TEXT-TO-SPEECH: the Linguistic Perspective

Graz University of Technology, 29th October 2003
Outline

- Speech Synthesis: 2 ways
- TTS: Basic Components
- TTS: Complexity of Analysis
- Linguistic Analysis
  - Lexical Analysis
  - Morphological Analysis
  - Word Context Analysis
  - Phonological & Accent Analysis
- Summing Up
- Resources
Speech Synthesis: 2 ways

- **Text-to-Speech (TTS)**
  - Input: orthographical text
  - Method: conversion to speech of any kind of text
    - General text
    - Document structure
    - Markup
  - Features:
    - Flexible
    - Exposed to error
  - Usage: general speech synthesis purposes

- **Concept-to-Speech (CTS)**
  - Input: linguistic representation
  - Method: conversion to speech of concepts
    - Semantics
    - Pragmatics
    - Discourse knowledge
  - Features:
    - Specific
    - Reliable (esp. prosody)
  - Usage: dialogue systems, machine translation, etc.
TTS: Basic Components

- Each component may consist of different modules
TTS: Complexity of Analysis

- “Bei der Wahl am 12.3.1998 gewann Tony Blair ca. 52% der Wählerstimmen.”

- st should be realized as [ʃt] and not as [st] (see “Erstimpfung”)
- Tony Blair should be recognised as foreign name entity
- 52%, ca., 12.3.1998 should be treated as regular words
- Punctuation (.) has here 3 different meanings:
  - 12.3.1998 part of date
  - ca. abbreviation
  - Wählerstimmen. sentence boundary

- **Text-to-Speech conversion is NOT a trivial task:**
  Linguistic Knowledge is necessary
Linguistic Analysis: Relevant Components

- Lexical Analysis

- Morphological Analysis (Derivation, Composita)

- Word Context Analysis
  - Syntactic Agreement
  - Syntactic Phrases/Sentences
  - Prosodic Phrases/Sentences
  - Sentence mode

- Phonological & Accent Analysis (Out-of-Lexicon words)
Lexical Analysis

- **Lexicon**: a dictionary. It contains, for each entry, relevant informations such as:
  - **Part of speech (POS)**: name, verb, adjective etc.
  - **Phonetic transcription**
  - **Relevant grammatical categories**: number, gender, etc.

- Different kinds of lexica:
  - **Whole form**: for each lexeme, all possible word forms are listed (e.g. “gehe, gehst, geht, ...”)
  - **Word stem**: for each lexeme, just the basic form and a general paradigm to be followed are listed (e.g. “geh-” + regular verbal flexion).
Lexical Analysis: Special Items

- **Word-stem lexica**: special *word lists* are created for:
  - Non-flectional words
  - Geographical nouns, proper nouns, foreign words and other special categories
  - Abbreviations, acronyms etc.
  - Numbers. They are associated with more or less complex linguistic models
Lexical Analysis: Lexical Entries

- Lexical entries: two examples

**Festival 1.4.0**

(“walkers” n ((( w oo ) 1) (( k @ z ) 0)) )
(“present” v ((( p r e ) 0) (( z @ n t ) 1)) )
(“monument” n ((( m o ) 1) (( n y u ) 0) (( m @ n t ) 0)) )

(“lives” n ((( l a i v z ) 1)) )
(“lives” v ((( l i v z ) 1)) )

**CELEX**

<table>
<thead>
<tr>
<th>Headword</th>
<th>PhonStrsCPA</th>
<th>MorphStructure</th>
<th>MorphC</th>
<th>Cla</th>
<th>Freq</th>
</tr>
</thead>
<tbody>
<tr>
<td>celebrant</td>
<td>‘sE.1.I_br@nt</td>
<td>((celebrate),(ant))</td>
<td>Vx</td>
<td>N</td>
<td>6</td>
</tr>
<tr>
<td>celebrated</td>
<td>‘sE.1.I_bre/.tId</td>
<td>((celebrated))</td>
<td>V</td>
<td>A</td>
<td>158</td>
</tr>
<tr>
<td>celebration</td>
<td>‘sE.1.I_bre/.Sn</td>
<td>((celebrate),(ion))</td>
<td>Vx</td>
<td>N</td>
<td>221</td>
</tr>
<tr>
<td>celibacy</td>
<td>‘<a href="mailto:sE.1.I_b@.sI">sE.1.I_b@.sI</a></td>
<td>((celibate),(cy))</td>
<td>Ax</td>
<td>N</td>
<td>13</td>
</tr>
<tr>
<td>celibate</td>
<td>‘sE.1.I_b@t</td>
<td>((celibate))</td>
<td>H</td>
<td>N</td>
<td>2</td>
</tr>
<tr>
<td>cell</td>
<td>‘sE1</td>
<td>(cell)</td>
<td>N</td>
<td>N</td>
<td>1216</td>
</tr>
<tr>
<td>cellar</td>
<td>‘sE.10@r*</td>
<td>(cellar)</td>
<td>N</td>
<td>N</td>
<td>225</td>
</tr>
<tr>
<td>cellarage</td>
<td>‘<a href="mailto:sE.10@.rIj">sE.10@.rIj</a>/</td>
<td>((cellar),(age))</td>
<td>Nx</td>
<td>N</td>
<td>0</td>
</tr>
<tr>
<td>cellist</td>
<td>‘T/E.1.Iist</td>
<td>((cello),(ist))</td>
<td>N</td>
<td>N</td>
<td>36</td>
</tr>
<tr>
<td>cello</td>
<td>‘T/E.10/</td>
<td>(cello)</td>
<td>V</td>
<td>N</td>
<td>0</td>
</tr>
</tbody>
</table>

START GOTO ZOOM HIDE COUNT PRINT SAVE QUERY
Morphological Analysis

- Relevant for **inflectional** languages (e.g. German) as well as **polysynthetic and agglutinative** languages

- **2 different processes:**
  - **Derivation:** word stem + adfixes
    
    Example: schlag / vor-ge-schlag-en
  
  - **Composition:** 2 or more word (stem)s are joined
    
    Example: Kopf + Hörer / Kopfhörer
Morphological Analysis: Composition

-Languages like German are very **productive in word composition** - no lexicon could include every possible realisation: morphological analysis is needed.

- Morphological analysis is **problematic**: more than one (more or less plausible) analysis is possible.
  - Example: Währerstimmen
  - Solution:
    - Statistical methods: use of language corpora
    - Consideration of the syntactic context
Word Context Analysis: Syntax

- **Syntactic agreement:** checks for grammatical congruence between linked words (e.g. Det + Adj + Nom)


- **Syntactic phrases/sentences:** phrase and sentence boundaries are individuated through punctuation and syntactic structure. The last one is usually realized in form of hierarchical representations
Word Context Analysis: Syntactic Analysis
Word Context Analysis: Prosody

- **Prosodic phrases/sentences:**
syntactical and lexical information is used to determine prosodic boundaries in order to build intonation and prominence models

- **Sentence modes:**
retrieved prosodic information will determine the sentence mode, basing on intonation and prominence
Word Context Analysis: Sentence Mode
Phonological & Accent Analysis

- In texts we oft encounter out-of-lexicon words, for which there's no pronunciation information available.

- For such words we need to build phonological rules, which associate certain phonemes to certain graphemes.

- In most languages word accent has also to be determined.

- In many languages morphological analysis of unknown words may help to find the right pronunciation.
Phonological & Accent Analysis: Phonological Rules

• For languages with a more or less 1:1 grapheme-phoneme relationship simple conversion rules may be sufficient

Example: spanish
(Festival 1.4.3)

```
(a [ a ] = a ) ( [ h ] = h ) (# 0.0 0.250)
(a [ e ] = e ) ( [ j ] = x ) (a 0.0 0.090)
(a [ i ] = i ) ( [ k ] = k ) (e 0.0 0.090)
(a [ o ] = o ) ( [ l l ] # = l ) (i 0.0 0.080)
(a [ u ] = u ) ( [ l l ] = ll ) (o 0.0 0.090)
(a [ " ' " a ] = a1 ) ( [ l ] = l ) (u 0.0 0.080)
(a [ " ' " e ] = e1 ) ( [ m ] = m ) (b 0.0 0.065)
(a [ " ' " i ] = i1 ) ([ ~ n ] = ny ) (ch 0.0 0.135)
(a [ " ' " o ] = o1 ) ( [ n ] = n ) (d 0.0 0.060)
(a [ " ' " u ] = u1 ) ( [ p ] = p ) (f 0.0 0.100)
(b [ b ] = b ) ( [ q u ] = k ) (g 0.0 0.080)
(b [ v ] = b ) ( [ r r ] = rr ) (g 0.0 0.100)
(c [ " ' " EI = th ] ( [ r ] = r ) (j 0.0 0.100)
(c [ c ] EI = th ) (LNS [ r ] = rr ) (k 0.0 0.100)
(c [ c ] = ch ) ( [ r ] = r ) (l 0.0 0.080)
(c [ c ] = k ) ( [ s ] BDGLMN = th ) (m 0.0 0.070)
(c [ d ] = d ) ( [ s ] = s ) (n 0.0 0.080)
(c [ f ] = f ) ( # [ s ] C = e s ) (ny 0.0 0.110)
(c [ g ] " ' " EI = x ) ( [ t ] = t ) (p 0.0 0.100)
(c [ g ] EI = x ) ( [ w ] = u ) (p 0.0 0.030)
(c [ g u ] " ' " EI = g ) ( [ x ] = k s ) (rr 0.0 0.080)
(c [ g u ] EI = g ) (AEO [ y ] = i ) (s 0.0 0.110)
(c [ g ] = g ) ( # [ y ] # = i ) (t 0.0 0.085)
(c [ h u e ] = u e ) ( [ y ] = ll ) (th 0.0 0.100)
(c [ h ie ] = i e ) ( [ z ] = th ) (x 0.0 0.130)
```
Summing Up

- TTS for most natural languages needs quite complex **linguistic analysis** to perform a good job: **linguistic models** help improving system's performance.

- **Linguistic components** of a standard TTS system include a **lexicon**, **morphological** and **context** rules (syntactic and prosodic) as well as **phonological** rules.

- Each linguistic **component** is **strictly correlated** with the others: they all concur to build a **complete linguistic representation** for prosodic and phonetic synthesis.
Resources


  http://cslu.cse.ogi.edu/HLTsurvey/ch5node2.html