



Abstract

The SpeakGoodChinese.org (SGC) system aids beginning students of Mandarin Chinese with practicing tone pronunciation. Students pronounce a word spelled in pinyin notation and receive feedback from our system on their production of the tones. The system uses synthetic reference tones produced from the pinyin notation. First results indicate:

- < 5% rejection rate for six words read by three reference speakers
- ≈ 15% rejection rate for reference speakers using a test application
- ≈ 15% *false accept* and *false reject* errors on six shadowed words
- Students appreciate an automatic application for practicing tones

Introduction

In the Netherlands, and elsewhere, mastering the production and recognition of tones is a major stumbling block in learning Mandarin Chinese. Direct interaction with a highly proficient speaker, mostly the teacher, is needed to practice tone pronunciation.

As a consequence:

- Classes must be kept small to allow for ample student-teacher interaction
- Speaking and listening proficiency improves very slowly
- High drop-out rates of demotivated students
- Speaking is neglected in favor of writing

SpeakGoodChinese.org is an automatic aid for practicing Mandarin tones. Features:

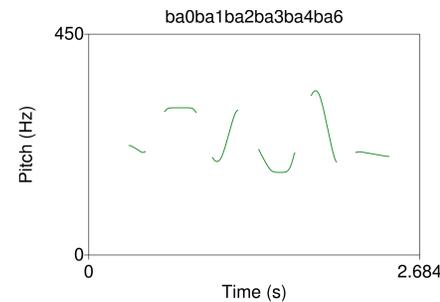
- Automatic Tone Recognition (implemented in *Praat*)
- Graphical Tone Presentation
- A written analysis of tone pronunciation.
- Hummed or pre-recorded examples of tone combinations
- Replaying recorded student pronunciation
- All mono- and bisyllabic words can be entered or read from precompiled word lists

Tone Recognition

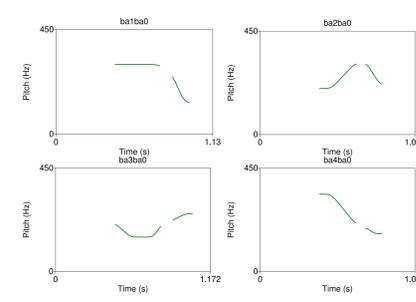
Simplest recognition set-up:

- Dynamic Time Warping (DTW) on pitch
- Require Citation style Clear speech
- Synthetic Tone Models for recognition
- Garbage model for L1 intonation (pseudo-tone 6)
- Mono- and Bisyllabic words only
- Pinyin-to-Pitch & Student Pitch level
- Reject speech with pitch too low, -3 st, or movements too narrow, -3 st
- Coarticulation/Assimilation rules, eg, neutral tone
- Post DTW rules for assimilation, eg, 2 ≈ 3 in context
- Bias recognizer towards accepting

Tone Models



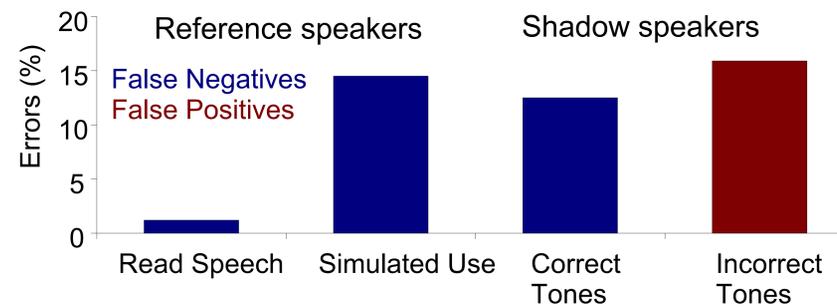
Neutral Tone assimilation



Mandarin Tone	Pitch	Duration	Student Errors
1	High, Flat	Long	Not high enough
2	Mid, Rising over 1	Shorter	Rise not high enough
3	Mid, Low, Rising (optional)	Long	Not low enough
4	High, Falling	Short	Fall not low enough
0 (neutral)	Mid*, no emphasis	Very short -	

*The neutral tone assimilates completely with the preceding tone

Recognition Results: False rejects and accepts



For *Read Speech* and *Simulated Use* from reference speakers, all rejections were counted. For *Shadowed speech*, rejections on *low pitch* and *narrow bandwidth* were not counted as false negatives.

Test Speech Materials

Correct Tones: Read Speech 3 Reference speakers read aloud 6 words: *cha2, dian4hua4, duo1shao3, gong1zuo4, jie2hun1, shi2jian1*, 83 tokens.

Correct Tones: Simulated Use 5 Reference speakers, Tcl/Tk RAD interface, free word choice, mislabeled or truncated items were removed, 358 tokens

Shadowed Speech 8 Speakers, shadowed prerecorded student utterances, both Correct, 160, and Incorrect, 320, tones, six words (see above)

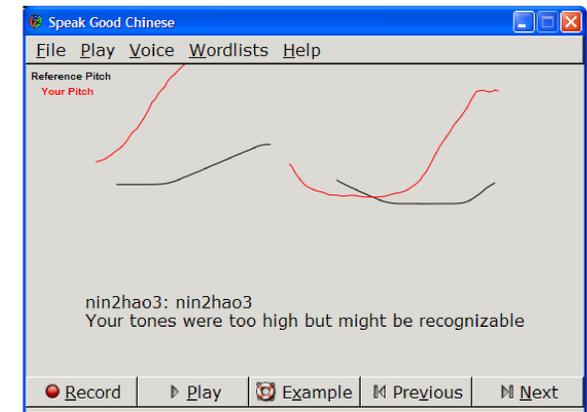
5 Reference Speakers • 3 Native Mandarin Speakers, 2F/1M

- 2 Advanced Mandarin Speakers (L1 is Dutch), 1F/1M

8 Shadow Speakers • 3 Reference Speakers 2F/1M

- 5 Low Proficient Students, 1F/3M

Application

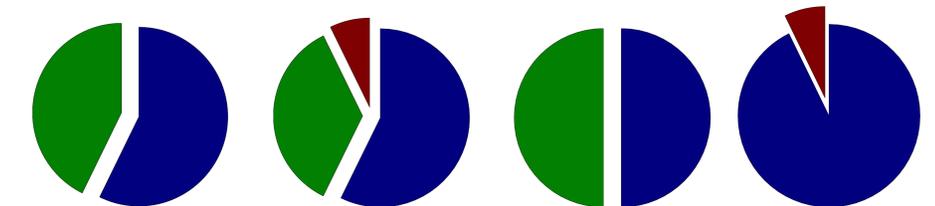


A simpler GUI with direct pinyin entry is also available. One student (F 13y) used the final application for 7 session, 1531 words total. Automatically determined error rate: 28% overall; 39% pre- and 24% post-practice ($p < 0.002$, X^2)

Student Questionnaire

14 Students tried a fully working Tcl/Tk RAD GUI application. 2 Students had used another CALL program before. Subjects were asked about the usefulness of (or grade from 1-10):

SGC application Synth. tone examples Graphic displays Will you use SGC†*



Legend: Not useful/No - Useful - Very useful/Yes

* One subject couldn't hear the tones clearly

† One subject preferred to practice with family members

Discussion

- For simple practice tasks, simple speech recognition (ASR) suffices
- Auditory, graphic, and verbal feedback should be combined
- For Mandarin tones, recognition from synthesis works
- Difficult to catch tone assimilation, students will have to learn that later
- Students can already work with a very basic interface
- Prerecorded audio examples might induce unsupervised learning of phonemic contrasts

Conclusions

- Simple ASR using DTW on synthetic tones works for Mandarin teaching
- 85% correct tone recognition in bisyllabic words is available
- Students are willing to work with a simple, Tone-only practice application
- Auditory examples and graphical feedback are appreciated by students