MODELING OF SPEECH SYNTHESIS OF STANDARD ARABIC USING AN EXPERT SYSTEM

Tebbi Hanane[1] and Azzoune Hamid[2]

[1,2] LRIA, Option: Representation of Knowledge and systems of Inference, USTHB, Algiers

[1]tebbi_hanane@yahoo.fr , [2]azzoune@yahoo.fr

ABSTRACT

In this work we present our expert system of speech synthesis based on a text written in Standard Arabic, our work is carried out in two great stages: the creation of the sound data base, and the transformation of the written text into speech (Text To Speech TTS). This transformation is done firstly by a Phonetic Orthographical Transcription (POT) of any written Standard Arabic text that to transform it into his corresponding phonetics sequence, and secondly by the generation of the voice signal which corresponds to the chain transcribed. We spread out the different steps of conception of the system, as well as the results obtained compared to others manners of works studied to realize TTS based on Standard Arabic.

KEYWORDS

Engineering knowledge, modelling and representation of vocal knowledge, expert system, (TTS Text To Speech), Standard Arab, PRAAT.

1. INTRODUCTION

Generate the voice is a complex work because of the variability intra and interlocutor of the voice signal. In computer science the difficulty of modeling the speech signal is for the reason that we don't know yet how to model very well the enormous mass of knowledge and information useful to the signal synthesis. Thus we have made a choice to use an expert system to modeling that knowledge to build a robust system which can really read a text written in a language chosen especially in Standard Arabic. To obtain a better organization of our work, we defined our direct aims. We divided our modeling into three essential stages; the signal analysis, the phonetic orthographical transcription (POT) and finally the synthesis of the textual representation written in Standard Arabic. So the finality considered here is that the user can understand the different phrases transcribed and synthesized which will be pronounced with a clear and high quality manner.

2. PREVIOUS WORKS

At the current time, we can judge that works emanated in the same context of ours are still not really colossal, and this is because of the complexity of the language itself. And if this some works exists, they are based on the same principle as transcribers of the other languages (French,

David C. Wyld et al. (Eds) : CCSIT, SIPP, AISC, PDCTA, NLP - 2014
pp. 469–476, 2014. © CS & IT-CSCP 2014 DOI : 10.5121/csit.2014.4241

English, etc.) [1], nevertheless, the efforts undertaken are encouraging and open a large window to follow research task in this field.

- TTS system MBROLA [2] which use the code SAMPA during the stage of transcription, in this case the user must respect the form of the SAMPA code which is not a universal code;

- Work of S. BALOUL [3] who represents a concrete example of transcription of words; based on morphological analysis, and on the studies of pauses to generate the pronunciation of the texts.

- SYAMSA (SYstème d'Analyse Morphosyntaxique de l'Arabe), realized by SAROH [4]. According to him, "the phonetisation of the Arabic language is based in particularly on the use of lexicons and on a morphological analyzer for the generation of the different forms of a word. In addition, they are the phenomena of interaction among the words (connection, elision, etc.) and the phenomena of assimilation which suggest the uses of phonological rules "[5]. This tool ensures for each word in entry, the root which correspond to it as well as the morphological and phonetic representations of the root.

- The GHAZALI [6] project which was carried out within the IRSIT (Institut Régional des Sciences Informatiques et des Télécommunications) of Tunisia, it is based on a transcription work which fits inside the framework of the realization of a TTS system. The characteristic of this system is shown in the use of a set of rules in the emphasis propagation.

- SYNTHAR+ [7] which is a tool studied by Z. ZEMIRLI within the NII (the National Institute of Informatics of Algiers), it ensures the transcription step for a TTS system so that it transmits the phonetic representation to the MULTIVOX synthesizer. It should be known that SYNTHAR+ is based on a morphological analysis before realizing the transcription.

3. STRUCTURE OF THE SYSTEM

The uses of the concept of code and the introduction of high levels in analysis (morphological, syntactic, and pragmatic.) makes the transcription task so difficult and requires deepened studies of the language itself. The difference in our work compared to all that exists is in the transcription using graphemes, i.e. the uses of the Arabic characters as basic units to transcribe directly the text, indeed modeled by an expert system.

Definition of an expert system:

An expert system is an informatics tool of AI (Artificial Intelligence), conceived to simulate the knowhow of an human expert, in a precise and delimited field, this thanks to the good exploitation of a set of knowledge provided explicitly by expert of the field [5].

The E.S is a system where the data (the knowledge database) is quite separated from program which handles it (the inference engine). In our case, we built our expert system basing on the two following components:

A knowledge database: It contains a database of facts and a database of rules, represents the knowledge (we speaks here about the sound database) and knowledge - to make (a set of

rewriting rules). The fact database integrate two types of facts: permanent facts of the field and deduced facts by the inference engine which are specific to the field considered (voice synthesis).

The inference engine: It is capable to reason starting from the information contained in the knowledge database and to make deductions. The inference engine uses the facts and rules to produce new facts.

Thus the role of an expert system is then to infer generally rules like if/then. To build this system our work targets the following steps:

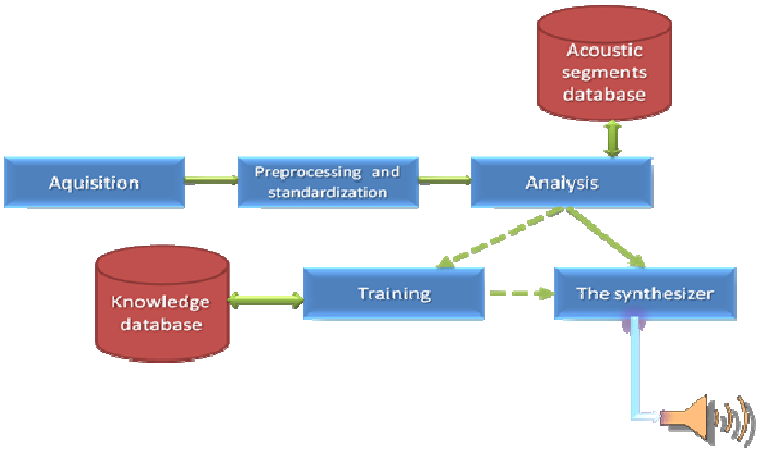

Figure 1. General schema of our system of speech synthesis

3.1. Acquisition

It is the first step of our process of voice synthesis; it plays the role of the interface intermediate between the user and the system. During this time the user have to enter his set of phrases to be pronounced by the system. After that the automatic process can begin.

3.2. Preprocessing and standardization of the text

In this phase any form of text is transformed to it literal, perfectly disambiguated form [1]. In particular, this module draft problems of formats, dates and hours, abbreviations, numbers, currencies, addresses, emails…etc. It was shown that majority errors of conversion grapheme/phoneme, for the best operational systems came from the proper names and the exceptions that they pose [8]. Here are some examples of the processing carried out in our case:

- The replacement of each composed character by its equivalents
  Example: لا ⟶ ل ا
- Consultation of the exceptions lexicon to eliminate the special words.
- The application of the transcription rules established for the language. This module must be able to process grammatical complexes tasks to identify the rules of transcription which will be used in a considered context.

3.3. Analysis (the extraction of the characteristics)

The aim of the analysis of the voice signal is to extracts the acoustic vectors which will be used in the stage of synthesis follows. In this step the voice signal is transformed into a sequence of acoustic vectors on the way to decrease redundancy and the amount of data to be processed. And then a spectral analysis by the discrete Fourier transform is performed on a signal frame (typically of size 20 or 30 ms) [9]. In this frame, the voice signal is considered to be sufficiently stable and we extract a vector of parameters considered to be enough for the good operation of the voice signal. In the speech synthesis, the characteristics extraction step, commonly known as the step of analysis, can be achieved in several ways. Indeed, the acoustic vectors are usually extracted using methods such as temporal encoding predictive linear (Linear Predictive Coding LPC) or Cepstrales methods as the MFCC encoding (Mel Frequency Cepstral Coding), as well as the process of segmentation, etc. This process delimits on the acoustic signal a set of segments characterized by labels belonging to the phonetics alphabet of the language under consideration.

At present, the segmentation completely automatic of a voice signal remains a fastidious task. Indeed, looking at the complexity acoustico-phonetics phenomena being studied, this activity requires often a manual intervention. Generally, the methods that perform the segmentation of the acoustic wave are divided into two great classes:

- The first include all the methods which allow segmenting a voice signal without a priori knowledge of the linguistics content of this signal. These methods split the voice signal into a set of zones homogeneous and stable;

- The second class includes all the methods which allow segmenting the voice signal basing on a priori linguistic description (typically on phonemes) of this signal. These methods of segmentation are revealed like methods with linguistically constraints. In our state, we have opted for the second technique using a tool of voice signal analysis which is PRAAT [10] so to slice manually the speech signal in a succession of segments, each one associated with an element acoustic unit (phoneme or diaphone).

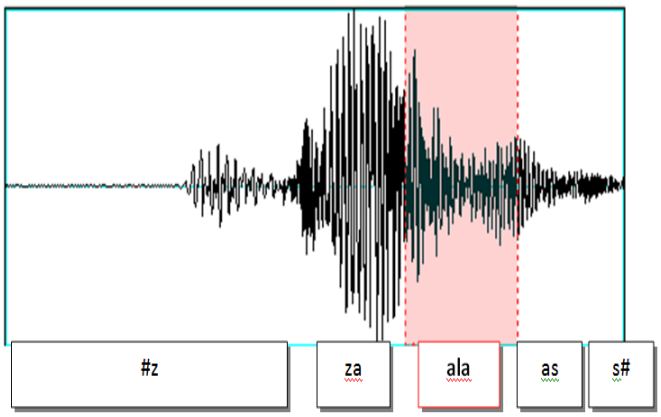

Figure 2. Decomposition in polysound of the word [جلس]

3.4. Modeling of the sound database

Majority of work curried out in the field of the spoken communication required often the recording, and the handling of corpuses of continuous speech, and that to curry out studies on the contextual effects, one the phonetic indices, and variability intra and inter-speaker. Our corpus is modeled by the diagram of class1 as follow:

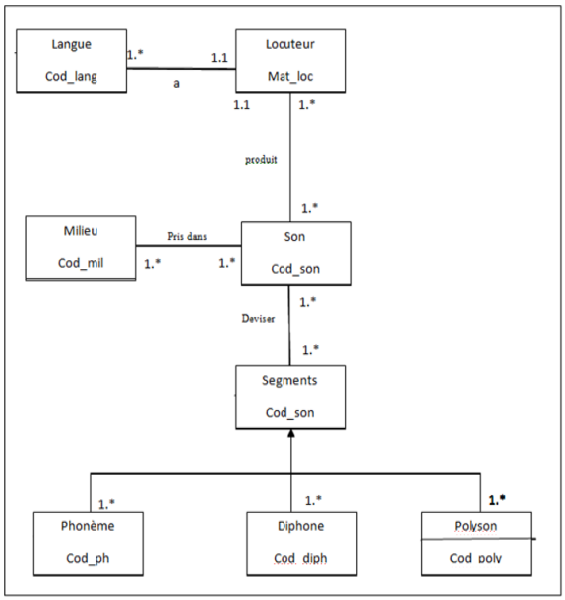

Figure 3. The diagram of class of our sound database

3.5. The training

Here we referred to a POT module to convert each grapheme in phoneme according to the context and that is possible using a set of rewriting rules. The principal advantage of this approach is the possibility to modeling the human linguistic knowledge by a set of rules which can be integrated in expert systems. Each one of these rules has the following form:

[Phoneme] = {LC (Left Context)} + {C (Character)} + {RC (Right Context)}
Here is concrete example of transcription rules [11]:

# is a beginning sign of the sentence,
$ is a sign of end of the sentence,
§ is extremity of a word,
C is a Consonant,
V is a Vowel,
SC is a Solar Consonant and LC is a Lunar Consonant
[uu]={SC}+{ُ}+ {و} [uu]= §+{ُ}+ {و}
[uu]={LC}+{ُ}+ {و} [uu]= $+{ُ}+ {و}

When the " و " is preceded by the vowel / ُ / and followed by a consonant, we obtain the long vowel [uu].

1: It is a diagram used in UML(Unified Modeling Language) which is used in object oriented modeling

When the " و " is preceded by the vowel / ُ / in final position, we obtain the long vowel [uu].

3.6. The synthesizer (the voice generation)

The generation of the voice signal is the real synthesis of speech. This operation consist of a transform of the phonetic sequence (which represents the pronunciation of the written text) resulting from the transcription step to its substance, i.e. to it acoustics realization. During this step we choose from our sound database the units (phonemes, diphones) most suitable to build the sentence to generate, this means that we will create a automatic function of reading, so that in the end, the user has only to listen to the synthetic sentences. To improve the quality of the generated synthesis, we increase each time the sound unit only. The general silhouette of our expert system can be as follow:

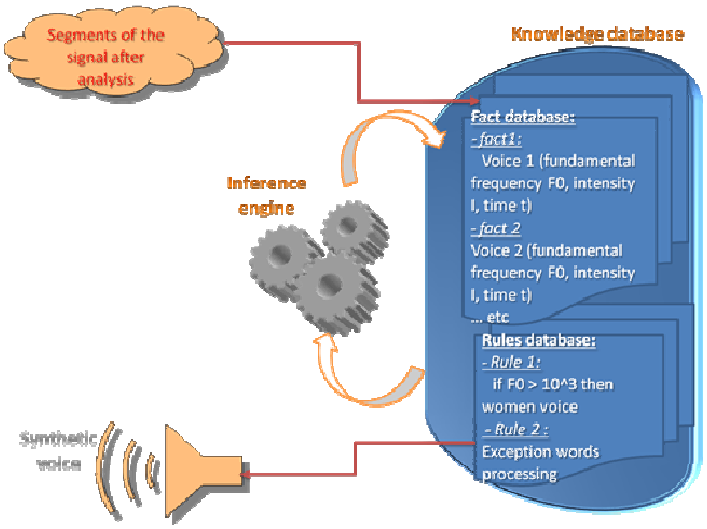

Figure 4. General diagram of our expert system

4. TESTS AND RESULTS

To test the performances of our TTS system based on Standard Arabic language, we have chosen a set of sentences which we judged like reference since they contain almost the different possible combinations specific to the language itself. To calculate the success rate (SR) associated with each sentence tested; we got the following formula:

$$SR = \frac{Numberofphrasewellpronounced}{Numberofphrasestested} * 100\%$$

The system present in general a SR of 96 % for the set of the sentences tested. Results obtained are summarized by the following table:

Table 1. Rate of success for a sample of selected sentences

Majority content	POT	Synthesis by Phonemes	Synthesis by Diphones
Short vowels	100%	95%	/
Long vowels	100%	95%	/
Solar consonants	100%	97%	/
Lunar consonants	100%	95%	/
Isolated words	100%	80%	90%
Sentences	100%	75%	85%
Numbers	90%	95%	100%
Exception Words	100%	/	/

5. CONCLUSION AND PROSPECTIVE

A system for synthesizing Arabic speech has been developed based on an expert system which uses a set of subphonetic elements as the synthesis units to allow synthesis of limited-vocabulary speech of good quality. The synthesis units have been defined after a careful study of the phonetic properties of modern Standard Arabic, and they consist of central steady-state portions of vowels, central steady-state portions of consonants, vowel-consonant and consonant-vowel transitions, and some allophones. A text-to-speech system which uses this method has also been explored. The input of the system is usual Arabic spelling with diacritics and/or simple numeric expressions. Synthesis is controlled by several text-to-speech rules within the rule database of the expert system, which are formulated and developed as algorithms more suited for computer handling of the synthesis process. The rules are required for converting the input text into phonemes, converting the phonemes into phonetic units, generating the synthesis units from the phonetic units, and concatenating the synthesis units to form spoken messages. The suitability of the method for synthesizing Arabic has been shown by realizing all its functions on a personal computer and by conducting understandability test on synthesized speech. So, we have detailed the different components which represents the basic blocks of our TTS system based on a written text in Standard Arabic and our modeling of it with the use of an expert system, this tool which is a fruits of the artificial intelligence remain less used in the field of automatic speech processing. This fact encouraged us to explore this world trying to give a little push to the research done in this multidisciplinary field. Like prospective to our work we show the following points:

- Improvement of the quality of voice generated by using methods of modification of the prosody;
- The use of others techniques of synthesis such as the Synthesis using the unit's selection;
- Make, if possible, the signal segmentation totally automatic.

REFERENCES

[1] Tebbi Hanane, « Transcription orthographique phonétique en vue de la synthèse de la parole a partir du texte de l'Arabe », Mémoire de magister en Informatique, Blida, Algérie, 2007

[2] http//:tcts.fpms.ac.be/ synthesis/ Mbrola.html.

[3] Baloul, « Développement d'un système automatique de synthèse de la parole à partir du texte arabe standard voyellé ». Thèse de doctorat d'université, Le Mans, France, 27 Mai 2003.

[4] SAROH Abdelghani, Base de données lexicales dans un système d'analyse morpho-syntaxique de l'arabe : SYAMSA, Toulouse 3, 1989.

[5] Guerti Mhania, Tebbi Hanane, La Conversion Graphèmes Phonèmes En Vue D'une Lecture Automatique de Textes en Arabe Standard, LANIA, Chlef, Algérie, 2007 S.

[6] Dichy J. & Hassoun M.O. (1998), Some aspects of the DIINAR-MBC research programme, in Ubaydly A., ed.,1998: 2.8.1-5.

[7] SYNTHAR+: Arabic speech synthesis under multivox. (SYNTHAR+: Synthèse vocale arabe sous Multivox.) (English) RAIRO, Tech. Sci. Inf. 17, No. 6, 741-761 (1998).

[8] Mehdi Braham, Oumaya Abbes, Maroua Trabelsi, Mehdi Dahmen, Rania Nouaari, « Intelligence artificielle : Diagnostic par système expert, Modèle d'analyse d'acteurs : MACTOR »

[9] P. Boula de Mareüil, « Synthèse de la parole à partir de courriers et évaluation de la conversion graphème-phonème ». LIMSI-CNRS http://www.limsi.fr/ Individu/ mareuil/

[10] Paul Boersma and David Weenink, "Praat: doing phonetics by computer" Phonetic Sciences, University of Amsterdam Spuistraat 210, 1012VT Amsterdam The Netherlands.

[11] T. Saidane, M. Zrigui, & M. Ben Ahmed, « La transcription orthographique phonétique de la langue Arabe ». RECITAL 2004, Fès, 19-22 Avril 2004.