Original Method for Romanian Text-to-Speech Synthesis Based On Syllable Concatenation

Ovidiu Buza, Gavril Toderean, Alina Nica
Technical University of Cluj-Napoca, Romania
Arpad Zsolt Bodo
Siemens PSE Program and System Engineering Ltd.
Budapest, Hungary
Objectives

- Build a rule-driven parser for Romanian capable to separate syllables from text and detect accentuation
  - LEX generator and two sets of phonetic rules
- Develop a TextToSpeech system prototype based on syllable concatenation
The Text Analysis and Synthesis principle

1. Text analysis stage
   - linguistic and phonetic rules -> appropriate units and prosody data

2. Unit matching stage
   - unit matching procedure -> data retrieval from units storage and construction of unit chains

3. Voice synthesis of unit chains
Text Analysis

Detection of linguistic units:

- sentences
- words
- syllables

Punctuation marks
Literal separators
Linguistic Rules
Text Analysis Modules

1. Text
2. Sintactical Rules
   - High Level Analysis
     - Sentences
     - Words
3. Lexical Analysis
   - Syllables
     - Numbers
     - Separators
   - Lexical Rules
4. Phonetic Analysis
   - Phonetic Rules
   - Stress position
Lexical Analyser

1. Extracts characters
2. Clusters chars into basic units
   - Alphabetical chars → Syllables
     m, a, m, a > ma - ma
   - Numerical chars → Numbers
     1, 2, 3 > 123
   - Punctuation marks → Separators
     ! ? , .
Phonetic Analyser

1. Gets the syllables between two separators
   SEP , car , te , SEP
2. Detects stress position
   SEP , car , te , SEP
1. Takes low-level information
   - basic units: syllables
   - basic prosody: stress

2. Constructs high-level structures
   - words
   - sentences
Lexical Analyzer for Syllable Detection

**AIM:**

Detection of **lexical units**: syllables, separators and numbers

**Text** → **Characters**

**Character** → **Digit**
  → **Separator**
  → **Alphabetic**

**Digits** → **Number**

**Alphabetic** → **Syllable**
Syllable LEX Rules

1. Basic Rule Set: general decomposition rules for Romanian

   syllable =\{CONS\}*\{VOC\} \quad \text{‘PA-SĂ-RE’} \quad (R1)

   syllable =\{CONS\}*\{VOC\}{CONS}/\{CONS\} \quad \text{‘CAN-DE-LĂ’} \quad (R2)

   syllable =\{CONS\}*\{VOC\}{CONS}/\{SEP\} \quad \text{‘BU-CU-RÂND’} \quad (R3)

2. Exception Rule Set: exceptions from basic set

   \{CONS\}*eu/\{SEP\} \quad \text{word-ending "eu"} \quad \text{‘ȘE-MI-NEU’}

   \{CONS\}*[ou]ri/\{SEP\} \quad \text{word-ending "ori"} \quad \text{‘CO-CORI’}

   \text{ia}{CONS}/\{CONS\} \quad \text{word-ending "uri"} \quad \text{‘GRU-PURI’}

   \text{ia}{CONS}/\{SEP\} \quad \text{word-beginning "ia"} \quad \text{‘IAR-NA’}

   \{CONS\}*e[çns]*[tli]/\{SEP\} \quad \text{word finished by a soften ‘i’} \quad \text{‘IS-TEȚI’}
1. Collects word phonems: $F_1, F_2, \ldots, F_k$
2. Detects stress position: $S_n, S_{n-1}, \ldots, S_{n-3}$
# Phonetic Rules for Accentuation

**1. General Rule:** Sn-1 syllable is stressed

\[
\text{(CHAR)}^+ / \text{(SEP)} \quad \text{stress} = \text{Sn-1} \quad \text{‘BǍ-IA-TUL’} \quad (R1)
\]

**2. Exception Rule Set:** exceptions from general rule

<table>
<thead>
<tr>
<th>Exception Type</th>
<th>Example</th>
<th>Stress Type</th>
</tr>
</thead>
<tbody>
<tr>
<td>verbs ended with ‘a’ (infinitives or past tense)</td>
<td>‘PI-CU-RA’</td>
<td>(Sn)</td>
</tr>
<tr>
<td>noun plurals</td>
<td>‘SCA-TII’</td>
<td>(Sn)</td>
</tr>
<tr>
<td>articulate or non-articulate nouns</td>
<td>‘PE-RE-TE-LE’</td>
<td>(Sn-2)</td>
</tr>
<tr>
<td>articulate nouns</td>
<td>‘FA-GU-RE-LE’</td>
<td>(Sn-3)</td>
</tr>
<tr>
<td>past tense verbs</td>
<td>‘O-RI-GI-NEA’</td>
<td>(Sn-3)</td>
</tr>
<tr>
<td>plural nouns</td>
<td>‘PLǍ-CEA’</td>
<td>(Sn)</td>
</tr>
<tr>
<td>adjectifs</td>
<td>‘MU-TǍRI’</td>
<td>(Sn)</td>
</tr>
<tr>
<td></td>
<td>‘CǍ-RUNŤI’</td>
<td>(Sn)</td>
</tr>
</tbody>
</table>
Prototype Implementation

Text Analysis Vocal database construction
Unit Matching Unit Concatenation and Synthesis

Text → Lexical analysis
     Syllable detect → Basic units: Syllables

Text → Phonetic analysis
     Accentuation → Prosodic info: Stress

Unit Matching → Vocal Database

Unit Concatenation and Synthesis
• Includes a subset of Romanian language syllables
• Acoustic units were separated from male speech and normalized in pitch and amplitude.
Unit Matching

(a) Syllable is matched in appropriate accentuated form -> acoustic unit directly used for concatenation

(b) Syllable is matched but not the accentuation -> unit is reconstructed from other syllables and phonemes which abide by the necessary accentuation

(c) Syllable is not matched at all -> constructed from other syllables and separate phonemes

<table>
<thead>
<tr>
<th>TEXT</th>
<th>DB</th>
<th>CHOISE</th>
</tr>
</thead>
<tbody>
<tr>
<td>már</td>
<td>már</td>
<td>már</td>
</tr>
<tr>
<td>cán</td>
<td>car</td>
<td>cá + r</td>
</tr>
<tr>
<td>sár</td>
<td>---</td>
<td>sá + r</td>
</tr>
</tbody>
</table>
Results

Text analysis:
• 85% correct syllable detection ⇐ more than 100 rules in BNF grammar
• 75% correct accentuation ⇐ almost 100 rules in BNF grammar
• reduced size of vocal database ⇐ a subset of Romanian language syllables (386 two-character, 139 three-character)

TTS prototype:
• synthesized speech is understandable and conserves the quality of recorded units
• normalization and correct concatenation

Future:
• F0 adaptive correction in concatenation points
• Completion of syllable database
Conclusion

⭐️ A Rule-driven approach for Text-to-Speech conversion using syllable concatenation

⭐️ Two sets of rules:
- detecting word syllables
- determining accentuation in each word

⭐️ Implementation of a Syllable-based Text-to-Speech prototype
Romanian Syllable-Based Text-To-Speech System

Text

Lexical analysis
Syllable detect

Basic units: Syllables

Phonetic analysis
Accentuation

Prosodic info: Stress

Unit Matching

Vocal Database

Unit Concatenation and Synthesis

Technical University of Cluj Romania