their dataset available publicly to help the budding NLP community in Ghana. Furthermore, since their NLP models are designed for businesses, their performance cannot be assessed externally.

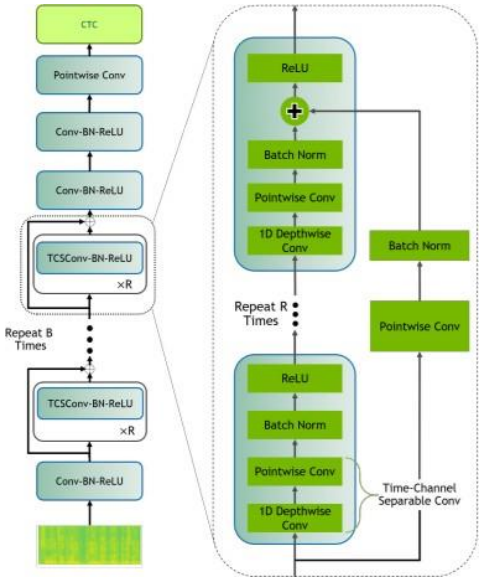


Fig. 2. QuartzNet Architecture [9]

### *C. Text To Speech*

The task of text-to-speech (TTS) is the reverse of automatic speech recognition. In TTS, we aim to map a piece of text to an acoustic waveform [5]. TTS has many applications in assistive technologies; however, most of these solutions have not yet been applied to African languages. The only country that seems to be developing a TTS system is South Africa, where Qfrency TTS voices have been integrated with an augmentative and alternative communication (AAC) system [13]. This system allows users to communicate by inputting a text in English, which is then translated into a specified South African language and read out by the Qfrency TTS voice [13]. Alternatively, the text can be directly translated from the South African language and read out. However, the researchers discovered that users had trouble pronouncing words in their native languages, which impacted the intelligibility of the system [13].

In addition to AAC, TTS solutions have been integrated with accessible Zulu books in South Africa. The system is capable of reading epub versions of books, but users have described the monotony of the Qfrency TTS voices as a downside.

## 4. Creative Imagination

ASR, NLP, and TTS technologies can provide accessible information for visually impaired citizens in Ghana who speak English or other local languages. The idea is to build a system that can caption an image taken by a visually impaired person and then read the caption back to the user in English or any local language of their choice. However, in this project, I will consider Twi since it is the Ghanaian language that has been extensively worked on. The idea lies in the area of computer vision, natural language processing, and text-to-speech.

The direct impact of such a project will be to aid visually impaired people to navigate their environment easily by creating a mental picture of what they hear through a captioned photo. This idea can also benefit Ghanaian schools for the blind, as teachers can use the system to depict accurate images of the world around them. For example, a teacher can point out an image from a textbook, take a picture of it, wait for it to be captioned, and read out for students. This idea can be achieved in the following steps.

The first step is to caption the image using convolutional neural networks (CNN) and a transformer model. The CNN will generate a feature representation of the image, which becomes an encoding for the image. This encoding is then passed to the transformer model to generate a word embedding, which is further passed to a decoder to obtain a caption. Hidden in the decoding stage is a language model that uses word embeddings to generate sentences in English.

The generated caption is passed to the TTS system, which has three phases: encoding, decoding, and vocoding. The encoding phase will first take the caption and transform it into character embeddings by passing it through CNNs and a bidirectional LSTM. The decoder takes this and predicts a log mel spectrogram for these final encodings, which is then sent to the final stage where the mel spectrum is passed through a WaveNet to get waveforms of the mel spectrum, which will be read to a user.

If a user prefers the captions to be read in Twi, it could be done in two ways. The first is to generate the captions in Twi directly before the TTS stage. The other is to translate the generated English captions before passing them as input to the TTS system, and the most efficient approach will be chosen.

## 5. Conclusion

While efforts to apply ASR, NLP, and TTS technologies to African languages do exist, the low resource levels of these languages present challenges for researchers in this area. Nevertheless, there have been various approaches applied to African languages, such as the ABENA model we discussed for Twi in NLP. In addition, the idea of providing image captioning for visually impaired Ghanaians could be of great benefit to the education sector. Teachers could use the system to accurately depict images from textbooks, for instance, by taking a picture and having it captioned and read out to students.

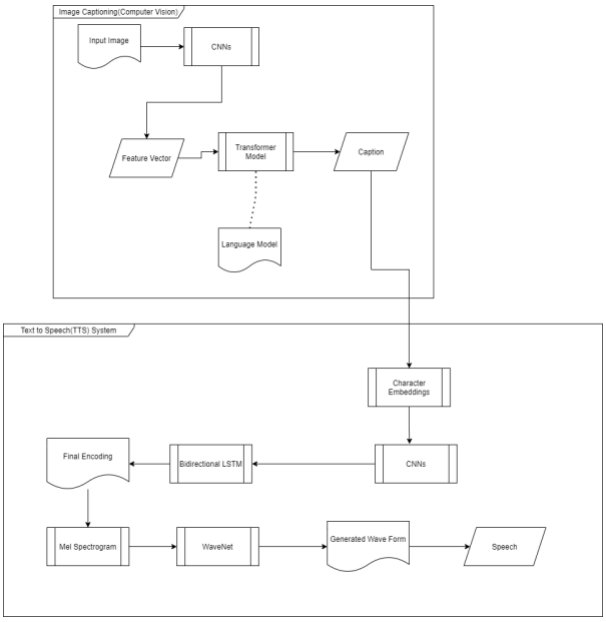


Fig. 3. High level structure of creative idea

## References


[1] L. Besacier, E. Barnard, A. Karpov, and T. Schultz, "Automatic speech recognition for under-resourced languages: A survey," Speech Communication, vol. 56, pp. 85–100, Jan. 2014, doi: 10.1016/j.specom.2013.07.008.

[2] David M Eberhard, Gary F Simons, and Charles D Fennig, Ethnologue: Languages of the worlds, Twenty Second Edition.

[3] J. Abbott and L. Martinus, "Benchmarking Neural Machine Translation for Southern African Languages," Proceedings of the 2019 Workshop on Widening NLP, Association for Computational Linguistics, Florence, Italy, pp. 98-101, 2019

[4] A. L. Dahir, "African languages are being left behind when it comes to voice recognition innovation," Quartz. https://qz.com/africa/1475763/african-languages-are-lagging-behind- when-it-comes-to-voice-recognition-innovations/ (accessed Apr. 14, 2021).

[5] D. Jurafsky and J. H. Martin, Speech and language processing: an introduction to natural language processing, computational linguistics, and speech recognition. India: Dorling Kindersley Pvt, Ltd., 2016.

[6] E. Gauthier, L. Besacier, and S. Voisin, "Automatic Speech Recognition for African Languages with Vowel Length Contrast," Procedia Computer Science, vol. 81, pp. 136–143, 2016, doi: 10.1016/j.procs.2016.04.041.

[7] D. Ajisafe, O. Adegboro, E. Oduntan, and T. Arulogun, "Towards End-to-End Training of Automatic Speech Recognition for Nigerian Pidgin," arXiv:2010.11123 [cs, eess], Oct. 2020, Accessed: Apr. 14, 2021. [Online]. Available: http://arxiv.org/abs/2010.11123.

[8] J. Li et al., "Jasper: An End-to-End Convolutional Neural Acoustic Model," arXiv:1904.03288 [cs, eess], Aug. 2019, Accessed: Apr. 14, 2021. [Online]. Available: http://arxiv.org/abs/1904.03288.

[9] S. Kriman et al., "QuartzNet: Deep Automatic Speech Recognition with 1D Time-Channel Separable Convolutions," arXiv:1910.10261 [eess], Oct. 2019, Accessed: Apr. 14, 2021. [Online]. Available: http://arxiv.org/abs/1910.10261.

[10] P. Azunre et al., "Contextual Text Embeddings for Twi," arXiv:2103.15963 [cs], Mar. 2021, Accessed: Apr. 14, 2021. [Online]. Available: http://arxiv.org/abs/2103.15963.

[11] "Nokwary Technologies: AI is here — Artificial Intelligence and What- sApp Banking in Ghana and Africa." https://nokwary.com/ (accessed Apr. 14, 2021).

[12] "Hey Google! Hey Alexa! Hey Nokwary! AI in Ghana – AI and NLP in Ghana and Africa." https://nokwary.com/blog/2020/05/25/hey-google- hey-alexa-hey-nokwary-ai-in-ghana/ (accessed Apr. 14, 2021). [

[13] G. I. Schlu¨nz et al., "Applications in accessibility of text-to-speech synthesis for South African languages: initial system integration and user engagement," in Proceedings of the South African Institute of Computer Scientists and Information Technologists on - SAICSIT '17, Thaba 'Nchu, South Africa, 2017, pp. 1–10, doi: 10.1145/3129416.3129445.