

Improving access to radio and television for deaf and hard of hearing people in Botswana through Free Automatic Speech Recognition Technology

Kagiso Jacob Sello
University of Botswana

The aim of this study was to propose the Automatic Speech Recognition (ASR) technology as a means of improving access to televised and radio-broadcast government communications in the English language for people who are deaf and hard of hearing (DHH) in Botswana. To gain better insight into the possibilities of using ASR for information accessibility for people who are DHH, this paper evaluated the accuracy rate of an open-source ASR technology in terms of Word Error Rate (WER) by comparing transcriptions generated by ASR software and transcriptions generated and edited by the researcher. The study results show that all the transcribed audio-visual documents scored a WER of less than 15%, indicating that transcriptions were of a very high quality. Some documents even scored an impressive WER of less than 5%, meaning that the transcriptions generated could be used and understood without any human intervention. The study concludes that freely available ASR technology is adapted as a means of improving and diversifying accessibility modes to televised and radio-broadcast government communications in the English language for people who are DHH in Botswana.

Keywords: automatic speech recognition, accessibility, deaf and hard of hearing, information, inclusion, marginalised populations

1. Introduction

Automatic Speech Recognition (ASR) as a field of technology has made enormous advancements in the past few years. This technology has rapidly progressed because of the increased demand in applications for in-meeting assistants, voice search, subtitling, virtual speech assistants, intelligent devices, or dictation tools (Ulasik et al, 2020: p.1) Nowadays ASR systems can achieve higher accuracy rates depending on the methodology applied and datasets used (Radzikowski et al., 2021). Given these results, there is an increasing interest in using ASR technology for the Deaf and Hard of Hearing (DHH) populations (Gambill et al. 2019; Kafle and Huenerfauth, 2019). ASR, also known as speech-to-text technology, uses artificial intelligence (AI) and machine learning (ML) capabilities to automatically transcribe raw speech into structured written text. This technology integrates grammar, syntax, structure, and composition of audio and voice signals to understand and process human speech¹. Several factors can impact the quality of the transcription, such as pronunciation, accent, pitch, volume, background noise, domain-specific and context-specific jargon, dialects, and multiple languages². AI and ML capabilities have drastically improved the accuracy of ASR technologies as applications are able to learn on their own from observational data. The ASR software itself is programmed to autonomously learn and adopt new words. By storing data from previous interactions, it constantly expands its vocabulary. Advanced versions of ASR

¹ <https://www.ibm.com/topics/speech-recognition>

² <https://omniscien.com/blog/speech-recognition-speech-synthesis-glossary-a-g/>

transcription technologies now incorporate what is known as Natural Language Processing (NLP). NLP can manipulate human language or data that resembles human language. Therefore, ASR could potentially improve communication for the DHH populations by providing transcriptions of speech in real-time.

At present, however, this technology is not optimized for all users. For instance, ASR services are usually trained for dominant varieties of pluricentric languages³ such as English. As such, most ASR services have developed recognition systems for Australian English, Canadian English, Ghanaian English, Indian English, Irish English, Kenyan English, New Zealand English, Nigerian English, Philippines English, South African English, Tanzanian English, United Kingdom English, and United States English. The technology is not optimized for non-dominant varieties⁴ of English such as Botswanan English. It is therefore not surprising to note that, according to numerous studies, ASR systems still struggle and have deficiencies when applied to speech produced by non-native speakers (Cumbal et al., 2021; Sullivan et al., 2022; Radzikowski, 2021). In a study by Ashwell and Elam (2017), the researchers evaluated the *Google Web Speech* service through reading of a set of 13 sentences by 44 learners of English as a foreign language (42 Japanese and 2 Chinese) and 2 native English speakers. The results of their study showed a score of 89,4% against 65,7% in favour of native speakers. Mulholland et al. (2016) also found that ASR accuracy for the task of large vocabulary spontaneous speech tends to be substantially lower for non-native speakers, with WER values around 30-35% for Gaussian Mixture Model (GMM)-based systems.

Nevertheless, a comparison of current versions of ASR systems with previous versions shows an improvement. McCrocklin et al. (2019) found that *Google Voice Typing* offered substantial increases in accuracy for non-native speakers as compared to earlier research (Derwing, Munro, and Carbonaro, 2000) in which the researchers reported a recognition accuracy of the non-native speech by the ASR system used, *Dragon Naturally Speaking*, of 71-73% even though the speakers who participated in the study were all highly proficient, educated speakers of English. The accuracy was much lower than the findings for human listeners (89-98%). McCrocklin et al. (2019) reported *Google Voice Typing* accuracy for non-native speech ranging from 88.61% (sentences) to 93.47% (free speech).

Although ASR technology has been mentioned in a study on the accessibility of people who are DHH to information in Botswana (Mukhopadhyay and Moswela, 2020), it has never been tested to evaluate its accuracy in transcribing a recorded or live audio document produced in Botswanan English for which the system was not trained. Therefore, the purpose of this study is to evaluate the accuracy rate of a free online ASR technology software to propose it as a means of improving access to televised and radio-broadcast government communications for people who are DHH in Botswana.

³ A pluricentric language is a language that is used in at least two nations where it has an official status as state language, co-state language, or regional language with its own (codified) norms that usually contribute to the national/personal identity, making the nation a norm-setting centre by the deliberate use of the norms native to this specific nation. (Muhr, Rudolf (2016): *The state of the art of research on pluricentric languages: Where we were and where we are now*. In: Rudolf Muhr, Kelen Ernesta Fonyuy, Zeinab Ibrahim, Corey Miller (eds.) (2016): *Pluricentric Languages and non-dominant Varieties worldwide*. Volume 1: *Pluricentric Languages across continents – Features and usage*. Wien et. al., Peter Lang Verlag. p. 9-32.)

⁴ The term “non-native speakers” which is usually used in research seems inappropriate to us given that the speech corpora used to train ASR system to recognise, for instance South African English, is not solely based on native South African English speakers but rather on English as spoken by all in the South Africa context. We therefore advise the reader to understand non-native speakers as people who speak languages or varieties of languages that are different from the models’ target languages, irrespective of whether they learnt that language on their mothers’ knees as Bloomfield (1933) would put it, or not.

2. Access to information for DHH populations in Botswana

In today's societies, the creation, distribution, use, integration, and manipulation of information is an important aspect of economic, political, and cultural activity (Echezona, 2007). There is a high and increasing dependency of individuals and institutions on information and communication to be able to function effectively in almost every sphere of activity (McQuail, 1992). Nevertheless, DHH populations all over the world are often excluded from the information-sharing processes. The same is true for DHH populations in Botswana. The government of Botswana has made considerable efforts to improve both access to and quality of public service through the development of the E-Government portal (Republic of Botswana, 2011). However, the E-Government portal does not address the question of digital exclusion, especially for those who are technologically illiterate, sensory impaired, and linguistic minorities, including sign languages. For instance, Chinkonono et al. (2020) note that the preferred mode of communication for the DHH population of Botswana is Botswana Sign Language (BSL), yet the E-Government portal is not customized to cater to this population.

Restricted access to information for marginalized or disadvantaged groups is also evident on radio and television, two traditional modes of information sharing process for the government of Botswana. While the use of the internet for information sharing by the government has grown to unprecedented levels in the past five years, the government still relies on the State radio stations, Radio Botswana 1 and 2 (RB1 and RB2), and the State television station, Botswana Television (BTV) as an effective means of mass communication to, among others, inform people about government programmes and policies, and issues of national interest. Radio and television therefore play an important role in shaping Botswana society, including rural populations. However, radio and television cannot benefit the DHH population because they heavily rely on audio mode of communication which is inaccessible to the DHH population. Audio is defined as anything related to sound in terms of its reception, transmission, or reproduction. Although some BTV programmes, news bulletins for instance, are simultaneously interpreted using BSL, most programmes on BTV are neither interpreted in sign language nor subtitled. It should also be noted that, just like with spoken languages, there is no "universal" sign language that is understood by all deaf communities around the world. Sign languages are unique to the regions and cultures of people who use them. Furthermore, contrary to the common misconception, sign languages are not versions of spoken languages. They are fully-fledged, complex, natural languages, with their grammar, vocabulary, and dialects. (Foltz, 2019). There are about 300 recorded living sign languages in the world today (United Nations, n.d.; National Geographic, n.d.). DHH population is therefore only minimally informed.

By over-reliance on audio to disseminate information, the government of Botswana excludes the DHH population from the information-sharing process, thereby creating stumbling blocks in their livelihoods because they miss out on issues of national interest. This not only denies this marginalised population access to information but equally jeopardizes their participation in both the information society and the knowledge-based economy. Indeed, Idowu et al. (2017: p.1) rightly view information as a "strategic resource and tool that all living beings regardless of their categories, communities, societies, and nations need to have access to and utilize at all times to develop and advance in any type of endeavour." Information enables people to participate effectively in the process of economic, social, and political activities of the society (Laloo, 2002). By being excluded from the information-sharing process, the DHH may miss issues of national interest. Information should therefore be disseminated multilingually and multimodally to ensure its free and equal accessibility by every member of a given society. Consequently, while the BTV's efforts to reach

out to the DHH population and their audience through BSL are commendable, other communication modes need to be put in place to cater to those who are still left out because they do not understand the BSL. ASR technology can aid in that regard.

3. Methodology

3.1 Approach. The purpose of this study was to evaluate the accuracy rate of ASR technology for Botswanan English to propose it as a means of improving access to televised and radio-broadcast government communications for people who are DHH.

3.2 Data collection. To gain better insight into the possibilities of using ASR for information accessibility for people who are DHH, the study used qualitative data. The idea was to measure the quality of an ASR transcription by comparing it to a human-transcribed and edited text (TER). The number of errors could presumably serve as a qualitative index of how the ASR software performs. The ASR software used for the study was *Speech Typing* (Speechtyping.com). The researcher opted for *Speech Typing* because the software is freely available and simple to use, even though it is only available on Google Chrome. *Speech Typing* is customized to transcribe English varieties from India, Australia, Canada, Kenya, Tanzania, Ghana, New Zealand, Nigeria, South Africa, the Philippines, the United Kingdom, and the United States. Given that the system has not been trained with Botswanan English, a one-minute audio was used to determine which of the different varieties would yield the best results. The results showed no or insignificant differences. Indian English was therefore used as a default setting.

The paper used samples of official government communications in English collected from Botswana Television's (BTV) Facebook page. Although a total of 10 audio-visual documents accounting for more than 5 hours of audio were transcribed, only 3 abstracts accounting for more than thirty minutes of audio will be used for demonstration purposes, since most documents had almost the same accuracy rate.

4. Analysis procedure

First, the audio documents were transcribed using the ASR software. Then the researcher manually transcribed and edited the same documents. During the editing process, the speakers' mistakes were not corrected but transcribed as is. Capitalization and punctuation of the text were not considered since the ASR produced a continuous raw text that was neither capitalized nor punctuated. The use of either American spelling or United Kingdom spelling was not taken into consideration. As such, if the ASR software rendered *honourable* and *honorable* in the same text, the word was not considered an error.

Then the accuracy rate, measured in Word Error Rate (WER), was obtained by comparing the texts transcribed by the ASR software (ASR-TT) to the transcription generated and edited by the researcher, TER. Amberscript Wer Tool (<https://www.amberscript.com/en/wer-tool/>) was used to calculate the WER. WER is measured as a percentage of errors for every 100 words. The number of errors in the ASR-TT is divided by the number of total words in the TER. Errors include substitutions, insertions, and deletions that occur in the ASR-TT. A substitution (sub) occurs when a word in the TER is replaced by another word in the ASR-TT, for instance *firmly* transcribed as *family*. An insertion (ins) occurs when a word that is not present in the TER is added to the ASR-TT. A deletion (del) occurs when a word in the TER is left out in the ASR-TT. For analysis purposes,

errors are transcribed in brackets [] but are not categorized as either substitutions, insertions, or deletions.

Research indicates that a more accurate ASR transcription will produce a lower WER score, whereas a less accurate ASR transcription produces a high WER score. A human transcribed text WER score is usually around 4%⁵. ASR transcription of WER of 5-10% is usually considered of very good quality. A transcription of a WER of between 10 to 20% is acceptable whereas a transcription of a WER of 30% or more is considered unacceptable⁶. It should be noted that WER does not account for the reason why errors occur. Factors that can affect WER, without necessarily reflecting the capabilities of the ASR technology itself, include recording quality, microphone quality, speaker pronunciation, background noise, unusual names, locations, and other proper nouns, technical or industry-specific terms.

5. Results and discussions

Document 1

In March 2023, HB Antwerp officially opened the doors of HB Botswana, the company's first branch outside Antwerp. The following is an abstract of the speech pronounced during the inauguration ceremony by the Honourable Lefoko Moagi, Minister of Minerals and Energy. The speech was recorded indoors, in the presence of invited guests and journalists. The extract covers four minutes and eleven seconds of a thirteen-minute-long address.

“thank you director of ceremonies or director of proceedings as you put it the excellency the president of the Republic of Botswana Dr **mokgweetsi [mokwetsi]** Eric **[keabetswe]** masisi and our first lady Mrs **Neo Jane [Neil Jain]** masisi **honourable [you're one of the]** vice president **rre [rani] [tsogwane]** member of parliament **Boteti west [with the US]** honorable minister **[mister] Matthias Diependaele [mathius dipendao]** minister of finance budget living and **immovable [movable]** heritage of Flanders **honourable [I will]** **peggy serame [pegaserame]** your **[or]** counterpart minister of finance here in Botswana **honourble [on our]** ministers here present and members of Parliament **honorable [ary pono moathodi]** the deputy speaker of the National assembly of Botswana **[and]** miss **[emma peloetletse]** permanent secretary to the president as well as your deputy **mme mmabokole [mem mabukule]** excellency **[mme mmasekgoa Masire]** **mwamba [members] [require] [Botswana]** ambassador **in [and]** Belgium and other excellencies of the **diplomatic [ripple] [corps]** here present miss Ellen Richard **madisa [madiza]** permanent secretary ministry **[minister]** of minerals and energy **deputy** chief of staff to **minister [minister diependaele]** miss **Naseem lahri [nasimlari]** managing director Lucara Botswana Mr pule founder **[found ceo]** of Africa lighthouse capital captains of **industries [industry]** here present members of The Fourth Estate lovely HB people Young team are you there ladies and gentlemen a very good morning to you all your excellency distinguished guests ladies and gentlemen I'm greatly honored **its only that [as long as]** I don't know how to show it and deeply excited to be part of this very special **occasion [location]** today the groundbreaking ceremony for **HB [HBO]** Botswana facility to be constructed here at the digital and Innovation hub **under [and]** our **diamond [time and]**

⁵ <https://smartaction.ai/blog/does-word-error-rate-matter/>

⁶ <https://learn.microsoft.com/en-us/azure/ai-services/speech-service/how-to-custom-speech-evaluate-data?pivots=speech-studio>

story diamonds for development this is a story well-known to the world over and today we're witnessing yet another **ground [grand]** breaking highlight in our government's relentless drive to make sure the story continues to be told for many years to come director of proceedings I'm delighted more so because today we're here to Witness the fruits of his **excellency's [excellencies]** drive to market Botswana to investors as the destination of choice HB **Antwerp [antrimp]** is one of those who have agreed to partner with Botswana and begin a journey that will go along when promoting diamond beneficiation driving economic diversification and the development of a private sector LED economy **distinguished guests [these things gets]** today we break ground on many Frontiers as **we seek to [with 60]** expand and grow meaningful participation in the entire diamond value chain we are breaking ground on the **frontier [front of]** participation in the value chain we will expand local and citizen participation in the midstream that is the cutting and polishing which for a long time stood at around 2% share of the global sector the Investment is a step in the right direction to **ensure [showing]** increase our stake in addition we are breaking ground in our participation in the downstream which currently has not much footprint in the country as you continue to **break [pray]** this ground and Frontiers of skills development our partnership with HB Antwerp **[and 12]** will through HB Botswana Foster deeper collaboration and knowledge sharing between **Antwerp [entreprene and]** Botswana **thus [that]** cementing **Botswana's [bosanna's]** reputation as a modern technology driven diamond manufacturing country”

The total word count of the extract is 493 words. 69 errors were identified in the extract. This corresponds to a WER of 14,1%. 42 out of 69 errors, accounting for 61%, were identified in the introductory part of the speech, where an important number of proper nouns were encountered, most of which were in Setswana. Indeed, proper nouns, which are part of out-of-vocabulary words, are known to have a strong impact on WER. Out-of-vocabulary (OOV) words are unknown words that appear in the testing speech but not in the recognition vocabulary (Qin, 2013). This, according to Laurent et al. (2014), is because there is much less predictability in how proper nouns may be pronounced than for regular words, depending on both the origin of the speaker pronouncing it and the origin of the proper noun itself. The more foreign the origin, the less predictable the pronunciation. This high variability induced by low predictability is a source of difficulty for ASR systems when dealing with proper nouns (*Ibid.*). Furthermore, the recognition errors made on proper noun errors also affect the recognition accuracy of the surrounding words (Qin, 2013). In other words, the influence of the predictability of proper nouns extends further than just the recognition of proper nouns themselves and can cause a WER of about 50% on the words to the left and to the right of the proper noun (Dufour, 2008). If proper nouns are excluded from the above calculation, then the WER falls below 7,0%. The fact that the speaker was at times turning his head away from the microphone could also explain some errors given that his voice was a bit indistinct, with background noise.

Document 2

The extract below is the transcript of the public address by the Botswana Minister of Labour and Home Affairs, Honourable Anna Mokgethi during the official launch of the use of national identity cards as travel documents between Botswana and Namibia at the Mamuno border gate. The speech was recorded outdoors and there was some background noise and a bit of wind. The extract covers four minutes and twenty-two seconds of a thirteen-minute-long address.

“to update you on the historic event that will take place at **Mamuno [mamono]** border this Friday where the Republic of Namibia and the Republic of Botswana will be signing a memorandum of understanding to allow for the use of the national identity cards by their citizens for cross border travel between the two countries the signing of this agreement will take place on Friday the 24th 2023 at Mamuno **Trans-Kalahari [transcalahari] border [for the]** post this is an initiative ladies and gentleman by the two heads of state being his excellency **[Mokgweetsi Masisi]** the president of the Republic of Botswana and his excellency doctor **[Hage Geingob]** the president of the Republic of Namibia way back in 2019 the use of the national Identity Card ladies and gentlemen for cross-border will enhance the momentum of social and economic and regional integration and further promote the safe and orderly migration between these two countries this will be impactful or more impactful on the people's lives as mamuno trans Kalahari border **[as Mamuno] Trans-Kalahari [trans calahari]** border post attains a fully-fledged status of One Stop border post operating on a 24-hour basis this financial year it also will **foster [first]** the social cohesion amongst the Citizens of these Two Nations who share a profound historical and cultural and economic bonds and in addition to that they share the values of **democracy [democracies]** self-direct determination as a people human rights peace rule of law and good governance **mamuno [mamuna]** border post will be the first among the Botswana Namibia borders to commence the use of the national identity as a travel document and as time goes on **kasane [kasani]** immigration office **mohembo [Mohammed]** and ngoma **border [but]** will follow suit if I may just mention ladies and gentlemen that at this point in time once this ones will follow first and then dobe border will follow later once we have installed all the infrastructure at dobe **[and]** we are expecting to do that during this financial year we anticipate that with the use of the national identity card as travel documents it will facilitate families to be brought together and the cost associated with cross **border travel [volatile] financially [financial]** and time-wise will be reduced and all this will be to the advantage of the Citizens of Botswana and Namibia now the use of this travel the identity document as a travel document will achieve the objective of the **SADC [saddak]** Treaty which called for **SADC [saddened]** member states to develop policies aimed at the progressive elimination of all barriers or obstacles to the free movement of people goods and services to this end ladies and gentlemen both countries have taken serious measures and precaution to secure this national document against fraud and misuse”

The total word count of the extract is 445 words. Despite the presence of background noises from the wind and attendees, only 23 errors were identified in the extract. This corresponds to a WER of 5,16%. If minor errors and non-English proper nouns are excluded from the calculation, then the WER falls to 2,7%. Minor errors are transcriptions that do not affect the comprehension of the text (*mamuna, mamono, kasani, transcalahari, financial, democracies*). It is also worth noting that the acronym SADC was mentioned on two occasions and on both occasions, it was transcribed differently (*saddak and saddened*). Acronyms form a specific set of OOVs and the difficulty in handling them has been documented (Mansikkaniemi and Kurimo, 2015; Miranda, Neto, and Black, 2012). If commonly used acronyms might be easily recognized by ASR systems, less used acronyms are not present in the dictionary, and their pronunciations do not really spell out all the letters that make up the acronym (Miranda, Neto, and Black, 2012).

Document 3

The following is the transcription of a five-minute speech pronounced by the President of the Republic of Botswana, Dr Mokgweetsi Eric Masisi, during the second Summit for Democracy initiated by US President Joe Biden, virtually held in March 2023.

“your excellencies and dear colleagues all dignitaries ladies and gentlemen a very good morning or afternoon to **you all [your]** the theme of this plenary session suggests that without democracy established and **firmly [family]** in place it would prove difficult to talk about economic growth or shared prosperity we know this to be the case in many parts of the world where the absence of democracy is reflected in armed conflicts human displacements economic collapse absence of freedoms violations of human rights and no shared prosperity over the years **Botswana [Madonna]** has received many accolades as Africa's economic **miracle [medical]** and the Beacon of peace and democracy to many little attention was given to a forgotten reality that at the time of our independence in 1966 Botswana was among the poorest countries in the world a tactless Wasteland with no hope as some **opined [applied]** indeed there was no prosperity to share and **little [liquid]** was known about a democracy which is founded on the principles already practiced in the **setswana [tzuana]** culture of **therisanyo [teresano]** or consultation our democratic ideals and the economic transformation have long been set on a strong Foundation of principles of democracy development unity and self-reliance and **botho [boruto]** which is humility our Founding Fathers were fully convinced that a constitutional and multi-party Democratic Republic had to be in place a **republic [at a public] premised [premise]** on the primacy of respect for and commitment to the rule of law was best suited to take the country forward after independence in 1966 we enacted legislative instruments to sustain our democratic ideals and management of all natural resources of the country the national **vision 2036 [visa in 2013]** the national development plans recognized sustainability as a development imperative to be achieved I can confidently **ascertain [as attain]** that on the basis of a people centered governance democracy delivers a pathway for the needed avenues to pursue sustainable Development Goals **your [you]** excellencies our practice of consultation is a major component of our democratic discourse centered around our **kgotla [cordless]** system our traditional Congress so to speak which promotes tolerance inclusion freedom of assembly and speech and collective decision making the reason **we are [why]** able to participate in this summit is primarily linked to our firm grip on democratic principles of over 5 decades I wish to point out that we do not take for granted what democracy has delivered for us as a country we have learned by experience that there is a parallel between improving economic conditions and consolidating democracy we hold a strong view and conviction that a participatory democratic system comprising strong and functional institutions not strong men or women will ensure that the primary shareholders that is our people **[are]** at the epicenter of the formation and functionality of the structures of governance and shaping their Own Future your excellencies it is against this background that the government of Botswana and the national democratic Institute together with other partners hosted the summit on constitutionalism and democratic consolidation in Africa from six to eight July 2022 **[to]** in **Gaborone [in haboroni]** this summit **fulfilled [for failed]** the pledge we made during the 2021 summit for democracy and its key outcome was the adoption of **[the Gaborone]** declaration which among other issues acknowledge African citizens support for constitutionalism and presidential term limits it is our Earnest hope that the **wide [white]** spread circulation and signing of the Declaration of principles

will secure prosperous democratic future for our continent allow me to reiterate that Botswana strong economic growth and shared prosperity has been shaped our well-founded belief in a strong functional and a more inclusive **participatory [participated]** democracy I therefore wish to urge other global leaders to strengthen our commitment to advancing global Democracy in order to secure a foundation for Greater human progress finally I wish to thank the co-hosts for organizing this very important summit thank you”

The whole speech has a total word count of 628, and only 30 errors were identified. This corresponds to a WER of 4,77%. Three minor errors were identified (*tzuana, haboroni and teresano*). If minor errors and non-English proper nouns are excluded from the calculation, then the WER falls to 3,8%. This is an exceptionally low WER indicating a very high quality of the ASR transcription. One other thing that could explain this very good score is the high quality of the recording. The speech was recorded indoors and there was no or very insignificant background noise. Indeed, research has demonstrated that when a good-quality recording is used ASR systems cope well, with the resulting transcript being usable and, for the most part, accurate (Loakes, 2022).

6. Conclusions

The purpose of this study was to evaluate the accuracy rate of the ASR technology with the aim of proposing it as a means of improving access to televised and radio-broadcast government communications for people who are DHH. The results of our study indicated that ASR technology is adequate as an alternative to improve the accessibility of radio and television broadcasts for people who are DHH in Botswana. The ASR software used in this study performed exceptionally well in all the audio-visual documents that were sampled, with some documents scoring a WER of less than 5%, prompting us to believe that this software can even be used as live captioning software. One document scored a WER of just above 14% which was the highest of all the documents but still largely within the acceptable limits. From these results, we can deduce that transcriptions of government communications that are entirely in English will need little or no human intervention.

The ASR technology is rapidly improving and is already offering a suitable alternative for personal transcription needs. ASR is rapidly approaching the quality levels of human transcriptions and is already creating opportunities to assist marginalized and disadvantaged individuals who are navigating life with disabilities. Two public radio stations in Washington, D.C., looking to improve the accessibility of their broadcasts for people who are DHH, are using automated speech-to-text technology to provide live captioning of their programming⁷. Existing research on ASR shows promising results that can lead to a further reduction of error rates, making ASR transcripts readily available without any human intervention. Further testing is needed and planned.

References

- Ashwell, Tim and Elam, Jesse R. (2017). “How accurately can the Google web speech API recognize and transcribe Japanese L2 English learners’ oral production?” vol. 13, 59–76.
- Chinkonono, Gorgeous Sarah., Namuli, Vivian., Atuhaire, Catherine., Massaquoi, Hamouda., Mukhopadhyay, Sourav., and Cumber, Samuel Nambile. (2020). Access to Health Care Services by Individuals Who Are Deaf or Hard of Hearing Dhh in Botswana: Francistown and

⁷ <https://current.org/2022/12/how-public-radio-stations-can-serve-deaf-and-hard-of-hearing-audiences/>

- Tati, 15 June 2020, PREPRINT Version 1 available at Research Square [https://doi.org/10.21203/rs.3.rs-29186/v1]
- Cumbal, Ronald., Moell, Birger., Lopes, José., and Engwall, Olov. (2021). You don't understand me!": Comparing asr results for l1 and l2 speakers of swedish. In Proceedings of INTERSPEECH 2021 pp. 4463-4467. ISCA. <https://doi.org/10.21437/Interspeech>, 2021-2140.
- Derwing, Tracey., Munro, Murray and Carbonaro, Michael. (2000). Does popular speech recognition software work with ESL speech? *TESOL Quarterly*, 343, 592-603.
- Dufour, Richard. (2008). From prepared speech to spontaneous speech recognition system: a comparative study applied to French language. *Proceedings of the 5th international conference on Soft computing as transdisciplinary science and technology, Cergy, France*, 595-599.
- Echezona, Roseline Ifeoma. (2007). The Role of Library in Information Dissemination for Conflict Resolution, Peace, Promotion and Reconciliation. *African journal of library and information science*, 172:pp143-151.
- Foltz, Anouschka. (2022, September 13). *Sign languages are fully-fledged, natural languages with their own dialects – they need protecting*. The Conversation. Retrieved April 24, 2023, from: <https://theconversation.com/sign-languages-are-fully-fledged-natural-languages-with-their-own-dialects-they-need-protecting-109388>
- Gambill, Evan; Kushalnagar, Raja S.; Rodolitz, Jason; Vogler, Christian and Willis, Brittany (2019). Accessibility of Voice-Activated Agents for People who are Deaf or Hard of Hearing. *Journal on Technology and Persons with Disabilities Santiago, J. Eds: CSUN Assistive Technology Conference*, 144-156.
- Idowu, Taiwo., Okunlaya, Rifqah., Emmanuel, S. O, Amalahu, C., and Ibrahim, R. O (2017). Relevance and Use of News Bulletin in Public Information Dissemination within Academic Communities: A Case Study of Three Universities in Ogun State, Nigeria. *New Media and Mass Communication, Vol.60*, 1-9
- Kafle, Sushant and Huenerfauth, Matt. (2016). "Effect of Speech Recognition Errors on Text Understandability for People who are Deaf or Hard of Hearing," in 7th Workshop on Speech and Language Processing for Assistive Technologies, INTERSPEECH.
- Laloo, Bikika Tariang. (2002) *Information Needs, Information Seeking Behaviour and Users*. New Delhi: ESS Publication.pp.3-5.
- Laurent, Antoine., Meignier, Sylvain., and Deléglise, Paul. (2014). Improving recognition of proper nouns in ASR through generating and filtering phonetic transcriptions, *Computer Speech & Language*, 284, 979-996.
- Loakes Debbie. (2022). Does Automatic Speech Recognition ASR Have a Role in the Transcription of Indistinct Covert Recordings for Forensic Purposes? *Front. Commun.* 7: 803452.
- Mansikkaniemi, André and Kurimo, Mikko (2015). "Adaptation of Morph-Based Speech Recognition for Foreign Names and Acronyms," in IEEE/ACM Transactions on Audio, Speech, and Language Processing, vol. 23, no. 5, 941-950.
- McCrocklin, Shannon., Humaidan, Abdulsamad., & Edalatishams, Idée. (2019). ASR dictation program accuracy: Have current programs improved? In J. Levis, C. Nagle, & E. Todey Eds., *Proceedings of the 10th Pronunciation in Second Language Learning and Teaching Conference*, ISSN 2380-9566, Ames, IA, September 2018 pp. 191-200. Ames, IA: Iowa State University.
- McQuail, Denis. (1992). *Mass Communication and the Public Interest*. New Delhi: Sage.
- Miranda, João., Neto, João Paulo and Black, Alan W. (2012). "Recovery of acronyms, out-of-lattice words and pronunciations from parallel multilingual speech," 2012 IEEE Spoken Language Technology Workshop SLT, Miami, FL, USA, 2012, 348-353.

- Mukhopadhyay, Sourav. and Moswela, Emmanue. (2020). Disability Rights in Botswana: Perspectives of Individuals with Disabilities. *Journal of Disability Policy Studies 2020, Vol. 311*, 46-56.
- Mulholland, Matthew., Lopez, Melissa., Evanini, Keelan., Loukina, Anastassia., and Qian, Yao. (2016). "A comparison of ASR and human errors for transcription of non-native spontaneous speech," *2016 IEEE International Conference on Acoustics, Speech, and Signal Processing ICASSP*, Shanghai, China, 2016, 5855-5859.
- National Geographic. (n.d). *Sign Language*. National Geographic Education. Retrieved April 25, 2023, from: <https://education.nationalgeographic.org/resource/sign-language/>
- Qin, Long. (2013). *Learning Out-of-Vocabulary Words in Automatic Speech Recognition*. PhD thesis, Language Technologies Institute, School of Computer Science, Carnegie Mellon University.
- Radzikowski, Kacper., Wang, Le., Yoshie, Osamu and Nowak, Robert. (2021). Accent modification for speech recognition of non-native speakers using neural style transfer. *J AUDIO SPEECH MUSIC PROC.* 2021, 11 .
- Republic of Botswana. (2011). Botswana e-government master plan - 2015-2021. Retrieved on 21 March 2020 from <https://www.ellipsis.co.za/wp-content/uploads/2016/07/Botswana.pdf>
- Sullivan, Peter., Shibano, Toshiko., Abdul-Mageed, Muhammad. (2023). Improving Automatic Speech Recognition for Non-native English with Transfer Learning and Language Model Decoding. In: Abbas, M. eds *Analysis and Application of Natural Language and Speech Processing. Signals and Communication Technology*. Springer, Cham.
- Ulasik, Malgorzata Anna., Hürlimann, Manuela., Germann, Fabian., Gedik, Esin., Benites, Fernando and Cieliebak, Mark. (2020). CEASR: A Corpus for Evaluating Automatic Speech Recognition. In *Proceedings of the Twelfth Language Resources and Evaluation Conference*, pages 6477–6485, Marseille, France. European Language Resources Association.
- United Nations (n.d.). *International Day of Sign Languages*. United Nations. Retrieved April 25, 2023, from <https://www.un.org/en/observances/sign-languages-day>.

Kagiso Jacob Sello
Department of French
University of Botswana
sellojk@ub.ac.bw