Smoothing of micromelodic effects in fundamental frequency contours

Raphael Winkelmann
raphael@phonetik.uni-muenchen.de

Institute of Phonetics and Speech Processing, Ludwig-Maximilians-Universität München

Hauptseminar Prosodie und Intonation II at Venice International University
2009-04-16
table of contents

introduction
  definition

outline
  motivation
  previous studies

our studies
  data used
    preprocessing (Kiel spont. Corpus)

approach
  Kiel Corpus: data evaluation
  Kiel Corpus: filtering
  Kiel Corpus: PCA, factor analysis
  Kiel Corpus: example
definition

- **prosody:**
  - prosody is the rhythm, stress and intonation of connected speech
- **macroprosody:**
  - intentional manipulation of duration, fundamental frequency and amplitude
- **microprosody:**
  - non-intentional segmental influences on duration, fundamental frequency and amplitude
definition

- microprosody can be subdivided into:
  - intrinsic pitch \((IF0)\)
    - intrinsic pitch is positively correlated with vowel height: \([i] > [e] > [a]\)
  - cointrinsic pitch \((CF0)\)
    - cointrinsic pitch studies indicate a higher F0 in the vicinity of voiceless as opposed to voiced obstruents
motivation

The goal is to separate micro- from makro-prosody and find factors in fundamental-frequency-contours that are responsible for IF0 and CF0. This method has to work without any knowledge about segmental information. After finding the factors that are responsible for IF0/CF0 these factors can be set to zero to then calculate the resulting “pure” macro-prosody.

tool requirements/goals

- developing a tool to filter micromelody effects while retaining macro-intonation
- pure bottom up
  ▶ tool has to work on any voiced segment independent of segmental information
- integrate tool in EMU Speech Database System
previous studies

Previous studies have shown that there are global differences in IF0 and CF0.

- Global comparisons of IF0 have shown that IF0 is positively correlated with vowel-height.
- Global comparisons of CF0 have shown that F0 is higher in the vicinity of voiceless obstruents as opposed to voiced obstruents.

previous applications

- With previous smoothing applications (e.g. Momel) it is unclear whether the smoothing only effects the IF0/CF0 and if it fully extracts all of the micro-prosodic effects.
A short introduction to Momel (MELodic MOdelisation)

- iterative approximation through parabolas
- windowing of F0-contour
- in every window a parabel $p$ gets iteratively fitted to the original F0 segment so that $p$ fits the contour best in an LMS sense.
- through every new fitting original-F0 values that deviate from the last fitted parabel by a certain measure are ignored: removal of micro-prosody and measurement errors
- the iteration procedure ends, when no more original-F0 values are found that deviate from the last fitted parabel by a certain measure
- extrema of the parabolas are used as anker-points for a spline function to smooth the F0-contour
data used/preprocessing

- Kiel spont. Corpus
  - (quasi-)spontaneous speech
  - turn-based file segmentation
  - hand-labeled
  - not controlled for macro- or microprosody

preprocessing (Kiel spont. Corpus)

1. data alignment
   - alignment of fundamental frequency information with hand-labeled transcription and syllable separators

2. calculate derived data
   - for further studies a few derived data values were calculated (e.g. average segmental F0, baseline values, time-normalized values, ...)

R. Winkelmann
Smoothing of micromelodic effects in fundamental frequency contours
Kiel Corpus: data evaluation

**global studies of the Kiel Corpus data:**

- mean over whole vowel contour (zeros omitted) in CVC sequences
- one-way ANOVA, factor: tongue height
- Tukey-Kramer posthoc
Kiel Corpus: data evaluation

**Figure: global analysis IF0**

R. Winkelmann

Smoothing of micromelodic effects in fundamental frequency contours
Kiel Corpus: data evaluation

IF0 results:

- ANOVA: means significantly different (p=0)
- posthoc: significant mean differences (alpha=0.05): low < mid, mid < high
Kiel Corpus: data evaluation

**CF0 studies**

- **motivation:**
  - determine the length of the scope of the vowel on intrinsic f0
  - determine the length of the scope of neighboring consonants on co-intrinsic f0

- **method**
  - dividing vowels in CVC sequences into 4 segments of equal size (overlap: 0.5*window length)
  - mean for each segment (zeros omitted)
  - three-way ANOVA for each segment, factors: tongue height, voice(C1), voice(C2)
  - Tukey-Kramer posthoc
Kiel Corpus: data evaluation

plots

(a) mean-F0 in CVCs  (b) mean-F0 C1 voiced vs. unvoiced  (c) mean-F0 C2 voiced vs. unvoiced

Figure: F0 contours with and without micro-melodic effects
Kiel Corpus: data evaluation

• results
  ▶ ANOVA, posthoc:
    ▶ significant mean differences in dependence of factor tongue height in all segments (p=0; same as for global means)
    ▶ voice(C1): **unvoiced** greater than **voiced** for segments 1–3 (p=[0.0000 0.0002 0.0306 0.1739])
    ▶ voice(C2): **unvoiced** greater than **voiced** for segments 1–4 (p=[0.0406 0.0107 0.0009 0.0000])

• conclusions
  ▶ intrinsic f0 related to vowel height in whole vowel
  ▶ co-intrinsic f0 of C1 related to voice in 1st-3rd quarter of vowel
  ▶ co-intrinsic f0 of C2 related to voice in whole vowel
Kiel Corpus: filtering

- idea:
  1. separation of F0-contour into macro- and micro-prosodic components
  2. in application: analysis-by-synthesis + smoothing due to setting the respective components to a value of zero

- separation: uniqueness of the approximation has to be given (as is the case with: polyfit, dct)

- the search for microprosodic-components
  ▶ required: segmental information (hence segmented corpus)
  ▶ question: are there components that only differentiate significantly according to segmental differences?
Kiel Corpus: PCA, factor analysis

• motivation:
  ▶ find principal components/ factors related to intrinsic or co-intrinsic f0

• method:
  ▶ orthogonalise f0 contour by PCA or factor analysis
  ▶ case FA: oblique promax rotation
  ▶ n-way ANOVA for all PCs or factors: vowel height, voice(C1), voice(C2), etc.
  ▶ identify PCs whose means differ significantly with the factors
  ▶ do the same after baseline subtraction

• results:
  ▶ no interpretable results, large differences for the majority of components and factors
Kiel Corpus: PCA, factor analysis

plots: factor analysis
Kiel Corpus: example

time normalised F0 contour for [iː] segment from Kiel Corpus
Kiel Corpus: example

perform dct to get the following values:

<table>
<thead>
<tr>
<th>number</th>
<th>DCTvalue</th>
<th>number</th>
<th>DCTvalue</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>446.4617</td>
<td>6</td>
<td>-0.5944</td>
</tr>
<tr>
<td>2</td>
<td>9.4029</td>
<td>7</td>
<td>-0.7546</td>
</tr>
<tr>
<td>3</td>
<td>-5.2135</td>
<td>8</td>
<td>0.4906</td>
</tr>
<tr>
<td>4</td>
<td>6.5519</td>
<td>9</td>
<td>-0.3732</td>
</tr>
<tr>
<td>5</td>
<td>-0.4011</td>
<td>10</td>
<td>0.2295</td>
</tr>
</tbody>
</table>
Kiel Corpus: example

perform dct to get the following values:

<table>
<thead>
<tr>
<th>number</th>
<th>DCTvalue</th>
<th>number</th>
<th>DCTvalue</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>446.4617</td>
<td>6</td>
<td>-0.5944</td>
</tr>
<tr>
<td>2</td>
<td>9.4029</td>
<td>7</td>
<td>-0.7546</td>
</tr>
<tr>
<td>3</td>
<td>-5.2135</td>
<td>8</td>
<td>0.4906</td>
</tr>
<tr>
<td>4</td>
<td><strong>6.5519</strong></td>
<td>9</td>
<td>-0.3732</td>
</tr>
<tr>
<td>5</td>
<td>-0.4011</td>
<td>10</td>
<td>0.2295</td>
</tr>
</tbody>
</table>

If we hypothesise that the fourth coefficient is responsible for the micro-melodic effects. We can then set its value to zero and resynthesize the new smoothed F0-contour.
Kiel Corpus: example

the smoothed contour:

![Graph showing smoothed FO-contour](image)
the end