# Which unit for acoustic and language modeling for Khmer Automatic Speech Recognition?

Sopheap Seng, Sethserey Sam, Viet-Bac Le, Brigitte Bigi, Laurent Besacier

# WHICH UNITS FOR ACOUSTIC AND LANGUAGE MODELING FOR KHMER AUTOMATIC SPEECH RECOGNITION?

*Sopheap Seng*'**, Sethserey Sam*'**, Viet-Bac Le*, Brigitte Bigi*, Laurent Besacier**

* LIG Laboratory, UMR 5217, BP 53, 38041 Grenoble Cedex 9, FRANCE
** International Research Center MICA, CNRS/UMI-2954, Hanoi, VIETNAM

e-mail: sopheap.seng@imag.fr

**ABSTRACT**

In this paper we present an overview on the development of a large vocabulary continuous speech recognition system for Khmer language. Methods and tools used for quick language resources collection for the development of an ASR system for a new under-resourced language are presented. Face with the problem of lack of text data and the word error segmentation in language modeling, we investigate how different views of the text data (word and sub-word units) can be exploited for Khmer language modeling. We propose to work both at the model level (by making hybrid vocabularies with both word and sub-word units) as well as at the ASR output level (by using a simple N-best list voting mechanism). For acoustic modeling, we use basic linguistic rules to automatically generate pronunciation dictionaries based on grapheme and phoneme. An experimental framework is setup to evaluate the performance of each modeling units.

***Index Terms*** *- ASR, Khmer, word and sub-word units, acoustic modeling, language modeling.*

## 1. INTRODUCTION

With respect to speech recognition, the Khmer language bears challenging characteristics: (1) the lack of language resources (text and speech corpora) in digital form, (2) the writing system without explicit word boundary, which calls for automatic segmentation approaches to make statistical language modeling feasible and (3) the acoustic and phonologic characteristics that are not yet well studied. The statistical nature of the approaches used in automatic speech recognition requires a great quantity of language resources in order to perform well. For under-resourced languages which are mostly from developing countries, those resources are available in a very limited quantity because of its economic interest and the lack of standardized automatic processing tools (standard character encoding, word processing software). In this situation, language data collection is a challenging task and requires innovative approaches and tools. Similar to Chinese and Thai, Khmer is written without spaces between words. A sentence in Khmer ពណ៌សម្តេចថាខ្មៅ could be segmented into ពណ៌|ស|ម្តេច|ថា|ខ្មៅ (color|white|why|say|black) or ពណ៌|សម្តេច|ថា|ខ្មៅ (color|king|say|black). A correct segmentation of a sentence into words requires the full knowledge of the vocabulary and of the semantics of the sentence. The automatic segmentation method which is generally based on a vocabulary can not give 100% of correct segmentation because of the ambiguities and performs worst when the out-of-vocabulary rate is high. This makes text data processing for word n-gram language modeling complicated and other modeling units must be investigated. Note that a text in Khmer could be segmented into syllables, character-cluster (group of inseparable characters) or characters. The character-cluster could be a potential modeling unit as its segmentation is trivial because of its non-ambiguities atomic structure.

In this paper we present an overview on the development of a large vocabulary continuous speech recognition system for Khmer language. We first describe our methodologies and tools for data collection and quick development of a new ASR system for under-resourced language. In section 3 we discuss how word and sub-word units can be used in language modeling for Khmer language. We propose to work both at the model level by making hybrid vocabularies with both word and sub-word units as well as at the ASR output level by using a simple N-best list voting mechanism. We present in section 4 the process of automatic generation of grapheme based and phoneme based Khmer pronunciation dictionaries for acoustic modeling. The experimental framework and the recognition results are presented in section 5. Section 6 concludes the work and gives some future perspectives.

## 2. LANGUAGE DATA ACQUISITION

### 2.1. Text data acquisition and processing

Creating a statistical language model consists in estimating from a text corpus the probability of word n-gram. A large amount of in domain text data (several hundred millions words) is needed in order to obtain accurate probability estimation. As our system is targeted to automatic broadcast

news transcription, the classic way to get in domain text data is to take content from newspapers. Method for text collection from the web is becoming more and more popular as the web allows obtaining freely and quickly a large quantity of text. Recently, several research works proposed techniques to exploit the resources from the web for natural language processing. In [1], a web robot that retrieves text from the Internet to build a text corpus is proposed. From some given starting points on the web, the robot can reach and retrieve recursively text documents and html pages. However, we must control the robot in order to get only the text in the target domain and language. Another approach in [2] consists in estimating words n-gram probabilities using the World Wide Web. The probabilities are estimated from the number of pages found using a given search engine. Those kinds of methods applied well to languages which have already a significant coverage on the Internet. For an under-resourced language like Khmer, the number of websites and the speed of Internet connections are often limited. There are only around 340,000 websites registered in Cambodian domain name *.kh* (results from Google) and most of them propose contents in English instead of Khmer.

In our case, retrieving the Khmer pages from some well selected news websites allows us to get big quantity of text more rapidly than using a robot to crawl many sites on the net as proposed in [1]. Once html pages are retrieved, further processing is needed in order to build a text corpus:

- Filtering in order to extract only text from html pages
- Converting legacy character encoding to standard Unicode encoding
- Segmenting text into sentences and word or sub-word units using automatic segmentation tools
- Converting special signs and numbers to text
- Normalizing the words

By using the *ClipsTextTK* [3], this process could be done rapidly by adapting the language dependent part of the toolkit for Khmer language. We developed tools for the conversion of encoding (from legacy ad-hoc code to Unicode), the conversion of special characters and numbers to text and the segmentation of text into sentences and words. The word segmentation tool is developed using an algorithm which segments a text into words based on a list of vocabulary of 18,000 words obtained from the official Khmer dictionary (*Chhoun Nat* dictionary) with an optimization criteria: *longest matching*. Our segmentation tool, estimated on some held-out data, gives 95% of correct word segmentation. Syllable and character cluster segmentation is done using rules created with linguistic knowledge. The syllable segmentation is not trivial and gives only 85% of performance while the character-cluster segmentation is 100% correct. A complete *ClipsTextTK-Khmer* version is added to the toolkit.

The collection of Khmer text from the Internet allows us to get 2,5130 (448Mb) *html* pages from 5 selected news websites. A text corpus of 15.5 millions words (249Mb) is obtained after applying the *ClipsTextTK* toolkit. In these 15.5 millions words, 15 % of OOV words are found compared to the original 18,000 vocabulary. A non negligible part of the OOV words is probably due to the word segmentation errors, while another part corresponds to real OOV words.

### 2.2. Speech data acquisition

To train the acoustic models for our system, a speech corpus is needed. Speech corpus can be created by recording a well prepared text read by professional readers in a studio. The recording task is however very time and resource consuming as we need to prepare the text data and scenarios and run the recording process. To obtain speech signal quickly and freely, we tried several techniques. The first consists in searching the websites that propose the radio broadcast news in Khmer language. Many organizations such as Voice of America, Radio Australia and Radio FreeAsia have broadcast program in Khmer language and put on their website the entire broadcast news for public download. Most of the time, the transcripts are also available. From those sites, we can retrieve quickly a big quantity of speech signal but with a poor quality (narrowband) because the signal is compressed. In order to obtain a good quality speech signal, with help from our partner, Institut de Technologie du Cambodge (ITC) in Cambodia, we built a recording system from basic equipments: a computer with a radio receiver card installed, a recording program that we scheduled to record several hours of broadcast news of different radio stations in Phnom Penh, Cambodia. From this operation, we got recordings of 30h of good quality speech signal of radio broadcast news in Khmer language. A manual transcription campaign of the recording speech signals was organized at ITC. Twenty volunteers (students at ITC) who were motivated to contribute to the development of the language resources for Khmer were recruited and trained to do the manual transcription. By using *Transcriber* [4], 6h30mn of speech signal were manually transcribed in Unicode Khmer script (only speech read in the studio was transcribed and without extra detailed information).
This 6h30mn of transcription contains 3200 phrases of 45200 words pronounced by 8 different speakers (3 women). 172 phrases (25mn of speech signal) are then extracted to serve as test corpus during the evaluation of our ASR system.

## 3. LANGUAGE MODELING

The statistical nature of the approaches used in language modeling requires a large quantity of text corpus in order to make accurate word probability estimation. Word, which is often defined as a sequence of characters separated by space, is traditionally the basic unit of modeling and work

well for languages like English and French. For languages which have a very rich morphology where prefixes and suffixes augment word stems to form words and for the languages without explicit word boundary, traditional word definition is not appropriate and leads to a high out-of-vocabulary (OOV) rate. In this case, language models must be estimated from error-prone word segmentation. In addition, when the text data available is limited, it will lead to poor estimates of the language model probabilities, and hence may hurt ASR performance. This is typically the case for languages like Khmer. Alternatively, we can make language model estimation at sub-word level (syllable or character). This has potentially bad consequences on the word coverage of the n-gram models but it allows more accurate probability estimation because sub-word vocabulary is smaller than word vocabulary.

Some previous works using sub-word units for language modeling have recently been published for Arabic, Turkish (morphological analysis). Data-driven or fully unsupervised [5] word decomposition algorithms were used like in [6, 7] as well as working on the character level for unsegmented languages like in [8]. For character-based language like Chinese, mixed vocabularies containing both characters and a set of frequent words (mostly 2 characters words) are used in language modeling [9].

Note that a text in Khmer could be segmented into word, syllables or character-cluster. The character-cluster could be a good modeling unit as its segmentation is trivial because of its non-ambiguities atomic structure. The aim of our work is to investigate how these different views of the data can be advantageously exploited in Khmer ASR. At the language model level, the general idea is that from the initial sub-word vocabulary (Khmer syllable or character-cluster vocabulary for example), we progressively add N most frequent words in the sub-word vocabulary. By increasing N, we have different hybrid sub-word/word vocabularies and different trigram language models (LMs) are trained with these vocabularies. We will discuss the performance of these different LMs in the experimentation section.

## 4. ACOUSTIC MODELING

### 4.1. Automatic pronunciation generation

The pronunciation dictionary provides the link between sequences of acoustic units and words as represented in the language model. Whereas text and speech corpora can be collected, pronunciation dictionary is generally not directly available. While a manually generated pronunciation dictionary gives a good quality ASR, this task is time consuming and requires extended knowledge on the acoustic and phonology systems of the language in question. There were several techniques found in the literature for generating a pronunciation dictionary. Among them we can mention [10] which proposed a modeling technique based on pronunciation rules. This method requires knowledge on the target language and also of its phonetic rules. Grapheme based modeling has been successfully addressed for different languages [11, 12]. It has the advantage of being straightforward and fully automatic. For Khmer language, we propose two methods to automatically generate pronunciation dictionaries: a grapheme based which is trivial and a phoneme based which uses basic linguistic grapheme-to-phoneme rules to generate pronunciation.

*4.1.1. Grapheme based pronunciation dictionary*

The process of pronunciation dictionary generation could be primarily done based on graphemes. For Khmer language the grapheme based dictionary is generated by converting the Khmer character in its Unicode representation to its Unicode name in Roman representation. There are totally 77 graphemes in modern Khmer alphabet: 33 consonants symbols, 16 dependents vowels symbols, 16 independent vowels and 12 diacritics and signs. In Unicode, each Khmer code has a name in Roman. The following table shows the correspondence between a Khmer word and its grapheme based pronunciation model.

| Khmer word | Grapheme based pronunciation |
|---|---|
| ចចក | Ca Ca Ka |
| ចតុមុខ | Ca Ta U Mo U Kha |
| ក្រោមដី | Ka COENG Ro OO Mo Da II |

*Table 1*: Example of Khmer grapheme based dictionary

*4.1.2. Grapheme-to-phoneme rules based pronunciation dictionary*

Khmer words are predominantly of one or two syllables. Khmer syllable structure is generally defined as `C[C]V[CF]` where `C is a` consonant, `V` is a vowel and `CF` is final consonant (`CF` is present if `V` is a short vowel) . `CC` is a double consonant also called consonant cluster. With the syllable structures definition as well as pronunciation rules and consonant and vowel phonemes inventories ( see table 2) described in Huffman Book [13], we create 20 basic grapheme-to-phoneme rules to recognize and phonetize Khmer syllables as shown in table 3.

| Type of phone | | Phone symbols |
|---|---|---|
| Initial Consonants | Single *CI* | k kʰ ŋ c cʰ ɲ d tʰ n t tʰ b p pʰ m j r s ʊ h l ʔ |
| | Consonant Cluster *CC* | 85 double consonants cluster possible. Please refer to [13] for a complete list |
| Vowels *V* | Short | i e ɨ ə a a u o |
| | Long | iː eː ɜː ɨː əː aː aː uː oː ɔː |
| | diphthong | iə ei ɨə əi ao uə ou ɔə eə̆ uə̆ oə̆ |
| Final Consonants *CF* | | k c t p h n ŋ ɲ m j l ʊ ʔ |

*Table 2*: Khmer Phones

To phonetize a monosyllable word $s$, we apply all rules sequentially from rule 1 to rule 20. When $s$ is recognized by a rule $R_i$ where $1 \le i \le 20$, the program will output the phonemes sequence using the linguistic knowledge corresponding to rule $R_i$. For example, the word គ្នាន [*CC*=គ្ន+*V*= ◌ា+*CF*=ន] will be detected by rule *R9: CCVCF* and the pronunciation generated is */km-i:-n/ (/CC-V-CF/)*

| | |
|---|---|
| R1 : CI<br>R2 : CI V<br>R3 : CI V CF<br>R4 : CI V CF BANTOC*<br>R5 : CI CF<br>R6 : CI CF BANTOC*<br>R7 : CC<br>R8 : CC V<br>R9 : CC V CF<br>R10 : CC V CF BANTOC*<br>R11 : CC CF | R12 : CC CF BANTOC*<br>R13 : CI SANYOK* CI<br>R14 : CC SANYOK* CI<br>R15 : CI CI CHOEUNG* CI V<br>R16 : CI CI CHOEUNG* CI V CF<br>R17 : CI CI CHOEUNG* CI V CF BANTOC*<br>R18 : CI CI CHOEUNG* CI CF<br>R19 : CI CI CHOEUNG* CI CF BANTOC*<br>R20 : CI CI CHOEUNG* CI SANYOK* CI<br>* BANTOC, SANYOK and CHOEUNG are Khmer special signs |

*Table 3*: Basic rules of Khmer monosyllables

A $n$-syllable word $w=s_1 s_2 \ldots s_n$ is recognized by rules formed by concatenating $n$ basic rules. So for bi-syllables words we have to generate totally 20x20=400 rules. At the moment, our tool generates only one pronunciation. Generating pronunciation variants is currently not possible and is part of future work. While a word can be recognized by more than one rule, only the first one is considered.

While these 20 basic rules can recognize and phonetize simple Khmer monosyllabic or polysyllabic words, the Khmer words lent from Pali and Sanskrit and some exceptional spelling can not be detected by these rules. In our reference Chhoun Nat dictionary, when the pronunciation of a word is not straightforward (basic pronunciation rules do not apply), there is a pronunciation guide using simple syllables spelling. Among the 18000 word in Chhoun Nat dictionary, only 11000 words can be recognized and phonetized by our rules. For the rest, we use the pronunciation guide given to phonetize. Finally, 350 remaining words had to be phonetized manually.

### 4.2. Acoustic modeling

The design factors of acoustic modeling include the number of modeling units which are suitable for the language coverage and the size of the speech training database. Several methods propose crosslingual acoustic modeling to take advantage from existing multilingual models when the size of the speech training database is limited in target language. This approach was successfully applied in ASR for Vietnamese in [14].

The most basic acoustic modeling for Khmer language is grapheme based modeling. We used our grapheme based dictionary which has 77 modeling units. In case of the phoneme based acoustic modeling, we use single phone as modeling unit. A consonants cluster is consider as a sequence of 2 single consonants, e.g., */pt/ → /p/ + /t/*. As the long and the short vowels have the same acoustic properties with different time duration (long vowels are generally 2 times longer than short vowels [17]), the long vowels are represented as the concatenation of 2 short vowels, e.g., */e:/ → /e/ + /e/*. In the same manner, the diphthongs are considered as a sequence of single vowels. This single phone modeling leads to 33 single phonetic units. Table 4 shows the single phonetics units used in Khmer acoustic modeling.

| Type of phone | Phone symbols |
|---|---|
| Consonant | k kʰ ŋ c cʰ ɲ d tʰ n t tʰ b p pʰ m j r s ʋ h l ʔ |
| Vowel | i e ɛ ɨ ə a ɑ a u o ɔ |

*Table 4*: Single Phone for Khmer

We used SphinxTrain [15] toolkit from Sphinx project for building Hidden Markov Models (HMMs) acoustic models. With our speech training database described previously, we train acoustic models based on grapheme and phoneme. Context-independent (CI) and context-dependent (CD with 1000 tied states) models are both considered. Finally, we obtain 4 acoustic models namely *Grapheme_CI, Grapheme_CD, Phoneme_CI and Phoneme_CD.*

## 5. EXPERIEMENTS AND RESULTS

### 5.1. ASR system

Sphinx3 decoder [15] is used to in our experiments. The model topology is a HMM with 8 Gaussian mixtures per state. The pre-processing of the system consists of extracting a 39 dimensional feature vector of 13 MFCCs, the first and second derivatives.

Our text corpus collected from the web is first segmented into words and we extract 20,000 most frequent words to be used as the test vocabulary. This word vocabulary and the corresponding LM training corpus are also segmented in syllables (8,800 syllables vocabulary) and character-cluster (3500 character-clusters vocabulary). The transcript from speech corpus is also used for language models as it is from the same broadcast news source as the test corpus. The language models used in our experiments are obtained by linear interpolation of web corpus LM and the speech corpus transcript LM. A development corpus is used to tune the interpolation parameters.

In addition to Word Error Rate (WER) measure, we use Syllable Error Rate (SER) and Character Cluster Error Rate (CCER) for evaluation. Since Khmer word and syllable segmentation is not a trivial task and segmentation errors may prevent a fair comparison of different ASR hypotheses, we believe that comparing the ASR hypotheses at the

character-cluster level gives a more accurate idea of the relative performance of different systems. The test is run on our test corpus which contains 172 utterances (around 20mn of speech signal).

### 5.2. Grapheme based Vs Phoneme based acoustic model

In this experiment, we want to compare the performance of our different acoustic models. The tests are run on the same test corpus using the same word based language model.

| Acoustic Model | WER | SER | CCER |
|---|---|---|---|
| Grapheme_CI | 64.9 | 39.9 | 33.6 |
| Grapheme_CD | 47.8 | 26.9 | 20.8 |
| Phoneme_CI | 57.9 | 38.2 | 31.9 |
| Phoneme_CD | 49.6 | 25.1 | 19.1 |

*Table 5*: Test results of different acoustic models

The results in table 5 show that the context-dependent model performs better than the context-independent model both in grapheme based and phoneme based models. The performances of grapheme based and phoneme based models are very comparable. This shows the potential of grapheme based acoustic modeling when a phoneme-based pronunciation dictionary is not available. The big difference between WER and CCER is partly due to the word segmentation errors that occur in output hypotheses and in the reference. This suggests that CCER gives a more accurate evaluation and should be used if we want to compare different systems using different segmentation units

### 5.3. Word and Sub-word language models

In this experiment, we train 3 trigram LMs by using respectively word, syllable and character-cluster as modeling units. We test the performance of these 3 LMs using both grapheme based and phoneme based acoustic models.

| Language Model | Acoustic Model | CCER |
|---|---|---|
| LMword | Grapheme_CD | 20.8 |
| LMsyl | Grapheme_CD | 25.0 |
| LMcc | Grapheme_CD | 32.3 |
| | | |
| LMword | Phoneme_CD | 19.1 |
| LMsyl | Phoneme_CD | 26.3 |

*Table 6*: Test results of word and sub-word units LMs

From the results in table 6, we can see that word unit is still the best unit of modeling. A Khmer word is in average composed of 3.2 syllables and 4.3 characters clusters. Thus, the effective span of a trigram model of syllables or character-cluster is shorter than that of a word trigram model, which leads to less accurate prediction. The advantage of these sub-word units may remain when there are high OOV rates in the test corpus.

### 5.4. Hybrid Word /Sub-word language models

We create hybrid language models by combining word and sub-word units. In this experiment we create hybrid model by adding progressively N most frequent words of our text corpus to character-cluster (Cc) vocabulary $V_0$. By increasing N from 0 to 20k, we create 6 different hybrid Word+Cc vocabularies (namely $V_0$, $V_{1k}$, $V_{5k}$, $V_{10k}$, $V_{15k}$ and $V_{20k}$). Different language models are trained using these 6 vocabularies with the same text corpus re-segmented each time to fit to each vocabulary. The recognition results with these 6 LMs using the Grapheme_CD acoustic model is given in figure 1.

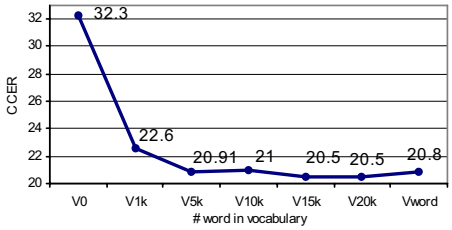


*Figure 1*: Comparison of hybrid word/sub-word language models

We can observe that when we progressively add words to character-cluster vocabulary the performance increase. The hybrid systems V15k and V20k are slightly better than the word based LM but the differences are not significant.

### 5.5. ASR outputs combination

To investigate the potential of word and sub-word units combination, we apply a simple N-Best list combination method and try to decode the best hypothesis. We decode 20-best hypotheses from different ASR system. A 40 Best hypothesis are obtained if we combine the output of 2 ASR systems. Similarly to ROVER [16] systems combination scheme, we use a voting algorithm based on the number of occurrence of sub-word in N-best list to decode the best hypothesis. The table 7 shows the results of the combination of word, syllable and character cluster based ASR systems as well as the Oracle CCER (Grapheme_CD acoustic model is used).

| ASR systems output | CCER | Oracle |
|---|---|---|
| 20 best LMword + 20 best LMsyl | 21.2 | 12.1 |
| 20 best LMword + 20 best LMcc | 23.3 | 11.8 |
| 20 best LMsyl + 20 best LMcc | 27.5 | 15.0 |
| 20 best LMword+20 best LMsyl+20 best LMcc | 23.5 | 10.8 |

*Table 7*: Test results of word and sub-word units LMs

While the oracle CCER of the combination hypotheses show the potential of this combination approach, the simple voting algorithm do not lead to improved performance. More advanced combination scheme like lattice combination and confusion network will be investigated in order to see the potential of word and sub-word combination.

## 6. CONCLUSION

This paper presents the development of an ASR system for an under-resourced language: Khmer language. Quick language data collection methodologies and tools were described. For acoustic modeling, we showed that the grapheme based modeling gives a comparable performance compared to automatic grapheme-to-phoneme rules based modeling. To deal with the problem of lack of text data and the word error segmentation in language modeling, we tried to exploit different views of the text data by using word and sub-word units. We used word, syllable, and character-cluster as modeling unit to make hybrid language models for Khmer language. The test results showed that word unit is still the best unit for language modeling. At the ASR system output level, the oracle error rate of the combined N-best hypothesis shows the potential of word and sub-word unit combination but a simple combination (N-bast voting) could not decode a better result. In our future work, more advanced combination scheme like lattice combination and confusion network decoding will be investigated in order to see the potential of word and sub-word combination.

## REFERENCES

[1] D. Vaufreydaz, “Modélisation statistique du langage à partir d'Internet pour la reconnaissance automatique de la parole continue”, *Thèse de doctorat de l’Université J. Fourier - Grenoble I*, France, 226 pages, January 2002.

[2] X. Zhu, R. Rosenfeld, “Improving Trigram Language Modelling with the World Wide Web”, *ICASSP'01*, pp. 533-536, Salt Lake City, USA, Mai 2001.

[3] www-clips.imag.fr/geod/User/brigitte.bigi/logiciel.html

[4] C. Barras et al, Transcriber: development and use of a tool for assisting speech corpora production, *Speech Communication* Vol 33, No 1-2, January 2000.

[5] M. Kurimo et al., “Unsupervised segmentation of words into morphemes - Morpho Challenge 2005: Application to Automatic Speech Recognition”, *Interspeech’06*, pp. 1021-1024, Pittsburgh, PA, 2006

[6] N. Abdillahi et al., “Automatic transcription of Somali language”, *Interspeech*, pp. 289-292, Pittsburgh, PA, 2006.

[7] M. Afify et al., “On the use of morphological analysis for dialectal Arabic Speech Recognition”, *Interspeech’06*, pp. 277-280, Pittsburgh, PA, 2006.

[8] E. Denoual, Y. Lepage, “The character as an appropriate unit of processing for non-segmenting languages”, *NLP Annual Meeting*, pp.731-734, Tokyo, Japan, 2006.

[9] L. Chen, et al, "Broadcast News Transcription in Mandarin," Proc. ICSLP'2000, Beijing, China, 2000

[10] X. Huang, et al, Spoken Language Processing – A Guide to Theory, Algorithm, and System Development, *Prentice Hall*, 2001.

[11] Billa J. et al, “Audio indexing of Arabic broadcast news”, In Proceedings of the *IEEE International Conference on Acoustique, Speech and Signal Processing*. 2002 Orlando, FL, PP. 5-8

[12] Bisani M., Ney H., “Multigram-based grapheme-to-phoneme conversion for LVCSR” In Proceedings of the *EUROSPEECH*. 2003 Geneva, Switzerland, pp.933-936

[13] Huffman, Franklin. “Cambodian System of Writing and Beginning Reader”. *Yale University Press* 1970

[14] Viet Bac Le, Laurent Besacier, Comparison of Acoustic Modeling Techniques for Vietnamese and Khmer ASR, *Interspeech-ICSLP*, pp. 129-132, Pittsburgh, PA, USA, 17-21 September 2006.

[15] http://cmusphinx.sourceforge.net/html/cmusphinx.php

[16] Fiscus, J.G. “A Post-Processing System to Yield Reduced Word Error Rates: Recogniser Output Voting Error Reduction (ROVER)”. *Proc. IEEE ASRU Workshop* 97 , pp. 347-352,

[17] S. Seng et S. Sam, “Traitement Automatique de la Langue Khmer”, *Rapport Scientifique de Projet AUF TALK 2ème tranche*”. May 2006.