

영어교육에 있어서의  
영어억양의 역할  
(The role of prosody in English education)

Korea Nazarene University

2006. 5.19

Kyuchul Yoon

English Division

Kyungnam University

# Contents

- A pop quiz.....3
- Acquiring prosody in language learning.....7
- Previous approaches.....8
- A new tool.....9
- Technical details.....10
- Implications.....23

## Pop quiz

Speech sound is composed  
of consonants and vowels.

Answer

**WRONG!!**

Speech sound = segments + suprasegmentals  
(consonants & vowels) (intonation, pause, length, loudness)

# Suprasegmentals (or Prosody)

(1) pause = phrase breaks

(2) intonation = fundamental frequency (F0 contour)

(3) length = segmental durations

(4) loudness = intensity (or loudness) contour

# Acquiring prosody in language learning

- Prosody as non-segmental features of speech
  1. phrase breaks
  2. intonation (F0) contour
  3. segmental durations
  4. intensity contour

# Previous approaches

- **Explicit teaching**  
of prosodic features such as the intonation contours, segmental durations, etc.
- **Audio aid**  
*Listen and repeat!*
- **Visual aid**  
Visual display of suprasegmentals  
(Chun, 89; Spaai & Hermes, 92).  
Dr.Speaking<sup>®</sup> : F0 contour comparison between  
native speaker and non-native speaker

# A new tool

- A new kind of audio aid  
in the form of a non-native speaker's utterance with the prosodic features of a native speaker's utterance
- How this works
  1. Software presents a native speaker's utterance
  2. A non-native speaker repeats the utterance
  3. Software records the non-native speaker's utterance
  4. Software imposes the native speaker's prosody onto the non-native speaker's utterance
  5. Software presents the processed non-native utterance

# Technical details

- Manipulation of
  1. segmental durations, including phrase breaks
  2. F0 contours
  3. intensity contours
- For 1 and 2

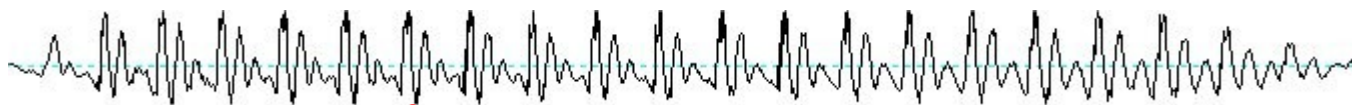
PSOLA (*Pitch Synchronous OverLap and Add*),  
developed by Moulines & Charpentier, 1990  
implemented in Praat
- For 3

Intensity swap in Praat

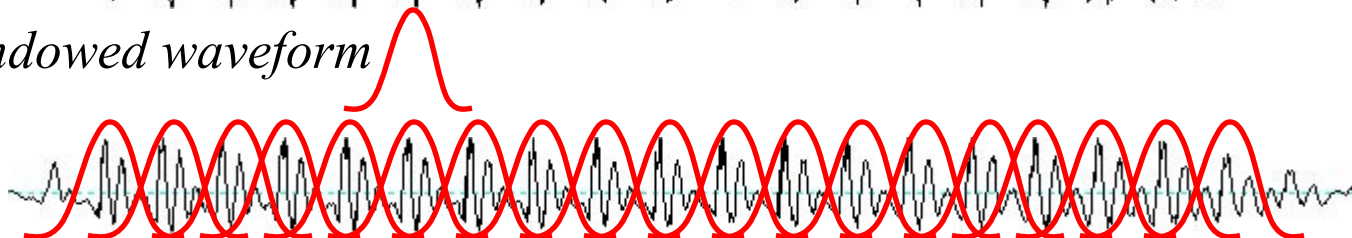
# Technical details

Moulines & Charpentier, 1990

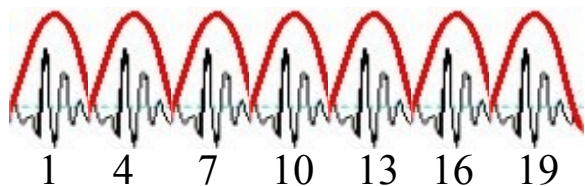
*original waveform*



*windowed waveform*

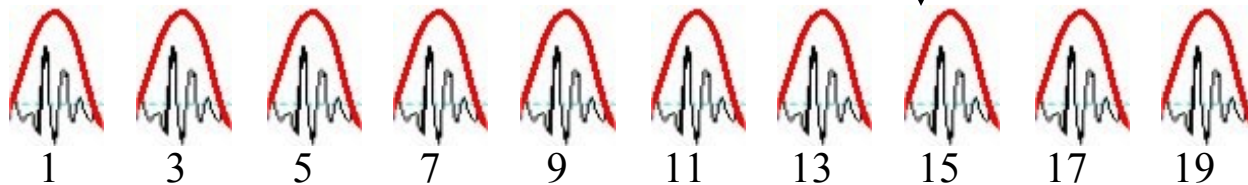


1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19



*shortened waveform*

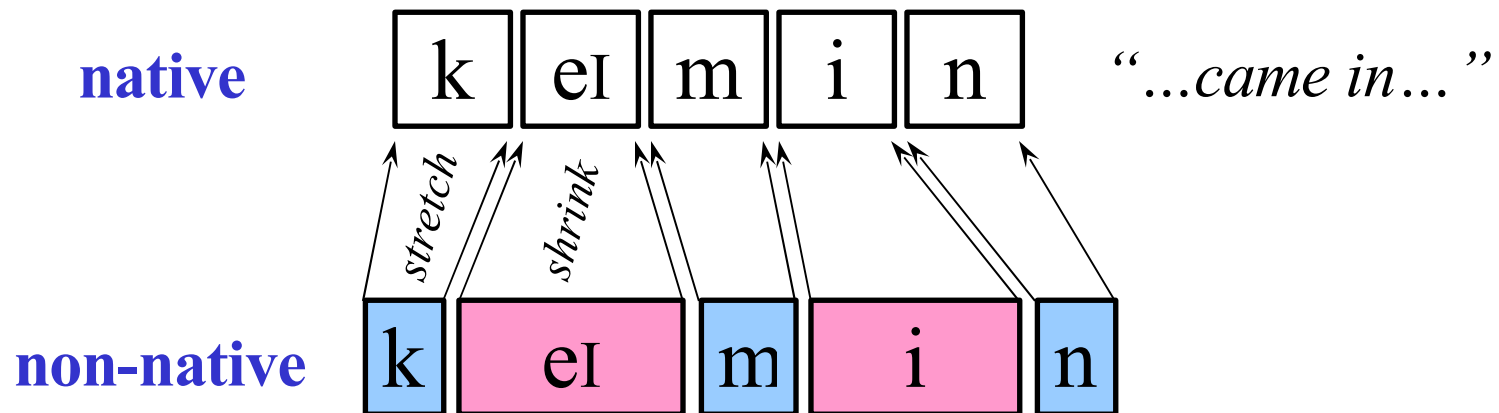
*waveform with lower F0*



# Technical details 1

## Segmental durations

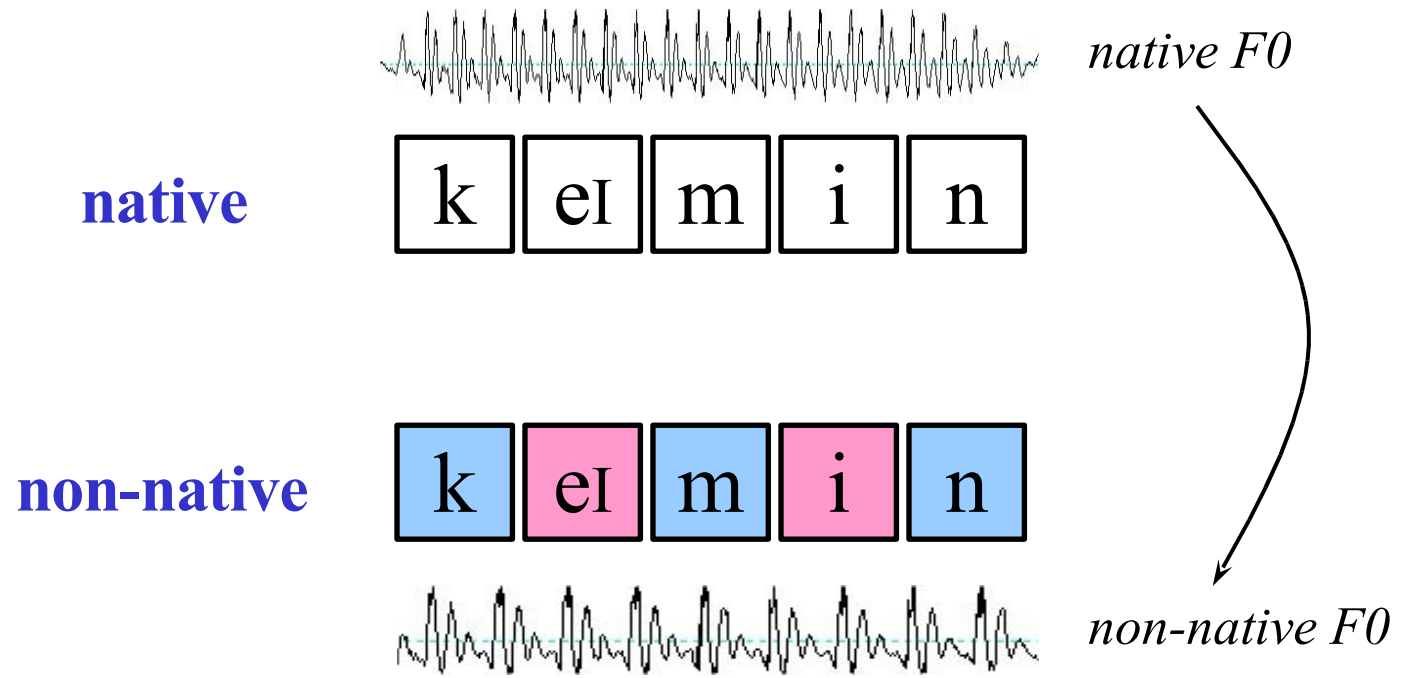
- Segment alignment & PSOLA processing of durations  
: Alignment can be manual  
or automatic (with the help of speech recognition)



# Technical details 1+2

## Segmental durations + F0 contour

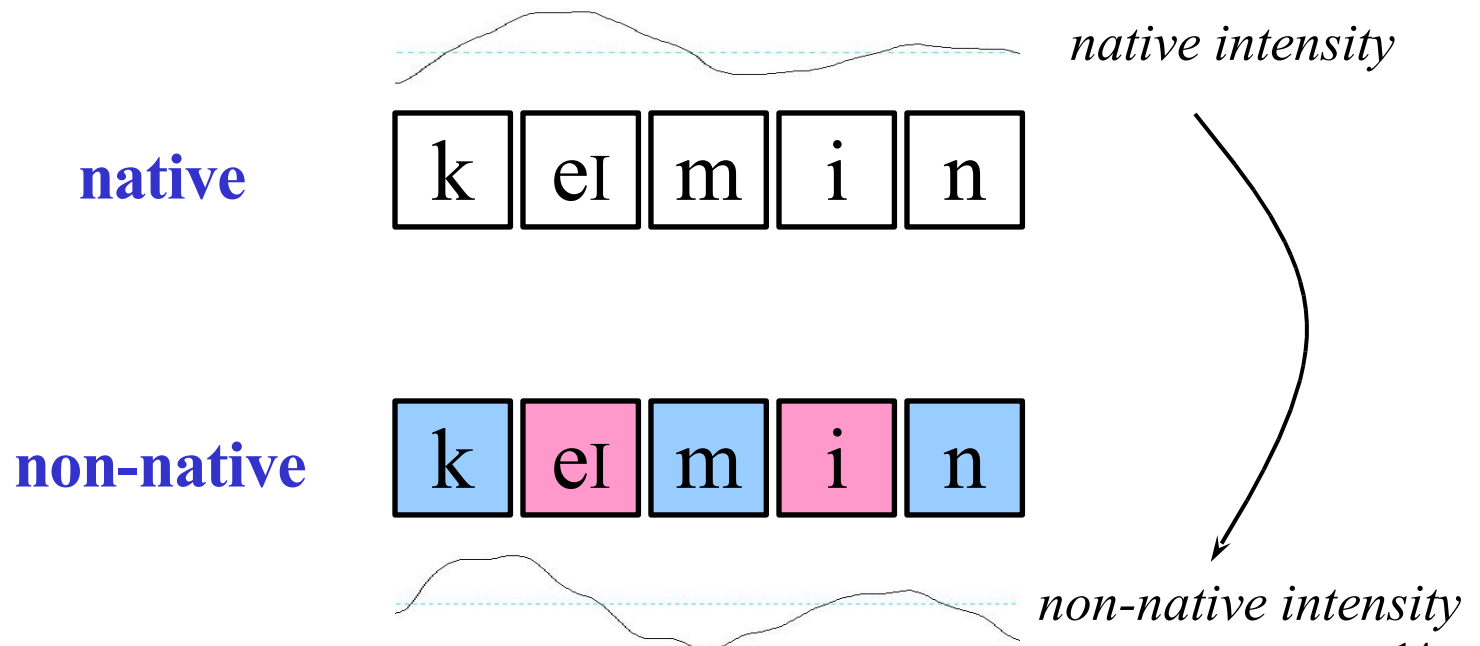
- PSOLA processing of F0 on duration-treated utterance



# Technical details 1+2+3

Segmental durations + F0 contour + intensity contour

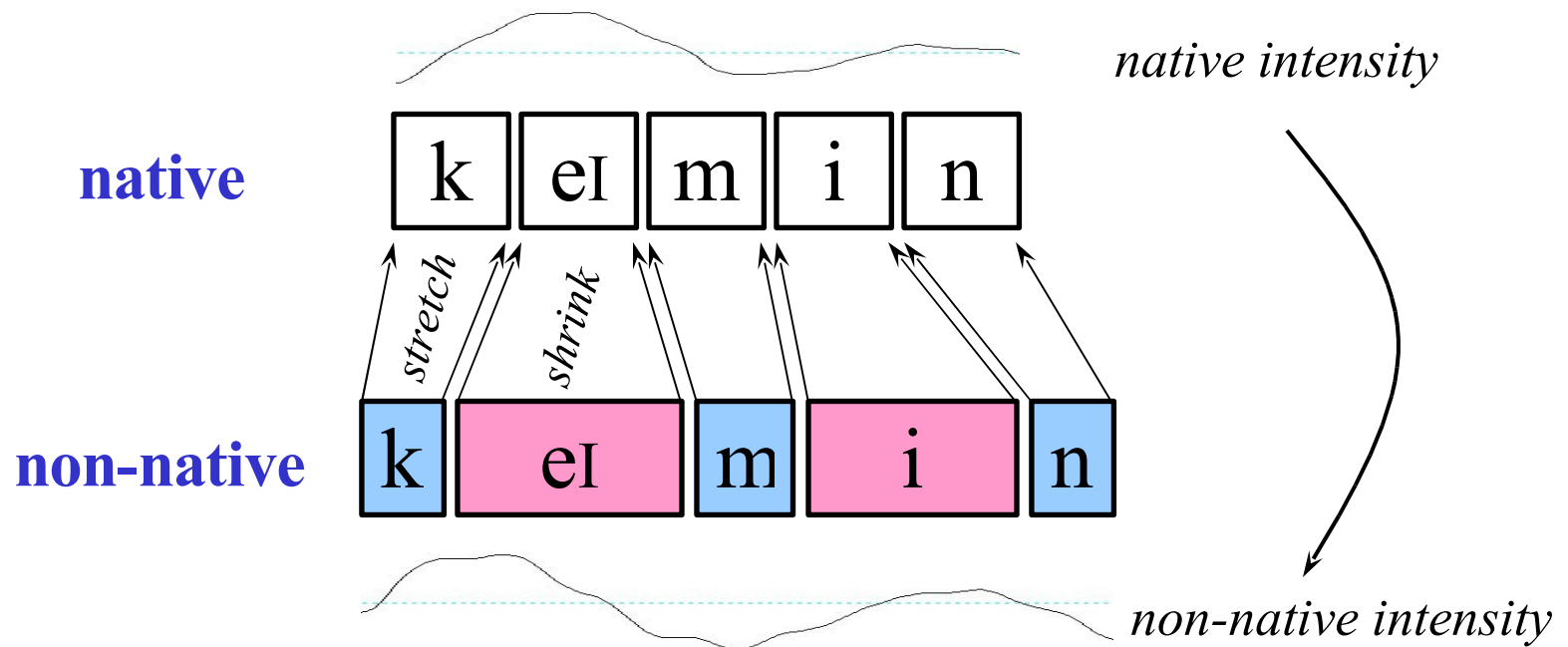
- Mathematically “neutralize” non-native speaker’s intensity contour and transfer native speaker’s intensity contour in Praat – Holger Miterer (personal communication)



# Technical details 1+3

## Segmental durations + intensity contour

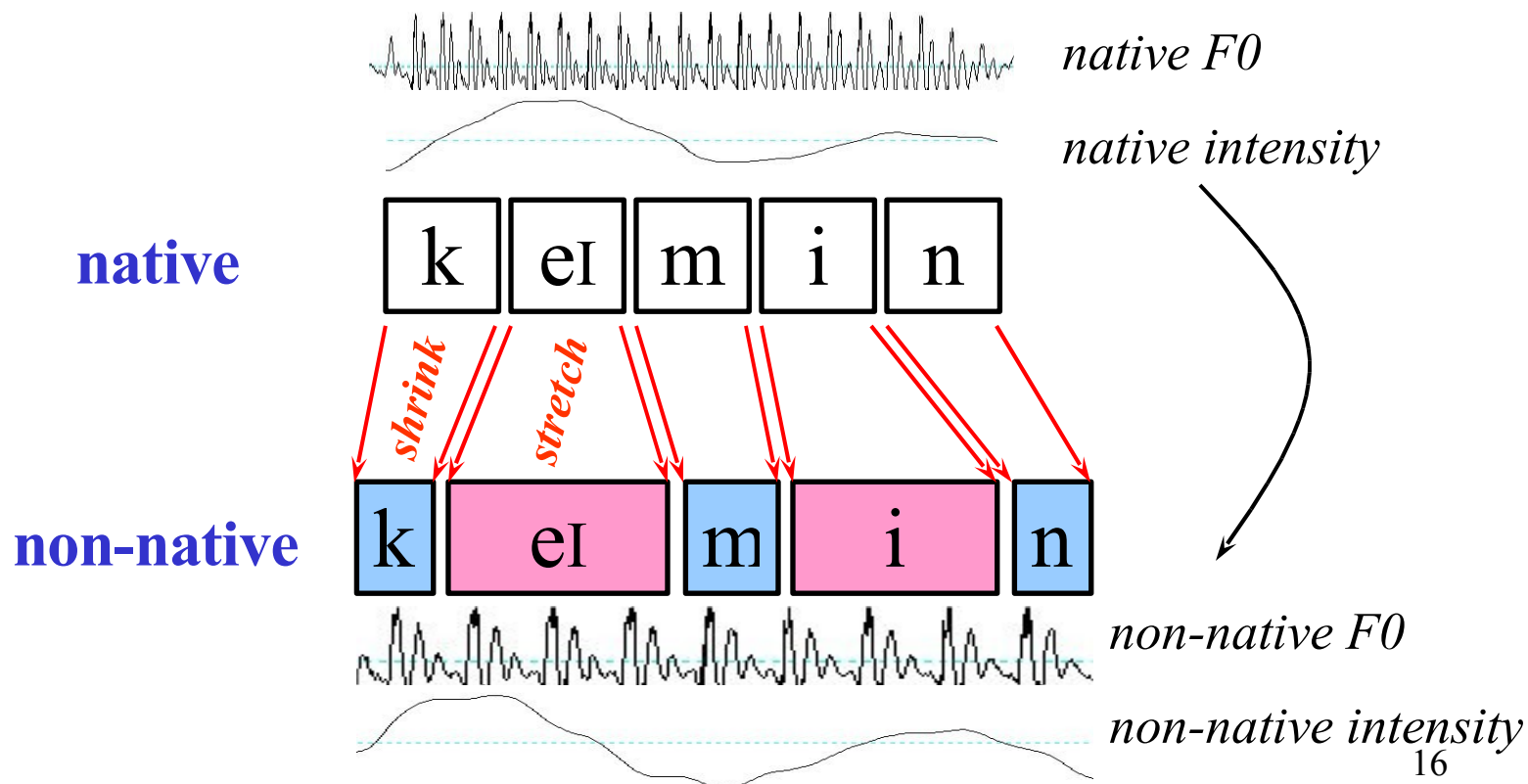
- Segment alignment & PSOLA processing of durations followed by intensity contour transfer



# Technical details 2+3

## F0 contour + intensity contour

- “Reverse” segment alignment & PSOLA processing of F0 followed by intensity contour transfer



# Technical details

- **Weakness**
  1. Voiceless segments can be made “voiced” in the windowing process (pitch-synchronous technique)
  2. Excessive handling results in unnatural synthesis
- **Segment alignment**

should be fine-tuned according to the voiced/voiceless status of the (sub-)segments for better results

# Technical details

## Examples

### Praat script

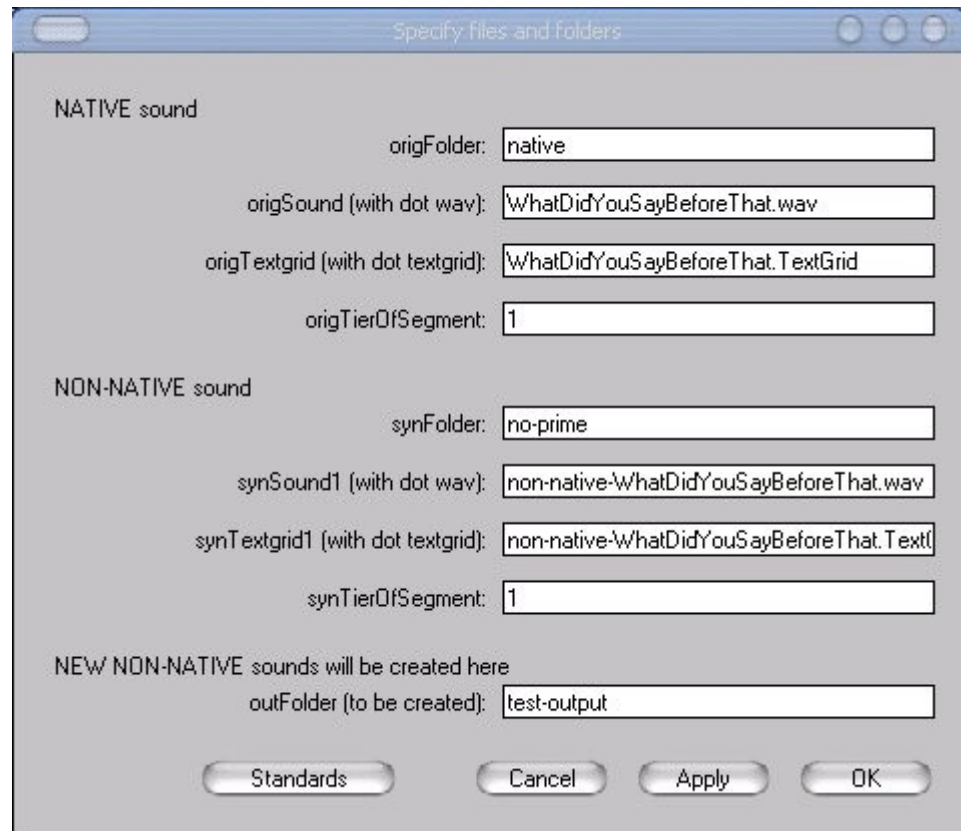
native utterance

non-native utterance

**synthetic non-native**  
(durations+F0+intensity)

**synthetic non-native**  
(durations+intensity)

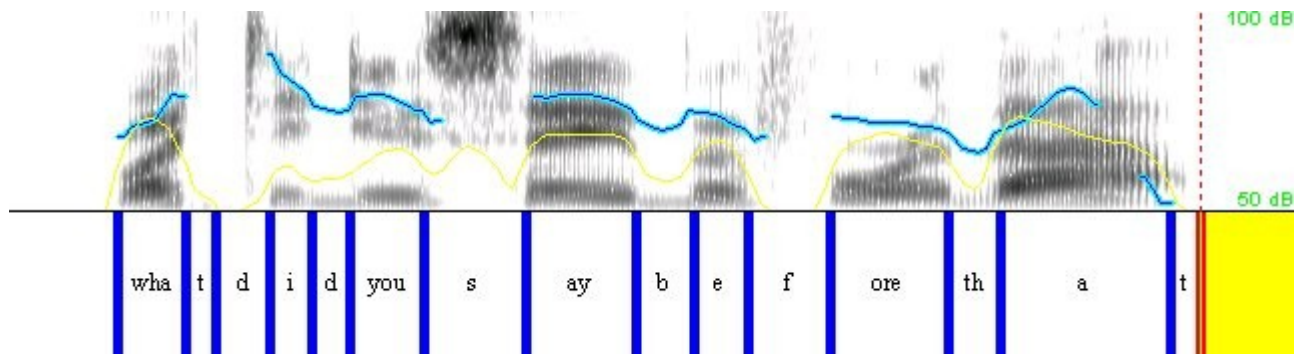
**synthetic non-native**  
(F0+intensity)



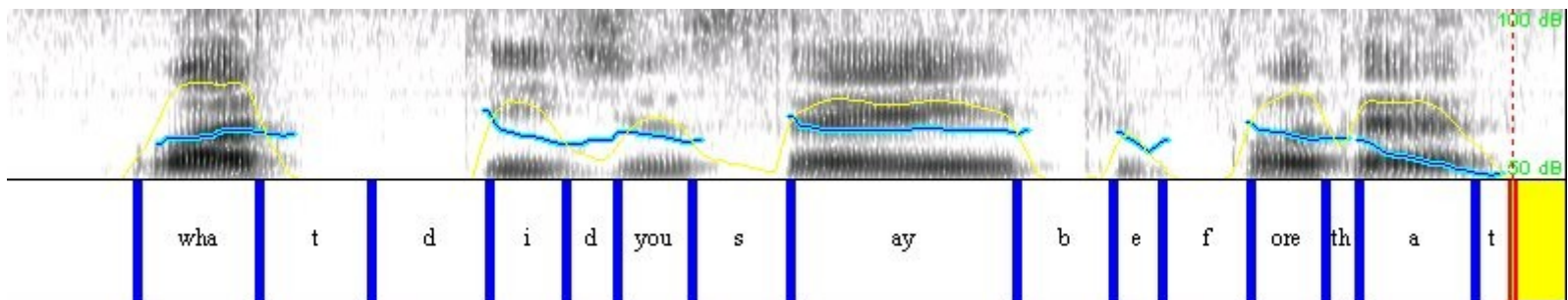
# Technical details

Comparison before synthesis – duration, F0 & intensity  
(blue & yellow)

native utterance



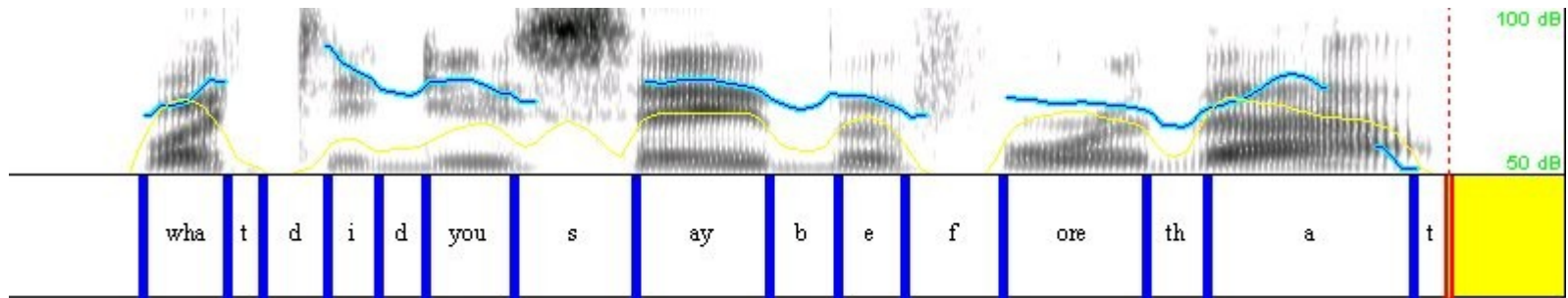
non-native utterance



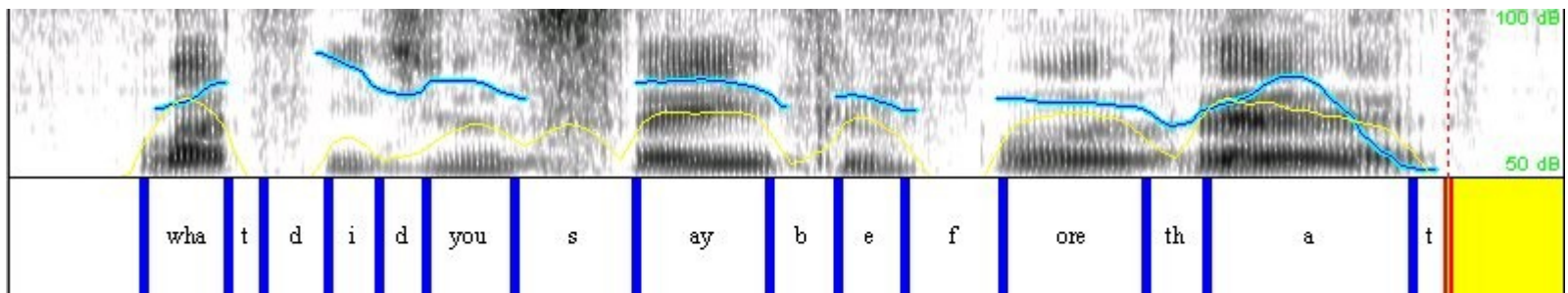
# Technical details

Comparison after synthesis – duration, F0 & intensity  
(blue & yellow)

native utterance



synthetic non-native

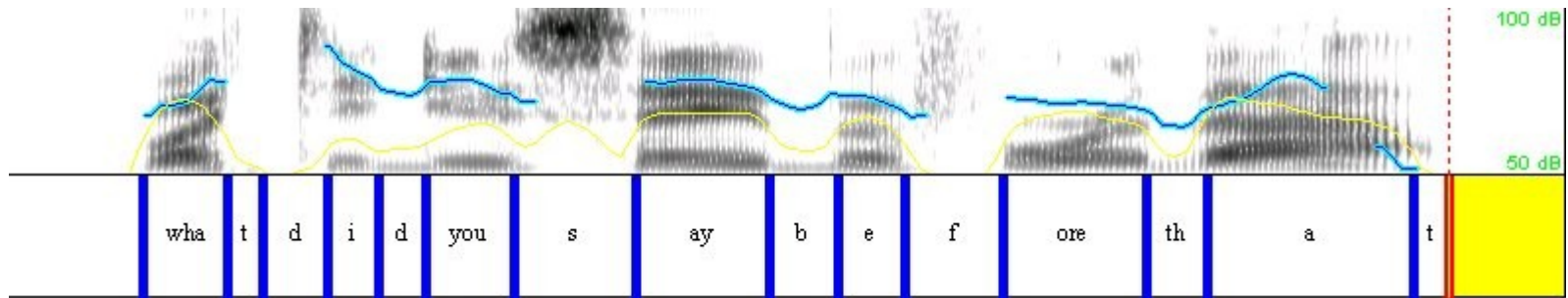


# Technical details

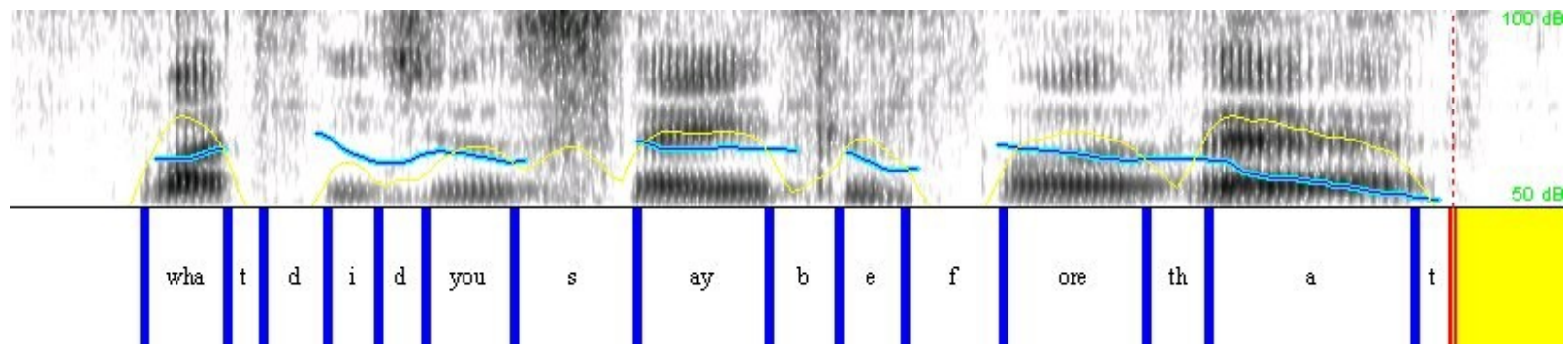
Comparison after synthesis – duration & intensity

(blue & yellow)

native utterance



synthetic non-native

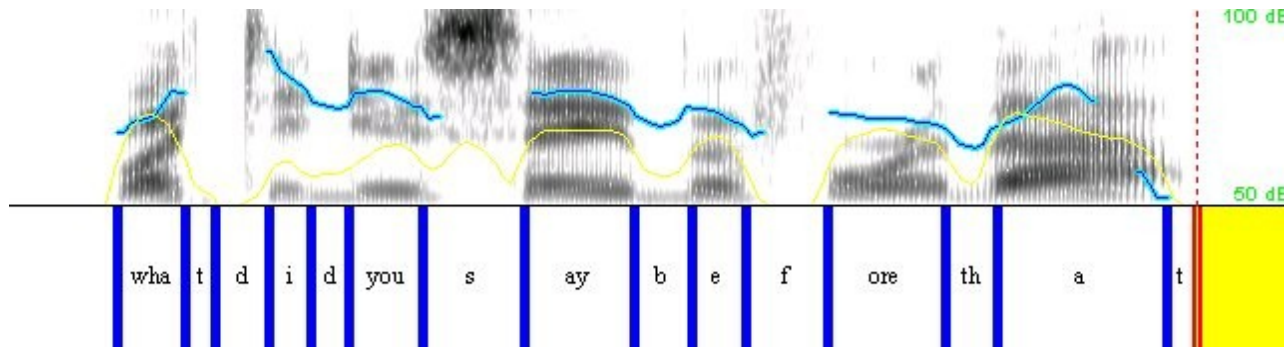


# Technical details

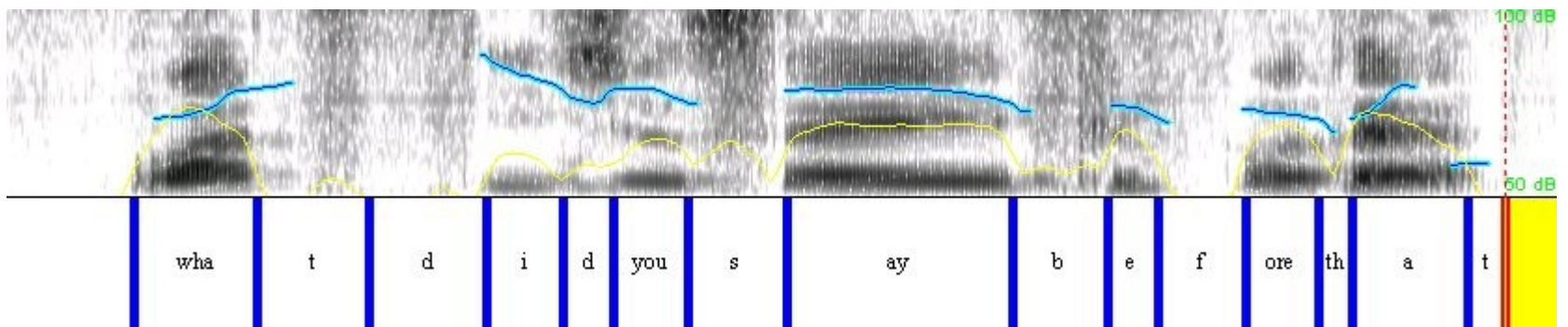
Comparison after synthesis –F0 & intensity

(blue & yellow)

native utterance



synthetic non-native



# Implications

- The technique could be used
  - (1) In second language education
    - to facilitate/motivate acquisition of the target language prosody
    - to emphasize the importance of prosody in achieving native speaker fluency
  - (2) For patients with vocal disorders
    - to help achieve the prosody of a normal voice
- ASR (Automatic Speech Recognition) can be employed to automate the segment aligning stage

# References

- E. Moulines and F. Charpentier (1990) “Pitch synchronous waveform processing techniques for text-to-speech synthesis using diphones” *Speech Communication* 9, 453-467.
- D. Chun (1989) “Teaching tone and intonation with microcomputers” *CALICO Journal* 6(3),21-47
- W. Spaai and D. Hermes (1992) “A visual display for the teaching of intonation” *CALICO Journal* 10(3), 19-30.