# A TEXT-TO-SPEECH SYNTHESIZER ADAPTED TO THE NEEDS OF MOTOR AND SPEECH HANDICAPPED PEOPLE

Pedro M. Carvalho INESC
Luís C. Oliveira IST/INESC
Isabel M. Trancoso IST/INESC
M. Céu Viana CLUL *

**INESC** R. Alves Redol, 9, 1000 Lisbon, Portugal

**ABSTRACT**

This paper describes the development of a vocal aid prototype for motor and speech handicapped people, based on text-to-speech synthesis, which may simultaneously serve as a learning tool for helping young children to associate sounds and letters. The system is primarily aimed at children suffering from cerebral palsy and has been developed on the basis of their specific needs. It is portable and was designed to assist children in different interactive situations at home, school, or rehabilitation centers.

**INTRODUCTION**

The main goal of the EDIFALA project described in this paper is the development of a text-editor with speech synthesis capabilities, aimed at assisting children suffering from cerebral palsy.

The design specifications of the system were the result of a joint effort of Engineers, Linguists and teachers and therapists working in this area. The system as a whole has not yet been subject to evaluation which will be done during the next school year with the close involvement of a group of such children using it in different interactive situations at home, school and rehabilitation centers.

Speech synthesis by itself is not sufficient to significantly reduce the serious communication problems of children with this type of multi-handicap, and similar systems developed in other countries for other languages have faced some difficulties. Some are caused by the handicap itself but others have a much broader scope and are related to the use of synthetic speech in interactive situations. In order to overcome these difficulties, the intelligibility and naturalness of synthetic speech had to be improved, and the man-machine interface must be adapted to motor-handicaps of different types and degrees so that the rate of message generation can be speed up and the number of typing errors reduced.

The project therefore comprises two main tasks. The first one is concerned with improving the text-to-speech synthesizer (DIXI) used as a basis in this project. We shall therefore devote the following section to the description of the DIXI system and the ongoing work to improve it. The second task, described in section 3, is related to the user interface and the main guidelines we followed to adapt it to the needs of children with motor and speech handicaps. Before concluding we shall discuss the main improvements we intend to introduce in the current system.

**THE DIXI TTS SYSTEM**

The DIXI TTS system is structured into four major modules: text normalization, linguistic and phonetic processing, generation of the synthesizer parameters and synthesis itself. It is basically a synthesis-by-rule formant synthesizer, using Klatt's model [1]. The DIXI architecture was designed to maximize flexibility and modularity, which makes it a very important research tool, since it allows the evaluation of the linguistic and phonetic theories of the language, and also serves as a test-bed for assessing different sound production models. Other design guidelines were its future extension to other varieties of Portuguese such as Brazilian Portuguese and varieties spoken in African countries and the feasibility of real-time implementation in the future, which was taken into account, namely, by using efficient coding and by limiting the dictionary size.

The system was primarily developed in a Unix platform. Much of the work done in this project from the engineering point of view involved porting the system to a PC running MS-DOS and, in the next couple of months, to a WINDOWS environment.

The second and third modules of DIXI were built using a multi-level rule-compiler (SCYLA - Speech Compiler for Your LAnguage), developed at CSELT [2], combined with a set of auxiliary functions written in the C language. The use of a rule-compiler has the advantage of imposing a more structured rule definition and of enabling the development of the system by non-programmers. SCYLA's choice was motivated by several features: its multi-level structure, which allows each procedure to simultaneously access all previous procedures; its ability to generate portable C code which can be optimized for the hardware where it is going to run; its powerful debugging tools; and, finally, its connection to a conventional

*in alphabetical order

procedural language for the operations more easily coded in this form.

In the development of the system a test set of about 25,000 different forms was used for testing procedures operating at the word level or bellow. This constitutes a frequency corpus collected by CLUL (PF), based on oral interviews carried out all over the country. It comprises citation and inflected forms corresponding to about 715,000 occurrences and about 3,030,000 segments.

Two main research lines are currently being followed to improve the quality of the system. The first one concerned the verification and adjustment of segment target values and rules for generating control parameters. In this context, the main limitation of the first version of DIXI was the realization of the consonants [ʎ], [ɲ] and [ɬ]. Another research line concerned prosody, namely in terms of syllabification and intra and inter-word coarticulation phenomena.

## THE USER INTERFACE

During the first phase of the project, the development of the the text-editor with the user interface adapted to the specific needs of the handicapped children was done in parallel with the improvement of DIXI. An existing DOS-based TTS system for English was thus adopted for the development of the editor in place of DIXI.

As is is well known, one of the most difficult problems of this target population is the message generation rate. The editor thus involves several acceleration procedures. One of them is the introduction of user defined abbreviations and the other the introduction of lexica which take into account the frequency of occurrence of the entries in the language. These lexica enable the user to complete a word with a single touch by selecting it from a menu of the most frequent 10 words starting with the same characters. By default, the system uses the PF lexicon mentioned above, but in alternative it is also possible to use user-adapted lexica, that change as a result of the system usage.

In order to reduce the effects of lack of motor coordination, the constant pressing of a key during a relatively long period of time is not interpreted as a sequence of repeated characters, but rather as a single character.

Major acceleration in interactive communication situations can also be achieved by the use of glossaries containing frequently used synthesized sentences that can either be the default ones prepared by the system developers or user-defined ones.

The system thus allows for the user intervention at different levels, being sufficiently modular to easily incorporate future expansions.

## FUTURE DEVELOPMENTS

The integration of the two parts of the EDIFALA project is the next goal to be achieved in the next couple of months in order to allow the test of the full system by children and their therapists in the next school year. Another future development concerns the autonomous access to other means of communication, namely the telephone, without demanding specialized equipment from other non-handicapped users.

## ACKNOWLEDGEMENTS

The EDIFALA project is supported by JNICT, in the scope of the ESTIMULO program. The consortium is formed by CLUL (Centro de Linguística da Universidade de Lisboa), INESC (Instituto de Engenharia de Sistemas e Computadores) and Centro de Paralisia Cerebral Calouste Gulbenkian.

*

References

[1] D.Klatt (1980), "Software for a cascade/parallel formant synthesizer", Journal of the Acoustical Society of America, 67, 971-995.

[2] S. Lazzaretto, and L. Nebbia (1987) - "SCYLA: Speech Compiler for Your Language", Proc. of the European Conf. on Speech Technology, Edimburgh, 2, 381-384.

[3] L. Oliveira, C. Viana and I. Trancoso (1991) - "DIXI - Portuguese Text-to-Speech System", Proc of the European Conf. on Speech Technology, Génova, pp.1239-1242.