International Journal of Engineering Innovation & Research
Volume 3, Issue 3, ISSN: 2277 – 5668

# Human-Computer Interaction System Based on Standard Arabic

**Tebbi Hanane**
Departement of CS, LRIA,
LRIA, USTHB, Algiers, Algeria
Email: tebbi_hanane@yahoo.fr;
htebbi@usthb.dz

**Hamadouche Maamar**
Departement of CS,USDB.
USDB, Blida, Algeria
Email: hamadouchemaamar@yahoo.fr

**Azzoune Hamid**
Departement of CS, LRIA,
LRIA, USTHB, Algiers, Algeria
Email: azzoune@yahoo.fr;
Hazzoune@usthb.dz

***Abstract* – In this paper, we are interested to works carried out in several fields linked to the Human-Computer interface especially the Automatic Natural Language Processing (ANLP). Our study is particularly dedicated to the voice synthesis more specifically to synthesis from text (Text To Speech; TTS).**

**TTS systems are positioned at the crossroads of computing (because the synthesizers are software), linguistics (each voice synthesis system is based on lexical analysis, syntactic, morphological and sometimes semantics, of a language), and signal processing (since the synthesized sounds are signals).**
**We demonstrate the real weight of studding the voice synthesis as well as its technological advancement and we highlight the products and technologies currently available.**

***Keywords* – About Human-Machine Interfacing, Phonetic Transcription, Standard Arabic, Voice Synthesis, Text To Speech, TTS.**

## I. Introduction

Highlight a section that you want to designate with a Human-Machine Interface (HMI) is the part of the machine that guarantees the Human-Machine Interaction. This interaction is usually based on keyboards, keypads and mice, but unfortunately in this world, there is a big part of the human race that cannot use easily those materials; visual unpaired, handicap persons, and others.

Speech was the first means of communication that existed between human being since the dawn of time, therefore the man has always thought about “teaching” this way of communication to the computer which became nowadays the most thing that a person prefer to communicate with. Starting from this fact, speech synthesis is a technology that is used in the human-Machine interaction in which we try to make the computer talk with his user basing on a text that he has entered.

Speech synthesis systems principally speech synthesis systems from text or Text To Speech (TTS) systems are one part of the great class of Automatic Speech Processing techniques. These techniques allow in particular the design of human-machine interfaces (HMI) in which the part of the interaction is done through the use of voice. These vocal systems are taking place into our daily life since they provide a simple and quick interaction way which does not require training. In addition, they remain not very expensive to be built.

Our aim in this modest work is the design and development of a TTS system based on a text written in the Standard Arabic language, we define our strategy of synthesis as well as the main modules that are necessary in providing and generating a high quality synthetic speech.

## II. Overview

Typical HCI dialog systems cover a well-defined application and perform several tasks within. Such system might be viewed as an interface between the user and the computer. It gathers user input and translates them into specific tasks [1].

Nowadays, several generations of speech synthesis systems already exists; some system are open source and they are compatible with the Windows voice; and other may be free but free for non-commercial use.

### *A. Systems in demonstration*

- HQ Acapela TTS interactive demo [2]
- KALI - voice synthesis [3]

### *B. Free systems*

- Festival [4]
- Espeak [5]
- FreeTTS [6]
- Sayz Me [7]

### *C. Commercial Freeware systems*

- The MBROLA project [8]
- Yread [9]
- DSpeech [10]
- TTSReader [11]
- Alive Text to speech 5.2.1.0 [12]
- Ivona MiniReader 1,010[13]
- NaturalReader [14]

### *D. Commercial shareware systems*

- ReadSpeaker [15]
- Dragon Naturallyspeaking [16]
- SnapVoice [17]
- Infovox Desktop [18]
- Ceptral [19]
- Infovox iVox [20]
- Proloquo [21]
- Speechissimo [22]

## III. Defining of Modules to be Considered During the System Design

When designing a system, two broad ways could be taken into account, the first one is to design the whole system using the known theories, and use it as it is designed, in the real conditions. An alternative way would

be to subdivide the system into modules that can be independently created and tested, to eventually be used in others systems to perform several functionalities.

Our TTS systems is developed using the modular way, it is based essentially on two principal parts; a front-end and a back-end. The front-end is composed of two modules, the first is for the sound database creation and the second is for the conversion text-to-phoneme or grapheme-to-phoneme.

The back-end part represents the speech generation module or in other words the synthesizer itself. So the different modules that compose the system are as follow:

- The sound database creation (segmentation): we have recorded a set of pieces of speech and store it in our database, this set is composed of phonemes and diphones which are the basic units utilized within the back-end module in order to generate voice using the concatenation method.
- The grapheme/phoneme conversion: before achieving this process, a text normalization or preprocessing operation has to be done. After that the module assigns to each word in entry it phonetic transcription, and then divides and marks the text into prosodic units like phrases or sentences. This process of assigning phonetic transcription to words is called text-to-phoneme or grapheme-to-phoneme conversion.

The output of the front-end module is a symbolic linguistic representation resulting from the phonetic transcription and prosody information together, which represents the input of the back-end module.

- The synthesizer: the back-end module uses information provided by the front-end to converts the symbolic linguistic representation to speech using a specific method. In literature, there are two kind of synthesis method; rule-based method and concatenative corpus-based method.

For our TTS system we have used the concatenative method of phonemes and graphemes previously stored in our sound database.

The general architecture of our system could be shown in figure 1 as follow:

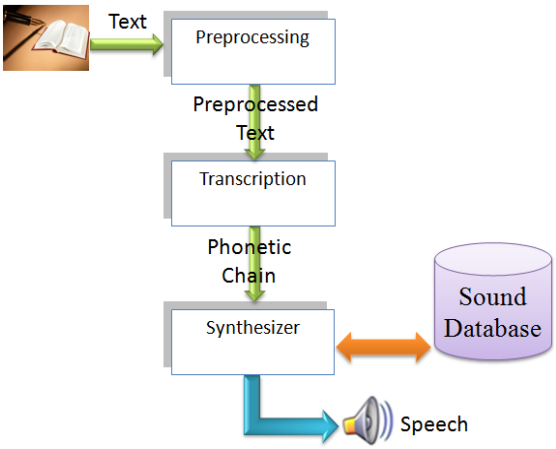


Fig.1. The general architecture of our system

## A. *The corpus description*

Most of the work carried out in the area of spoken communication often requires the registration, and the manipulation of corpus of continuous speech, and this to carry out the studies on the contextual effects, on the phonetic indices, and on the intra- and inter-speakers variability. We have created two corpuses:

- The first contains phonemes: It is composed of a set of basic sounds (which consists of the phonemes corresponding to the 28 consonants and 6 vowels, and other additional (corresponding to the three sounds of tanwiin ( [an], [a], [in] ), and the silence) character.
- To improve the quality of the words synthesized by the method of concatenation of phonemes, and to reduce the effects of co-articulation, the solution is to record the transition that exists between phonemes instead of recording the phonemes themselves; diphones which are an adjacent pair of phones.

Indeed, the transition (diphones) is the bearer of a significant quantity of acoustic information in relation to the phoneme itself. Each transition or diphone also varies from the stable part of a phoneme up to the stable part of phoneme that follows.

We give an example of an Arabic sentence « الحياة زهرة» (“life is a flower”), decomposed into diphones like follow:

{ "begin_alif_lam_h.wav", "h_ fatha.wav", " fatha _y.wav", "y_ fatha.wav", "fatha_t.wav", "t_dama.wav", "dama_silence.wav", "silence_z.wav", "z_fatha.wav", "fatha_H.wav", "H_sekoune.wav", "sekoune_r.wav", "r_fatha.wav", "fatha_t.wav", "t_on.wav", "on_silence.wav", }

## B. *The phonetic and orthographical transcription "POT"*

Transcription provides a phonetic text from the alphabetic text. To accomplish this, it must apply to many pronunciation rules. French language has a few thousands of basic rules; English language has tens of thousands of rules. The words of borrowing and the names own around the world require in each language an additional dictionary of several thousands of words; therefore in the passage of a written language to a language spoken. Two approaches can be used which are: the lexicon-based approach by and the rule-based approach [23] [24];

- *The use of rules*

In this approach each grapheme is converted to phoneme depending on the context and this is thanks to the use of a set of rewriting rules [25]. The main advantage of this approach is the ability to model the linguistic knowledge of human beings by a set of rules that can be incorporated in expert systems. Each of these rules has the following form:

[Phoneme] = {LC (Left Context)} + {C (Character)} + {RC context)}

Our transcription module grapheme-phoneme is based on a set of rules;

The rule of tanwin, al madd, etc… Prioritized, and organized in the form of a tree list. Each rule is written in the graphics context in which it is applied.

Here is a concrete example of transcription rule "The rule of Tanwin"

```
If (grapheme[char]== 'Tanwin' )
{ If  (API[position][0][ = = ' ') =
Phoneme = phoneme + "an";
Else
{ If (API[position][0][ = = ' ') =
Phoneme = phoneme + "in";
Else
Phoneme=phoneme+ "a"; }
}
```

- *The use of the lexicon*

In this case we must assign to each word in entry the pronunciation which corresponds to it without taking into account its context. The speed, flexibility and simplicity are the main advantages of this approach. It has been shown that the majority errors occurring during the conversion grapheme/phoneme, for the best operating systems came from the proper names and the exceptions which they cause [26]. The exceptions words are words that do not follow a certain pronunciation rule; therefore they represent the famous example of this approach.

### *C. The acoustic generation module "synthesizer"*

The main techniques used in speech synthesis design are Articulator synthesis, Formant synthesis, and Concatenative synthesis [27]. Articulatory synthesis attempts to model the human speech production system directly. Formant synthesis, which models the pole frequencies of speech signal or transfer function of vocal tract based on source-filter-model. Concatenative synthesis, which uses different length pre-recorded samples derived from natural speech.

In our case, we have used the concatenation method for the synthesis implementation which represent, in our opinion, the method that produce a synthetic voice the most natural and intelligible compared to the others. This result came from the fact of using a set of recording units pronounced by a real speaker, priory collected and embedded within our sound database. The reading function implemented in is shown below:

```
Position= seek (grapheme[ ig] ,API ) ;
If((grapheme[ig] == 'ا')&& (grapheme[ig+1) == 'ل))
  {
    MP2- >FileName= "C: \\son_hanane\\alif.wav";
    MP2- >Open( );
    MP2- >Wait=true;
    MP2- >Play( );
   IG=ig+2;
   Position=seek (grapheme[ig] ,PLC);
   If(API[position] [ 1] == ')
   {
     MP2- >FileName= "C: \\son_hanane\\l.wav";
     MP2- >Open( );
     MP2- >Wait=true;
     MP2- >Play( );
     MP2- >FileName=API[position] [ 2));
     MP2- >Open( );
     MP2- >Wait=true;
     MP2- >Play( );
    IG++;
   }
Else
    {
      If(API[position] [ 1] == 'S'
      {
       MP2- >FileName=API[position] [ 2));
       MP2- >Open( );
       MP2- >Wait=true;
       MP2- >Play( );
       IG++;
      }
    }
  }
```

## IV. 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:

$$SR = \frac{NB\_WELL\_PRONOUNCED\_SENTENCES}{NB\_TESTED\_SENTENCES} \times 100\%$$

The system present in general a SR of 96 % for the set of the sentences tested. Results obtained are summarized in the following table (Table 2):

Table 2: The success rate for a sample of phrases selected

| Majority Content | The 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% |
| Phrases | 100% | 75% | 85% |
| Numbers | 90% | 95% | 100% |
| Exception words | 100% | / | / |

## V. CONCLUSION

The primary objective of this research is to dissect a TTS system in its main stages. To that end, we have detailed our TTS system as well as the results obtained. The system presents, using the two corpuses that we have recorded a success rate of speech synthesis quite honorable and acceptable.

After a rapid assessment on the voice synthesis based on a text written in standard Arabic, it has been noted that this area is particularly broad and that there is no miracle product capable of responding to all applications. The voice synthesis stills a compromise between the size of the vocabulary, its possibilities multi-speaker, its rapidity, intelligibility of the speech generated, etc... The power of the current calculating tools and the integration capabilities of systems have caused a resurgence of

interest in the recent years among the industrials. In effect, they see in the voice synthesis, "the more commercial ", allowing making the difference with the competition.

As regards the future prospects, the optimism is more measured than in the past. Without risk, we can say that the general problem of the automatic processing of the voice signal will probably not rule before the middle of the next century.

At the end we give the advantages and disadvantages of our TTS system based on a text written in Standard Arabic in the following table (Table 3):

Table 3: Advantages and disadvantages of our system.

| Advantages | Disadvantages |
|---|---|
| • An unlimited size of vocabulary is used<br>• Can be used in the field of the visually handicapped, and unpaired because of its rapidity and simplicity<br>• No constraint are required for the Arabic texts reading<br>• It provides an acceptable quality of sentences synthesized.<br>• It is flexible since it can be integrated in any interactive system where the voice is the means of communication<br>• Can be improved with less coast | • The synthetic speech still not really natural<br>• product some errors while pronunciation of some specific words<br>• High quality of the voice generated requires a good segmentation<br>• Difficulty to develop the set of phonological rules and the models of grammar that are used in the transcription stage<br>• The group of exceptions words which is used in transcription based-lexicon is not complete (non-finalized) such that at any time we can insert new exceptions. |

## REFERENCES

[1] Rytis M., "The Evaluation of Spoken Dialog Management Models for Multimodal HCIs", The International Arab Journal of Information Technology, Vol. 11, No. 1, January 2014.
[2] http://www.acapela-group.com/voices/demo/, Last access time: June 7th, 2014.
[3] http://www.crisco.unicaen.fr/Demonstration-de-Kali.html, Last access time: May 1st, 2014.
[4] The Centre for Speech Technology Research, Informatics Forum, 10 Crichton Street, Edinburgh, EH89AB, Tel: +44 131 650 4434, Fax: +44 131 650 6626, email: admin@cstr.ed.ac.uk , http://www.cstr.ed.ac.uk/projects/festival, Last access time: May 11th, 2014.
[5] http://espeak.sourceforge.net, Last access time: May 11th, 2014.
[6] http://freetts.sourceforge.net/docs/index.php, Last access time: May11th, 2014
[7] http://www.fayar.net/east/sayzme.html, Last access time: May 10th, 2014
[8] The MBROLA Project, "Towards a Freely Available Multilingual Speech Synthesizer", http://mambo.ucsc.edu/psl/mbrola, Last access time: May 11th, 2014.
[9] http://www.spacejock.com/yRead3.html, Last access time: May 12th, 2014.
[10] http://dimio.altervista.org/eng/, Last access time: May 12th, 2014.
[11] http://www.sphenet.com/TTSReader/Voices.html, Last access time: May 11th, 2014.
[12] http://alive-text-to-speech.en.softonic.com/, Last access time: May 12th, 2014.
[13] http://www.ivona.com/en/mini-reader/, Last access time: May 12th, 2014.
[14] http://www.naturalreaders.com/, Last access time: May 12th, 2014.
[15] http://www.readspeaker.com/, Last access time: May 12th, 2014.
[16] http://www.nuance.com/dragon/index.htm, Last access time: May 13th, 2014.
[17] http://www.snapvoice.com, Last access time: May 13th, 2014.
[18] http://infovox-desktop. software. informer. com/2.2 /, Last access time: May 13th, 2014
[19] http://www.cepstral.com/, Last access time: May 13th, 2014.
[20] http://www.orin.com/access/infovox_ivox/, Last access time: May13th, 2014.
[21] http://www.orin.com/access/proloquo/, Last access time: May 13th, 2014.
[22] http://www.speechissimo.com/, Last access time: May 13th, 2014.
[23] Dragicevic P., "a model of interaction in input for interactive systems multi-devices highly configurable ", Ph.d. thesis from the University of Nantes, the National College of Industrial Technology and Mines of Nantes, France, March 09, 2004
[24] http://www.crisco.unicaen.fr/description-des-differentes.html, last acess time : April 24th, 2014
[25] Boula P., "Synthesis of the floor from couriers and evaluation of conversion grapheme-phoneme ". LIMSI-CNRS
[26] Tebbi H. and Guerti M., "The Conversion Graphemes Phonemes In view of an Automatic reading of texts in Arabic Standard", LANIA'2007, national seminar on the natural language and artificial intelligence, Chlef/Algeria, November 20-21, 2007.
[27] Othman O., Khalifa M., Obaid Z., Naji A. and Jamal I., "A Rule-Based Arabic Text-To-Speech System Based On Hybrid Synthesis Technique", Electrical and Computer Engineering Department, International Islamic, University Malaysia Gombak, P.O Box 10, 50728 Kuala Lumpur, Malaysia, Australian Journal of Basic and Applied Sciences, 5(6): 342-354, 2011.

## AUTHOR'S PROFILE

**Tebbi Hanane**
was born in Algiers, Algeria, 1981. She received her B.Sc (engineer) degree from University of Blida, Algeria in 2004, and her M.Sc. degrees from University of Saad Dahleb de Blida, Algiers, Algeria, in 2007. She is currently the Ph.D. student in department of Computer Science, University of Science and Technology of Houari Boumediene de Bab Ezzouar, Algiers, Algeria. Her research interests include Expert Systems, Natural Language Processing and Systems Engineering.

**Hamadouche Maamar**
Was born in Chlef, Algeria, 1981. He received his B.Sc. (Engineer) from University Hassiba Benbouali of Chlef, Algeria, in 2004, and M.Sc. degrees from University of Saad Dahleb de Blida, Algeria, in 2008.
His research interests include Pattern Recognition, Natural Language Processing, Systems engineering and Data Base.

**Dr. Azzoune Hamid**
was born in Algiers, Algeria, 1959. he received his B.Sc (engineer) degree in Computer Science from University of Science and Technology of Houari Boumediene de Bab Ezzouar, Algiers, Algeria, in 1984, and his DEA from ENSIMA, Grenoble, France in 1985, and his Ph.D from INPGrenoble, France in 1989. Presently working as Researcher Professor in Department of Computer Science at University of Science and Technology of Houari Boumedienne of Bab Ezzouar, Algiers, Algeria, since 1990. His research interests include: AI, DB, logic, CLP and web service.