Automatic Speech Recognition for Kreol Morisien: A Case Study for the Health Domain

Nuzhah Gooda Sahib-Kaudeer[(⊠)], Baby Gobin-Rahimbux,
Bibi Saamiyah Bahsu, and Maryam Farheen Aasiyah Maghoo

Department of Software and Information Systems, Faculty of Information,
Communicatio and Digital Technologies, University of Maurtius,
Reduit, Mauritius
{n.goodasahib, b.gobin}@uom.ac.mu, {bibi.bahsu,
maryam.maghoo}@umail.uom.ac.mu

Abstract. Automatic Speech Recognition (ASR) has revolutionized human-machine interactions as it allows the use of speech as an input modality. Speech is easy, natural and it is a skill that most people possess in their respective languages. Therefore, speech technology contributes to the usability and inclusivity of applications. ASR in languages such as English is extensively developed as there are large amounts of relevant resources available such as audio or transcribed data. For languages which are under-resourced, such as Kreol Morisien, ASR is a monumental task. In this paper, an attempt at developing an ASR system in Kreol Morisien is described. The ASR system was developed for the health domain to enable the automatic recognition of medical symptoms in spoken Kreol. The data collection process included the manual creation of a list of 848 symptoms along with 4000 audio files. Using the created corpus, the acoustic model for Kreol recognition was built and trained. This paper also describes a user evaluation which was conducted in different environments. Findings showed that the accuracy of the acoustic model was mainly affected by the level of noise. The gender of the speaker and the pronunciation style (depending on the region where the speaker originates from) did not cause any significant difference in the performance of the acoustic model.

Keywords: Automatic Speech Recognition · Kreol Morisien · Under-resourced languages

1 Introduction

Automatic Speech Recognition (ASR) has been the subject of research for many decades. However, with the recent popularity of technologies such as Amazon Alexa and Apple Siri, ASR has received a new surge of interest [1]. The worldwide technological advancements in terms of mobile devices such as smartphones and tablets have also highlighted the need for speech-based interactions [2] as speech is the primary means of human communication. Speaking is faster and more natural, therefore increasing the usability of many applications. Speech-based applications are also more inclusive [3] as they provide access to non-standard populations such as the elderly, the low-literacy group or the visually impaired.

Creating speech-based applications in well-resourced languages such as English and French is not a big task, since text-to-speech systems are already available for these languages. On the other hand, creating speech-based applications for languages that do not offer the resources for Human Language Technologies (HLT) is a monumental task. ASR in such cases require large amounts of transcribed data for the training process and very often, for these languages, there are no existing corpus of data that can be used. Generating this required transcribed data is an expensive process in terms of both manpower and time [4].

In Mauritius, to the best of our knowledge, there is only one previous research [5] on ASR in Kreol. It is most likely due to the absence of a corpus of text and audio data in the language. Yet, there are many possible applications of ASR in the Mauritian context since Kreol Morisien is spoken by the majority of the population [7]. For example, despite English being the official language, Kreol is used extensively in schools, the workplace and in most public institutions such as hospitals. In this paper, a first attempt at ASR in Kreol Morisien is presented whereby the authors describe their approach to building an acoustic model that is able to recognize spoken medical symptoms being experienced by patients. The health domain has been chosen only because of the authors' previous work in developing smart health applications for Mauritius [6]. The rest of this paper is structured as follows: Sect. 2 provides a literature review on Kreol Morisien and Automatic Speech Recognition. Section 3 describes the implementation of the acoustic model for Kreol recognition. In Sect. 4, the user evaluation process is outlined along with findings and discussions. We conclude the paper in Sect. 5.

2   Literature Review

2.1   Kreol Morisien

According to Ethnologue[1] (Accessed April 2019), the Kreol language, also known as Kreol Morisien, is the de facto language of national identity in Mauritius and is spoken by 1,339,200 around the world. Kreol can be defined as a French-based language including a number of words from English and from the African and South Asian languages spoken in Mauritius [6]. The status of Kreol Morisien has been the subject of an ongoing debate since Mauritius attained independence from the British in 1968. However, it is only in recent times that efforts have been made by the Government to formalize the language: In 2010, Akademi Kreol Morisien (AKM) was created and different committees were set up to define and standardize the spelling, syntax, pronunciation and grammar of the Kreol language. In 2012, the Government of Mauritius introduced the language in the curriculum of primary education.

2.2   Automatic Speech Recognition for Under-Resourced Languages

Automatic Speech Recognition (ASR) is an important technology for the most natural human-computer interaction, given that speech is a skill that the majority of people

---

[1] https://www.ethnologue.com.

have [1]. Speech technology can address barriers in human-human interactions (two people speaking different languages can use ASR to communicate seamlessly) as well as human-machine interactions (applications such as Voice Search [8] and Personal Digital Assistants [9]). ASR has already changed the way people live and work as speech becomes the input modality of human-machine interactions [1]. This is especially true for established languages such as English and French, for which a large amount of resources is available.

However, the same cannot be said for languages from developing countries which have so far received a lot less attention [12]. Yet, the need for speech technology in these languages is high as speech-based interactions are easy and thus accessible to a wider population including the low literate, the elderly and people with certain impairments [3]. The challenge for ASR in such languages is the limited availability of resources which has led to these languages being termed as 'under-resourced'. The concept of under-resourced language was introduced by [10] and [11]. In a survey for ASR in the context of under-resourced languages, [12] summarized the concept as a language with some or all of the following: "lack of a unique writing system or stable orthography, limited presence on the web, lack of linguistic expertise, lack of electronic resources for speech and language processing, such as monolingual corpora, bilingual electronic dictionaries, transcribed speech data, pronunciation dictionaries, vocabulary lists, etc."

In the context of Kreol Morisien, it can be considered as an under-resourced language mostly for the lack of electronic resources required for speech processing. In this paper, a first attempt at developing an ASR system in Kreol Morisien is described. The ASR system, through its acoustic model, aims to recognize spoken symptoms from patients using a health diagnosis tool. Thus, the conversation patients may have with a nursing staff while describing their symptoms is being simulated (A snapshot of such a conversation can be found in Table 1). Since, the focus of this paper is ASR, only the speech recognition part of this work is described, omitting details on health diagnosis.

Table 1. Examples of medical symptoms in Kreol Morisien and English

Kreol Morisien	English
Mo latet p fermal	*I have a headache*
Mo p gagne la fievre	*I have a fever*
Mo p tousser	*I am coughing*

3   Implementation of Acoustic Model

3.1   Data Collection

Since there are no existing corpus for Kreol Morisien, the implementation of the acoustic model included the data collection process during which both text and audio data was manually created.

Text Corpus. Since there are no corpus available for Kreol Morisien, the implementation of the acoustic model included the data collection process. A list of 848 commonly

used words to describe symptoms in Kreol was created and based on these words, a list of 2989 sentences was manually created to be used for language modelling.

Audio Recording. The audio for each word and sentence was recorded using Audacity[2] and saved as .wav files. Four different speakers (two males and two female) recorded 1000 audio files each. Therefore, a total of 4000 audio recordings was obtained. The absence of noise was ensured during the recording process as noise would cause interferences during the training of the acoustic model. Presence of noise would cause the amplitude of the audio to increase and therefore, it was ensured that the amplitude remained between −0.5 and 0.5.

3.2 Building of Phonetic Dictionary

A template dictionary of the list of 848 Kreol symptoms was constructed using the Lexicon tool[3] to understand the phonetic representation of each word (known as phoneme). Different pronunciations for the same word were catered for (see Fig. 1) to boost efficiency of the recognition model since the Kreol language is articulated differently by different individuals. The dictionary was built using the French phones since they are closer to Kreol pronunciation than English. For example, 'a' is represented as 'AE' in English phones whereas in French, it is represented as 'aa'.

```
acoz aa kk oo zz
adapter aa dd aa pp tt ei
adolessan aa dd oo ll ai ss an
adult aa dd uu ll tt
adilt aa dd ii ll tt
azitasion aa zz ii tt aa ss yy on
agitasion aa jj ii tt aa ss yy on
azitasian aa zz ii tt aa ss yy an
azitashion aa zz ii tt aa sh yy on
agitashion aa jj ii tt aa sh yy on
```

Fig. 1. Snapshot of Phonetic Dictionary

3.3 Building of Language Model

The Lexicon tool was used to generate the language model in order to calculate the probabilistic occurrence of words. A total of 2989 sentences and 784 words was used to build the language model.

3.4 Preparation of Transcript Files

The transcription files were manually created based on the audio recordings from the data collection process. Both Kreol_train.transcription and Kreol_test.transcription

---

[2] https://www.audacityteam.org/.

[3] www.speech.cs.cmu.edu/tools/lmtool.html.

have been prepared, one for training and one for testing respectively. Each word and sentence in the files were allocated a unique identifier. The transcription files was updated each time new audio recordings became available. This was an effort intensive task that required in depth revisions since mistakes could lead to failure in training.

3.5   Training the Acoustic Model

CMU Sphinx[4] was used to train the acoustic model with 80% of the audio recordings corresponding to 3.2 h of audio data. A phoneset file of all phones in the dictionary was created and a context dependent model was used for training. The details of the final version of the acoustic model are described in Table 1.

4   User Evaluation

A user evaluation was conducted to determine the accuracy of the acoustic model in correctly recognising the symptoms spoken by users in continuous speech. There were two main parts of the user evaluation, referred to as User Study 1 and User Study 2 for the rest of this paper. A set of 50 sentences in Kreol Morisien, that did not occur in the train and the test sets, was created to conduct the user studies. Bothe studies used the same sentences to ensure that while other variables such as level of noise were changing, the complexity of the speech was the same across studies.

4.1   User Study 1

The aim of this study was to determine the accuracy of the acoustic model in varying environments in order to simulate circumstances in which people may be using such an application in real-life settings. The participants and the methodology are described in the following.

Participants. Ten participants were involved in User Study 1 and they were divided into two groups (A and B) such that two different participants were assigned the same group of sentences. Additional demographic information about the participants which was collected through a questionnaire can be found in Table 2.

Table 2. Demographic information of participants in User Study 1.

	Group A	Group B
Age	34.5	21
Gender	Male (3) Female (2)	Male (3) Female (2)
Region	Urban (2) Rural (3)	Urban (2) Rural (3)
Level of Kreol	Native Speaker (5)	Native Speaker (5)

Methodology. The sentences were split in 5 sets of 10 sentences (S1 to S5) and each participant in Group A and Group B were assigned one set of sentences to speak. For comparison purposes, it was ensured that each set of sentences were assigned to speakers of the same gender from both groups. However, different speakers from each group tested the acoustic model in different environments in terms of noise levels. The participants spoke the sentences using the same hardware and the acoustic model output the transcribed speech for evaluation purposes.

Findings and Discussion. The ability of the acoustic model to recognize speech in Kreol Morisien is evaluated based on Word Error Rate (WER). WER is calculated as the total number of insertions, deletions and substitutions in the output of the acoustic model divided by the total number of words in the reference sentence. For each user study, the Sentence Error Rate (SER) is also provided. SER is the proportion of the sentences which have an error in them. In this paper, all reported WER and SER values have been calculated using the Python module for ASR evaluation[5].

The Word Error Rate for User Study 1 was 17.91%, that is, the overall accuracy of the acoustic model across all participants was 82.09%. In Fig. 2, WER for each participant from both Group A and Group B are displayed. Statistical testing was carried out at $p < 0.05$ using a two-sample t-test for unequal variances. There was no significant difference between Group A and Group B ($p = 0.07$). The regions from which the participants originated (Urban or Rural) and the gender did not cause any significant difference in the performance of the acoustic model ($p = 0.26$ and $p = 0.17$). The SER value was 57% across the sentences spoken by the participants.

In this user study, the authors did not control the environment with respect to noise level. Therefore, it was performed in mixed environments with some speakers inside a room with background noises like a running fan and some in open air with people
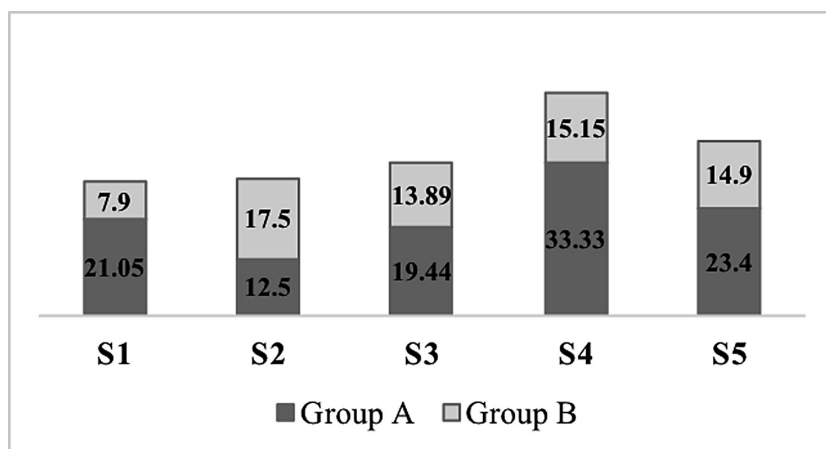

Fig. 2. WER of speakers in User Study 1

---

[5] https://github.com/belambert/asr-evaluation.

talking and moving nearby. The average accuracy is 82.09% for all the sentences across all speakers. The biggest differences in accuracy are between speakers 1A (21.05%) and 1B (7.9%) and speakers 4A (33.33%) and 4B (15.15%), despite each pair speaking the same sentences. This difference may have arisen because as per data gathered in the questionnaire, despite being a native creole speaker, speaker 1A speaks French on a daily basis and thus her accent is different from speaker 1B who speaks Kreol Morisien regularly. Speaker 1A was also in a noisier environment. The difference between speakers 4A and 4B may also have resulted due to the difference in environments.

4.2   User Study 2

Following User Study 1 in mixed environments where the accuracy of the acoustic model in different levels of noises was studied, User Study 2 was conducted with 10 participants in two different environments. The aim of this user evaluation was to study how the acoustic model performed in two different environments: a noisy environment as well as a quiet environment. For the noisy environment, an open corridor with people talking and laughing, sounds of doors opening and closing and people walking loudly was chosen. There was also a car park nearby and thus, there was also vehicle-related noises in the background. The quiet environment was indoors, in a classroom with closed doors.

Participants. Ten participants, who were all students from the University of Mauritius took part in this study. They were divided into two groups (A and B) such that two different participants were assigned the same group of sentences for each environment. Additional demographic information about the participants are given in Table 3.

Methodology. The same set of sentences as in User Study 1 were used whereby each participant in Group A and Group B were assigned one set of sentences (S1 to S5) to speak, irrespective of their gender. For comparison purposes, the environment was kept constant throughout the study, that is, for the first part all participants were in the noisy environment and for the second part, in the quiet environment. For example, speaker 1A spoken sentence set S1 in both the noisy and the quiet environments.

Table 3. Demographic information of participants in User Study 2.

|	Group A	Group B
Age	20	20.5
Gender	Male (3) Female (2)	Male (2) Female (3)
Region	Urban (2) Rural (1) Rodrigues (2)	Urban (3) Rural (2)
Level of Kreol	Native Speaker (5)	Native Speaker (5)

Findings and Discussion. As expected, WER for the quiet environment was 13.70% whereas for the noisy environment, it was 37.01%. Statistical testing was carried out at $p < 0.05$ with a paired t-test and the difference between the two environments was statistically significant ($p = 0.000004$). In the noisy environment, insertions and substitutions are more likely given the background noises and this significantly affected the

WER and the overall accuracy of the acoustic model. For the noisy environment, there was no statistically significant difference in the performance of the acoustic model for gender ($p = 0.30$) and region ($p = 0.24$). The SER value for the noisy environment was 90% while for the quiet environment it was 42%. Gender and Region did not cause statistically significant differences in the quiet environment ($p = 0.46$) and ($p = 0.12$).

For User Study 2, there were two participants (3B and 4B) from Rodrigues. Rodrigues is an autonomous outer island of the Republic of Mauritius and their style of Kreol can be different from people in the main island. Statistical testing was performed between participants from Mauritius and Rodrigues for the same sentences using a paired t-test at $p < 0.05$. Between participants 3A (from Mauritius, Rural region) and 3B, no statistically significant differences were observed for the ten sentences of S3 in both the noisy ($p = 0.11$) and the quiet environments ($p = 0.63$). Similarly, there were no statistically significant differences between participants 4A (from Mauritius) and 4B for the ten sentences of set S4 in the noisy environment ($p = 0.18$) and the quiet environment ($p = 0.94$) (Table 4 and Fig. 3).

Table 4. WER for participants in User Study 2

Sentence set	WER			
	Noisy environment		Quiet environment	
	Group A	Group B	Group A	Group B
S1	32.21	34.21	15.79	7.90
S2	30.00	40.00	15.00	17.50
S3	44.44	30.56	13.89	11.11
S4	54.55	42.42	15.15	15.15
S5	29.79	31.92	12.77	12.77
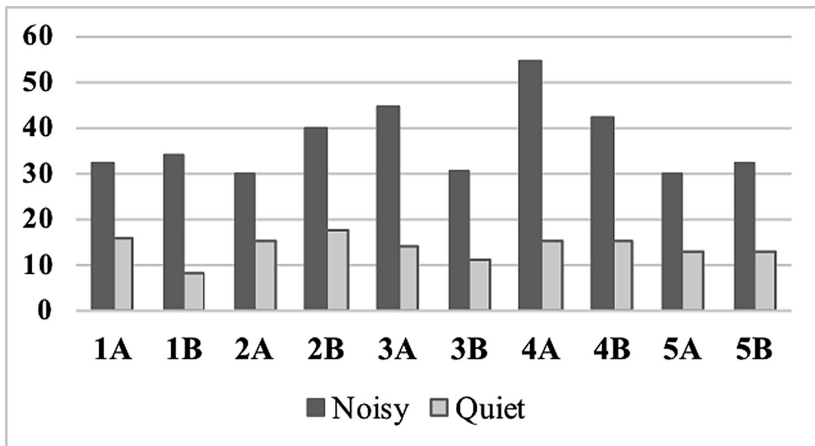
Fig. 3. WER for speakers in User Study 2

5   Conclusion and Future Work

In this paper, an initial investigation regarding Automatic Speech Recognition (ASR) in Kreol Morisien was presented. The context under study was the health domain whereby the aim of the ASR system was to be capture patients' symptoms as described through speech. Given the lack of a corpus in Kreol Morisien, the data collection process included the manual creation of both audio and transcribed data which was then used for training an acoustic model to recognize the language.

Given the widespread use of Kreol in Mauritius, speech technology can undoubtedly have a significant impact. However, given its under-resourced status with regards to the lack of resources for speech processing, the challenge is to investigate potential approaches for generalized ASR in Kreol without having to start from scratch as discussed by [12]. Future work will focus on how existing corpus for English and French can be used as a starting point in order to decrease the extensive efforts required to build a corpus for a new language from scratch.

References

 1. Yu, D., Deng, L.: Automatic Speech Recognition. Springer, London (2016)
 2. De Vries, N.J., et al.: A smartphone-based ASR data collection tool for under-resourced languages. Speech Commun. **56**, 119–131 (2014)
 3. Neerincx, M.A., Cremers, A.H., Kessens, J.M., Van Leeuwen, D.A., Truong, K.P.: Attuning speech-enabled interfaces to user and context for inclusive design: technology, methodology and practice. Univ. Access Inf. Soc. **8**(2), 109–122 (2009)
 4. Lamel, L., Gauvain, J.L., Adda, G.: Lightly supervised and unsupervised acoustic model training. Comput. Speech Lang. **16**(1), 115–129 (2002)
 5. Noormamode, W., Gobin-Rahimbux, B., Peerboccus, M.: A speech engine for Mauritian Creole. In: Satapathy, S.C., Bhateja, V., Somanah, R., Yang, X.-S., Senkerik, R. (eds.) Information Systems Design and Intelligent Applications. AISC, vol. 863, pp. 389–398. Springer, Singapore (2019). https://doi.org/10.1007/978-981-13-3338-5_36
 6. Aubeeluck, M., Bucktowar, U., Gooda Sahib-Kaudeer, N., Gobin-Rahimbux, B.: A smart mobile health application for mauritius. In: Satapathy, S.C., Bhateja, V., Somanah, R., Yang, X.-S., Senkerik, R. (eds.) Information Systems Design and Intelligent Applications. AISC, vol. 863, pp. 333–343. Springer, Singapore (2019). https://doi.org/10.1007/978-981-13-3338-5_31
 7. Baker, P.: Kreol: a description of Mauritian Creole. C. Hurst, London (1972)
 8. Wang, Y.Y., Yu, D., Ju, Y.C., Acero, A.: An introduction to voice search. IEEE Signal Process. Mag. **25**(3), 28–38 (2008)
 9. Milhorat, P., Schlögl, S., Chollet, G., Boudy, J., Esposito, A., Pelosi, G.: Building the next generation of personal digital assistants. In: 2014 1st International Conference on Advanced Technologies for Signal and Image Processing (ATSIP), pp. 458–463. IEEE (2014)
10. Krauwer, S.: The basic language resource kit (BLARK) as the first milestone for the language resources roadmap. In: Proceedings of SPECOM 2003, pp. 8–15 (2003)
11. Berment, V.: Méthodes pour informatiser les langues et les groupes de langues peu dotées. Doctoral dissertation, Université Joseph-Fourier-Grenoble I (2004)
12. Besacier, L., Barnard, E., Karpov, A., Schultz, T.: Automatic speech recognition for under-resourced languages: A survey. Speech Commun. **56**, 85–100 (2014)