Transforming F0 Contours

Citation for published version:

Link:
Link to publication record in Edinburgh Research Explorer

Document Version:
Peer reviewed version

Published In:
Eurospeech 2003 - Interspeech 2003

General rights
Copyright for the publications made accessible via the Edinburgh Research Explorer is retained by the author(s) and / or other copyright owners and it is a condition of accessing these publications that users recognise and abide by the legal requirements associated with these rights.

Take down policy
The University of Edinburgh has made every reasonable effort to ensure that Edinburgh Research Explorer content complies with UK legislation. If you believe that the public display of this file breaches copyright please contact openaccess@ed.ac.uk providing details, and we will remove access to the work immediately and investigate your claim.
To transform the F0 contour of some speech from a source speaker, such that listeners believe it to have been uttered by some target speaker.

Goal

Applications

- voice transformation
  - also need a method for transforming voice quality - see poster by King & Gillett in session PWeBe
- speech synthesis
  - as a way of adapting existing intonation models (trained on one speaker) to a new speaker, without having to annotate much more data

Standard method

Usual method of normalising F0 is to use this mapping:

\[ M_X(x) = \left(\frac{x - \mu_x}{\sigma_x}\right) \times \sigma_y + \mu_y \]

New method

A nonlinear mapping, \( M_{PL} \), composed of piecewise linear sections between F, L, H and S

Transformed contours

Histograms

Audio examples

Perceptual experiment

25 subjects presented with speech from target speaker and speech with transformed F0 contours (“imitator speech”) in XABX format. Asked to judge which of A or B was most like X. A and B varied between: actual target F0, normalised F0 (standard method), transformed F0 (new method).

Speaker pairs classified as “similar” (\( S_{\text{sim}} \)) or “different” (\( S_{\text{diff}} \)).

Results

<table>
<thead>
<tr>
<th>Preference</th>
<th>Mean (%)</th>
<th>Std. Dev.</th>
<th>( \alpha )</th>
<th>t</th>
</tr>
</thead>
<tbody>
<tr>
<td>Preference for ( M_{PL} ) over ( M_F ) for ( S_{\text{sim}} )</td>
<td>67</td>
<td>10</td>
<td>( &lt; 1 \times 10^{-5} )</td>
<td>-8.71</td>
</tr>
<tr>
<td>Preference for ( M_{PL} ) over ( M_F ) for ( S_{\text{diff}} )</td>
<td>54</td>
<td>8</td>
<td>( \sim U )</td>
<td>-2.49</td>
</tr>
<tr>
<td>Preference for target over mapped contours</td>
<td>73</td>
<td>9</td>
<td>( &lt; 1 \times 10^{-5} )</td>
<td>-13.8</td>
</tr>
</tbody>
</table>

After Patterson

- sentence-initial high (S)
- non-initial accent peaks (H)
- post-accent valleys (L)
- sentence-final low (F)

What next?

Apply the method to full voice transformation or speech synthesis.

See also...

- at this conference:
  - poster by Gillett & King in session PWeBe (voice quality transformation)
- www.cstr.ed.ac.uk
  - for latest progress on voice transformation and speech synthesis
- www.camelaudio.com
  - for musical instrument transformation and morphing