Automatic Detection of Rhythmic Patterns in Native and L2 Speech: Chinese, Japanese, and Japanese L2 Chinese

Shuju Shi, Yanlu Xie, Xiaoli Feng, Jinsong Zhang

Speech Acquisition and Intelligent Tech. lab.,

Beijing Language and Culture University, Beijing 100083, P.R. China

1. Overview

Aim:

To explore possible contribution of speech rhythm to the perception of foreign accent and to realize automatic detection.

Method:

Building classification models based on acoustic measures of speech rhythm; Conducting perceptual experiment using low-pass filtered stimuli.

Results:

High detection rate of the classification models and high consistency between the detection and perceptual results.

2. Rhythm

Rhythm (isochrony):

perception of temporal regularities in speech (stress timing, syllable timing and mora timing)

Languages in this study:

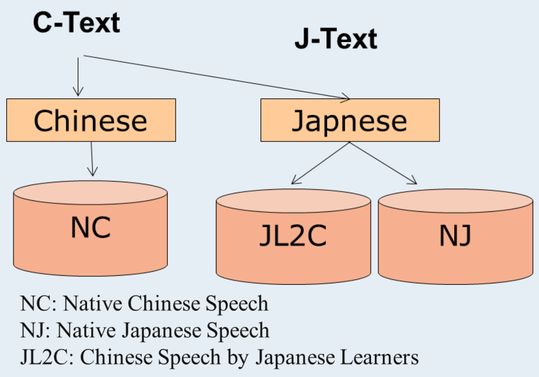
> native Chinese —> syllable-timing

> native Japanese —> mora-timing

> Chinese by Japanese L2—> ?

3. Corpus and features

Corpus Description



NC: Native Chinese Speech
NJ: Native Japanese Speech
JL2C: Chinese Speech by Japanese Learners

- Speakers:
  - 6 Chinese (F)
  - 6 Japanese (F)
- Text:
  - 301 Chinese conversation
  - Its Japanese translation
  - Total: 3048 utterances
- Segmentation
  - Force Alignment :HTK
  - Manual Check: 12 phonetics graduates

Acoustic features:

>Interval-based measures

Name	Description
ΔC	Standard deviation of consonantal intervals.
VarcoC	Standard deviation of consonantal intervals divided by mean and multiplies 100.
ΔV	Standard deviation of vocalic intervals
VarcoV	Standard deviation of vocalic intervals divided by mean and multiplies 100.
rPVI-C	Mean of the difference between successive consonantal intervals.
nPVI-C	Normalized PVI. Mean difference divided by mean of successive consonantal intervals.
rPVI-V	Mean of the difference between successive vocalic intervals.
nPVI-V	Normalized PVI. Mean difference divided by mean of successive vocalic intervals.
%V	Percentage of vocalic duration in an utterance.

>Amplitude-based measures: Fourier transform was applied to the envelope of the speech. Two metrics based on the spectrum are :

>>Spectral Band Power ratio (SBPr): calculated by defining relatively low and high frequency bands, then computing the ratio of the power in the lower band to the higher band.

>>Power-spectral Centroid (centroid): computed by summing all of the frequencies multiplied by their associated spectral power and then dividing by the sum of all spectral power.

4.Experiment

Classification Model:

>SVM was trained using LIBSVM with kernel function RBF, and was optimized using mesh optimization method.

> MLP was trained using sigmoid perceptron, one hidden layer and eleven nodes. The optimization process was based on gradient descent algorithm.

Perceptual Experiment:

> To limit the influence of segment and semantic meaning on listeners, the stimuli are low-pass filtered withcut-off frequency being 500 HZ.

>Five graduate students participated in the perceptual experiment

>The experiment was done using Praat and the participants were asked to identify the language type after they hear a sentence.

5. Result

Acoustic analysis of the features:

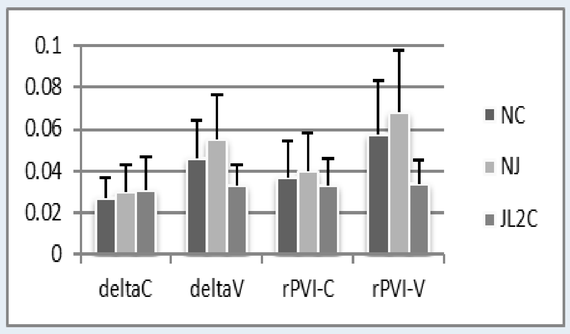


Figure 1: *Mean of deltaC, deltaV, rPVI_C and rPVI_V.*

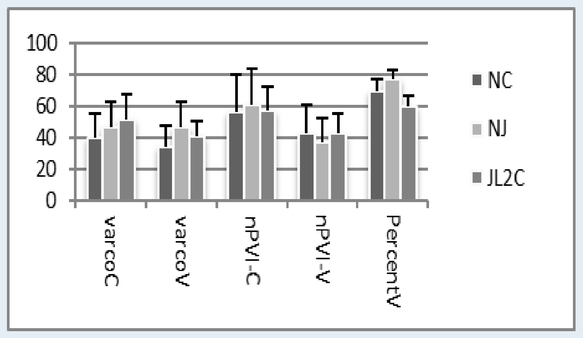
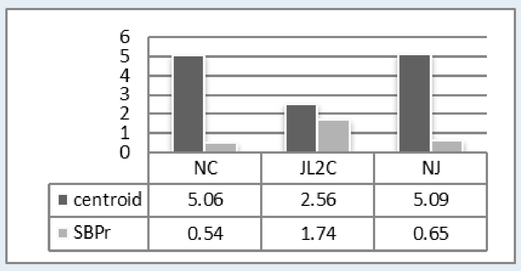


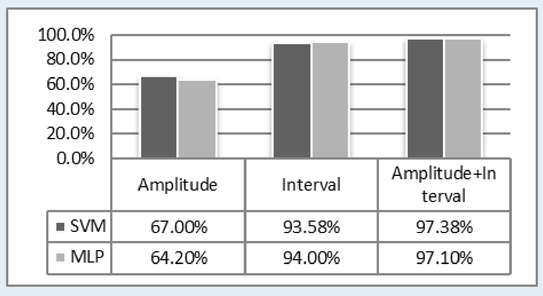
Figure 2: *Mean of PercentV, varcoC, varcoV, nPVI-C and nPVI-V.*



	NC	JL2C	NJ
centroid	5.06	2.56	5.09
SBPr	0.54	1.74	0.65

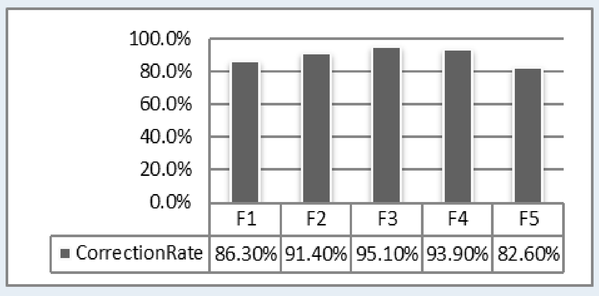
Figure 3: *Mean value of centroid and SBPr for each corpus*

Classification and perceptual results:



	Amplitude	Interval	Amplitude+Interval
SVM	67.00%	93.58%	97.38%
MLP	64.20%	94.00%	97.10%

Figure 4: *Classification Result of Statistical Models.*



	F1	F2	F3	F4	F5
CorrectionRate	86.30%	91.40%	95.10%	93.90%	82.60%

Figure 5: *Detection Result of the Participants.*

6. Conclusion

>NC, NJ and JL2C are different in both types of measures and this difference is statistical significant in most measures;

>High correction rate of the statistical models showed that measures in this study are effective in characterizing rhythmic difference;

>High detection rate by the participants showed the objective measures characterized by the measures are also subjectively perceptible.