

# CONSTRUCTION OF A SYLLABLE-BASED TEXT-TO-SPEECH SYSTEM FOR ROMANIAN

OVIDIU BUZA, GAVRIL TODEREAN,  
JOSZEF DOMOKOS and ARPAD ZSOLT BODO

We present in this article our experience in building a text-to-speech system for Romanian through concatenation method. Main stages of this work were following: voice signal analysis; voice signal segmentation; vocal database construction; text analysis: pre-processing, unit detection, prosody retrieval; unit matching; unit concatenation and speech synthesis. In our approach we consider word syllables as basic units and stress indicating intrasegmental prosody. A special characteristic of current approach is rule-based processing of both speech signal analyse and text analyse stages.

*Key words:* text-to-speech, syllable approach, rule-based processing.

## 1. INTRODUCTION

In the last decades many methods have been developed for generating acoustical parameters requested for a high quality voice synthesis. Researches proved that among methods with best results are those methods which store the real acoustic waveform uttered by a human speaker. These methods achieve voice synthesis through concatenation of acoustic units, so they are called concatenation methods [9],[14].

The authors have worked on this line of attaining a voice synthesis complying with quality parameters of natural, human speech. Our researches led into projecting a voice synthesis method specifically adapted to Romanian language, and also into a working approach for constructing an automated speech synthesis system.

Using syllables as basic units, the projected method is integrated into high quality methods category, based on concatenation. We propose here an original approach based on rules that apply in the most important stages of projecting a speech synthesis system: construction of the vocal database and text processing stage.

In building our text-to-speech system, we have followed two directions (Fig. 1):

A. Vocal database construction flow (off-line process), including: voice signal analysis, speech segmentation and vocal database construction;

B. Text to voice processing flow (on-line process), including: text pre-processing, unit detection, prosody prediction, unit matching, unit concatenation and voice synthesis.