# Construction of a Finite Grammar for Enhancing the Language Model for the Recognizer of Northern Sotho.

*T.M Zvenyika, H.J Oosthuizen, M.J.D Manamela*
Department of Computer Science,
University of Limpopo (Turfloop Campus),
Private Bag X1106,
Sovenga, 0727
Tel: +27 72 157 0502 / +27 15 268 2995
E-mail: tapiwazv@gmail.com

**Abstract - This research paper focuses on the progress in the construction of a finite grammar for enhancing the language model for the Northern Sotho speech recognizer. This automatic speech recognizer will play an important role in enabling a human-computer interface technology suited for domain-specific user applications. To enable this process, we collect nearly all possible speech utterances that can exist in a specific domain and construct a finite grammar suitable to cover most common utterance instances in that domain before testing the speech recognizer performance with that finite grammar. In this project we restrict our grammar to the birth, death and marriage registration systems domain of the Department of Home Affairs where queries are posed by applicants wishing to register birth, death or marriage events with appropriate responses telephonically provided by the system. This system approximates to an IVR system focused on a specific domain of information retrieval for event registration in Northern Sotho.**

## 1. Introduction

In the past few decades, general-purpose rule-based automatic speech recognition (ASR) systems have been developed [1]. Unfortunately these systems cannot be effectively adapted to small and medium vocabulary and domain-specific voice-enabled applications because of their generality. Already aviation systems, car computers, medical transcriptions, telephones as well as many other devices rely heavily on speech recognition and synthesis in a context-dependent environment. Many general day-to-day appliances may be more likely candidates for this technological revolution that may drastically reduce the use of keyboards and other computer input/output peripheral devices [2]. Most emerging technologies in speech recognition are domain-specific and/or context-dependent to capture all possible utterances that may exist within a defined context. These technologies have not been widely deployed as far as native African languages are concerned.

The first Northern Sotho automatic speech recognizer that was constructed was a very general ASR baseline system [3]. This was then enhanced by a language model (LM) to increase its recognition rate. This project then seeks to enhance the baseline ASR system with its language model by taking into consideration its finite grammar specifications to develop a simple IVR application. The main idea is to investigate the effect of restricting the speech recognition grammar on the performance of the speech recognizer. In general, the purpose of a speech grammar is to define what the user can say. However, our main aim is to improve the recognition rate of the baseline system [1] with a Language Mode which obtained a sentence recognition average rate of 51.1% and a word recognition rate of 84.4% using speaker adapted acoustic models. The second section briefly refers to similar projects on domain-specific automatic speech recognition applications. The third section outlines the approach to finite grammar specifications. In the fourth section, we outline our data collection procedure and lastly we give concluding remarks on the state of the project.

## 2. Similar Projects

One of the most notable domains for the commercial applications of speech recognition has been the health care domain and in particular the work of the medical transcriptionists. With the average processor now at and above Pentium III and RAM levels at 500MB and up, accuracy levels have reached 95% and better with transcription speeds at over 160 words per minute [4]. Such speech recognition systems will only transcribe utterance that are contained in the medical database and reject the out-of-vocabulary words.

In the last decade substantial effort has been devoted to the test and evaluation of speech recognition in fighter aircraft. Of notable instance is the program in France on installing speech recognition systems on the Mirage [5] aircraft platform. In both these major speech recognition advancements, only very limited, constrained vocabularies have been used successfully.

With the emergence of the mobile telephone technology, the technical challenge is to provide mobile users with embedded speech recognition tools and/or services. Since the mobile device has limited computing resources to perform the speech recognition, a remote server with ample resources can perform the speech recognition in mobile devices. Mobile devices become

easier to operate on the road with these speech recognition and synthesis interfaces.

### 3. Finite-grammar Specification

A syntactic model defines a set of sentences that are acceptable. This set of acceptable sentences can be specified in terms of a finite-state grammar, a context-free grammar or a unification grammar [6], [7]. These grammars have a fixed number of allowed utterances. The sentences used in the prompt sheets were chosen to cover most frequently used queries on birth, death or marriage registration systems in general.

The chosen grammar is subsequently integrated into the recognizer so that only legal sequences and partial sequences are hypothesized and evaluated. Finite-state grammars are a class of Markov models which can easily be implemented in an HMM-based speech recognition system [8]. In this class only the word sequence that complies with our restricted grammar can be recognized.

Finite grammars have a set of rules which delimits the search space that the speech recognizer will search to determine what the caller said. If the caller clearly utters something within this search space, then it should be recognized with a high probability. If they utter some something similar to a permissible phrase, it should be recognized as the similar phrase, though with a lower probability that reflects the small variation in these phrases. If the caller says something that is outside the search space and it doesn't sound anything closer to an allowed phrase no matching phrase will be found. Finite grammars work well for "command and control" voice-enabled applications [9].

In this project, we focus on the case of the Department of Home Affairs helpdesk system and its Frequently Asked Questions (FAQs) to get and cover a reasonable number of typical queries that a user can ask. Our envisaged system has to be close to reality to achieve a high and acceptable recognition rate within this domain. The speech recognizer is to be developed using the Hidden Markov Model Toolkit (HTK). A distinction in Automatic Speech Recognition is often made between "artificial syntax systems" which are usually domain-specific, and natural language processing systems which are usually language specific - each of which presents its own particular goals and challenges.

### 4. Data Collection

An existing telephone Northern Sotho speech database from [1] will be used to train an ASR engine. The testing data for this project will be created by recording utterances from recruited volunteers who are first language speakers of Northern Sotho. These microphone recordings will be made by cutting frequencies above 4000Hz to be compatible with the telephone recorded speech database. The testing data is a set of recorded raw speech samples from 30 chosen Northern Sotho speakers to counter the variability in speech styles like voice tone, production and speed. The specially prepared prompt text is used to elicit speech material from recruited participants. Most of these participants were recruited from our University of Limpopo community which is located in the midst of Northern Sotho speaking communities. The DictMaker [10] tool for setting-up and creating dictionaries will be used to append pronunciations of words available in the prompt text and not in the pronunciation dictionary created in [1]. A sound-proof room is used for recordings to ensure noise-free speech data. All recordings were collected using a cordless Sennheiser-Freeport noise cancellation microphone.

### 5. Concluding Remarks and Future Work

In this paper we presented our current work regarding the construction of an automatic speech recognizer that recognizes domain-specific utterances based on a finite-grammar. We have solicited recordings from 30 different volunteering subjects, in a sound proof room to eliminate the noises effects. Our future work will be to prepare the recorded testing speech data set to test the existing ASR system using the HTK toolkit to see the effects of incorporating a finite grammar into the speech recognizer.

### 6. References


[1] T.M Modiba, "Aspects of Automatic Speech Recognition (ASR) with Respect To Northern Sotho", University of Limpopo, 2005.

[2] C Becchetti and L.P Ricotti, "Speech Recognition Theory and C++ Implementation", John Wiley & Sons, Rome, 1999, pp 4.

[3]http://www.meraka.org.za/hlt_Sepedi.htm (Last accessed on 01/06/2008)

[4]http://archive.healthmgttech.com/archives/0205/0205is_speech.htm (Last accessed 20/05/2008).

[5] D.B Stockton, "Evaluating Syntactic Constraint to Speech Recognition in a Fire Fighter Aircraft Environment", Air Inst of Tech Wright-Patterson AFB OH School of Engineering, USA, 1984.

[6] B.T Lowerre and D.R Reddy, "The HARPY Speech Understanding System", Trends in Speech Recognition, Speech Science Publications, Minno, 1980.

[7] C Hemphill, J Picone, "Robust Speech Recognition in a Unification Grammar Framework", In IEEE International Conference on Acoustics, Speech, and Signal Processing.

[8] A. Waibel, K. Lee," Language Processing For Speech Recognition", Morgan Kaufmann Publishers, Palo Alto.

[9]http://www.voice-push.com/voice-push/grammar_design.php (Last accessed on 01/05/2008).

[10] http://www.hostalive.net/dictmaker.html (Last accessed on 24/05/2008).



**Tapiwa Zvenyika** holds an Honors degree in Computer Science from the University of Limpopo (Turfloop Campus) and is currently reading for his Masters degree in Computer Science at the same university. He is working in the Telkom Centre of Excellence for Speech Technology, with focus on Automatic Speech Recognition.