Accent groups, character contours, and prosodic phrasing

Uwe Reichel[1] & Katalin Mády[2]

[1]Institute of Phonetics and Speech Processing, University of Munich
[2]Research Institue for Linguistics, Hungarian Academy of Sciences

3rd June 2014

Joint work with Štefan Beňuš, Nitra, and Alexandra Markó, ELTE, Budapest.

Speech Prosody 2014, Dublin

Mády, Reichel & Beňuš: Accentual phrases in Slovak and Hungarian.
Reichel, Markó & Mády: Parameterization and automatic labeling of Hungarian intonation.

OLINCO 2014, Olomouc
Beňuš, Reichel & Mády: Modelling accentual phrase intonation in Slovak and Hungarian.

Contents

- The accent group as a **prosodic phrase**
  - ...as indicated by its deviation from the intonational phrase,
  - ...as indicated by its general intonational patterns.
- The accent group as a domain for **character contours**.

Contents

- **The accent group as a prosodic phrase**
  - **. . . as indicated by its deviation from the intonational phrase,**
  - . . . as indicated by its general intonational patterns.
- The accent group as a domain for character contours.

Prominence in prosodic words

Lexical stress: the position of the prominent syllable within the word is variable and is marked by pitch, duration, intensity and/or segmental features (e.g. English, German, Italian).

Postlexical stress: emphasis on the sentence level is realized on the stressed syllable of the prominent word. Languages with fixed stress (French, Bengali, Hungarian) have postlexical, but not lexical stress.

Not all languages have lexical or postlexical word-level stress (e.g. Korean, Tamil).

Prominence in intonational phrases

Intonational phrase (**IP**): nuclear accent (the most heavy accent of the prosodic phrase), usually phrase-initial or phrase-final.

The intonational phrase is a universal prosodic phrase present in all languages.

IP-final nuclear accent: right-headed prosody (Germanic & Romance languages), IP-initial nuclear accent: left-headed prosody (Korean, Hungarian).

Prominence in accent groups

- Accentual phrase (**AP**): phrasal stress at the beginning or end of the phrase. It is often found in languages with fixed postlexical stress, but seldom in languages with lexical stress (e.g. Farsi).
- In an AP, pitch contours show a regular pattern → rising, falling, or rising-falling (Jun & Fletcher, 2014).
- Postlexical pitch can mark prominence (English), demarcate a prosodic boundary (Japanese) or both (French).

Accent groups in Hungarian and Slovak

- Slovak and Hungarian: stress fixed to the left-most syllable of a prosodic word.
- Hungarian: stress is word-initial, Slovak: stress can be shifted to the preposition preceding the lexical word, e.g. *HOry* 'hills', *DO hory* 'to the woods'.
- Hungarian: pitch accent = left boundary of a character contour (Varga 2002) or phonological phrase (Hunyadi 2002).
- Do Hungarian and Slovak make use of accentual phrases?

Speech data

50 Slovak and Hungarian spontaneous utterances

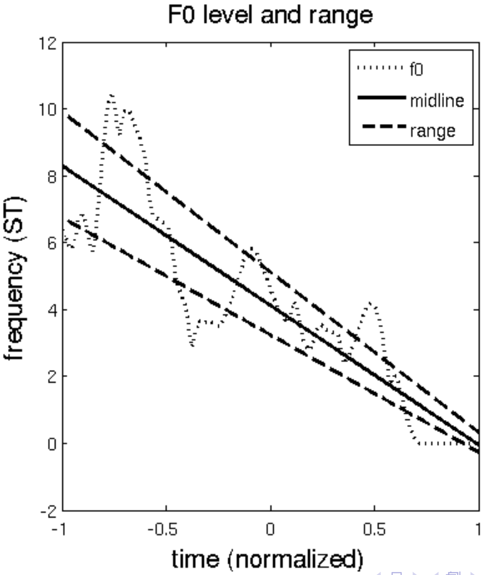
- forming a single intonational phrase (**IP**),
- with at least two pitch accents (manual labelling),
- with a low IP-final boundary tone,
- 5 utterances of 10 speakers in each language.

F0 level and range stylization

For IP and AGs:

- F0 medians calculated within each window (length 200 ms, window shift 10 ms)
    - of the values $<$ 10th percentile for the baseline,
    - of the values $>$ 90th percentile for the topline, and
    - of all values for the midline.
- Fit line to each of the three median sequences.
- **Advantages:**
    - robustness,
    - no fuzzy peak and valley detection required.

F0 level and range stylization



Measuring AG–IP deviation and AG prominence

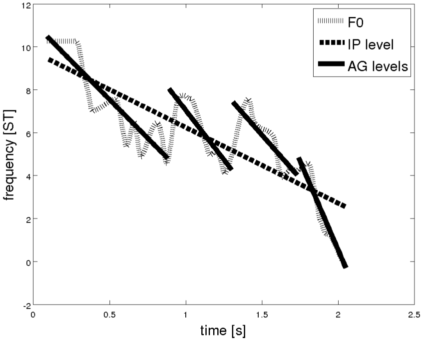
Motivation

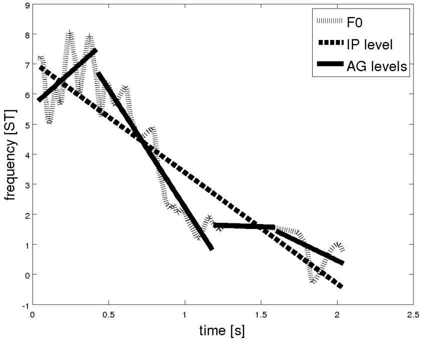
acoustic correlates for the presence of accentual phrases are

- prominent F0 movements (large AG range),
- considerable local level deviations between the AG and the IP.

Measuring AG–IP deviation and AG prominence

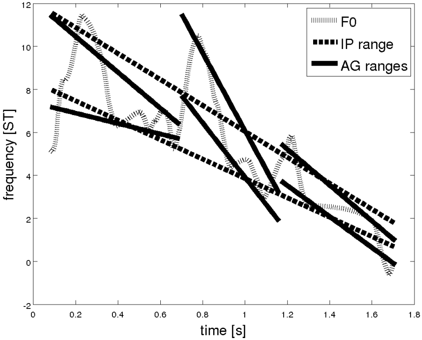
AG and IP level

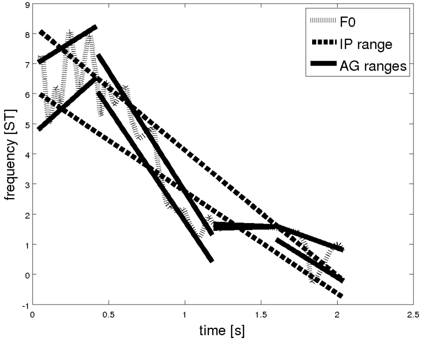




Measuring AG–IP deviation and AG prominence

AG and IP range





Measuring AG–IP deviation and AG prominence

Features

Level-related features	
mlSlope	slope of AG midline
mlSlopeDiff	absolute slope difference of AG and IP midline
mlRms	mean squared deviation of the AG line from the corresponding section of the IP line
mlInitDiff	(initial F0 value of AG midline) − (corresponding IP midline value)
mlFinDiff	(final F0 value of AG midline) − (corresponding IP midline value)
Range-related feature	
rangeRms	RMS between AG base- and topline

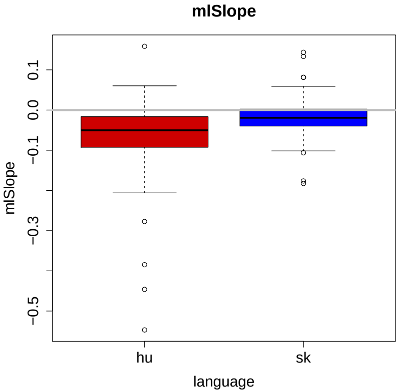
Statistical analysis

t-tests:

- the mean of each data set was compared to 0,
- the means of Hungarian and Slovak data were compared to each other.

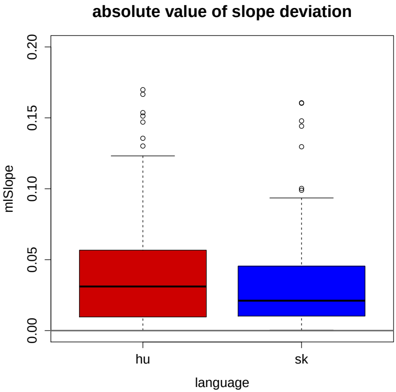
A significant difference from $mean = 0$ refers to a steeper rise or fall of the AG pitch compared to the IP pitch.

Midline slope



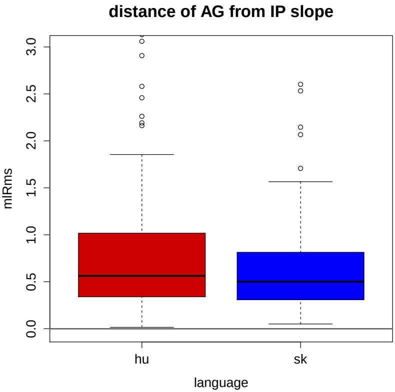
Hungarian and Slovak differ significantly from 0 and from each other ($p < 0.0001$). Slopes are more negative in Hungarian.

Absolute slope difference of midlines



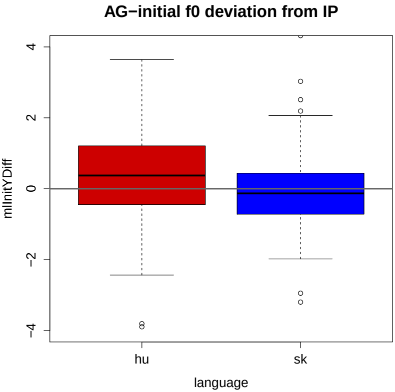
Languages differ significantly from 0 and from each other. AG and IP midlines differ more in Hungarian.

Root mean squared deviation between AG and IP midline



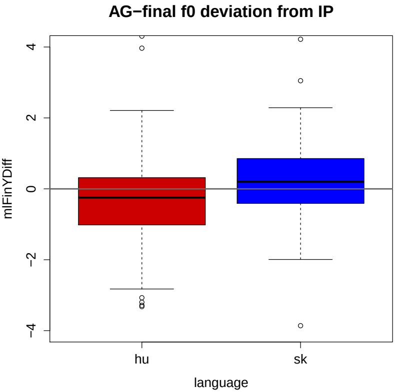
Difference between languages: $p = 0.06$. Tendency for Hungarian to have a larger deviation between AG and IP midlines.

AG-inital deviation from IP



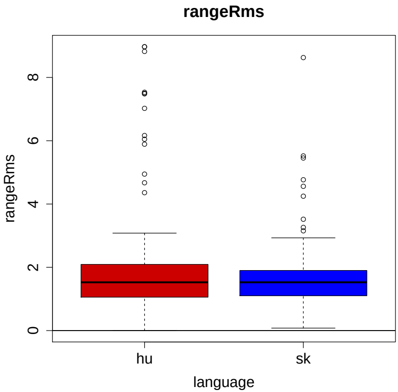
Hungarian differs significantly from *mean* $= 0$ and from Slovak. Slovak values are normally distributed around 0.

AG-final deviation from IP



Both languages differ significantly from *mean* $= 0$ and from each other. Phrase-final pitch is lower in Hungarian than the corresponding IP value, but not in Slovak.

Root mean square between AG base- and topline



Both Hungarian and Slovak differ significantly from 0. No significant difference between the two languages.

Discussion

- AG contours were typically falling ones in both languages. The tendency was stronger in Hungarian.
- Initial f0 of Hungarian AGs was higher than the corresponding f0 of the IP declination line, whereas the final f0 was lower $\Rightarrow$ AGs are better modelled as a separate prosodic unit.
- Linear slopes in Slovak are rather flat without a clear orientation.
- APs show a falling tendency in Hungarian, no falling or rising tendency in Slovak.

Problem: rising-falling pattern cannot be captured by linear (1st order) stylization. Next step: quadratic (2nd order) stylization.

Contents

- **The accent group as a prosodic phrase**
  - ...as indicated by its deviation from the intonational phrase,
  - **...as indicated by its general intonational patterns.**
- The accent group as a domain for character contours.

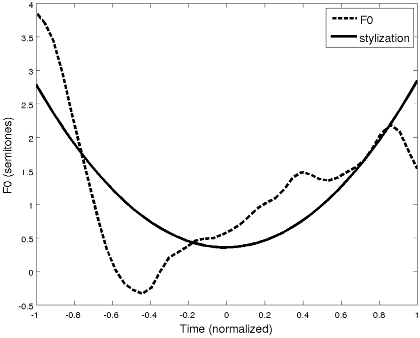
Measuring general AG intonational patterns

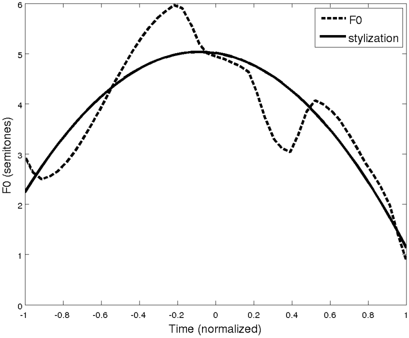
Method

- Within each AG a 2nd order polynomial was fitted to the F0 contour: $f0 = c_0 + c_1 \cdot t + c_2 \cdot t^2$.
- Curvature of the F0 contour quantified in terms of the quadratic polynomial coefficient c_2.
- c_2 negative: **concave (rising–falling)**.
- c_2 positive: **convex (falling-rising)**.

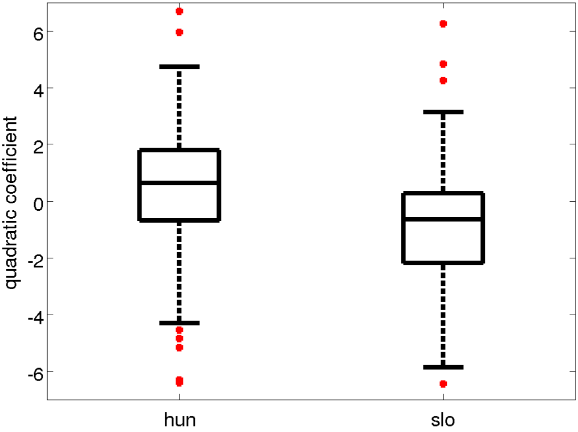
Measuring general AG intonation patterns

Parabolic AG intonation patterns





Results



Discussion

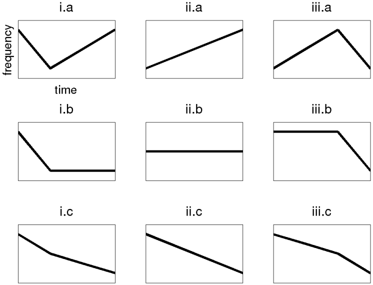
- **Slovak:** consistent tendency for **rising-falling** F0 patterns.
- **Hungarian:** no such tendency.
- **Support:**
  - mean quadratic coefficient c_2 differs significantly (2-tailed t-test, $p < 0.001$): -0.99 for Slovak, 0.59 for Hungarian.
  - Slovak c_2 **negative** $\longrightarrow$ **concave** contour.
  - **Higher absolute value** of Slovak c_2 $\longrightarrow$ curvature more **pronounced**.
  - For Slovak the maximum of the polynomial is significantly **more often contained within the time interval** of the accent group (χ^2 test, $p < 0.001$) $\longrightarrow$ **less homogeneous shapes in Hungarian** (parabolic and linear).

Contents

- The accent group as a prosodic phrase
  - ... as indicated by its deviation from the intonational phrase,
  - ... as indicated by its general intonational patterns.
- **The accent group as a domain for character contours.**

Character contours

- **Character contour** (**CC**, Varga 2002) "*discrete, meaningful speech melody*" with a "*characteristic shape*".
- **Domain:** accent group (AG).
- **Shapes:**



- **meanings:** *self-contained, forward-pointing, yes-no interrogative.*

Aims, data, processing steps

Aims

- find appropriate linear stylization to map F0 contours within AGs to CCs,
- predict CC labels from the stylization parameters for automatic CC annotation.

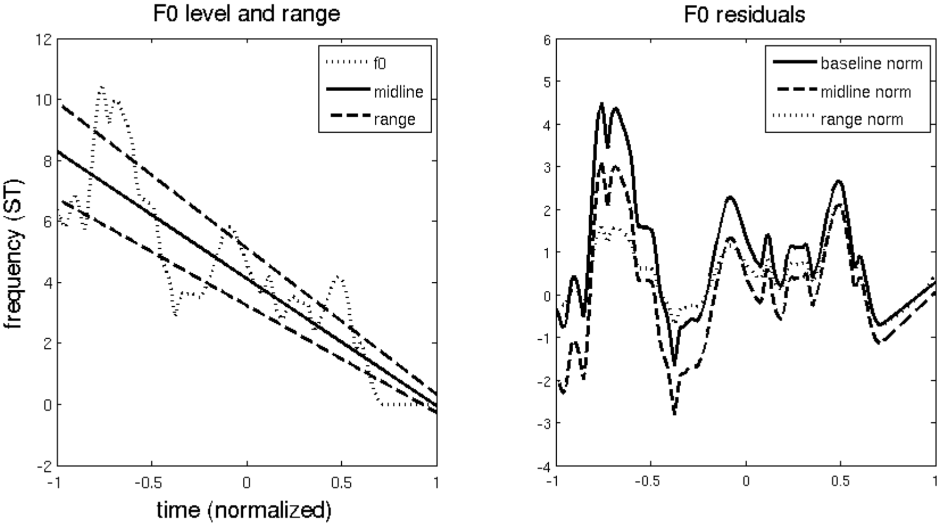
Data

- 50 Hungarian spontaneous dialog speech utterances by 10 speakers.
- Manual segmentation and CC labeling of AGs (140).
- Only CCs i.b, i.c, ii.b, ii.c contained.

Processing steps

1. F0 normalization to reduce register effects.
2. F0 stylization $\longrightarrow$ CC representation.
3. Classifier training for CC prediction from the stylization parameters.

Normalization



Normalization

F0 residuals

- **level residuals** by base-, resp. midline subtraction,
- **range residual** by normalizing F0 to the range between base- and topline.

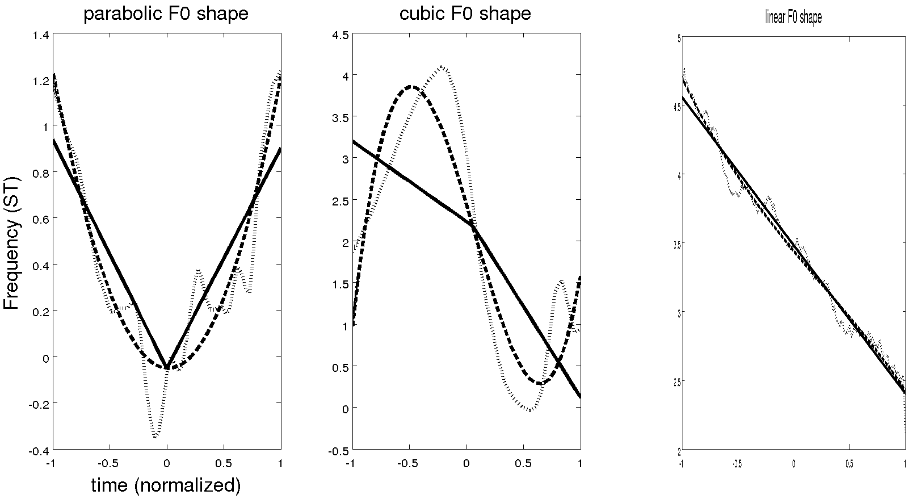
Parameterization

- of F0 and its 3 residuals,
- CC represented by 1 or 2 lines.

Joint determination:

- third order polynomial fit,
- localization of extreme values and turning point within AG,
- both absent $\longrightarrow$ single line,
- extreme value only $\longrightarrow$ parabolic shape, joint at extreme value,
- turning point present $\longrightarrow$ cubic shape, joint at turning point.

Parameterization



CC classification

Features

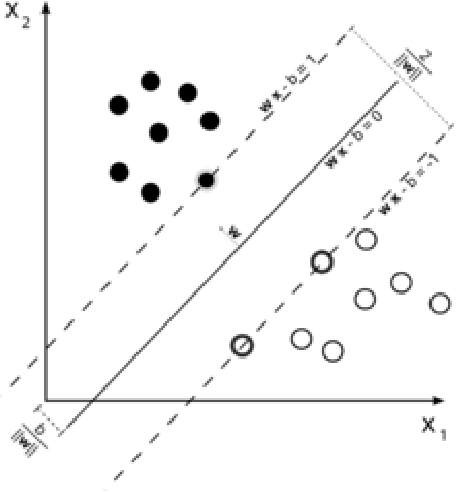
slope of first line
slope of second line
time position of joint

Classifiers

- classification trees (CART) and support vector machines (SVM) with a linear Kernel function,
- applied separately for parabolic and cubic contours.

CC classification

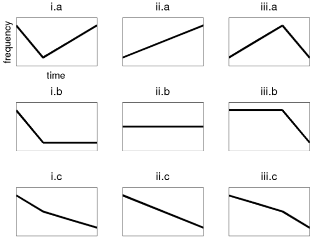
Support vector machine

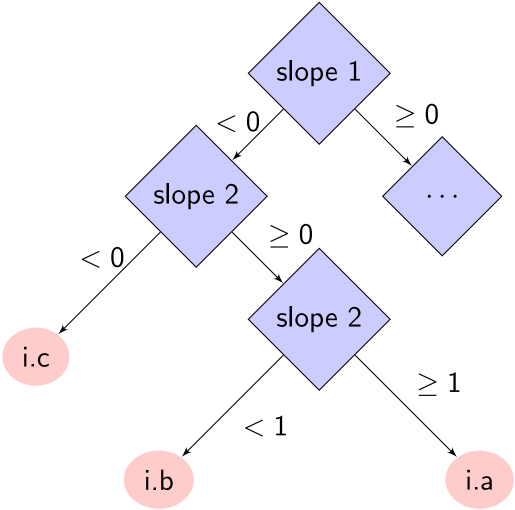


(Wikipedia)

CC classification

Classification tree





CC classification

Results

- Leaving-one-out validation

	CART		SVM	
	parabolic	cubic	parabolic	cubic
none	69.23	29.91	69.23	34.18
baseline residual	76.92	30.77	69.23	36.75
midline residual	**84.62**	**43.59**	69.23	35.04
range residual	62.50	35.96	**75.00**	**41.22**

Automatic label correction

- **correction** = revealing the most probable hidden label sequence $C = c_1 \dots c_n$ (**expert annotation**) that underlies the **classifier output** $O = o_1 \dots o_n$

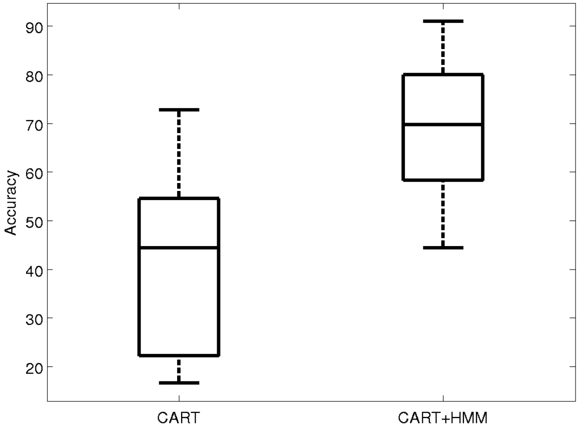
$$
\begin{aligned}
\hat{C} &= \arg\max_{C}[P(C|O)] \\
&= \arg\max_{C}[P(C)P(O|C)] \\
&\approx \arg\max_{c_1 \dots c_n}[\prod_{i=1}^{n} P(c_i|c\text{-context}_i)P(o_i|c_i)]
\end{aligned}
$$

- **transition probabilities** $P(c_i|c\text{-context}_i)$: linear interpolated uni- and bigram model
- counts smoothed by Good-Turing discounting
- expert annotation $\hat{C}$ retrieved by Viterbi algorithm

Automatic label correction

Results

- 10-fold cross-validation



Discussion

Application of the CC framework on spontaneous speech

- Non-uniform CC distribution (see also Varga, 2002).
- CC parameters overlap to a high extent across different contour classes
  $\longrightarrow$ difficult to relate a realized contour to its underlying prototype,
  $\longrightarrow$ non-trivial relation between auditory expert judgments and the signal.
- Gain from F0 normalization, but not in a uniform way for all classifiers.

Discussion

Gain from HMM postprocessing of the classifier output

- Significant performance improvement.
- Postprocessing accounts for the left label context in terms of transition probabilities $\longrightarrow$ context plays a role for label assignment.
- Systematic correspondences between the classification output and reference labels can be used for automatic label correction.

Summary

1. Hungarian AG declination deviates clearly and consistently from IP declination.
2. Slovak AGs show a consistent rising–falling intonation pattern.
3. Stylization and prediction of character contours in AGs is a challenging task.

Acknowledgements

Hungarian Scientific Research Fund (PD 101050)
Momentum program of the Hungarian Academy of Sciences