World Academy of Science, Engineering and Technology
International Journal of Electrical and Computer Engineering
Vol:1, No:11, 2007

SySRA: A System of a Continuous Speech Recognition in Arab Language

Samir Abdelhamid, and Noureddine Bouguechal

Abstract**—**We report in this paper the model adopted by our system of continuous speech recognition in Arab language SySRA and the results obtained until now. This system uses the database Arabdic-10 which is a corpus of word for the Arab language and which was manually segmented. Phonetic decoding is represented by an expert system where the knowledge base is translated in the form of production rules. This expert system transforms a vocal signal into a phonetic lattice. The higher level of the system takes care of the recognition of the lattice thus obtained by deferring it in the form of written sentences (orthographical Form). This level contains initially the lexical analyzer which is not other than the module of recognition. We subjected this analyzer to a set of spectrograms obtained by dictating a score of sentences in Arab language. The rate of recognition of these sentences is about 70% which is, to our knowledge, the best result for the recognition of the Arab language. The test set consists of twenty sentences from four speakers not having taken part in the training.

Keywords—Continuous speech recognition, lexical analyzer, phonetic decoding, phonetic lattice, vocal signal.

I. INTRODUCTION

SIGNIFICANT progress in the field of the continuous speech recognition was made these last years [4]. A system of continuous speech recognition rests primarily on a system of phonetic decoding reliable which permit to the syntaxic analyzer to determine, by means of comparison, the text (sentences) dictated by the speaker. The techniques of decoding used are:

- The Dynamic Comparison
- Stochastic Models (Hidden Markov Model)
- The Connexionist Models (founded on a modeling of the Neuronal Networks).

These methods are judged techniques of forms recognition which prove often badly adapted to the recognition of phonetic vectors presenting the nonfixed characters and consequently nonstable. The solution used nowadays takes as a starting point the techniques of the artificial intelligence holding account of the phonetic features and the context. This technique adopts a strategy of expert system which rests on the knowledge translated in the form of rules [10]. The SySRA system that we propose uses this strategy of expert system to ensure a reliable decoding of the vocal signal in a phonetic lattice. The user, however, is subjected to some constraints such as: careful elocution, punctuation of the dictation by making pauses between the syntaxico-semantic groups, etc...

The object of such a system is the integral recognition of the text i.e. the various sentences dictated by making of course the difference between the dictated text and the commands of punctuation. We currently work on the assets of systems of realization of prototypes of automatic dictating machine conceived for other languages other than the Arab language. We are interested in the two aspects:

- Text-editing
- Correction of the seized documents

II. ARCHITECTURE OF THE SYSTEM SySRA

The recognition is carried out by an iterative process by which we build a base of chains candidates. This base gives a partial interpretation of the phonetic lattice. We associate to each chain a rate of recognition which will be used thereafter to select the chains which give a complete interpretation of the lattice. The process of recognition continues until the base of the partial chains is empty.

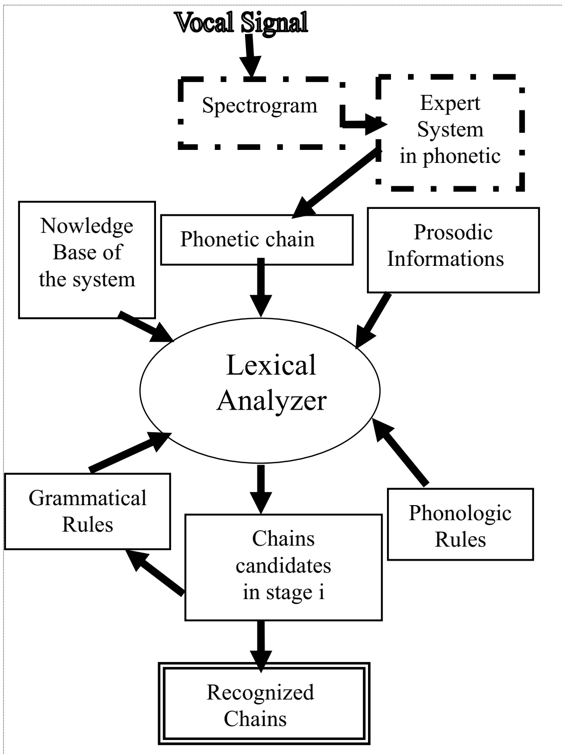

Fig. 1 Architecture of the system SySRA

Samir Abdelhamid is with Institute of computer science, University of BATNA, 05000, Algeria (e-mail: S_Abdel@hotmail.com).
Noureddine Bouguechal is with Institute of Electronic, University of BATNA, 05000, Algeria (e-mail: Bouguechal@univ-batna.dz).

III. The Phonetic Decoder

Our system contains an expert system in phonetic decoding. The idea comes from the fact that there are human experts able to read spectrograms obtained starting from a vocal signal. The knowledge base is in the form of production rules. The syntax of the rules used was selected in order to facilitate the task with the expert phonetist in the event of modification of rules.

R_x IF Right_Context = Liste_of_phonemes
Then Liste_of_phonemes
EndIF

Example [6] :

R_{025} IF NO(pitch_as)
Then PHONEMES[t 50 k 50 A 50 q 50]
EndIF

Let us note that on the level of phonetic decoding, the goal is to determine the phonetic class of the segments of the signal used. In our case, the tolerance of recognition on the level of each segment is opened to no more than three phonemes in order to enabling us to obtain the correct phoneme without creating a too strong indeterminism for the higher levels. The expert system receives, in input, the spectrogram of the vocal signal of the speaker, it interprets it and delivers, at output, a phonetic lattice which will be used as bases of search and recognition.

IV. The Lexicon

Arabdic-10 is a corpus of words read by ten speakers. The texts used are extracted from the Algerian daily newspapers in Arab language in order to provide a corpus of continuous speech average vocabulary (about 5.000 words). This corpus will also enable to acquire knowledge acoustico-phonetics on the Arab language. The dictionary is, for automatisation needs, was structured as follows:

```
LEXICON
 < CLASS >
   .
   .
   VERB
     Type
       masc, fem, duel, pluriel : Boolean
      Enter
   yaEkoulou [ masc, 3sing]
   yaEkoulani    [ masc, 3duel]
   yaEkoulou [ masc, 3pluriel]
   taEkoulou  [ fem, 3sing]  ta/ ?/ku/lu
   yaEkoulani    [ fem, 3duel]  ya/ ?/ku/la/ni
   taEkoulna  [ fem, 3pluriel]   ta/ ?/ku/l/na
   taEkoulou  [ masc, 2sing]     .
   taEkoulani [ masc, 2duel]         .
   taEkoulou  [ masc, 2pluriel]      .
   taEkouli        [ fem, 2sing]       .   taEkoulani [ fem, 2duel]       .
   taEkoulna  [ fem, 2pluriel]       .
   aEkoulou       [ --, 1sing ]?/ ?/ku/lu
   naEkoulou [ --, 1pluriel ] na/ ?/ku/lu
   .
   .
   Orthographic Syntaxic        Phonetic
   Form     prorieties       representation

 ENDCLASS
.
.
ENDLEXICON
```

This example illustrates the conjugation of the Arab verb "to eat" أكل [5], conjugated with the fourteen persons permitted by the Arab language:

V. The Lexical Analyzer

A. ntroduction

The phase in which we are interested now is that of the interpretation of the phonetic lattice. Indeed, after phonetic decoding, we trie to propose one or more interpretations of the phonetic chain thus obtained by using the whole of the linguistic knowledge acquired until now. The adopted solution, used before successfully for the French language [5], consists in seeking an optimal path in the phonetic lattice whose interpretation corresponds to a coherent succession of words.

Pronounced sentence: ' Dear Mister ' أيها السيد

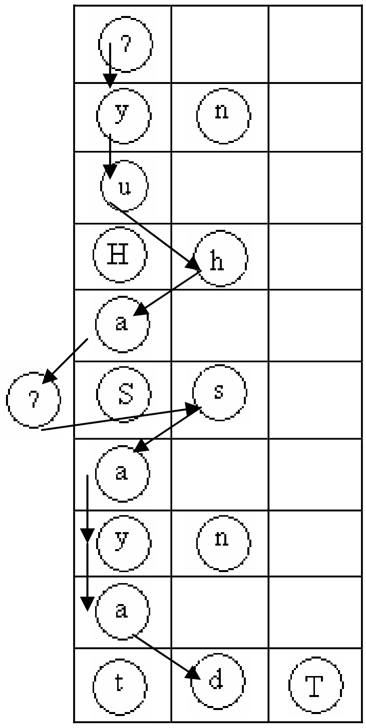

Fig. 2 Phonetic chain

The errors of segmentation and identification of the decoder can involve omissions of phonemes (see Fig. 2) this encourages us to put in assumption missing phonemes.

B. The exical Analysis

The solution adopted by our system is inspired from the method of resolution used in the system MYRTILLE I, developed by the research center in data processing of Nancy (CRIN) - FRANCE It consists in seeking a path in the lattice of phoneme corresponding to the phonetic description of the word to identify. In this approach, the evaluation of the rate of dissimilarity is simple since a local score is associated to an arc of the graph of reference as soon as this one is validated in the phonetic lattice. The lexical analysis, in this case, takes into account six properties with knowing:

1. Ideal Case

T(j,*) = GRAPH (i , 1)
T(j,*) = GRAPH (i , 2)
$j \leftarrow j + 1$

i ← i + 1
RDC ← RDC +1
2. Insertion envisaged
T(j,*) = GRAPH (i , 3)
j ← j + 1
RDC ← RDC + RINS
3. Elision envisaged
GRAPH (i , 1) < 0
i ← i + 1
RDC ← RDC + REL
4. Assumption of elision
i ← i + 1
RDC ← RDC + RHEL
5. Assumption of substitution
j ← j + 1
i ← i + 1
RDC ← RDC + RHSUB
6. Assumption of insertion
j ← j + 1
RDC ← RDC + RHINS
7. Other case
QUIT

NOTATIONS

- Graph(i, 1): standard phoneme
- Graph(i, 2): phoneme substituted
- Graph(i, 3): phoneme inserted
- RDC : Rate of Dissimilarity Cumulated
- RINS: Rate of insertion
- REL: Rate of Elision
- RHEL: Rate of Assumption of Elision
- RHSUB: Rate of Assumption of Substitution
- RHINS: Rate of Assumption of insertion

GRAPH (i, m) and T (j, N) are respectively indicating the form of reference and the phonetic lattice. The dynamic comparison between these two forms consists in seeking in the graph Graph (X , T) an optimal path (see Fig. 3). What meaning that we seek among all the possible functions of retiming that which is optimal within the meaning of a certain metric.

Conclusion and Perspectives

Our system of recognition of continuous word in Arab language, re-enters within the framework of a research project approved by the Algerian ministry of the higher education and scientific search under the code B 0501/01/02. The originality of our work lies in the choice of the language where few research tasks were carried out until now. We could establish an automated model describing the Arab language while taking as a starting point the the models established for the languages English and French. The experiments have highlited some of the main problems with the Arab language, particulary the difficulties with homophones and liaisons. We hope to have brought more to the Arab language and which we wish to have contributed to his development. Our future efforts will relate primarily to the two research orientations to knowing:

- Co-operation with the phonetic level
- The definition of strategies of powerful analyses

where it remains much to bring.

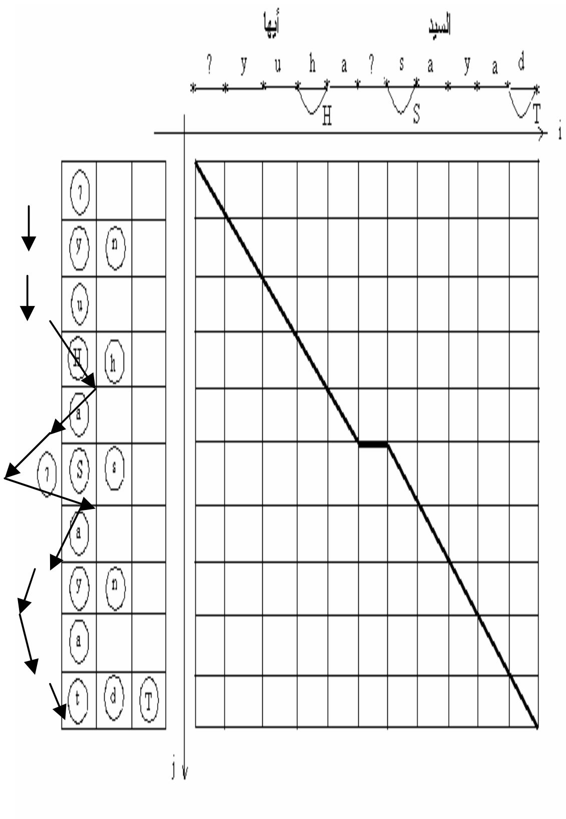

Fig. 3 Retiming Function of the sentence:أيها السيد

References

[1] Young, S.J. and al, multilingual vocabulary speech recognition, Compuet Speech and Language 0-73-89-0, 1997.
[2] J.P. Haton, D. Fohr et M. Djoudi : Un système expert pour le décodage acoustico-phonétique pour l'Arabe standard. Conférence Maghrébine, Septembre 1989.
[3] Y. Belkaid : Les voyelles de l'Arabe littéraire moderne. Analyse spectrographique Rapport N° 16, travaux de l'institut de phonétique de Strasbourg, 1984.
[4] O. Deroo, C. Ris: Hybrid HMM/ANN Systems speaker independent continuous speech recognition in French. Travaux de l'école Polytechnique de MONS Belgique, 2000.
[5] S. Abdelhamid : Contributions à l'étude et à la réalisation d'une machine à dicter en Français. Thèse de Magister de l'institut d'informatique de l'université de Batna, Algérie, 1994.
[6] M. Djoudi ET J.P. Haton: The SAPHA Acoustic Phonetic Decoder System for standard Arabic. International conference on spoken language processing, 1990.
[7] M. Guerti : Contribution à la synthèse de la parole en Arabe standard. Actes des 16ème journées d'études sur la parole. Hammamet, Tunisie 1987.
[8] A. Giannini et M. Petorino. The Emphatic Consonants in Arabic. Giadini Editori, 1982.
[9] Benhamouda : Morphologie et syntaxe de la langue Arabe. Nationale Edition, 1983.
[10] N. Carbonell, J. P. Haton, D. Fohr. Aphodex: design and implementation of an acoustic phonetic decoding expert system. IEEE: International conference on Acoustics, speech and signal processing, 1986.
[11] V. Barreaud : Reconnaissance automatique de la parole continue: compensation des bruits par transformation de la parole. Thèse de l'université de Nancy1, 2004.