Contents

- Challenges
- Currently Used Technologies
  - NLP
  - DSP
- Currently Available
  - Non-commercial
  - Commercial
- Applications
- Conclusion:
  A Bright (Corpus-based) Future?

You idiot!
Speech synthesis needs a corpus-based solution

I tell you:
Speech synthesis will be solved by rules!

Von Kempelen’s talking machine (1791)

Mouth
Nostrils
Small bellows
'S' pipe
'S' lever
'Sh' pipe
'Sh' lever

(J.S. Liénard, LIMSI)

Von Kempelen’s machine, revisited (2002)

http://www.eng.kagawa-u.ac.jp/~sawada/

Just for fun (2002)
1. The Formant Synthesizer, (1964)

2. Diphone concatenation (1977)
2. Diphone concatenation (1977)

- Diphone
- Database
- Prosody
- Modification

- Smooth joints

2.1 The LPC synthesizer (1977)

- Frame-by-frame (typ.: N=30ms/L=10ms)
- Duration, F0 modification: easy
- Smoothing: on PARCORs or LARS or LSPs

2.1 LPC Synthesis (cont’d)
2.2 PSOLA (1988)

Limsi (Paris, 1992)

2.3 MBROLA (1993)

Based on the same OLA principle as PSOLA, but using edited diphones: Constant phases, Constant F0
- Similar overall quality as PSOLA
- Same computational load
- Completely automatic!

⇒ can be used to create lots of compatible synthesizers


How to get the best sequence of units for a given utterance? **Viterbi search**

- **Target cost?**
  How to predict which units will sound as they would naturally connected? (should be perceptual)

- **Concatenation cost?**
  How to predict which sequences of units will sound naturally connected? (should be perceptual)
3. Automatic unit selection

**Very Large corpus**

![Diagram of unit selection]

- Concatenation cost \(cc(i-1,i)\)
- \(=0\) in case of successive units

**3. Automatic unit selection**

**Which one to choose? Engineering considerations**

<table>
<thead>
<tr>
<th></th>
<th>Rule-based</th>
<th>Diphone Concatenation</th>
<th>NUU</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Database preparation</strong></td>
<td>not fully automatic</td>
<td>Automatic, easy</td>
<td>Automatic (\Rightarrow) MBROLA project</td>
</tr>
<tr>
<td><strong>Database size</strong></td>
<td>30kb</td>
<td>100kb</td>
<td>5Mb</td>
</tr>
<tr>
<td><strong>Computational load at synthesis time</strong></td>
<td>70 operations per sample</td>
<td>70 operations per sample</td>
<td>7 operations per sample</td>
</tr>
</tbody>
</table>

→ **PSOLA/MBROLA for user products**
→ **NUU for telecom applications**


**Engineering issues**
Currently 100s of Mb; Real time on fast computers

Towards passing the TURING test?

Copyright (c)2002 Faculté Polytechnique de Mons - T. Dutoit