Automatic Mandarin Prosody Boundary Detecting Based on Tone Nucleus Features and DNN Model

Ju Lin, Yanlu Xie, Wei Zhang, Jinsong Zhang

Speech Acquisition and Intelligent Technology (SAIT) Lab, Beijing Language and Culture University

1. Overview

Problem:

Automatic detection of Mandarin prosody boundary.

Method:

Detecting prosody boundary based on tone nucleus features and Deep Neural Network (DNN) model.

> This method firstly calculated the boundary-related parameters by applying the tone nucleus features.

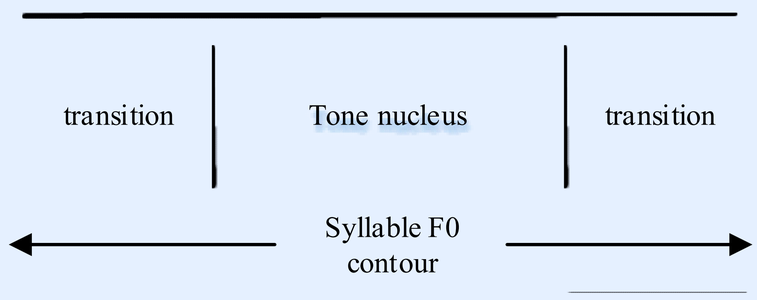
> Then, the parameters were modeled by DNN.

Conclusion:

Considering interaction between adjacent tones, the method of using tone nucleus model to extract boundary-related parameters is effective.

2.Tone Nucleus Model

2.1 Tone nucleus



1)Tone nucleus denotes the F0 target and serves as the primary acoustic cue for tone perception.
2)Articulatory transitions are the F0 variations occurring as the transitions to or from the tonal targets.

2.2 Tone nucleus and prosody boundary

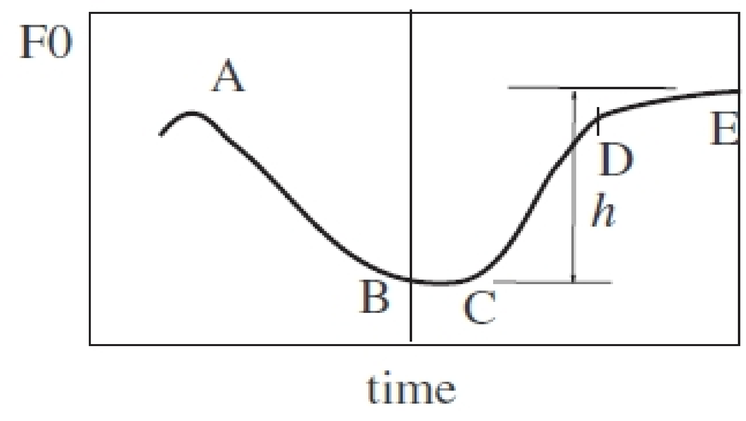
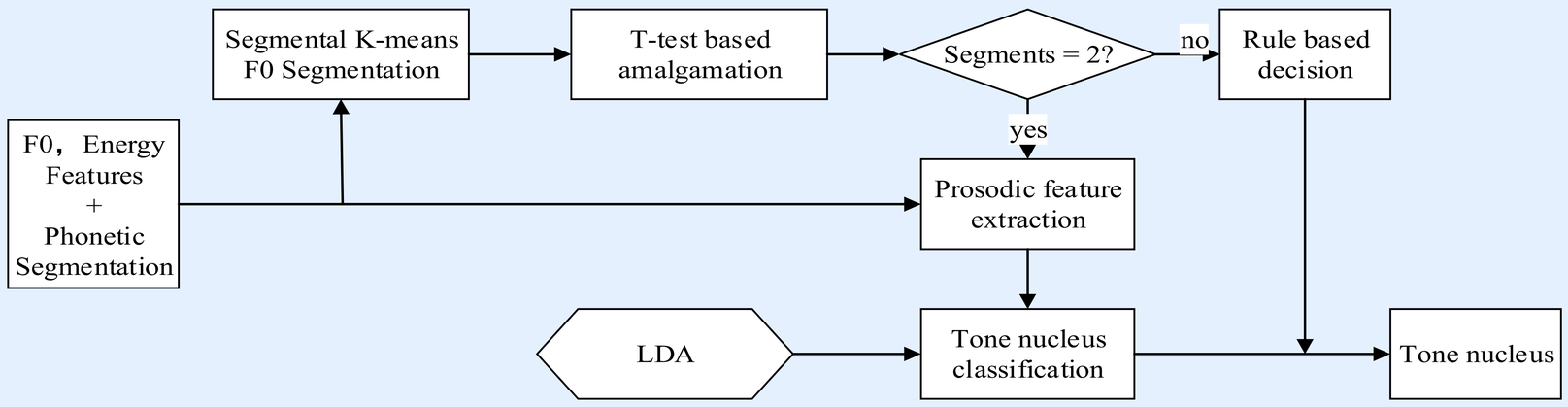
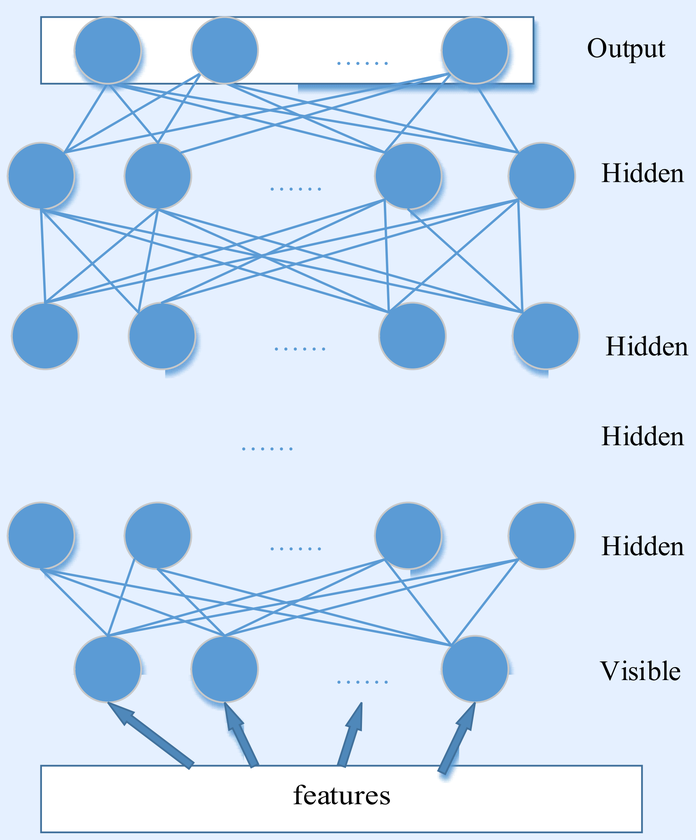


Illustration of an F0 contour of adjacent “Tone 4 Tone 1”, with a prosodic phrase boundary between them

2.3 Tone nucleus detection



3. DNN Model



4. Acoustic features

Duration

1) SilD_f: The duration of silence after the current syllable.
2) SylDur: The duration of the current syllable.
3) SylDurRatio_foll: The ratio of the duration of the current syllable and that of its following syllable.
4) SylDurRatio_pre: The ratio of the duration of the current syllable and that of its preceding syllable.

Pitch

1) Fitting the pitch contour of current tone nucleus with f(x)=a+bx+cx2 and {a, b, c} was used to represent the pitch contour feature.
2) PMax: The maximum F0 value of the current tone nucleus.
3) PMin: The minimum F0 value of the current tone nucleus.
4) PRange: The F0 range of the current tone nucleus.
5) PMean: The mean F0 value of the current tone nucleus.
6) PMRatio: The ratio of current syllable’s PMean and the following syllable’s PMean.
7) PRatio: The ratio of the F0 of the last point in current tone nucleus and that of the following syllable.
8) Delta_Max: The difference between the current syllable’s PMax and the following syllable’s PMax.
9) Delta_Min: The difference between the current syllable’s PMin and the following syllable’s PMin.

Energy

1) EgMax: The maximum energy value in the current tone nucleus.
2) EgMin: The minimum energy value in the current tone nucleus.
3) EgRange: The energy range of the current tone nucleus.
4) EgMean: The mean energy value of the current tone nucleus.
5) EgRatio: The ratio of current syllable’s EgMean and the following syllable’s EgMean.

5. Experiments and Results

5.1 Corpus: ASCCD

Total	B0	B1	B2	B3	B4
87628	61518	16334	8442	7449	4072
100%	62.9%	16.7%	8.6%	7.6%	4.2%

5.2 Experiment Setup

1) A 20-unit input layer.
2) 3 hidden layers, each layer consists of 512 sigmoid units.
3) An output layer consists of 5 softmax units.

5.2 Experiment results

System	B0	B1	B2	B3	B4
SY-CART	90.9	48.6	50.9	80.8	61.7
SY-DNN	96.1	21.2	44.1	75.3	74.6
TN-DNN	96.1	22.6	47.6	83.6	75.1

System	SY-CART	SY-DNN	TN-DNN
Accuracy	78.3	76.5	77.34

SY: syllables, TN: tone nucleus

6. Conclusion

This paper proposed a method of automatic prosody boundary detecting based on acoustic features of tone nucleus and DNN model, the experimental results demonstrate that the method is effective.