Journal of Theoretical and Applied Information Technology
30th June 2015. Vol.76. No.3

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

MODELING OF SPEECH SYNTHESIS USING INTELLIGENT SYSTEM BASED ON PROLOG

[1]TEBBI HANANE, [2]HAMADOUCHE MAAMAR AND [3]AZZOUNE HAMID
[1]M.Sc, Department of Computer Science, LRIA, USTHB
[2]M.Sc, Department of Computer Science, USDB
[3]Dr, Department of Computer Science, LRIA, USTHB
E-mail: [1]tebbi_hanane@yahoo.fr,htebbi@usthb.dz, [2]azzoune@yahoo.fr, hazzoune@usthb.dz, [3]hamadouchemaamar@yahoo.fr

ABSTRACT

In this work, we present our Expert System for the automatic reading of a text written in Standard Arabic based on logic programming. To accomplish this, we have gone through three stages:

Firstly, the creation of a knowledge base that contains different rules which determine the pronunciation of Arabic words and a set of the exceptions words;

Secondly, the creation of a sound database that contains specific sounds (phonemes, diphones) to that language, spoken by a female speaker;

Finally, the implementation of the system itself that will ensure the transformation of the written text into a spoken text.

This transformation is performed by two manners; linguistic processing for converting graphemes phonemes "produce a phonetic text of the written text" and another acoustic treatment for the generation of the sound signal "play a sound corresponding to the resulting phonetic chain of the transcription step». This last step is achieved using a logic programming language which is less used nowadays, but represents in our opinion the cornerstone of logic programming languages, especially for natural language processing; Prolog.

Keywords: *Intelligent System, Artificial Intelligence, Speech Synthesis, Standard Arabic, Prolog, Logic Programming.*

1. INTRODUCTION

Artificial intelligence (AI) is a field, which studies systems that perceive their environments and formulate actions based on observations. AI research is highly technical and specialized, resulting in subfields that work in isolation from one another. This discipline, which has become more benefits in our daily lives through the several contributions it offers, remains the heartbeat of what we call now intelligent systems in areas that rely on computer technology. It is becoming common as well as computer skills are concerned with human activities. Intelligent systems exist now among our daily life, based on the numerous applications that use the paradigms of artificial intelligence.

We can define intelligent systems as systems that include process based on several theories to reproduce some human behavior in order to achieve a set of tasks. The design of such systems in artificial intelligence, need several tools and programming paradigms, but remains one of the main tools of logic programming that allows us to free up some programming constraints. Indeed, implementing instructions are not to describe since they are already available with the language as a theorem prove or inference engine. This type of programming is particularly suited to the needs of the AI [1].

Our work makes the report a few years of scientific activities that our efforts have focused on the design and development of a speech synthesis expert system based on a text written in Standard Arabic; whose goal is to produce speech the likely

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

to the human voice, in terms of both intelligibility and naturalness.

The automatic speech generation is a complex task due to the variability intra and interlocutor of the voice signal. In computer science, the difficulty lies in the fact that we do not know so far how to model the huge amount of knowledge and information useful for the synthesis of signals that are as close as possible to natural speech of humans.

As the Arabic language is one of the most complex languages (involves more rules and exceptions without talked about his dialect) so we make the choice to use an expert system to model this knowledge to build a strong system which can really read a text written in standard Arabic.

2. DEFINITIONS

One of the ultimate goals of AI is to simulate cognitive and behavioral activities of a human being with conscience and feelings. These activities include reading a written text.

Speech synthesis from a text exists where the word can replace or supplement an existing interface to help the machine to communicate an information using an artificial speech. It plays an important role in the achievement of human-machine interfaces (HMI), where part of the interaction is assured using voice. Speech synthesis has several advantages; it is more natural for the public, offering comfort in the exchange of information.

Indeed, the user can express himself whenever he wants without waiting for an eye contact on one hand. Also it is faster than in the case of a written message by getting the possibility of exchanging with any interlocutor sharing the same language while keeping the free field of vision to make another task, what makes the emission (and understanding) of the message faster, especially if the code and the tool are completely mastered, in the other hand.

The two main criteria for automatic speech synthesis are the intelligibility and naturalness. If today, the first criterion is met, the second is still under development. In fact, if synthesizers reproduce a completely understandable voice, the intonations and the expressiveness are not ended in the optimum yet.

3. PROBLEMATIC

The main goal of the automatic natural language processing (NLP) is the design of a system capable of processing data expressed in a natural language, through several steps in order to extract the necessary information.

The aim of our work is the realization of an expert system capable of reading: the Arabic alphabet, words in Arabic, then a sentence with a concatenation of words, based on reading rules that depend on the nature of each grapheme according to its neighborhood and the nature of each word according to its position (verb, subject, adjective, etc.).

To complete this task, we will face technical problems, some particular language include the case of pronunciation at different levels; as the pronunciation of new names. Another problem is the lack of vowels in most Arabic texts, which may generate some ambiguity at two levels (Meaning of the word, and the difficulty in identifying its function in sentence) and, finally, the famous problem of the expression of emotions in speech generated (sad, happy, affirmative, exclamatory, etc.).

The main objective of TTS systems is to equip the computer with the ability to read texts aloud. Despite the advances made in recent years in these areas, progress remains to be done to increase the comfort of use of current systems.

We aim the design of a computer system with a voice synthesizer, able to transform the written form (graphemes) to the oral form (phonemes) of a text written in Standard Arabic, so we can consider the following reasons:

- The Arabic writing is very little used, the use of computer systems speaking Arabic (interactive terminals, querying of data banks, etc.) must at present incorporate an oral answer.
- The dialect of Arabic is not yet completely fixed, it seems interesting to provide to linguists who are working on those problems a tool of reading texts in order to test the modes of representation of the phonemes of the language.
- Provide interactivity and ease of use, since our work is intended to experts, and non-experts especially for blind who must understand and be able to use our system ;

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

- The user will be able to understand the different synthesized sentences that must be clear and pronounced with an acceptable quality.

4. STATE OF THE ART

At present we can judge that works done in the same context as ours are still not really colossal and sufficient (the best known in the industrial world working on the speech synthesis of Arabic language are Acapela-group [2] and Sakhr [3]), and this because of the complexity of the language itself. In addition, works that exist are based on the same principle transcribers of other languages (French, English, etc.), however, efforts are encouraging and open a large door for research in this area.

In the following statement, we describe some previous works:

- 2005: TTS system (Text To Speech) MBROLA [4] that uses the SAMPA code during the step of transcription, in this case, the user must follow the form of the SAMPA code that is not a universal code
- 2003: The work of S.Baloul [5] which represent a concrete example of the transcription of word on the basis of the morphological analysis, and studies of break to generate the pronunciation of texts
- 1998: The GAZALI project, which was achieved in the RICST (Regional Institute of the Computing Sciences and Telecommunications) of Tunisia. It concerns the transcription part of a TTS system that has a particularity of using a set of propagation rules of emphasis. [6]
- 1998: SYNTHAR + [7] a tool studied by Z. ZEMIRLI in the NIC (the National Institute of Computing of Algiers), he assures the transcription step for a TTS system which will transmit the phonetic representation of the text to the MULTIVOX synthesizer. It is necessary to know that SYNTHAR + is based on a morphological analysis before realizing the transcription.
- 1989: SYAMSA [8] (SYSTEM of Morphosyntactic Analysis of the Arab), which was realized by SAROH. According to him, «the Arabic phonetization is based in particular on the use of lexicons and a morphological analyzer for the generation of the various forms of a word. Besides, it is the phenomena of interaction between the words (connection, elision, etc.) and the phenomena of assimilation that suggest the use of phonological rules". This tool assures for every word in entry, the root that corresponds to it, as well as its morphological and phonetic representations. Consequently, the operation of transcription in this system is in fact a comparison of words in entrance with those stemming from the morphological analysis. If there is a correspondence between the word in entry and one of the resultant words of the morphological analysis, the system provides directly its phonetic representation.

These days and with the emergence of Internet and the intensification of research in the field of information technology, business and more independent startup created their own voice synthesis systems, examples are given in the following paragraph:

- **Free Software**
  Festival [9] Espeak (Linux and Windows) [10], FreeTTS [11] Sayz Me (Windows) [13]
- **Owners**
  Free
  MBROLA (voice synthesizer) [4], yRead (Windows) [14] DSpeech [15] TTSReader [16]
  Pay
  Voice Reader (Windows) [17] Naturally Speaking (Windows) [18], Cepstral Swift (Windows Linux), [19] SnapVoice (Windows) [20], Speechissimo [21] Proloquo [22] GhostReader [23] , Zyyne voice synthesis papers (online) [24], Software AIACA Speech Kali USB key [25]

With all these examples, we can say that the realization of intelligent systems that use the word begins to take an unavoidable aspect of our society.

The special feature of our work in relation to all that exists is the use of expert systems in the Arabic grapheme transcription phase, what gives very acceptable results. One component of this innovative technology, we will find out in the next section, is the logic programming.

Journal of Theoretical and Applied Information Technology
30th June 2015. Vol.76. No.3

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

5. LOGIC PROGRAMMING

To build applications, Logic programming is a programming paradigm founded on formal logic. It means that it use a set of sentences in logical form, expressing facts and rules about some problems in a specific domain. Major logic programming language, like Prolog, use rules written in the form of clauses like:

C: - F1,F2,…, Fn.

Which means that C is true if F1 and F2 … and Fn are true. Here C is called the head of the clause and F1,…Fn is called the body. For the case of facts, they are simple clauses without body in the form of C.

In the simplest case in which C, F1, …, Fn are all atomic formulae, these clauses are called definite clauses or Horn clauses. However, there exist various extensions of this simple case, the most important one being the case in which conditions in the body of a clause can also be negations of atomic formulae.

Logic programming languages that include this extension have the knowledge representation capabilities of a non-monotonic logic [26].

6. GENERAL ARCHITECTURE OF OUR SYSTEM

Our speech synthesis system is divided into two big parts:

1. A linguistic or symbolic part, which allows from a written text generating a labelled phonetic text, this passage of phonetic spelling transcription (PST) includes three stages:
   - The preprocessing of the text
   - The consultation of lexicon of the exceptions
   - The application of the rules of transcription established for the Arabic language.
2. An acoustic part (the vocal synthesizer) which is in charge of generating speech corresponding to the resultant phonetic chain of the operation of transcription.

The following diagram describes the general schema of our system:

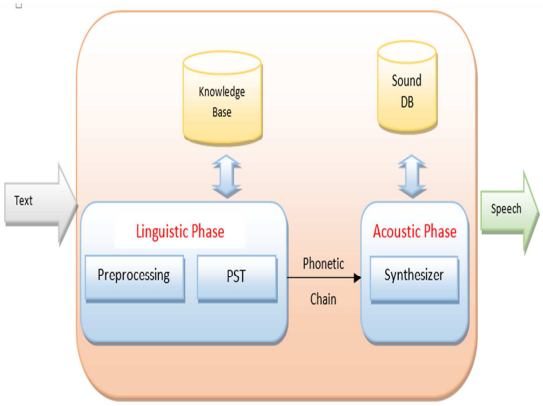

Figure 1. General architecture of our system

7. THE PROJECT IMPLEMENTATION

7.1. Functional Specifications

In this work, we aim the design and implementation of a system capable of converting written text into voice. The required characteristics are as follows:

- The programming language : Prolog;
- The language used: Standard Arabic;
- Target Platform: Windows;

The system provides to the user an interface with which he will communicate, the user writes the text and the system generates the appropriate corresponding speech.

7.2. Feasibility study and technological choices

During the implementation phase, we chose Prolog as a logic programming language, to create our knowledge base, and we have used a sonagraphique analysis by using the analysis tool of the signal voice PRAAT to obtain spectrogram. The result spectrogram provide a three-dimensional representation of the signal voice through three axes:

- Vertical axe represents the sound frequency in Hz.
- Horizontal axe represents the temporary evolution of the sound.
- The degree of blackening which represents intensity or sound energy in dB (decibel).

After that, we manually cut the speech signal into a sequence of segments; each one is associated with a single acoustic element (phoneme or

Journal of Theoretical and Applied Information Technology
30th June 2015. Vol.76. No.3

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

diphone). These elements are the components of our basic sound database.

7.3. The use case diagram "use case"

During the modeling phase of our system, we make the choice to use the UML (Unified Modeling Language) language; it is based on a set of diagram. The first diagram used is the use-case diagram (Fig2).

In the following diagram, a use case is describing how a user interact with the system by writing the text, and listening to the sound file, which was opened by the system.

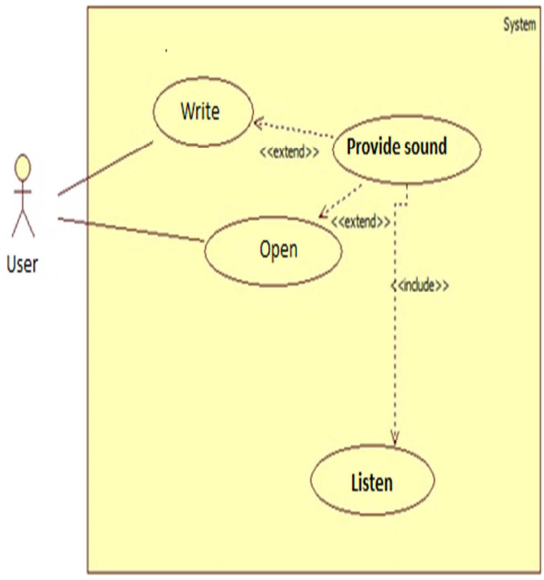

Figure 2. Use case diagram

7.4. The activity diagram

In this diagram, we expose the mechanism of working of the system, the point of beginning is the text entered, and after that, we apply successive tests on it one by one. The last point of the diagram is either, a synthetic voice corresponding to the text entered, or a warning message to inform the user that he made a syntactic error. The fig 3, describes the activity diagram of the system.

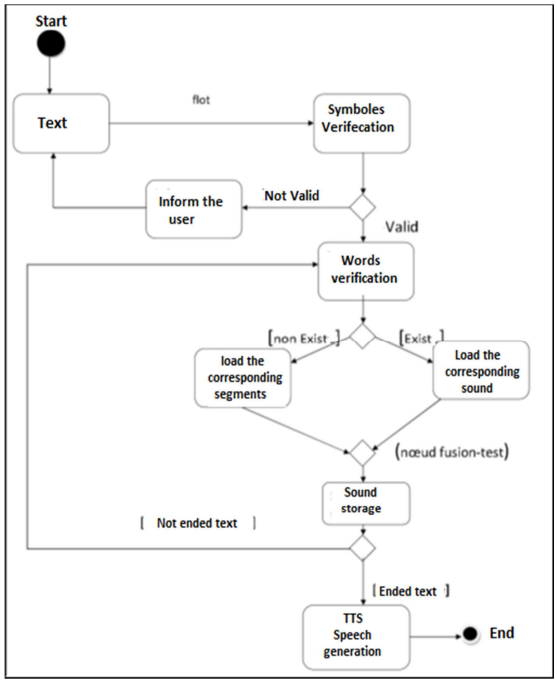

Figure 3. The activity diagram

7.5. The sequence diagram

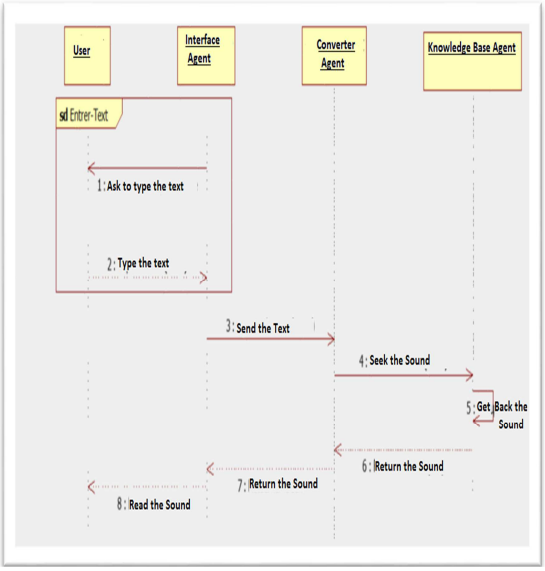

Figure 4. The sequence diagram

7.6. Our expert system

The expert system is a computer tool of artificial intelligence, designed to simulate the expertise of a specialist in a specific and well-defined area through the operation of a number of explicit knowledge provided by experts in domain [27]. To build our system, we need the following basic elements:

Journal of Theoretical and Applied Information Technology
30th June 2015. Vol.76. No.3

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

1. **The inference engine:**

An inference engine (the verb "inferred" which means: "deducted") is a program corresponding to a simulation of a deductive reasoning algorithm. It allows experts systems to conduct logical reasoning and conclusions derived from a base of facts and a knowledge base .

The inference engine may implement:

- a formal logic of order 0 (propositional logic) 0+ order, or first order (predicate logic)
- a forward chaining , backward chaining or mixed chaining [28]

In our case we have used SWI-Prolog (Logic Programming) as the inference engine, it is a descriptive programming language (facts and relations) and prescriptive (inference); it is based on the logic of the first order. Alain Colmerauer invented it in the early 70s in Marseille (France), in order to be able to do the processing of natural language. The software can answer a request from the user to trigger a reflection defined by its rules of inference (deduction) that will use the knowledge base.

2. **Creation of the knowledge base**:

A knowledge base includes knowledge specific to a given specialized domain, under an exploitable form by a computer. It can contain rules, in this case, we speak about rules database, or facts thus we speak about facts database [29]. Our knowledge base is divided into two blocks:

- The facts database: it concerns the phonology of Arabic letters (Al horof: الحروف), It includes three categories of predicate written as: speak (X, / Y /). We have used 62 predicates to model the 28 consonants and 6 vowels (short, and long) specific to the Standard Arabic, here is some concrete examples of these predicates:

lettre(أ,alif). lettre(ب,ba).
lettre(ت,ta).
api(أ,a,'g:/sons/alif fatha.wav').
api(ب,bh,'g:/sons/ba fatha.wav').
api(ت,t,'g:/sons/t fatha.wav').
fatha(a,[a]). damma(ou,ou).
kasra(i,i).

- The rules database: it contains rules written as following:

```
Pronounce ([], []).
Pronounce ([Head|Tail], Result):-api (Head, R1), // api is the international phonetic alpabet

Pronounce (Tail, R2), Concate (R1, R2, Result).
```

If {أ} + {ل} + {الحروف الشمسية} Then {أ} + {الحروف الشمسية} + {ّ}
الشمس ➔ اشّمس
If {أ} + {ل} + {الحروف القمرية} Then {أ} + {ل} + {الحروف القمرية}
القمر ➔ ألقمر
If {الحروف} + {َ} + {ّ} Then {الحروف} + {ْ} + {الحروف} + {َ}
If {الحروف} + {ُ} + {ّ} Then {الحروف} + {ْ} + {الحروف} + {ُ}
If {الحروف} + {ِ} + {ّ} Then {الحروف} + {ْ} + {الحروف} + {ِ}
اشّمس ➔ اَشْشَمس
If {الحروف} + {ً} Then {الحروف} + {َ} + {ن} + {ْ}
If {الحروف} + {ٌ} Then {الحروف} + {ُ} + {ن} + {ْ}
If {الحروف} + {ٍ} Then {الحروف} + {ِ} + {ن} + {ْ}
لبنٌ ➔ لبنُنْ
If {الحروف} + {َ] + {ى} + { } Then {الحروف} + {َ} + {ا}
عصَى ➔ عصَا
If {أ} + {ل} + {ل} + {َ} Then {أ} + {ل} + {َ}
If {أ} + {ل} + {ل} + {ُ} Then {أ} + {ل} + {ُ}
If {أ} + {ل} + {ل} + {ِ} Then {أ} + {ل} + {ِ}
ألَلَبَنُ ➔ اَلَبَنُ
If {ؤ} + {ُ} Then {أ} + {ُ}
If {ئ} + {ِ} Then {إ} + {ِ}
If {ء} + {َ} Then {أ} + {َ}
أسَاءَ ➔ أسَاَا
If {أ} + {َ} + {أ} + {ْ} Then {آ}
If {أ} + {ُ} + {أ} + {ْ} Then {أ} + {ُ} + {و}
If {إ} + {ِ} + {أ} + {ْ} Then {إ} + {ِ} + {ي}
أأَمَنَ ➔ آمَنَ
If {ة} + {َ} + { } Then {ه}
If {ة} + {ُ} + { } Then {ه}

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

If {ة} + {ِ} + { } Then {ه}
فَاطِمَةُ ➔ فاطِمَه
If {ن} + {ْ} + {ب} + {ب} Then {م} + {ْ}
مِثْبَرٌ ➔ مِمْبَرٌ

If {ض} + {ْ} + {ط} Then {ط} + {ْ} + {ط}
اِضْطَرَ ➔ اِطْطَرَ
If {ض} + {ْ} + {ت} Then {ط} + {ْ} + {ط}
نَهَضْتُ ➔ نَهَطْطُ
If {د} + {ْ} + {ت} Then {ت} + {ْ} + {ت}
عَبَدْتُ ➔ عَبَتْتُ
If {ه} + {ُ} + { } Then {ه} + {ُ} + {و}
لَهُ ➔ لَهُو
If {الحروف} + {ً} + {ا} + { } Then {الحروف} + {َ} + {ن} + {ْ}
عَصَا ➔ عَصَنْ

As well as to model the exceptions of the following proper nouns:

اَللَّتِي ➔ الـتِي
اَللَّذِي ➔ الذّي
ذالِكَ ➔ ذَلِكَ
كَذالِكَ ➔ كَذَلِكَ
غَرامْ ➔ غ
كيلوغرام ➔ كغ
هَاؤُلاءِ ➔ هَؤُلاء
هَاذَا ➔ هَذا
هَاذِهِ ➔ هَذِهِ
هَاذِه ➔ هَذَانِ
هَاذَيْن ➔ هَذَيْنِ
اَللَّذِي ➔ اَلْذِي
مِثْرٌ ➔ م
لِثْرٌ ➔ ل
أنَا ➔ أنَ
إذاْ ➔ إذَنْ

8. CONCLUSION

To conclude this research, we can declare that the face value of an intelligent system that can read textual resources aloud is great. Reading aloud is a crucial element of emergent and early reading and literacy efforts, of course, but it is also an important part of teaching methods targeted to struggling readers for example. Balajthy and Lipa-Wade [30] identified several teaching strategies concerning reading aloud that would be useful for students who are severely challenged in reading, including the neurological impress method [31], echo reading [32], and comprehension for decoding [33].

In addition, professional advice to continue reading aloud to students even as they reach adolescence is common in the literature [34]. Zirinsky and Rau (2001) mention that [35].

As texts become more difficult and more obscure nowadays, maybe oral reading would give a voice to texts that silent reading alone would render inaudible. To extend natural reading practices, students for example, could be helped by hearing their works read aloud.

Artificial Intelligence is a science that studies technics and methods to transform computers to intelligent machines, in order to brake the barrier of communication between man and machines, by exploiting better its capacities. There is no doubt that IA will increasingly be an integral part of information systems future [36].

At the end, we can say that lot of people in many situations are assisted by being able to listen to text as well as read themselves. Some of them use it for the accessibility of it, and others because they do not have a choice; they are obliged. TTS systems offer an alternative (or complementary) way to text that can offer students what they need to understand and remember. This will be, in our opinion, the future of education and Learning.

REFERENCES

[1] Mohamed Bouanane, "Modeling of a graphical tool to help in the understanding of solving logic programs," Memory Master of Science (M.Sc.), Quebec, Canada, 2014

[2] http://www.acapela-group.com/

[3] http://sakhrusa.com/

[4] http://tcts.fpms.ac.be/synthesis/mbrola.html

[5] S. Baloul (27 May 2003) Development of an automatic speech synthesis system from the standard Arabic text vowelized, PhD thesis, Le Mans, France.

[6] Tebbi Hanane "phonetic spelling for the speech synthesis Transcription from the Arabic text," magister thesis in Computer Science, Blida, Algeria, 2007

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

[8] SAROH Abdelghani, lexical database in a morphosyntactic analysis of the Arab system: SYAMSA, Toulouse 3, 1989.

[9] http://www.cstr.ed.ac.uk/projects/festival/

[10] http://espeak.sourceforge.net/

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

[11] http://freetts.sourceforge.net/docs/index.php
[12] http://sayzme.sourceforge.net/
[13] http://tcts.fpms.ac.be/synthesis/mbrola.html
[14] http://www.spacejock.com/yRead.html
[15] http://dimio.altervista.org/eng/
[16] http://www.sphenet.com/TTSReader/index.htm
[17] http://www.linguatec.net/products/tts/voice_reader
[18] http://www.nuance.com/dragon/index.htm
[19] https://www.cepstral.com/store/
[20] http://www.snapvoice.com/
[21] http://www.speechissimo.com/
[22]http://www.assistiveware.com/product/Proloquo
[23]http://www.convenienceware.com/product/GhostReader
[24] http://www.zyyne.com/site/
[25] http://www.aiaca.fr/vente.php
[26]http://fr.wikipedia.org/wiki/Programmation_logique
[27] Mehdi Braham, Oumaya Abbes, Maroua Trabelsi, Mehdi Dahmen, Rania Nouaari, "Artificial Intelligence: Diagnostic expert system analysis model of actors: MACTOR"
[28] http://fr.wikipedia.org/wiki/Moteur_d%27inf%C3%A9rence
[29] http://fr.wikipedia.org/wiki/Base_de_connaissance
[30]Balajthy, E., & Lipa-Wade, S. (2003). Struggling readers: Assessment and instruction in grades K-6. New York: Guilford.
[31] Heckelman, R.G. (1969). A neurological impress method of remedial reading instruction. Academic Therapy, 4, 277-282.
[32] Anderson, B. (1981). The missing ingredient: Fluent oral reading. Elementary School Journal, 18, 173-177.
[33] Lipa, S. (1990). Comprehension for decoding. Journal of Reading, Writing, and Learning Disabilities International, 6, 93-105.
[34] Trelease, J. (1989). The new read-aloud handbook. New York: Penguin.
[35] Zirinsky, D., & Rau, S.A. (2001). A classroom of teenaged readers: Nurturing reading processes in senior high English. New York: Longman.
[36] John Paul HATON, Artificial Intelligence, Text of the 263rd meeting University of all knowledge given on 19 September 2000.