SpeakGoodChinese: Learn to speak the tones of Mandarin Chinese

D.J.M. Weenink and R.J.J.H. van Son

Institute of Phonetic Sciences
ACLC
University of Amsterdam

Guangqin Chen, Zongyan Chen, Stefan de Konink, Eveline van Hagen, Dennis Vierkant

`http://www.SpeakGoodChinese.org/`

Introduction

Mandarin Chinese is the official administrative language of China
Studying Chinese is popular in the West (and elsewhere)

Problems Teaching Madarin

- Mandarin Chinese is a tone language
- Every syllable in a word has one of 4 (5) tones which determines the meaning of the word
- Using the wrong tone makes a word incomprehensible (cf, English *bad* and *bat*, Dutch *boot* and *bot*)
- Mastering the production and recognition of tones is a major stumbling block in learning Mandarin Chinese
- Direct interaction with a highly proficient speaker, usually the teacher, is needed to practise tone pronunciation

Introduction

Mandarin Chinese is the official administrative language of China
Studying Chinese is popular in the West (and elsewhere)

Problems Teaching Madarin

- Mandarin Chinese is a tone language
- Every syllable in a word has one of 4 (5) tones which determines the meaning of the word
- Using the wrong tone makes a word incomprehensible (cf, English *bad* and *bat*, Dutch *boot* and *bot*)
- Mastering the production and recognition of tones is a major stumbling block in learning Mandarin Chinese
- Direct interaction with a highly proficient speaker, usually the teacher, is needed to practise tone pronunciation

Introduction

Mandarin Chinese is the official administrative language of China
Studying Chinese is popular in the West (and elsewhere)

Problems Teaching Madarin

- Mandarin Chinese is a tone language
- Every syllable in a word has one of 4 (5) tones which determines the meaning of the word
- Using the wrong tone makes a word incomprehensible (cf, English *bad* and *bat*, Dutch *boot* and *bot*)
- Mastering the production and recognition of tones is a major stumbling block in learning Mandarin Chinese
- Direct interaction with a highly proficient speaker, usually the teacher, is needed to practise tone pronunciation

Introduction

Mandarin Chinese is the official administrative language of China
Studying Chinese is popular in the West (and elsewhere)

Problems Teaching Madarin

- Mandarin Chinese is a tone language
- Every syllable in a word has one of 4 (5) tones which determines the meaning of the word
- Using the wrong tone makes a word incomprehensible (cf, English *bad* and *bat*, Dutch *boot* and *bot*)
- Mastering the production and recognition of tones is a major stumbling block in learning Mandarin Chinese
- Direct interaction with a highly proficient speaker, usually the teacher, is needed to practise tone pronunciation

Introduction

Mandarin Chinese is the official administrative language of China
Studying Chinese is popular in the West (and elsewhere)

Problems Teaching Madarin

- Mandarin Chinese is a tone language
- Every syllable in a word has one of 4 (5) tones which determines the meaning of the word
- Using the wrong tone makes a word incomprehensible (cf, English *bad* and *bat*, Dutch *boot* and *bot*)
- Mastering the production and recognition of tones is a major stumbling block in learning Mandarin Chinese
- Direct interaction with a highly proficient speaker, usually the teacher, is needed to practise tone pronunciation

Introduction

Consequences of problems with tones

- Classes must be kept small to allow for ample student-teacher interaction
- Teachers are scarce
- Speaking and listening proficiency improves very slowly
- High drop-out rates of demotivated students
- Speaking is neglected in favor of writing
- $\Rightarrow$ Use speech technology to help students practise

Introduction

Consequences of problems with tones

- Classes must be kept small to allow for ample student-teacher interaction
- Teachers are scarce
- Speaking and listening proficiency improves very slowly
- High drop-out rates of demotivated students
- Speaking is neglected in favor of writing
- $\Rightarrow$ Use speech technology to help students practise

Introduction

Consequences of problems with tones

- Classes must be kept small to allow for ample student-teacher interaction
- Teachers are scarce
- Speaking and listening proficiency improves very slowly
- High drop-out rates of demotivated students
- Speaking is neglected in favor of writing
- $\Rightarrow$ Use speech technology to help students practise

Introduction

Consequences of problems with tones

- Classes must be kept small to allow for ample student-teacher interaction
- Teachers are scarce
- Speaking and listening proficiency improves very slowly
- High drop-out rates of demotivated students
- Speaking is neglected in favor of writing
- $\Rightarrow$ Use speech technology to help students practise

Introduction

Consequences of problems with tones

- Classes must be kept small to allow for ample student-teacher interaction
- Teachers are scarce
- Speaking and listening proficiency improves very slowly
- High drop-out rates of demotivated students
- Speaking is neglected in favor of writing
- $\Rightarrow$ Use speech technology to help students practise

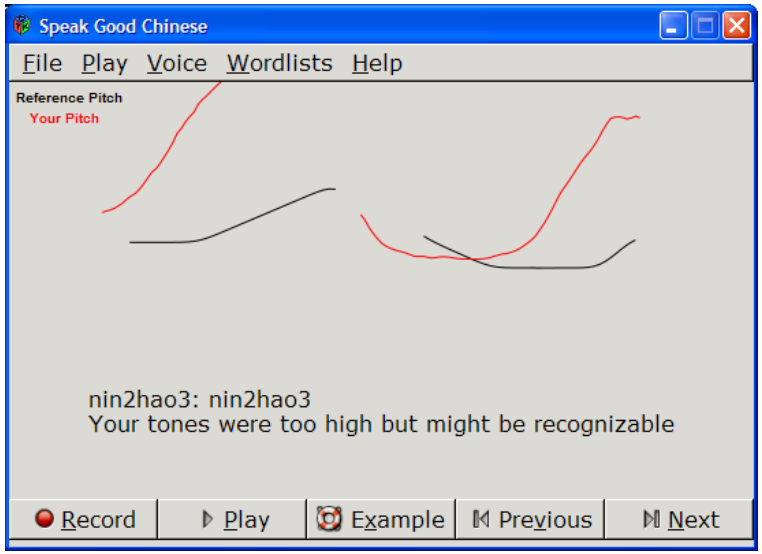
Introduction

Consequences of problems with tones

- Classes must be kept small to allow for ample student-teacher interaction
- Teachers are scarce
- Speaking and listening proficiency improves very slowly
- High drop-out rates of demotivated students
- Speaking is neglected in favor of writing
- $\Rightarrow$ Use speech technology to help students practise

SpeakGoodChinese

An aid for practising Mandarin tones.

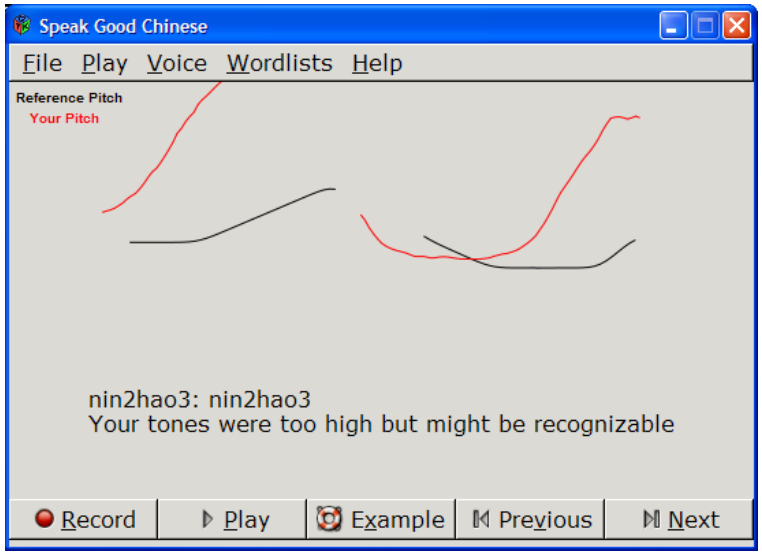


`http://www.SpeakGoodChinese.org/`

- All mono- and bisyllabic words
- Automatic Tone Recognition
- Graphical Tone Presentation
- A written analysis of tone pronunciation.
- Hummed (TTS) or pre-recorded examples
- Replaying recorded student pronunciation
- Automatic student evaluation (hidden)

SpeakGoodChinese

An aid for practising Mandarin tones.

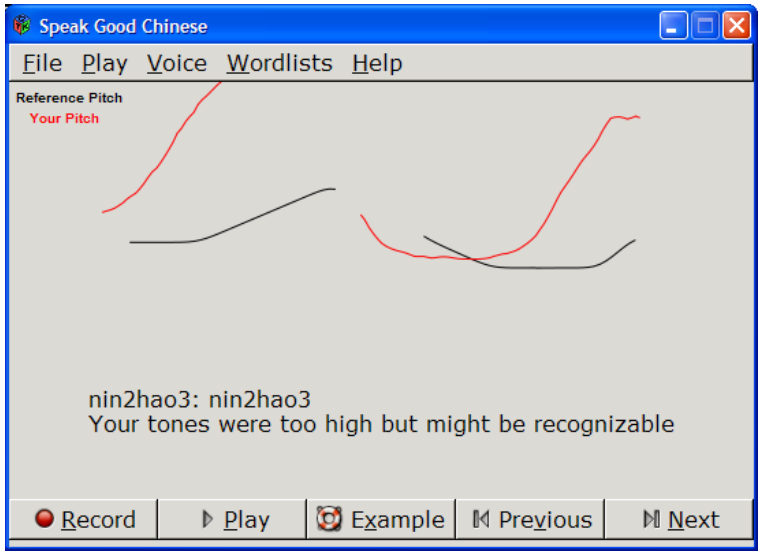


`http://www.SpeakGoodChinese.org/`

- All mono- and bisyllabic words
- Automatic Tone Recognition
- Graphical Tone Presentation
- A written analysis of tone pronunciation.
- Hummed (TTS) or pre-recorded examples
- Replaying recorded student pronunciation
- Automatic student evaluation (hidden)

SpeakGoodChinese

An aid for practising Mandarin tones.

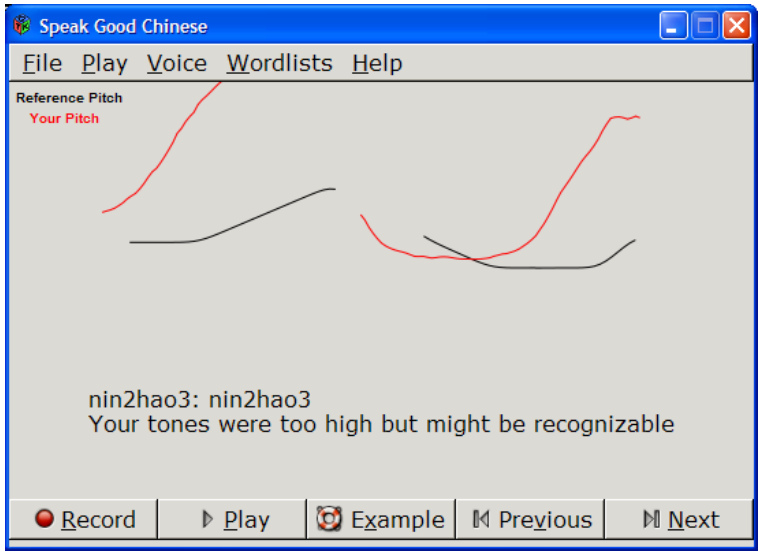


`http://www.SpeakGoodChinese.org/`

- All mono- and bisyllabic words
- Automatic Tone Recognition
- Graphical Tone Presentation
- A written analysis of tone pronunciation.
- Hummed (TTS) or pre-recorded examples
- Replaying recorded student pronunciation
- Automatic student evaluation (hidden)

SpeakGoodChinese

An aid for practising Mandarin tones.

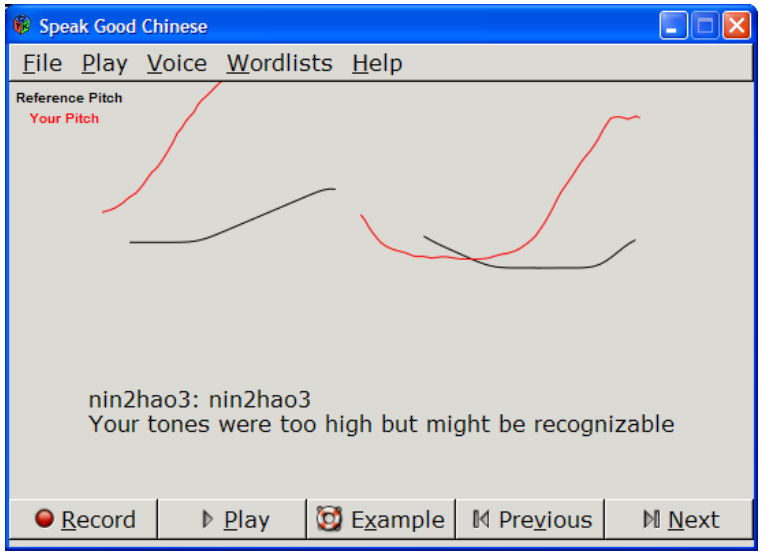


`http://www.SpeakGoodChinese.org/`

- All mono- and bisyllabic words
- Automatic Tone Recognition
- Graphical Tone Presentation
- A written analysis of tone pronunciation.
- Hummed (TTS) or pre-recorded examples
- Replaying recorded student pronunciation
- Automatic student evaluation (hidden)

SpeakGoodChinese

An aid for practising Mandarin tones.

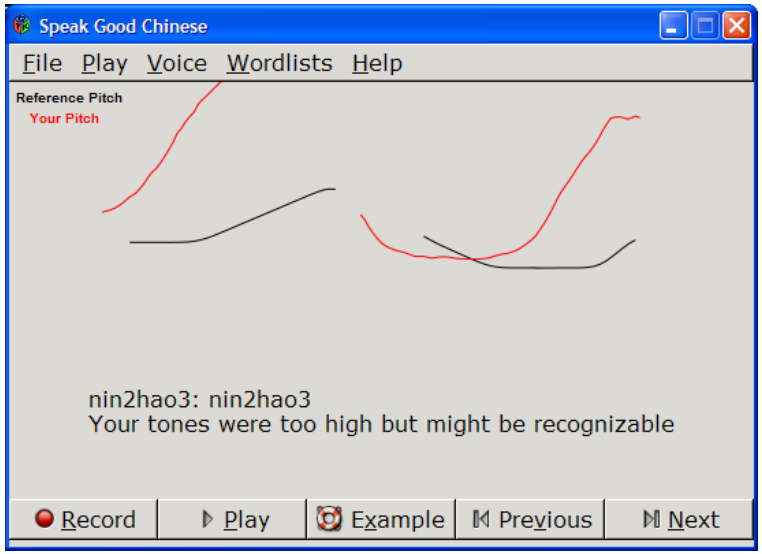


http://www.SpeakGoodChinese.org/

- All mono- and bisyllabic words
- Automatic Tone Recognition
- Graphical Tone Presentation
- A written analysis of tone pronunciation.
- Hummed (TTS) or pre-recorded examples
- Replaying recorded student pronunciation
- Automatic student evaluation (hidden)

SpeakGoodChinese

An aid for practising Mandarin tones.

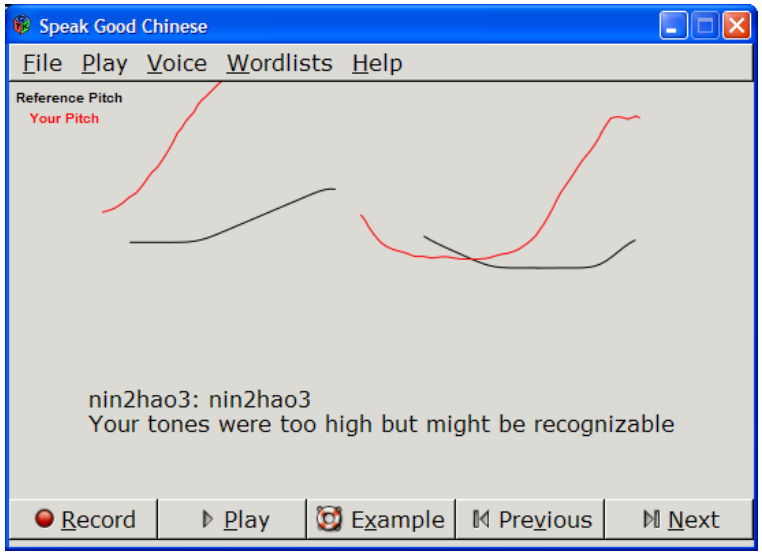


http://www.SpeakGoodChinese.org/

- All mono- and bisyllabic words
- Automatic Tone Recognition
- Graphical Tone Presentation
- A written analysis of tone pronunciation.
- Hummed (TTS) or pre-recorded examples
- Replaying recorded student pronunciation
- Automatic student evaluation (hidden)

SpeakGoodChinese

An aid for practising Mandarin tones.



http://www.SpeakGoodChinese.org/

- All mono- and bisyllabic words
- Automatic Tone Recognition
- Graphical Tone Presentation
- A written analysis of tone pronunciation.
- Hummed (TTS) or pre-recorded examples
- Replaying recorded student pronunciation
- Automatic student evaluation (hidden)

SpeakGoodChinese

Pinyin to Tone synthesis as TTS

- Pinyin phonetic transcription system (eg, *ni3hao3*)
- Each syllable has a number 1-4 or the neutral tone 0
- Split pinyin word into syllables (on tone number)
- Split pinyin syllable into Unvoiced initial and voiced final
- Tone contour is realized on voiced part only

SpeakGoodChinese

Pinyin to Tone synthesis as TTS

- Pinyin phonetic transcription system (eg, *ni3hao3*)
- Each syllable has a number 1-4 or the neutral tone 0
- Split pinyin word into syllables (on tone number)
- Split pinyin syllable into Unvoiced initial and voiced final
- Tone contour is realized on voiced part only

SpeakGoodChinese

Pinyin to Tone synthesis as TTS

- Pinyin phonetic transcription system (eg, *ni3hao3*)
- Each syllable has a number 1-4 or the neutral tone 0
- Split pinyin word into syllables (on tone number)
- Split pinyin syllable into Unvoiced initial and voiced final
- Tone contour is realized on voiced part only

SpeakGoodChinese

Pinyin to Tone synthesis as TTS

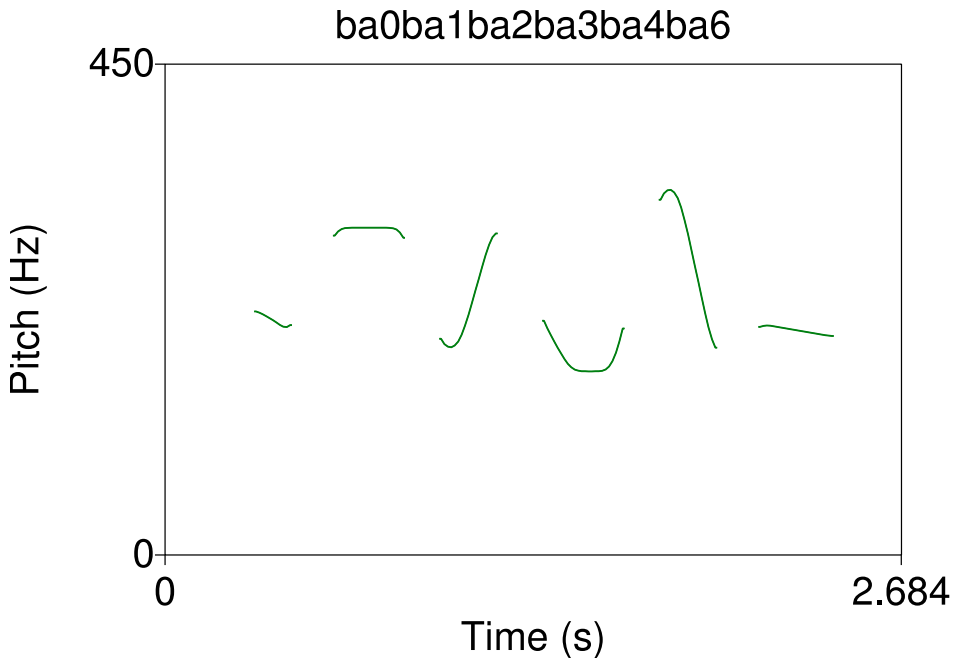
- Pinyin phonetic transcription system (eg, *ni3hao3*)
- Each syllable has a number 1-4 or the neutral tone 0
- Split pinyin word into syllables (on tone number)
- Split pinyin syllable into Unvoiced initial and voiced final
- Tone contour is realized on voiced part only

SpeakGoodChinese

Pinyin to Tone synthesis as TTS

- Pinyin phonetic transcription system (eg, *ni3hao3*)
- Each syllable has a number 1-4 or the neutral tone 0
- Split pinyin word into syllables (on tone number)
- Split pinyin syllable into Unvoiced initial and voiced final
- Tone contour is realized on voiced part only

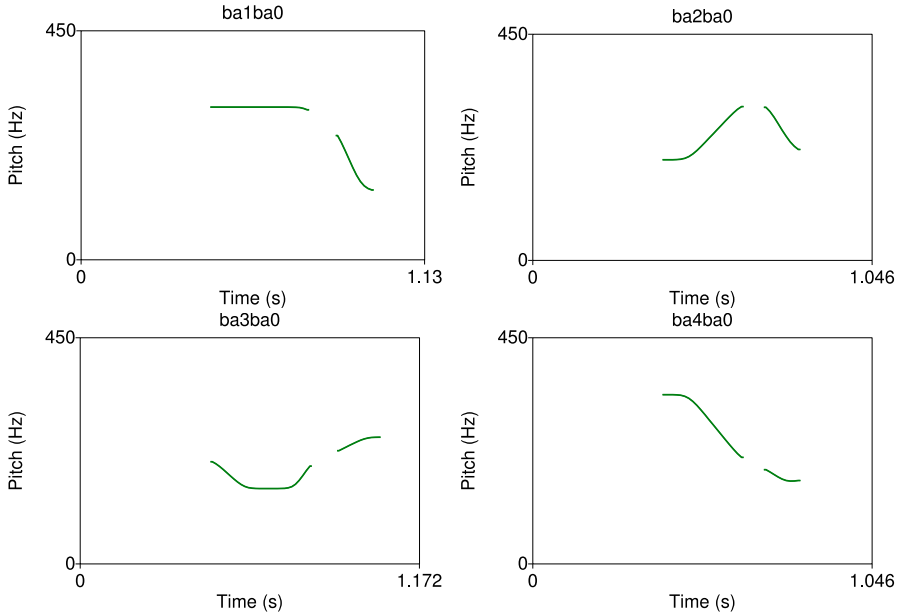
Tone models: All tones



SpeakGoodChinese tone models

- Neutral tone, 0, tones 1-4, and garbage model 6
- Tones change in "context"

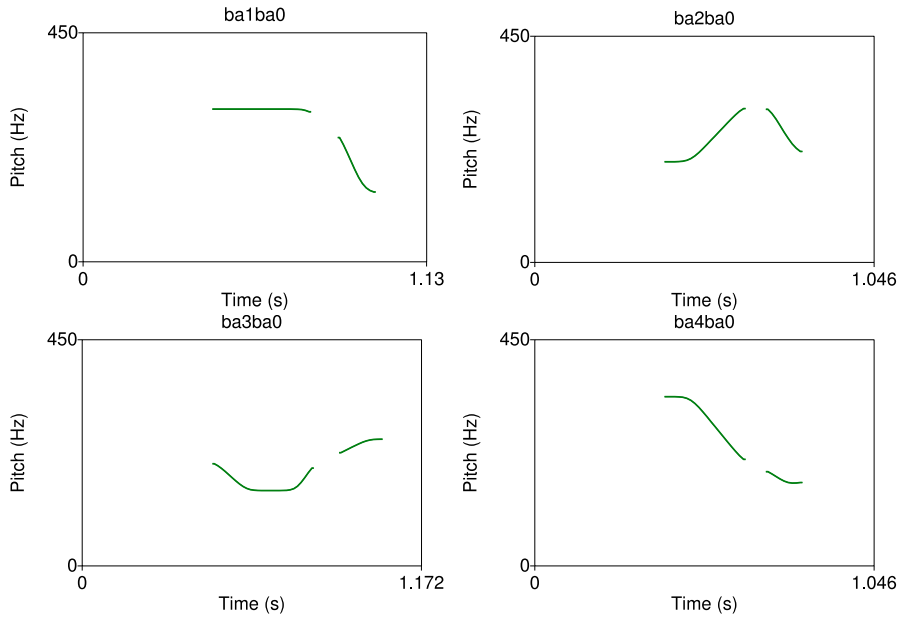
Tone models: Assimilation of neutral tone



Examples

- Neutral tone continues from previous tone
- Returns to "neutral" position
- Fourth tone seems exception

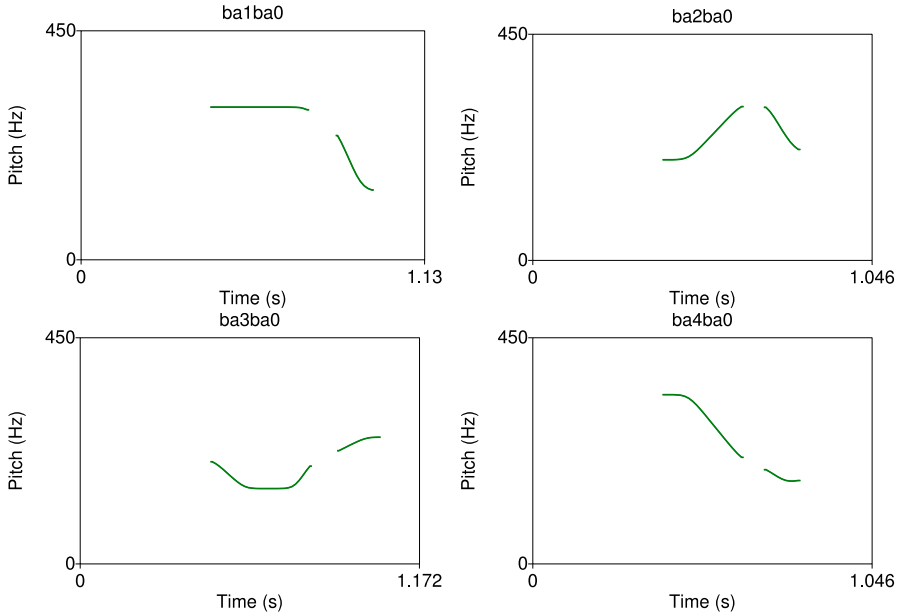
Tone models: Assimilation of neutral tone



Examples

- Neutral tone continues from previous tone
- Returns to "neutral" position
- Fourth tone seems exception

Tone models: Assimilation of neutral tone



Examples

- Neutral tone continues from previous tone
- Returns to "neutral" position
- Fourth tone seems exception

Tone synthesis: Initials and finals

	a	ei	ong	ia	iong	uan
b	ba	bei				
d	da	dei	dong			
zh	zha		zhong			zhuan
r						
j				jia	jiong	
g	ga	gei	gong			guan

Durational model

- Syllable: Optional Initial (*zh*) + Obligatory Final (*ong*)
- Tones are realized on the voiced part of the syllable
- Estimate durations of Initial and Final
- Crude model: Fixed duration + $\delta\cdot$ number of symbols (iao=3)
- Adapt duration to tone: $3 > 1 > 2 \approx 4 \gg 0$

Tone recognition

Tone recognition: Was student correct?

- Extract utterance pitch contour (F_0)
- Pinyin-to-Tone synthesis for all tones (correct and *in*correct)
- Compare student utterance to all possible tone contours using Dynamic Time Warping
- Pick best matching model $\Rightarrow$ Recognition
- Construct possible countours from theoretical tone model
- Limited to two syllables (combinatorial explosion)
- Student pitch register must be known

Tone recognition

Tone recognition: Was student correct?

- Extract utterance pitch contour (F_0)
- Pinyin-to-Tone synthesis for all tones (correct and *in*correct)
- Compare student utterance to all possible tone contours using Dynamic Time Warping
- Pick best matching model $\Rightarrow$ Recognition
- Construct possible countours from theoretical tone model
- Limited to two syllables (combinatorial explosion)
- Student pitch register must be known

Tone recognition

Tone recognition: Was student correct?

- Extract utterance pitch contour (F_0)
- Pinyin-to-Tone synthesis for all tones (correct and *in*correct)
- Compare student utterance to all possible tone contours using Dynamic Time Warping
- Pick best matching model $\Rightarrow$ Recognition
- Construct possible countours from theoretical tone model
- Limited to two syllables (combinatorial explosion)
- Student pitch register must be known

Tone recognition

Tone recognition: Was student correct?

- Extract utterance pitch contour (F_0)
- Pinyin-to-Tone synthesis for all tones (correct and *in*correct)
- Compare student utterance to all possible tone contours using Dynamic Time Warping
- Pick best matching model $\Rightarrow$ Recognition
- Construct possible countours from theoretical tone model
- Limited to two syllables (combinatorial explosion)
- Student pitch register must be known

Tone recognition

Tone recognition: Was student correct?

- Extract utterance pitch contour (F_0)
- Pinyin-to-Tone synthesis for all tones (correct and *in*correct)
- Compare student utterance to all possible tone contours using Dynamic Time Warping
- Pick best matching model $\Rightarrow$ Recognition
- Construct possible countours from theoretical tone model
- Limited to two syllables (combinatorial explosion)
- Student pitch register must be known

Tone recognition

Tone recognition: Was student correct?

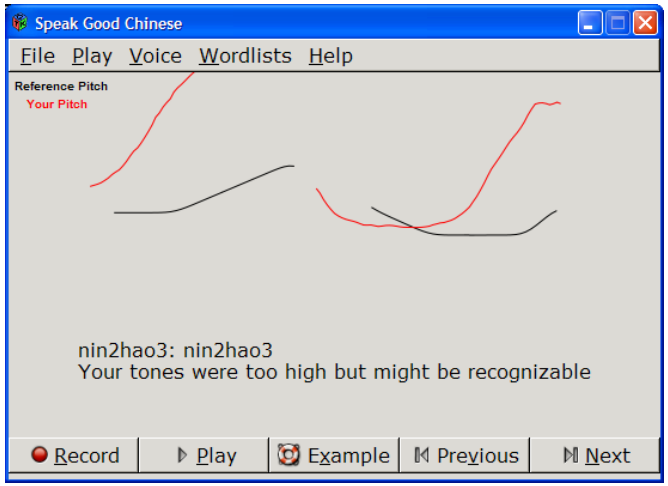
- Extract utterance pitch contour (F_0)
- Pinyin-to-Tone synthesis for all tones (correct and *in*correct)
- Compare student utterance to all possible tone contours using Dynamic Time Warping
- Pick best matching model $\Rightarrow$ Recognition
- Construct possible countours from theoretical tone model
- Limited to two syllables (combinatorial explosion)
- Student pitch register must be known

Tone recognition

Tone recognition: Was student correct?

- Extract utterance pitch contour (F_0)
- Pinyin-to-Tone synthesis for all tones (correct and *in*correct)
- Compare student utterance to all possible tone contours using Dynamic Time Warping
- Pick best matching model $\Rightarrow$ Recognition
- Construct possible countours from theoretical tone model
- Limited to two syllables (combinatorial explosion)
- Student pitch register must be known

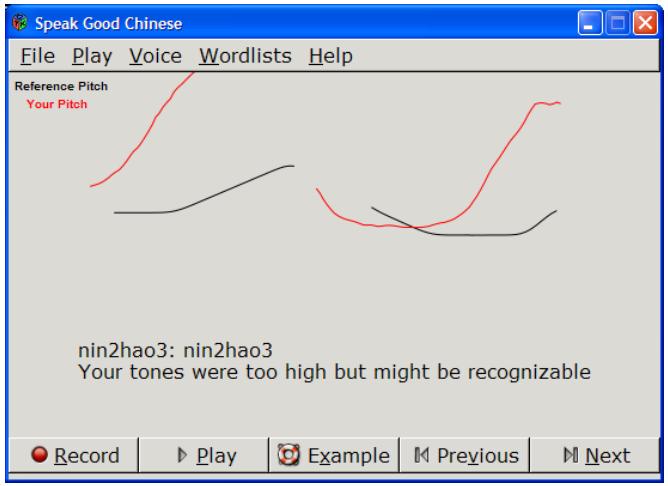
Tone recognition: Pitch height and movements



A good tone has correct pitch height and movements

- If *top pitch* deviates from model, flag an error
- If *pitch range* deviates from model, flag an error
- Students will exagerate tones, punish exagerations less
- Flag error if 3 semitones too low or too narrow
- Flag exageration if 6 semitones too high or too wide

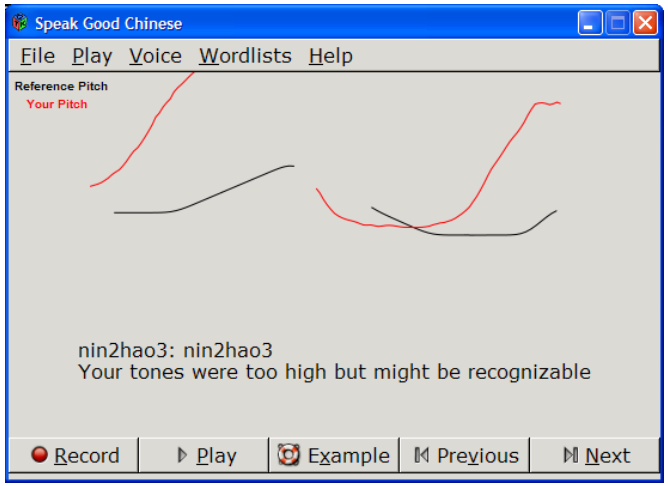
Tone recognition: Pitch height and movements



A good tone has correct pitch height and movements

- If *top pitch* deviates from model, flag an error
- If *pitch range* deviates from model, flag an error
- Students will exagerate tones, punish exagerations less
- Flag error if 3 semitones too low or too narrow
- Flag exageration if 6 semitones too high or too wide

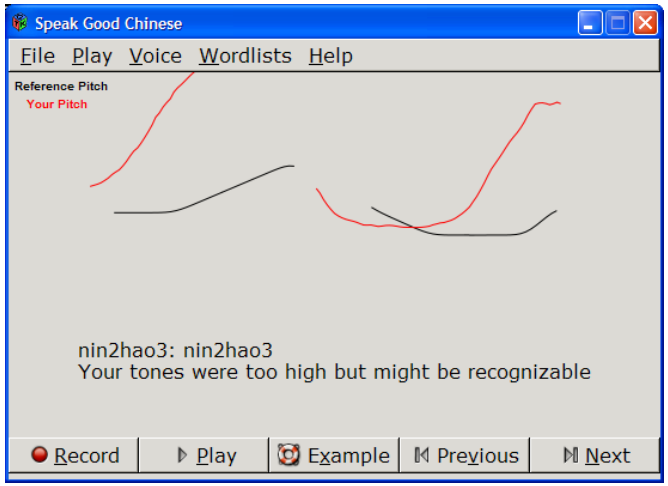
Tone recognition: Pitch height and movements



A good tone has correct pitch height and movements

- If *top pitch* deviates from model, flag an error
- If *pitch range* deviates from model, flag an error
- Students will exagerate tones, punish exagerations less
- Flag error if 3 semitones too low or too narrow
- Flag exageration if 6 semitones too high or too wide

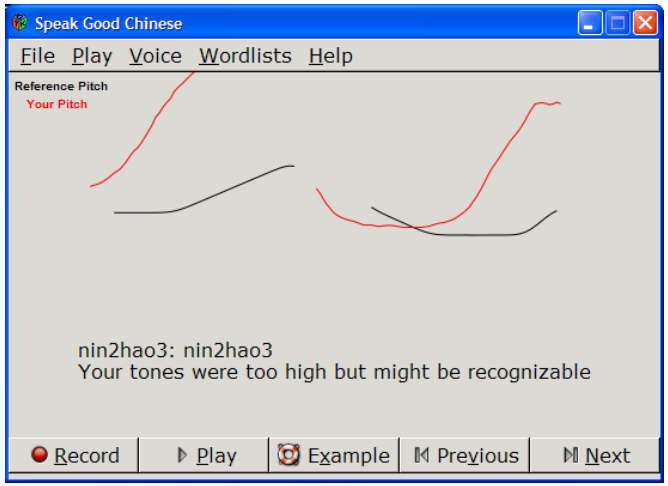
Tone recognition: Pitch height and movements



A good tone has correct pitch height and movements

- If *top pitch* deviates from model, flag an error
- If *pitch range* deviates from model, flag an error
- Students will exagerate tones, punish exagerations less
- Flag error if 3 semitones too low or too narrow
- Flag exageration if 6 semitones too high or too wide

Tone recognition: Pitch height and movements



A good tone has correct pitch height and movements

- If *top pitch* deviates from model, flag an error
- If *pitch range* deviates from model, flag an error
- Students will exagerate tones, punish exagerations less
- Flag error if 3 semitones too low or too narrow
- Flag exageration if 6 semitones too high or too wide

Tone recognition: Heuristic rules

Model tones do not model enough variation

- Duration rules currently very bad
- Current tone models do not capture variation
- Use "heuristic" rules to capture common confusions
- Eg, tones 2 and 3 merge before another tone 2 or 3
- Eg, tones 2 and 4 often misidentified as tone 0 in DTW but tone 0 would have been flagged by tone height and movement

Tone recognition: Heuristic rules

Model tones do not model enough variation

- Duration rules currently very bad
- Current tone models do not capture variation
- Use "heuristic" rules to capture common confusions
- Eg, tones 2 and 3 merge before another tone 2 or 3
- Eg, tones 2 and 4 often misidentified as tone 0 in DTW but tone 0 would have been flagged by tone height and movement

Tone recognition: Heuristic rules

Model tones do not model enough variation

- Duration rules currently very bad
- Current tone models do not capture variation
- Use "heuristic" rules to capture common confusions
- Eg, tones 2 and 3 merge before another tone 2 or 3
- Eg, tones 2 and 4 often misidentified as tone 0 in DTW but tone 0 would have been flagged by tone height and movement

Tone recognition: Heuristic rules

Model tones do not model enough variation

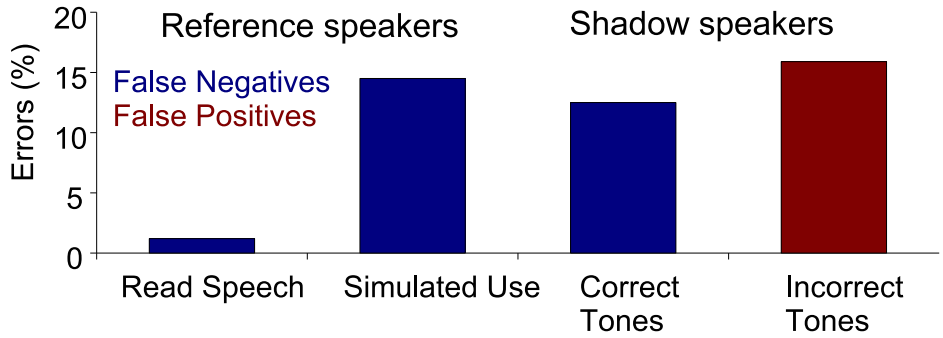
- Duration rules currently very bad
- Current tone models do not capture variation
- Use "heuristic" rules to capture common confusions
- Eg, tones 2 and 3 merge before another tone 2 or 3
- Eg, tones 2 and 4 often misidentified as tone 0 in DTW but tone 0 would have been flagged by tone height and movement

Tone recognition: Heuristic rules

Model tones do not model enough variation

- Duration rules currently very bad
- Current tone models do not capture variation
- Use "heuristic" rules to capture common confusions
- Eg, tones 2 and 3 merge before another tone 2 or 3
- Eg, tones 2 and 4 often misidentified as tone 0 in DTW but tone 0 would have been flagged by tone height and movement

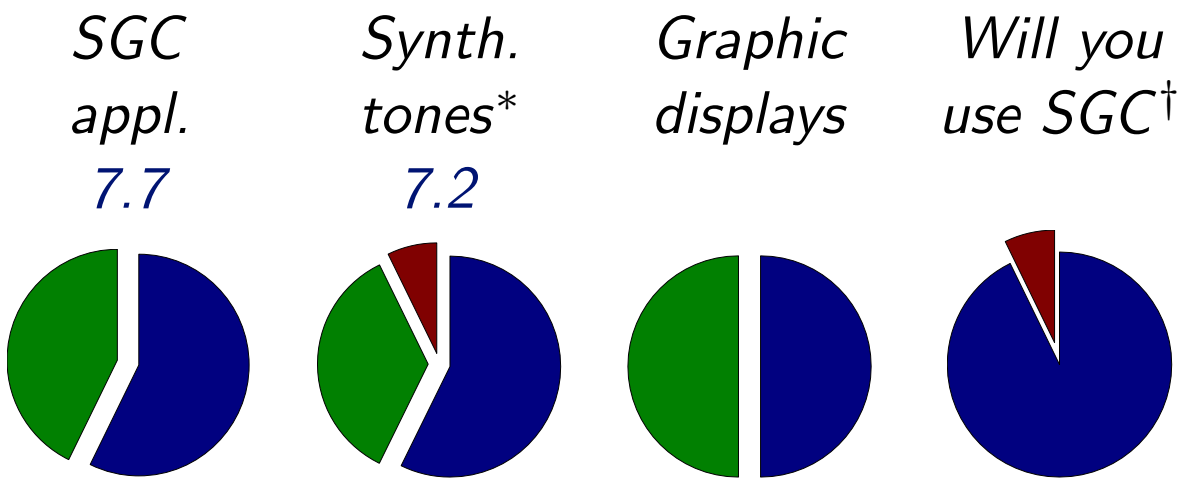
Evaluation: Recognizer False *rejects* and *accepts*



Reference speakers and Students

- Correct Tones
  - Read Speech: R read aloud 6 words: *cha2, dian4hua4, duo1shao3, gong1zuo4, jie2hun1, shi2jian1*, *83* tokens.
  - Simulated Use: R free word choice, *358* tokens
  - Shadowed Correct Speech: R and S shadowed 6 words, *160* tokens
- Incorrect tones
  Shadowed Incorrect Speech: R and S shadowed 6 words, *320* tokens

Evaluation: Usefulness and grade 1-10



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

* One subject couldn't hear the tones clearly

† One subject preferred to practice with family members

Questionnaire to 14 students

- Tested RAD Tcl/Tk GUI with functional recognition
- Responses used to design User Interface

Evaluation: Usage data

Does SpeakGoodChinese improve tone pronunciation?

- Single Female student (13)
- Tried out SpeakGoodChinese in 7 sessions of a few hours
- In total she utterred 1531 words
- Each session started and ended with test runs without audio feedback
- Pretest and Posttest $\approx$ 30 words
- Practise $\approx$ 83-389 words
- Automatically determined error rate ($^{*}p < 0.002, X^2$)
  - Overall: 28% (including practise)
  - Pretest: 39% *
  - Posttest: 24% *
- Real progress awaits human judgment

Conclusions

Computer Assisted Language Learning of Mandarin tones

- A Free, GPL, CALL applications based on Praat
- Real-time tone recognizer with an estimated 15% error rate
- Recognition can be improved upon
- The method is biased for acceptance of one- and two syllable words
- Besides tone, it can be extended to word pronunciation
- Students seem to like it

Conclusions

Computer Assisted Language Learning of Mandarin tones

- A Free, GPL, CALL applications based on Praat
- Real-time tone recognizer with an estimated 15% error rate
- Recognition can be improved upon
- The method is biased for acceptance of one- and two syllable words
- Besides tone, it can be extended to word pronunciation
- Students seem to like it

Conclusions

Computer Assisted Language Learning of Mandarin tones

- A Free, GPL, CALL applications based on Praat
- Real-time tone recognizer with an estimated 15% error rate
- Recognition can be improved upon
- The method is biased for acceptance of one- and two syllable words
- Besides tone, it can be extended to word pronunciation
- Students seem to like it

Conclusions

Computer Assisted Language Learning of Mandarin tones

- A Free, GPL, CALL applications based on Praat
- Real-time tone recognizer with an estimated 15% error rate
- Recognition can be improved upon
- The method is biased for acceptance of one- and two syllable words
- Besides tone, it can be extended to word pronunciation
- Students seem to like it

Conclusions

Computer Assisted Language Learning of Mandarin tones

- A Free, GPL, CALL applications based on Praat
- Real-time tone recognizer with an estimated 15% error rate
- Recognition can be improved upon
- The method is biased for acceptance of one- and two syllable words
- Besides tone, it can be extended to word pronunciation
- Students seem to like it

Conclusions

Computer Assisted Language Learning of Mandarin tones

- A Free, GPL, CALL applications based on Praat
- Real-time tone recognizer with an estimated 15% error rate
- Recognition can be improved upon
- The method is biased for acceptance of one- and two syllable words
- Besides tone, it can be extended to word pronunciation
- Students seem to like it

http://www.SpeakGoodChinese.org/

Thank you very much
Any questions?

Acknowledgement

The SpeakGoodChinese project was made possible by grant 6046 from the Digital University