Development of a BOSS unit selection module for tone languages
Exemplification in Ibibio

Arne Bachmann & Stefan Breuer
University of Bonn

6th ISCA Workshop on Speech Synthesis
23rd August 2007
Ibibio corpus sample / synthesis

/ana ekop dindem odo ojOhO ke ifuuro uwem/

Original

Synthesized
Ibibio
Nigeria
Ibibio

State map
Ibibio Tones

- High $H$
- Downstepped high $D$ (!H)
- Low $L$
- Rising $R$ (LH)
- Falling $F$ (HL)
## Ibibio

### Tonal contrasts

<table>
<thead>
<tr>
<th>óbóŋ</th>
<th>mosquito</th>
</tr>
</thead>
<tbody>
<tr>
<td>óbòŋ</td>
<td>cane</td>
</tr>
<tr>
<td>sé</td>
<td>look</td>
</tr>
<tr>
<td>áà-sèè-hè</td>
<td>one who looks</td>
</tr>
<tr>
<td>áà-!ké-séé-hé</td>
<td>one who looked</td>
</tr>
<tr>
<td>áà-!dî-sé</td>
<td>one who will look</td>
</tr>
</tbody>
</table>
Ibibio

Start-up effect

#owo ibarake ikaN ikeene daNa udVikikOt OJOON#

<table>
<thead>
<tr>
<th>Time (s)</th>
<th>Pitch (Hz)</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.95</td>
<td>0</td>
</tr>
<tr>
<td>2.67</td>
<td>100</td>
</tr>
</tbody>
</table>

\begin{tabular}{|c|c|c|c|c|c|c|c|c|c|c|c|c|}
\hline
\texttt{owo \#} & \texttt{ibarake} & \texttt{ikaN} & \texttt{ikeene} & \texttt{daNa} & \texttt{udVikikOt} & \texttt{OJOON} \\
\hline
\texttt{P} & \texttt{H} & \texttt{H} & \texttt{H} & \texttt{H} & \texttt{H} & \texttt{H} & \texttt{H} & \texttt{H} & \texttt{L} & \texttt{H} & \texttt{P} \\
\hline
\texttt{$p$} & \texttt{owo} & \texttt{i} & \texttt{ba} & \texttt{ra} & \texttt{ki} & \texttt{kaN} & \texttt{i} & \texttt{kee} & \texttt{ne} & \texttt{$p$} \\
\hline
\texttt{P} & \texttt{HH} & \texttt{HHHHHH} & \texttt{HLH} & \texttt{P} \\
\hline
\texttt{$p$} & \texttt{owo\%} & \texttt{ibarakikaN} & \texttt{ikeene} & \texttt{$p$} \\
\hline
\end{tabular}
Ibibio Downdrift

Pitch (Hz)

<table>
<thead>
<tr>
<th>Time (s)</th>
<th>0</th>
<th>100</th>
<th>200</th>
<th>300</th>
<th>400</th>
<th>500</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.03</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2.76</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

dOktO desi widisin

<table>
<thead>
<tr>
<th>d</th>
<th>O</th>
<th>k</th>
<th>t</th>
<th>a#</th>
<th>d</th>
<th>e</th>
<th>s</th>
<th>i#</th>
<th>widisin</th>
</tr>
</thead>
<tbody>
<tr>
<td>P</td>
<td>H</td>
<td>L</td>
<td>H</td>
<td>L</td>
<td>H</td>
<td>H</td>
<td>L</td>
<td>P</td>
<td></td>
</tr>
</tbody>
</table>

$pdOktadesiwidisin$ $pdOktadesiwidisin$

A. Bachmann & S. Breuer BOSS-IBB
BOSS
Bonn Open Synthesis System

- BOSS source code:
  sourceforge.net/projects/boss-synth

- BOSS website:
  www.ikp.uni-bonn.de/boss/
Goals

▶ Adaptability
▶ Extensibility
▶ Simplicity
▶ Universality
Goals

- Adaptability
- Extensibility
- Simplicity
- Universality
Goals

- Adaptability
- Extensibility
- Simplicity
- Universality
Goals

- Adaptability
- Extensibility
- Simplicity
- Universality
Stylization of F0
Linear model
Stylization of F0

Representing syllable contours by quartic polynomials

<table>
<thead>
<tr>
<th>Time (s)</th>
<th>Pitch (Hz)</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.518073</td>
<td></td>
</tr>
<tr>
<td>100</td>
<td></td>
</tr>
<tr>
<td>200</td>
<td></td>
</tr>
<tr>
<td>300</td>
<td></td>
</tr>
<tr>
<td>400</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Pitch (Hz)</th>
</tr>
</thead>
<tbody>
<tr>
<td>100</td>
</tr>
<tr>
<td>200</td>
</tr>
<tr>
<td>300</td>
</tr>
<tr>
<td>400</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Order 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>ka</td>
</tr>
<tr>
<td>ba</td>
</tr>
<tr>
<td>Nu</td>
</tr>
<tr>
<td>duak</td>
</tr>
<tr>
<td>e</td>
</tr>
<tr>
<td>di</td>
</tr>
<tr>
<td>ja</td>
</tr>
<tr>
<td>hu</td>
</tr>
<tr>
<td>tom</td>
</tr>
<tr>
<td>$p$</td>
</tr>
<tr>
<td>u</td>
</tr>
<tr>
<td>mIN</td>
</tr>
<tr>
<td>N</td>
</tr>
<tr>
<td>kpO</td>
</tr>
<tr>
<td>$p$</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Time (s)</th>
<th>3.67064</th>
</tr>
</thead>
<tbody>
<tr>
<td>Order 4</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>H</th>
</tr>
</thead>
<tbody>
<tr>
<td>H</td>
</tr>
<tr>
<td>H</td>
</tr>
<tr>
<td>H</td>
</tr>
<tr>
<td>L</td>
</tr>
<tr>
<td>L</td>
</tr>
<tr>
<td>P</td>
</tr>
</tbody>
</table>

A. Bachmann & S. Breuer

BOSS-IBB
Stylization of F0

Basic value data representation

\[ f(x) = 2x^4 - 3x^2 + 0.6 \]

- \( c_4 = 2 \)
- \( c_3 = 0 \)
- \( c_2 = -3 \)
- \( c_1 = 0 \)
- \( c_0 = 0.6 \)

- \( bv_0 = -0.4 \)
- \( bv_1 = -0.025 \)
- \( bv_2 = 0.6 \)
- \( bv_3 = -0.025 \)
- \( bv_4 = -0.4 \)
Vector quantization
Vector quantization

Code vectors 32 - 47 out of 64
Prediction

Requirements
Prediction

Most important parameters of a test tree

- Position of the syllable in the phrase
- Position of the word in the phrase
- Number of phones in the syllable
- Syllable structure (e.g. C, V, N)
- 4th left neighbouring tone
- Remaining
Prediction
Most important parameters of a test tree

- Position of the syllable in the phrase
- Position of the word in the phrase
- Number of phones in the syllable
- Syllable structure (e.g. C, V, N)
- 4\textsuperscript{th} left neighbouring tone
- Remaining

- 62.96 %
- 
- 67.82 %
- 71.89 %
Prediction

Example from CART

((sylphrase < 15.7)
 ((sylphrase < 8)
   ((sylphrase < 3.4)
     ((ltone2 is P)
      (ltone1 is P)
      ((phonessyl < 1.3)
       (50)
       (53)
       (50)
     (f < 0.1)
     ((wordphrase < 1.2)
      (45)
      (catsylword is i)
      (45)
      (50)
Prediction

Tone template classification results

On 10 different test sets (10 % held-out data):

- Correct code vectors (out of 64): 38.55 % - 59.04 %
- Correct code vector classes (out of 16 in second codebook): 42.77 % - 68.07 %
Unit selection
Conclusion
Conclusion

To do

- Get hold of a bigger corpus
- Perform subjective listening tests
- Adapt the module to another tone language
- Try the approach for an accent language
Conclusion

To do

- Get hold of a bigger corpus
- Perform subjective listening tests
  - Adapt the module to another tone language
  - Try the approach for an accent language
Conclusion

To do

- Get hold of a bigger corpus
- Perform subjective listening tests
- Adapt the module to another tone language
- Try the approach for an accent language
To do

- Get hold of a bigger corpus
- Perform subjective listening tests
- Adapt the module to another tone language
- Try the approach for an accent language
Thank you!