Towards a Multilingual Recognition System Based on Phone-clustering Scheme for Decoding Local Languages

Edward Lebese[1], Jonas Manamela[2] and Nalson Gasela[3]
The Centre of Excellence for Speech Technology
Department of Computer Science
University of Limpopo (Turfloop Campus), Private Bag X1106, Sovenga 0727
Tel: +27 15 268 2751, Fax: +27 15 268 3487
email: kalevratic@gmail.com[1],{jonas.manamela[2], nalson.gasela[3]}@ul.ac.za

***Abstract*-This paper presents the development of a multilingual speech recognizer for Northern Sotho mixed with English. This multilingual style communication has inspired South African research groups affiliated to speech and language technology (SLT) to work towards the development of multilingual speech recognition systems that accept, accommodate and handle mixed speech encountered in daily communication episodes. The multi-pass and one-pass recognition frameworks are two reported schemes for development of a mixed speech recognition systems. To obviate problems of language boundary detection and language identification (LID) involved in the former scheme, the one-pass recognizer with a multilingual language and acoustic models is preferred. These models are used to support clustering of similar speech sounds across the targeted languages so as to reduce the recognition system training data and thereby contribute to the performance improvement of one-pass recognizer. The preliminary results of an HMM-based one-pass recognizer are presented.**

Index Terms—code-mixing, cross-lingual, phone-clustering algorithm, ML-ASR and HMMs.

I. INTRODUCTION

The migration of one or more monolingual socio-cultural groups from one region to the next seeking places to settle, rural-urban migration and also multilingual diversity of South African population encourage code mixing in conversations. Most native speakers often use non-native words in their formal and informal daily conversations. This phenomenon is generally found to be on the increase especially in our modern information and communication technology (ICT) era. In many developing countries people tend to move from their respective rural areas to urban areas. This urbanization tendency encourages the cosmopolitan use of more than one spoken language among speakers willing to achieve mutual understanding.

Code mixing refers to the intra-sentential switching of two different languages in an utterance [1] and code switching or inter-sentential code-alternation occurs when a bilingual speaker uses more than one language in a single utterance above the clause level to appropriately convey his/her intents [2]. In bilingual societies, code mixing and code switching are common phenomena [1][3]. As they are encountered in day-to-day communication actions, the two closely linked concepts serves as tools for enhancing understanding among speakers. As a result, this steers South African human speech technology (HLT) research interests towards a new direction within speech recognition systems, i.e., multilingual automatic speech recognition (ML-ASR). In South Africa mother tongue speakers of any African language tend to revert to the English language each time they are referring to *numeric numbers* of any kind, when referring to *time* stipulations (be it in a rephonologized form, e.g., "u-half past six" or "u-quarter to one"), and when using the *alphabet* (including foreign ones such Greek alphabet) [4]. It is true that modern telephone conversations and/or computer-based interactions are facilitated through a mixed language speech. As a result, it is necessary to develop automatic speech recognition for African indigenous languages such as Northern Sotho (L1) mixed with English (L2) one of the global and widely used languages of South Africa. The inevitable changing trends and dynamics of local language usage in modern ICT oriented society also calls for this spoken language processing development avenue.

This paper is outlined as follows: Section II start with a basic architecture of speech recognition systems. Section III describes an overview of the background of architectural structure to code-mixing in multilingual automatic speech recognition systems. Section IV gives the proposed code-mixing ML-ASR structure with the methods and the protocols used in the design. The preliminary results of our mixed language recognition are presented and discussed in Section V. Section VI presents concluding remarks on this research project.

II. BASIC STUDY OF SPEECH RECOGNITION SYSTEMS

For the given acoustic observation $\boldsymbol{X} = \mathbf{X}_1,\mathbf{X}_2,\ldots,\mathbf{X}_n$, the goal of speech recognition is to find out the corresponding word sequence $\hat{\mathbf{Y}} = \mathbf{Y}_1,\mathbf{Y}_2,\ldots,\mathbf{Y}_m$ that has the maximum posterior probability $P(\mathbf{Y}|\mathbf{X})$ [5].

$$\hat{Y} = \underset{w}{argmax}\, P(Y|X) = \underset{w}{argmax} \frac{P(Y)P(X|Y)}{P(Y)} \quad \ldots\ldots\ldots\ldots (1.1)$$

In a mixed languages recognition, a speech signal may be viewed as a sequence of observation symbols: $\boldsymbol{Z} = \boldsymbol{O} \cup \boldsymbol{Q} = \boldsymbol{O}_1, \boldsymbol{O}_2,\ldots,\boldsymbol{O}_T \cup \boldsymbol{Q}_1, \boldsymbol{Q}_2,\ldots,\boldsymbol{Q}_R$ that represents a string composed of elements of alphabets V and P of symbols belonging to languages LL_V and LL_P respectively. If, in addition, we have a vocabulary $(V \cup P)$ of all the mixed language words w_i, $1 \leq i \leq |V|+|P|$, which can be uttered. Then mathematically speaking the speech recognition problem comes down to finding the mixed word sequence $\hat{W} = W_1, W_2, \ldots, W_m$ that have the maximum probability of being spoken, given the cross-lingual acoustic evidence Z. Thus we have to solve the following equation:

$$\hat{W} = \underset{W}{argmax}\, P(W|Z) \qquad \ldots\ldots\ldots\ldots(1.2)$$

It is unfortunate we cannot directly and easily compute equation (1.2) since the numbers of possible observed sequences are many. But Bayes formula, which shows relationship between conditional probabilities, reduces to:

$$P\,(W|Z) = \frac{P(W)P(Z|W)}{P(Z)} \qquad \ldots\ldots\ldots\ldots(1.3)$$

where P (**W**), is called the cross-lingual language model, is the probability that the word string **W** will be uttered and P (**Z**|**W**) is the probability that when word string **W** is uttered the super acoustic evidence **Z** will be observed; the latter is called the super acoustic model. The probability P(**Z**) is usually not known but for a given utterance it is of course just a normalizing constant and can be ignored. Thus to find a solution to formula (1.2) we have to find a solution to:

$$\hat{W} = \underset{W}{argmax}\, P(W)P(Z|W) \qquad \ldots\ldots\ldots\ldots(1.4)$$

The extraction of feature vectors forms the first phase of data preparation before training a speech recognizer. There are several techniques used to extract speech signal feature vectors such as Linear Prediction Coefficients (LPC), Mel Frequency Cepstral Coefficients (MFCC), and Perceptual Linear Prediction Coefficients (PLP). We use MFCC to extract speech acoustic feature vectors from raw waveform data when training a multilingual speech recognizer to develop a limited vocabulary continuous speech recognition system using hidden Markov models (HMMs) for decoding speech data drawn from L1 and L2 languages (mainly focusing on phrases encompassing digits strings) within predefined communication domain. A Hidden Markov Model Toolkit (HTK) is used for experimentation purposes in this research work [6].

III. MULTILINGUAL ASR SYSTEMS DESIGN

The recognition of mixed language speech is still deemed to be in its initial stages of research [7]. Some semblance of code-mixing can be traced back to the African Speech Technology (AST) project [4] - a South African-based speech technology project dealing with telephone-based information systems implementation. The project included a language monitoring module to allow more elegant recovery in case the two different languages are encountered. There are two reported frameworks for developing mixed language speech recognition systems: multi-pass and one-pass [7][8]. The former approach contrasts with the latter approach in that multiple monolingual speech recognition systems are used. They are selected by language identification (LID) process depending on the input language, while the latter approach is motivated by an integrated system to allow monolingual speech recognition of more than one language by a single speech recognition system [8].

A. Multi-pass recognition framework environment

In this framework the challenge is to accept, identify and divide a mixed speech sentential form into its distinct segments belonging to each of the participating language. Segmenting of such an utterance into segments of different languages is crucial step towards the development of a LID module [9]. The main challenge is to identity the language that is being spoken in an utterance [7][9].

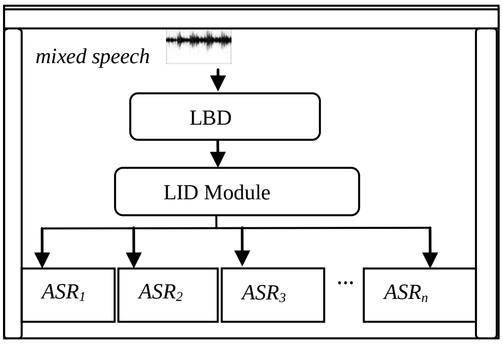

Figure 1. Multiple monolingual systems with language identification

Figure 1, represents components involved from language identification to speech recognition. The overall recognition performance of a multi-pass approach depends on (a) the performance of the language boundary detection (LBD), (b) the language identification block and (c) the actual performance of the language dependent ASR. So a poor performance by any one of the three blocks affects the overall performance of the multilingual recognizer [7]. This framework inhibits the possibility of sharing of data across the languages. As a result previous research studies of code-mixing and code-switching discourage the development approach using multi-pass pass environment[7][8][9].

B. One-pass recognition design

The speech recognition through a one-pass recognizer design attempts to avoid the possible drawbacks that may be introduced by a multi-pass recognition framework. This approach has benefits over explicit integration of monolingual recognizers, as the single recognizer makes less demand on memory and processor usage [8]. A one-pass recognition design is usually divided into two protocols namely:

1) Single recognizer with a multilingual language model and monolingual acoustic models.
2) Single recognizer with a multilingual language model and a multilingual acoustic model.

The common advantage of these two protocols is that the performance degradation introduced through LID is avoided. The first protocol supports different acoustic model from the languages in participation. The two are motivated by an integrated system to allow monolingual recognition in several languages by a single recognition engine [8]. The second protocol design supports sharing of similar sounds across the participating languages offered by a multilingual acoustic model. It requires building the multilingual pronunciation dictionary, cross-lingual language model and multilingual/super acoustic model that encompass all the languages in the mixed language episodes. It is common in many countries that similar sounds across languages can be are shared. Figure 2 illustrates one-pass recognition framework with multilingual acoustic model.

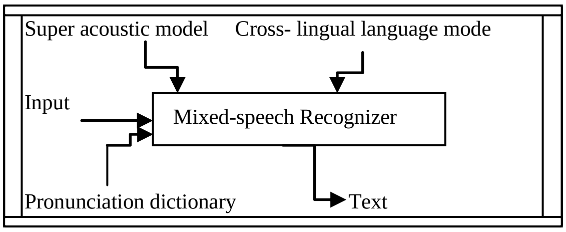

Figure 2. Single mixed recognizer with a multilingual language model and a multilingual acoustic model (adapted from [7])

A super acoustic model is generated for the phone set (PS) of the targeted languages in a mixed language setting. It is observed that a super or cross-lingual acoustic model may require a large collection of speech data. Thus, to build a robust super acoustic model-access to a collection of mixed language speech and text corpus-similar sounds across the two languages may be clustered to reduce the size of the bilingual pronunciation dictionary.

The selection between a multi-pass recognition and a one-pass recognition frameworks depends also on correct identification of phonemes set from the participating languages, domain restriction and the class of a speech recognizer project at hand. The recognition based on one-pass framework with multilingual acoustic model offers the advantage of sharing data across languages [7][8]. Similar sounds across participating languages are shared using international phonetic inventories like Speech Assessment Methodology Phonetic Alphabet (SAMPA) or International Phonetic Alphabet (IPA) to assist in similarity detection. Tracing back in the field of South African LST, there are no resources developed for the purpose of creating a code-mixed ASR. As a result we identify a research challenge as being to solve the problem of code-mixing with its LID feature and acoustic similarity in order to support the growth of code-mixing in South Africa as one of the novel linguistic idea in a bilingual society.

IV. CODE-MIXING ASR EXPERIMENTAL DESIGN

The implementation of code-mixed speech recognition systems is currently in high demand especially in multilingual commonly speaker population across the world. The variation in implementation designs differ according to: research problem, speech data, and toolkit and implantation platform. Figure 3 depicts commonly used design of one-pass recognition based on a multilingual acoustic model.

A. Is code-mixing a reality in South Africa?

In order to establish the prevailing position with regard to mixed language scenarios among under-resourced languages of South Africa, we needed to verify the common usage of our target languages. We have conducted a short survey consisting of five questions asked in L1. These questions were designed with the aim of observing the reality of code-mixing in South Africa particularly on L1-speakers. The prompt sheets were handed to twenty individuals' selected based on balanced gender, education background and age group: (19-40). The selection was at random from University of Limpopo Turfloop Campus, half of whom were not guided on which language to respond with and the other half were given guidance on which language to respond with using our L1 language. The questions on a prompt sheet included:

A_1: Re fe leina la gago le nomoro ya mogala wa gago /Give us your name and phone number/
A_2: O belegwe neng? /When were you born?/
A_3: Re fe nomoro ya kamora /Give us your room number/
A_4: Toropo ya kgauswi le ga geno ke efe? /What is your nearest home town?/
A_5: O jele eng lehono? /What did you eat today?/

We calculated code-mixing rates (CMR) as follows:

$$\text{CMR} = \frac{No.\ of\ L2\text{-}words}{Total\ no.\ of\ words}$$

Table 1: A short code-mixing survey among L1-speakers.

L1-QUESTIONS	A1	A2	A3	A4	A5
CMR-Guided (%)	0	40	10	0	5
CMR–Unguided (%)	100	50	100	0	85

From Table 1, it is observed that unguided respondents used their languages of choice in order to provide answers to the questions. The guided people tried to stick to L1 as instructed on a prompt sheet, but they reached a point where they could not pronounce other L1-words. Most of these guided people used L2 language construct, especially when citing numerals and other L2 common words. From these limited observations, one can conclude that code-mixing is real, unavoidable and/or uncontrollable among indigenous languages users of South Africa.

The speech recognition systems of mixed speech codes involve recognition of more than one language. Figure 3 outlines our proposed systems architectural structure for the development of mixed language speech.

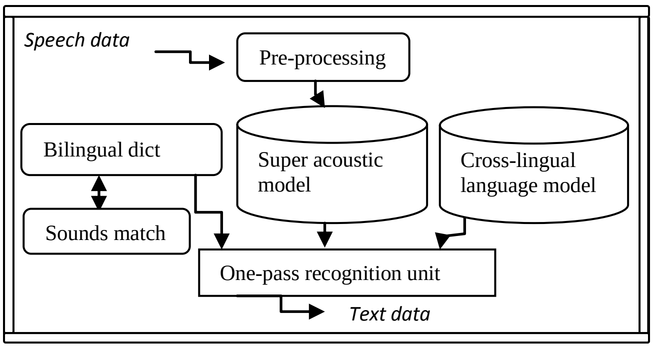

Figure 3. A block diagram of one-pass recognizer

The sound similarity relationships between the languages in participation are assed and these similarities are implemented in necessary model: a bilingual pronunciation dictionary, cross-lingual language model and super acoustic model.

B. Speech data and domain

The Northern Sotho language contains about *44* phonemes and English language contains approximately *44* phonemes [10]. To overcome this burdensome mission of balancing phonetic coverage we restricted our vocabulary to a *call center domain*. We use existing telephone speech data from the University of Limpopo Telkom Centre of Excellence (CoE) for Speech Technology speech data that consist of L1 sentential forms that are mixed with L2. The addition of a mixed languages speech data drawn from L1 and L2 languages was recorded from twenty individuals with ages ranging from 19-40 years. These recorded speech data in the training corpus was drawn with the following factors in mind in order to avoid bias and achieve expected relative balance: genders, level of education, and different socio-economic backgrounds.

C. Analysis of similar cross-lingual acoustics models

Speech consists of set of different sounds or phonemes-defined as the smallest distinctive sound unit. The IPA symbols are used to uniquely represent phonemes. It is necessary to use similar sounds/phonemes representation across languages [7]. Similar sounds will be clustered, to reduce the size of the pronunciation dictionary, language and acoustic models with concomitant aim to boost the performance of the recognizer. The phone clustering methods start at the monophone to triphone levels matched according to a clustering algorithm. This clustering algorithm outputs the list of triphones that are similar enough to be equated across the languages and the unlisted triphones remain language specific [11]. The clustering algorithm outlined in Figure 4, is applied in multilingual systems with large collection of speech data. Consider the two phones that sound the same drawn from Northern Sotho phonemes and English phonemes *L1-*h_i and *L2-*h_j respectively. One sound <*LL-*h_k> is trained for the recognition of the two phones from the two languages. In the case where L1 and L2 have similar phonemes that sound different, then two different phonemes have to be used for acoustic model training.

--Clustering Algorithm--

A group of triphones is equated if an average distance among all triphones from the group is less than a predefined threshold T. Average distance among M triphones was defined as:

$$S(\psi_1\psi_2 K \psi_m) = \frac{\sum_{n=1}^{m} \sum_{k=1}^{m} S(\psi_k\psi_t)}{\sum_{k=1}^{m} k}$$

$\psi_k\psi_t \in (\psi_1\psi_2,\ldots,\psi_m)$, $k \neq 1$

Where ψ_k *denotes the triphone* l_k-c_k+r_k , $(\psi_1\psi_2,\ldots,\psi_m)$ *is the group of triphones,* $S(\psi_1\psi_2,\ldots,\psi_m)$ *is the average distance among all triphones from the group* $(\psi_1\psi_2,\ldots,\psi_m)$.

Figure 4. A clustering algorithm based on triphone distance measure (adapted from [11])

D. Grammar and feature extraction

A grammar defines set of words to be recognized wherein those words that are not covered by the definition will be rejected by the speech recognizer. A sample of a mixed language grammar defined in *Backus Normal Form (BNF)* for generation of legal sentences is briefly provided by Figure 5 drawn from both L1 and L2 languages. The L1 names and the L1 surnames are sourced from [12]. The L2 surnames from the definition are British popular surnames.

```
$digits = ZERO|ONE|...|NINE|OH|LEFELA|PEDI|...SENYANE;
$letters = [A-Z];
$titles = MR|MRS|MISS|DR|PROF|SIR;
$initials = $letter[$letter][$letter];
$L1_names = LESIBA|LERATO|MAHLATSE;
$L1_surnames = MABOTJA|MASHALA|MAROPOLA;
$L2_names = SMITH|JOHNSON|LEWIS;
$L2_surnames = MIDDLETON|TURPIN|POYNTER;
$ phone_no = ($digit$digit$digit$dgit) | ($digit$digit$digit-$digit$digit$digit$dgit);
$names = ($L1_names [$L1_ surnames]) | ($L2_names [$L1_ surnames]);
Sent-start ( dial ($phone_no) | (phone|call) $initials $names) sent-end)
```

Figure 5. A mixed language grammar definition

Following the basic architectural structure of automatic speech recognition system as depicted in figure 6. A repertoire of HTK commands are used for the development of a one-pass recognizer [6].

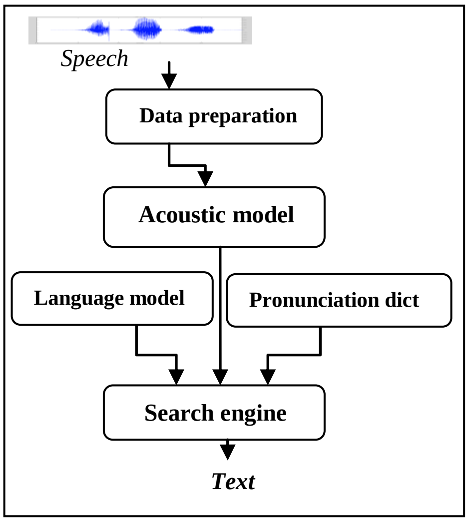

Figure 6. A typical block diagram of a speech recognizer

E. Super acoustic and cross-lingual language models

A super acoustic model ***P (Z)*** is a statistical component of a speech recognizer that estimates the probability that a certain phoneme has been uttered in a recorded audio segment. The phone level transcriptions of all different words defined in a code-mixed grammar are used to construct multilingual acoustic model. Usage frequency of each phone across all the languages is also taken in consideration. By so doing, we attempt to achieve a mixed languages speech recognition system that will be embedded within real-time speech enabled systems or agents. A cross-lingual language module defines constraints between the two languages depending mainly on the code-mixed grammar and its content. The n-grams language models offer a robust technique for modelling natural languages [13]. In cross-lingual language modelling the aim is to find best possible estimate for $\boldsymbol{P(W_i)}$, which is the probability for the word $\boldsymbol{W_i}$ to occur, where a word $\boldsymbol{W_i}$ belongs to $(L1 \cup L2)$. Table 2 depicts an extract from a super phone set where PS represents phone sets, L represents language and SPS represents super phone set.

Table 2. The super phone set used to create super acoustic model.

PS	L	SPS
a	L1	-
ax	L2	a
s	L1	-
s	L2	s
hl	L1	hl
iy	L2	iy
. . .		

Similar sounds phonemes across L1 and L2 are detected using IPA representation.

As shown in Figure 6, a pronunciation dictionary, a language model and an acoustic model constitute the core element of any speech recognition system; as a result much effort has to be based on working towards perfecting the development of these models *in a multilingual setting*.

F. *Data realignment*

A bilingual pronunciation dictionary (with one or more pronunciations of each word) created by utilizing British English dictionary *Beep 1.0* freely downloadable [14]. It is used to handle English pronunciations when the system encounters English words. An existing University of Limpopo Northern Sotho limited domain pronunciation dictionary created using word/pronunciation pairs and word-to-pronunciation rules was used in the experimental design. The HTK recognition tool *HVite* is used together with these two lexical modules to realign the training data and create new transcriptions.

V. EXPERIMENTAL RESULTS AND DISCUSSIONS

A. *Training of a domain specific one-pass recognizer*

The mixed language recognition systems involved words from the L1 and L2. For English language only numerals were trained for recognition. An overall corpus of the training data consisted of 90% to 95 Northern Sotho language words and 5% to 10% English language words. This percentage distribution requirement was calculated based on the complete phone sets and words percentage across the entire list of words to be recognized. The phones involved with English numerals returned a ratio of 1:9 compared to the Northern Sotho ones. This distribution was assessed to avoid bias over one language and add neutrality to the system. The HTK command tool *HSGen* was used to randomly generate a single set of 205 sentences based of the BNF grammar. These generated sentences were used to train the limited domain HMM-based bilingual speech recognizer. For testing we randomly generated 20 sentences that are recorded under the same conditions as the training data. The generated test set contained few new words that were not part of the training data.

B. *The preliminary results of domain specific one-pass recognizer*

The preliminary mixed speech recognition results based on a small training speech data set are presented in Figure 7. The test set consisted of twenty randomly generated sentences to observe the behavior of the recognizer.

```
=====================
HTK Results Analysis
=====================
  Date: Mon May 21 09:13:44 2012
  Ref : testref.mlf
  Rec : recout.mlf
------------------------ Overall Results --------------------------
SENT: %Correct=60.00 [H=12, S=8, N=20]
WORD:%Corr=83.05,Acc=76.27 [H=49, D=0, S=10, I=4, N=59]
```

Figure7. The results of HTK-based mixed language recognizer.

The results show the performance of a recognizer. Under the heading *Overall Results*, the first line preceded by SENT, shows that out of 20 testing sentences only 60% of these sentences were correctly recognized by the developed limited domain HMM-based bilingual speech recognizer. The second line with a WORD parameter, states that out of N (59) words, 83.05 where correctly recognized. The overall word accuracy returned 76.27(Acc). We had no deletion errors, 10 substitution errors and 4 insertion errors. This results show the initial success rate of a code-mixed speech recognition system that is constrained a predefined limited vocabulary domain.

VI. CONCLUSION

While code-mixing research in multilingual speech recognition system is still at its infancy stage, it is becoming common to the extent that it is becoming predictable. As a result, it is necessary to continue developments of resources (including mixed speech data, methods and toolkits) that assist in the development of code-mixing and code-switching speech technology systems. It is also assumed that having a large mixed speech data corpus used and applying sound similarity concept may improve the robustness of domain specific ML-ASRs.

VII. FUTURE WORK

It appears inevitable that mixed speech data will need to be expanded to make a modern multilingual ASR system more robust. Access to sizeable mixed speech data will enable us to objectively measure the ML-ASR properly within some kind of multilingual setting that uses the clustering algorithm to identify similar triphones to be equated across two or more languages.The multilingual set of triphones produced by the clustering algorithm can improve the performance of a parallel multilingual recogniser based on language-specific set of triphones.

VIII. ACKNOWLEDGEMENT

We acknowledge the sponsorship from University of Limpopo Telkom CoE with its funding partners – Telkom SA and the National Research Foundation (NRF) through the THRIP programme. An appreciation of our collaborators, the Meraka Institute of the Council of Scientific and Industrial Research (CSIR), is proper for their contribution with regarding specialized systems and application software training.

REFERENCES

[1] Chan J.Y.C, Cao H., Ching P. C. and Lee T. (2009): "Automatic Recognition of Cantonese-English Code-Mixing Speech". *The Association for Computational Linguistics and Chinese Language Processing*, Vol. 14, No. 3, pp. 281-304.

[2] Cárdenas-Claros M.S. and Isharyanti N. (2009): "Code switching and Code mixing in Internet Chatting: Between 'yes','ya', and 'si' - a Case Study", *Jaltcalljournal*, Vol. 5, No. 3, pp. 67-78.

[3] Ayeomoni M. (2006): "Code-switching and Code-mixing: Style of Language Use in Childhood in Yoruba Speech Community", *Nordic Journal of African Studies*, *15*(1), 90-99.

[4] Roux J.C., Botha E.C. and Du Preez, J.A. (2000) "Developing a Multilingual Telephone Based Information System in African Languages". *In Proc. of 2nd Intl. Language Resources and Evaluation Conf. Athens, Greece.*, Vol. 2. pp 975-980.

[5] Huang X, Acero A, and Hon H. (2001): "Spoken Language Processing: A Guide to Theory, Algorithm, and System Development", Prentice Hall PTR.

[6] Yared Glauco. (2003): "HTK-Hidden Markov Model Toolkit guide based on HTK Tutorial Version 3.1", *State University of Campinas-Unicamp.*

[7] Bhuvanagiri K.K. and Kopparapu S.K. (2010): "An Approach to Mixed Language Automatic Speech Recognition", *Oriental COCOSDA*, Nepal.

[8] Schultz I.T. and Fung P. (2009): "Detecting Code-Switch Event Based on Textual Features". *Karlsruhe Institute of Technology*, Retrieved from *http://cburgmer.nfshost.com/files/DetectingCodeSwitchEvents.pdf* on 15 May 2011.

[9] Chiang Y.C., Hsu C.N., Lyu D.C. and Lyu, R.Y. (2006): "Speech Recognition on Code-Switching Among Chinese Dialects", *In Proc. ICASSP*, France, pp. 1105-1108.

[10] Ted Power English Language Learning and Teaching Retrieved on *http://www.btinternet.com/~ted.power/folkchart.htm* on 14 February 2012.

[11] Imperl B. (2000): "Clustering of Context Dependent Speech Units for Multilingual Speech Recognition, *University of Maribor*.

[12] Modipa, T., Oosthuizen, H. & Manamela, M. (2007): "Automatic Speech Recognition of Spoken Proper Names", Unpublished Dissertation, University of Limpopo: Sovenga.

[13] Mansikkaneiemi A., Sams M. and Kurimo M. (2010): "Acoustic model and language model adaptation for a mobile detection service", Master's Thesis, *Aalto University School of Science and Technology*.

[14] Beep-1.0 British pronunciation dictionary, Retrieved on 26 March 2011. *ftp://svr-ftp.eng.cam.ac.uk/pub/comp.speech/dictionaries/beep.tar.gz*

Edward S Lebese received his undergraduate BSc degree in 2009 from the University of Limpopo. He is presently studying towards his Master of Science degree at the same institution. His research interests include Speech Technology, Security in Wireless Networks and Biometrics.