Towards a Free Multilingual Speech Synthesis Software for the Vocally Handicapped

Thierry Dutoit, Xavier Ricco

Faculté Polytechnique de Mons, MULTITEL-TCTS Lab
Parc Initialis
7000 Mons Belgique
{ricco,dutoit}@tcts.fpms.ac.be
URL : http://tcts.fpms.ac.be/synthesis

Abstract

The goal of the W project, launched in 1999 by the TCTS Lab of the Faculté Polytechnique de Mons, is to let people with speech disability benefit from recent developments in terms of speech synthesis systems. HOOK has been developed in the context of this project; it is a software that hooks keystrokes in MS-Windows, performs abbreviation expansion, and automatically shows, in real time and through a popup window, the proposed expanded form. Expansion rules are based on multi-level regular rewrite rules. An implementation of Grade II Braille abbreviations is provided for French and English. When used in combination with the EULER text-to-speech system (or with any other TTS system), HOOK makes it possible for people with a speech handicap to communicate through the MBROLA speech synthesizer (included in EULER). HOOK, MBROLA, and EULER are distributed for free, for non-commercial, non military use.

This paper describes the current state of these projects.

Introduction

People working in the area of Text-to-Speech (TTS) synthesis frequently receive calls from handicapped people, or associations thereof, requesting information on the availability of TTS technologies for the blind and for people with a speech impairment. As a matter of fact, speech synthesis can, in theory, provide invaluable help in such cases, by making it possible for the blind to access text documents without the need to use Braille, and providing the speech impaired with a means of (synthetic) oral expression. While significant efforts have been made these last years towards shortening delays between research and production in this area, one must admit that people with speech disabilities are frequently left aside. One of the reasons for this is undoubtedly that real-time interfaces to TTS systems remain a problem: even a trained secretary cannot type as fast as one speaks.

It turns out, however, that many people (other than researchers or people involved in companies in the field) could take part in the developments of these tools (provided they receive help when needed), namely the handicapped people themselves, or benevolent people willing to spend some of their time on such a project. The MBROLA, EULER, and W projects were designed, among other goals, with this view in mind.

Main Part

The MBROLA project (http://tcts.fpms.ac.be/synthesis/mbrola/) was launched in 1995. It is a phonetic-to-speech synthesis engine which is compiled on 21 combinations of computer/OS, and now speaks 28 languages. Its use is free for non-commercial, non-military purposes : among other things, it is free for the private use of handicapped people.

As such, however, MBROLA cannot read texts : it needs additional tools for computing the phonetic transcription of the text to be read, as well as its intonation. EULER (http://tcts.fpms.ac.be/synthesis/euler/)

was launched in 1999 to bridge this gap. It is a complete text-to-speech system, currently available for MS-Windows and Linux (Mac port targeted for September 2001), and speaks French, Arabic, and (rudimentary) English, Spanish, and Dutch (for these last languages, intonation is still flat). The EULER license is very similar to that of MBROLA : handicapped people can use it for free privately.

But having free TTS tools available does not mean they are really usable. One needs to provide adapted interfaces. Existing solutions, namely sentence assembly (using a specially designed hardware that replace traditional keyboards, or software tools to emulate such hardware), word prediction, and abridged languages, have pro's and con's.

Word prediction has been (and still is) widely studied, a large number of software tools have been produced. They all use the same idea : knowing previous keystrokes and some lexical (sometimes also morphological and syntactic) information, the tool predicts the word with highest likelihood (or a list of *n*-best candidates). KTH, which has been a major player in this field for more than 15 years, has reported on the relative inefficiency of such prediction tools (Magnusson, 1994) if they are to be used for real-time communication, mostly because people cannot predict themselves what will really be proposed by the tool. As a result, they have to validate its proposals by inspecting the screen (which takes time), and possibly correct mistakes, (which also consumes time).

Abridged languages do not suffer this impediment : they can be learned, so that validation and correction can be minimized with time. In the context of the W project (http://tcts.fpms.ac.be/synthesis/w/), we have chosen to explore abridged languages and more particularly Grade II Braille for that purpose. Grade II Braille has been in use since 1829 to help blind people read and write texts in a concise way. It differs markedly from Grade I Braille, in which each character is mapped to 6 dots (no contraction is used). For English, for instance, Grade II Braille provides 189 contractions, including contractions for sequences of characters that are frequently encountered (i.e., contractions which should be seen as rules, rather than as simple substitutions). It provides contraction ratios of 30 to 50 percent, depending the type of text involved. This result is similar to that provided by word prediction algorithms, but contractions are much easier to use than word prediction systems : contractions are systematic enough for human users to learn them by heart, which has been done extensively by blind people for the past hundred and seventy years. Additionally, Grade II Braille progressive learning methods are already available (for French : Kommer, 1993).

In the W project, abbreviation expansion is described in a regular grammar, in terms of rewrite rules. More particularly, we use a Multi-Level Rewrite Rules (MLRR) parser. MLRR rules have the general form :

A1 / L1 _ R1 ; A2 / L2 _ R2 ; … ; An / Ln _ Rn ---> B

Such a rule means that symbol (or a sequence thereof) *A1* in the main input string is transduced into symbol (or a sequence thereof) *B* in the output string when it is surrounded by *L1* and *R1* in the main input string, and when symbols (or sequences thereof) *A2, … , An* (which are synchronized with *A1* but belong to other input layers) are respectively surrounded by symbols (or sequences thereof) *L2* and *R2*, … , *Ln* and *Rn* in their respective layers.

The design of the MLRR transducer was motivated by problems encountered in speech synthesis, where the use of standard regular transducers (for phonetization, for instance) leads to complex, hardly readable rules. RULESYS (Carlson, Granström, 1976), developed some 25 years ago, is still used today, including commercially, and provides an excellent example of the use of such standard regular transducers. Complexity mostly originates in the fact that information of various kinds (typ : morphological, syntactic, graphemic, etc.) tend to be stacked in the input/output string. On the opposite, TTS systems now tend to store linguistic information in data types composed of several layers, with links between related data in different layers (see for instance Van Leeuwen, te Lindert, 1993). MLRR implements this principle.

Tableau 1 gives a sample of the content of an MLRR rule database for W. Classes can be defined, so as to shorten rules and make them more readable. Some rules correspond to simple lexical substitutions, while others are more context dependent.

```
 [STRUCTURE]
WABREV = ABREVIATION -> RESULTAT
[CLASS]
…
CLASS <S> (^|<W>|<P>|[\'])                    # left word delimiter
CLASS <T> ($|<W>|<P>)                         # right word delimiter
CLASS <X> (<T>|s<T>)                          # allow « s » if plural
…
[WABREV]
...
[<S>](bê)[<X>]->(bête)
[<S>](bêm)[<T>]->(bêtement)
[<S>](bf)[<X>]->(bienfait)
[<S>](bfc)[<X>]->(bienfaisance)
[<S>](bf9)[<X>]->(bienfaiteur)
…
 [](blt)[<T>]->(bilité)
 [<C>](m)[<T>]->(ment)
 [(<S>|ce|dé|er)](pd)[<F>]->(pendant)
…
```

Table 1 – Sample of the MLRR database for French.

The MLRR transducer is now practically made available through HOOK, a (free) software component which captures keystrokes from any text-enabled application in MS-Windows, and proposes expanded words in real time, in a popup window. Each time a key is pressed, a new expansion is proposed. Validation is assigned to the space bar, so that trained users need not really check proposed expansions.

Conclusion

The combination of MBROLA/EULER/W/HOOK is an important effort towards free multilingual speech synthesis software for the vocally handicapped. These projects, however, are far from being completed. MBROLA is now in its cruise speed. We receive (and co-produce) about 10 complete new voices per year (among which about 5 covering new languages). EULER is still growing. Collaborations for extending it, merging it with other free tools, and developing new language modules, are very welcome. HOOK still needs to integrate contractions for more languages (possibly not Grade II Braille : it is compatible with any kind of rule-based contraction system). The Multi-level aspect of MLRR itself is not really exploited yet in the current version of HOOK. It should come to use in the context of the IST5 FASTY project ("Faster Typing for Disabled Persons", IST-2000-25420, http://www.fortec.tuwien.ac.at/reha.e/projects/fasty/fasty.html), started in January 2001. Fasty aims at using a combination of intelligent prediction and abbreviation expansion for increasing typing speed. In this context, morphological and syntactic layers could be used to store information on previously types words, making it possible for the expansion process to account for morphological features when proposing expanded words.

References

Magnusson T. (1994), "Evaluation of Predict", Quaterly Progress Scientific Report, KTH Department of Speech, Music and Hearing.

Carlson, R., Granström, B. (1976), "A Text-to-Speech System Based Entirely on Rules", Proceedings of ICASSP 76, Philadelphia, pp. 686-688.

Kommer E. (1993), "La méthode d'abrégé braille en noir", available from Association Valentin Hauy, Paris.

Van Leeuwen, H.C., te Lindert, E. (1993), "Speech Maker: a Flexible and General Framework for Text-to-Speech Synthesis, and its Application to Dutch", Computer, Speech and Language, n°2, pp. 149-167.