Constructing Asian English Speech Corpus for Universal Purposes

Mariko Kondo[1,2] & Yoshinori Sagisaka[1,3]
[1]Language and Speech Science Laboratory
[2]SILS & [3]GITI
Waseda University

AESOP (**A**sian **E**nglish **S**peech c**O**rpus **P**roject)

- Research Consortium of Asian L2 spoken language
- Aiming at
  - Commonly sharable L2 spoken corpus building
  - Trans-disciplinary spoken language research promotion
  - Close communication among researchers in Asia
- Founded in Sept. 2008
- Current stirring committee members:

  Waseda U. (Japan), Academia Sinica (Taiwan),
  CUHK (Hong Kong), NECTEC (Thailand),
  CASS (China), Wonkwang U .(Korea)

Trans-disciplinary studies on L2 spoken language

- Linguistics & Phonetics
  - Derive phonetic properties common to all Asian Englishes
  - Discover phonetic features particular to individual languages
  - Controlling principles of languages
- Psychology & Education
  - Perceptual modeling
  - Scientific evaluation and assessment of language acquisition and learning
- Speech Science & Information Engineering
  - Controlling mechanisms
  - Objective evaluation
  - CALL & WELL tool development

AESOP2010 International Workshop at Kathmandu

- Date: 23 November 2010
- Venue: Hotel Paradise Plaza, Kathmandu, Nepal
- Meeting agenda

(1) Presentations on current status and research
+ Report on L2 spoken language WS (Sept. ’10)
(2) Presentations on the L2 related research
(3) Future planning and Open forum
in particular to related Nepali researchers
(4) Next year plan
Next meeting in Taiwan (25 Oct. ’11)

Why Asian English Corpora ?

- English as a Lingua Franca
- Asia is the largest English speaking community & market:

  Outer circle & Expanding circle

  (native-, L2-speakers) (learner-speakers)
- Rich variation in pronunciation, lexicon and grammar, blended with local languages (world Englishes)

  (Piat et al., 2008, Nariai et al. 2008)
- Study of variation present in Asian English dialects
  - Look for features that are particular to individual varieties of Asian English
  - Conduct cross-linguistic comparative analyses
- Develop technologies to cater for the requirements of Asian speaker populations

GOAL! Establish a new reference model of Asian Englishes for all research purposes

How does AESOP work?

- Each research team uses a common recording platform and shares an experimental task set.
- Each team will develop a common, open-ended annotation system.
- AESOP-collected corpora are an open resource

Common recording platform for AESOP (1/2)

1. Reading written materials (read speech: natural & instructed)
   - lists of words & phrases, text

2. Picture description Task (semi-spontaneous speech)
   – syntactic structures in word strings differentiated using prosody

3. Computer prompted dialogue (semi-spontaneous speech)

4. Full range of segmental phonemic contrasts in English

5. Comprehensive coverage of stress types and syllabicities
   – Word selection based on overall frequency count
   < The CMU dictionary database

Common recording platform for AESOP (2/2)

- Common EFL Problems among Asian language speakers covered in AESOP data (Visceglia et al. 2008)

(1) Phoneme level:
- Full range of English phonemes: Cs & Vs

(2) Syllable structure:
- 2-4 syllable words, different syllable structures

(3) Suprasegmental:
- Various phonological units (word, phrase, utterance)
- Phrase boundary features:
  e.g. declarative falls and interrogative rises
- Syllabicity
- Rhythm and timing
- Lexical stress and phrase accent
- Location of pitch accents; broad and narrow focus

(4) Phonological rules:
- Sound change
  e.g. assimilation, coalescent, allophonic variation
- Elision
- Consonant epenthesis
- Syllabic consonant

Prosodic Context (1/2)

(1) Target Words in Narrow Focus

e.g) We have to finish the project ***overnight***, not over the weekend.

e.g.) I don't think you stole the ***money***, but you probably stole the car.

(2) Stressed and Unstressed Function Words

e.g.) I can [ə] run faster than you can [æ].

e.g.) *He went to a fancy dress party as*[ə] *a guest, but what did he dress as*[æ]*?*

Prosodic Context (2/2)

(3) Prosodic disambiguation

e.g.) When Alice leaves, Tom will be upset.

When Alice leaves Tom, we'll be upset.

e.g.) He [washed and brushed] his hair.

He [washed (himself)] and [brushed his hair].

(4) Alphabetic strings and number sequences

e.g) 'CN'N 'BB'C 'AB'C '1 2'3

'Samuel L 'Jackson 'J F 'Kennedy

'OEC'D 'RSPC'A '194'2

'38'7- '546'8

Language Specific Features (L1 Japanese) (1/4)

(1) Consonants

Japanese consonants

	bilabial	labio-dental	dental	alveolar	post-alveolar	alveolo-palatal	palatal	velar	glottal
plosives	p b			t d				k g	
nasals	m			n			ɲ	ŋ	
fricatives	ɸ			s z		ɕ	ç		h
affricates				ts dz		tɕ dʑ			
lateral									
approxi-mants	(w)						j		(w)
Tap/flap				ɾ					

Language Specific Features (L1 Japanese) (2/4)

English consonants

	bilabial	Labio-dental	dental	alveolar	post-alveolar	palato-alveolar	palatal	velar	glottal
plosives	p b			t d				k g	(ʔ)
nasals	m			n				ŋ	
fricatives		f v	θ	ð s z	ʃ	ʒ			h
affricates				ts			tʃ dʒ		
lateral				l					
approxi-mants	(ʍ) (w)			ɹ			j		(ʍ) (w)

Non-phonemic in Japanese

No phonemic contrast in Japanese

Contextual problems for Japanese speakers

Language Specific Features (L1 Japanese) (3/4)

(2) Vowel system

Japanese	vs.	English
5 vowels		GenAm: 15 vowels
		RP: 19 vowels
		AU: 18-19 vowels
		NZ: 17-19 vowels
		CAN: 14 vowels

- Timing: vowel sequences vs. diphthongs
  vowel weakening (←stress & rhythm)

Language Specific Features (L1 Japanese) (4/4)

(3) Syllable structures

Japanese	vs.	English
(C)(j)(V)(V)(C)(C)		(C)(C)(C)V(C)(C)(C)

- *CC
- No deletion rule → vowel epenthesis

stranged /stɹeɪndʒd/ => |sutore(i)Nzido|

u o i o

strengths /stɹeŋθs/ => |sutoreNgusu|

u o u u

Target words

	2-1	3-1	3-2	3-3	4-1	4-2	4-3	4-4	LH	RH
Y-N (rise)	money	wonderful	apartment	overnight						white wine
WH (fall)					elevator	available	information	misunder-stand	supermarket	
Cont. (rise)					January	experience	California	Vietnamese	department store	
Decl. (fall)	morning	video	tomorrow	Japanese						afternoon
Narrow focus	money morning	wonderful video	apartment tomorrow	overnight Japanese	elevator January	available experience	information California	misunder-stand Vietnamese	supermarket department store	white wine afternoon

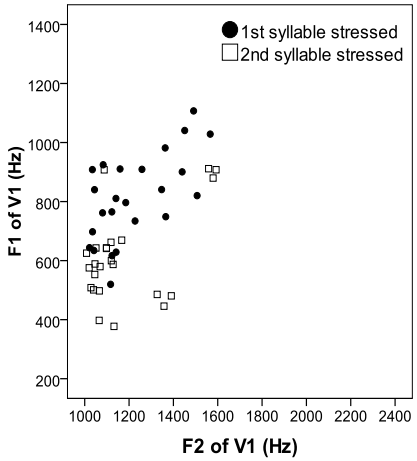
Note: The 2-2 (2-syllable final stress) type has been excluded as this type is expected to yield very similar data to that of 3-3 and 4-4.

Strong vs. Weak vowels

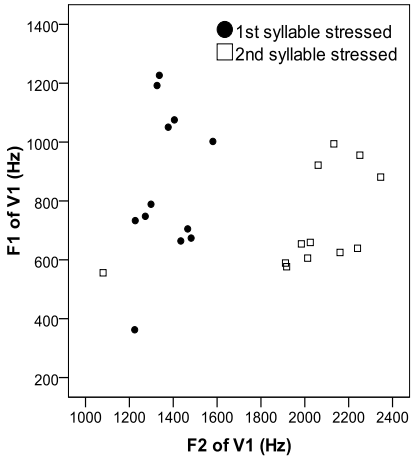
elevator	/ˈeləveɪtə(r)/
apartment	/əˈpɑː(r)tmənt/
tomorrow	/təˈmɔːroʊ/təˈmɒrəʊ/
Japanese	/ˌdʒæpəˈniːz/
(white wine	/waɪt ˈwaɪn/)

F1 & F2 of stressed and unstressed vowels by Japanese and native American English speakers (Kondo, 2009)

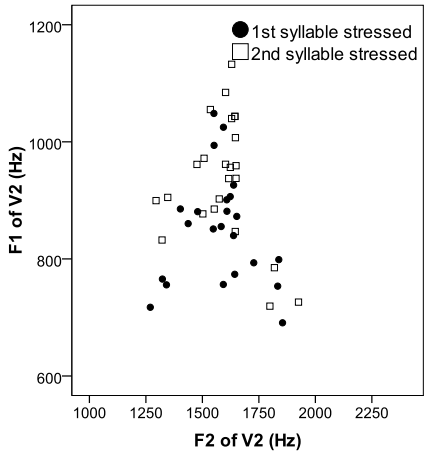
Japanese: F1 & F2 of V1 of *contract*



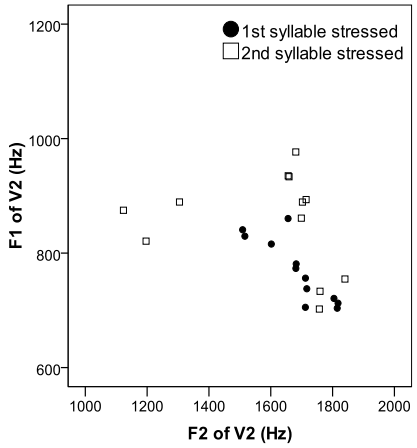
American: F1 & F2 of V1 of *contract*



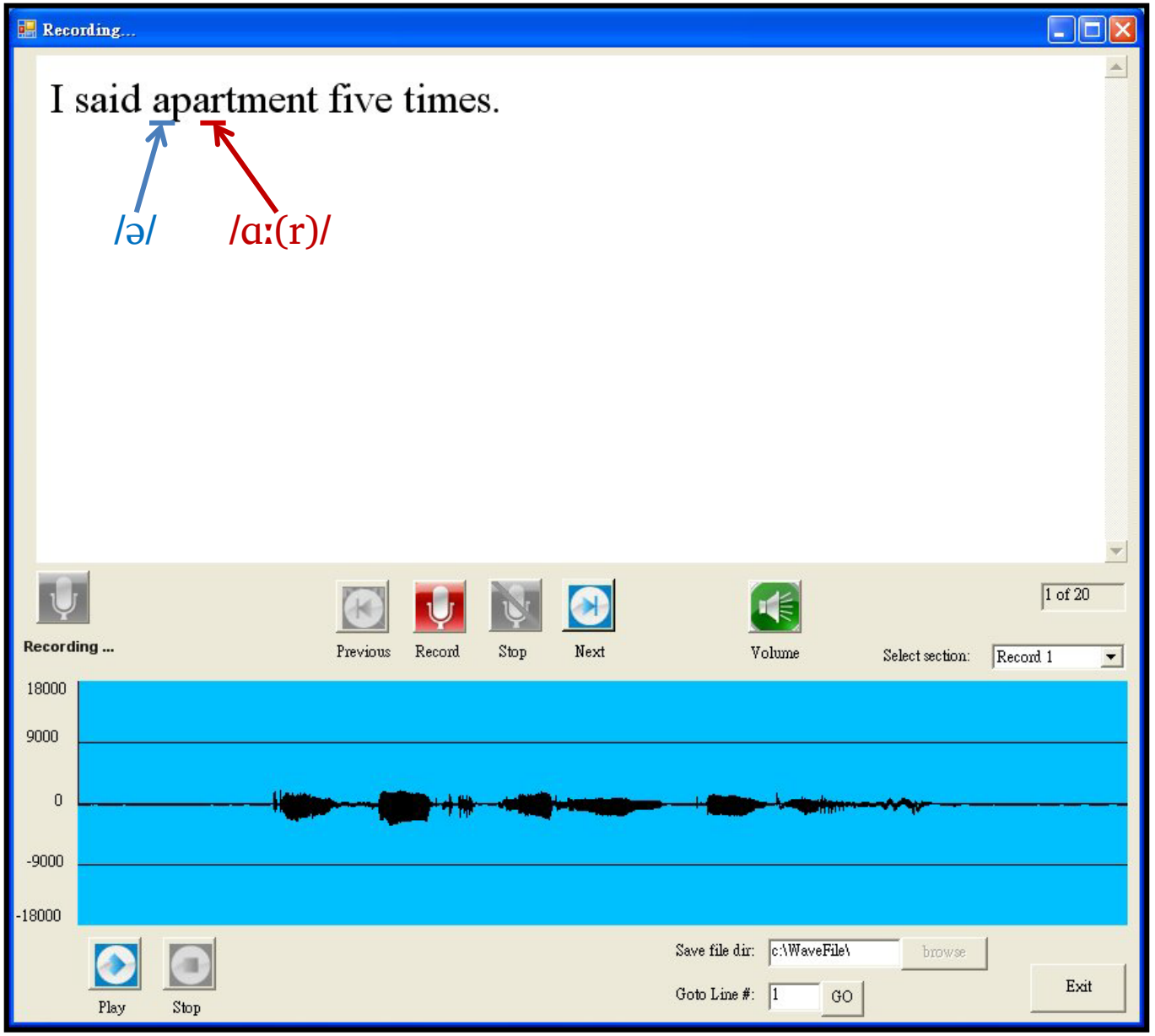
Japanese: F1 & F2 of V2 of *contract*



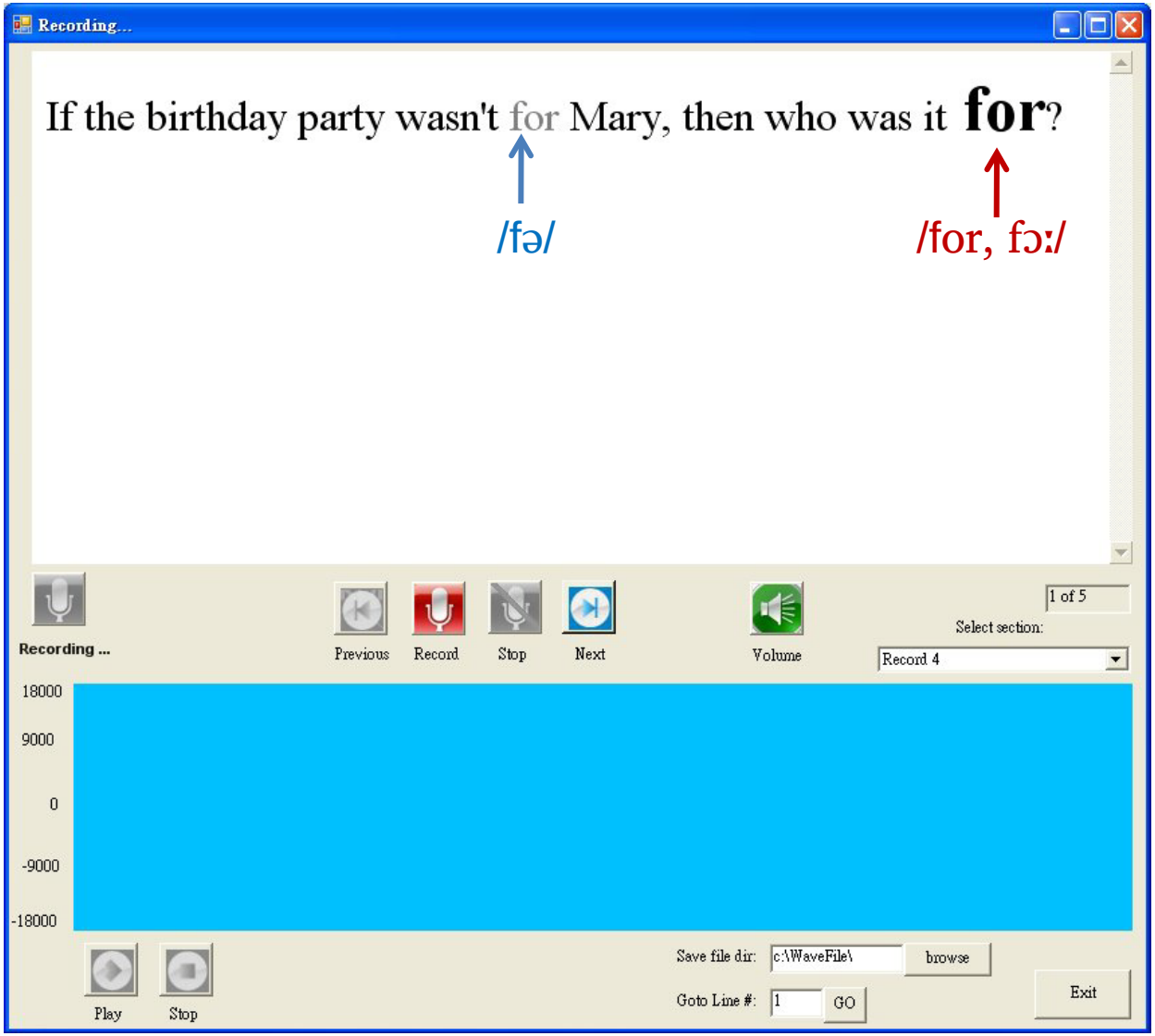
American: F1 & F2 of V2 of *contract*



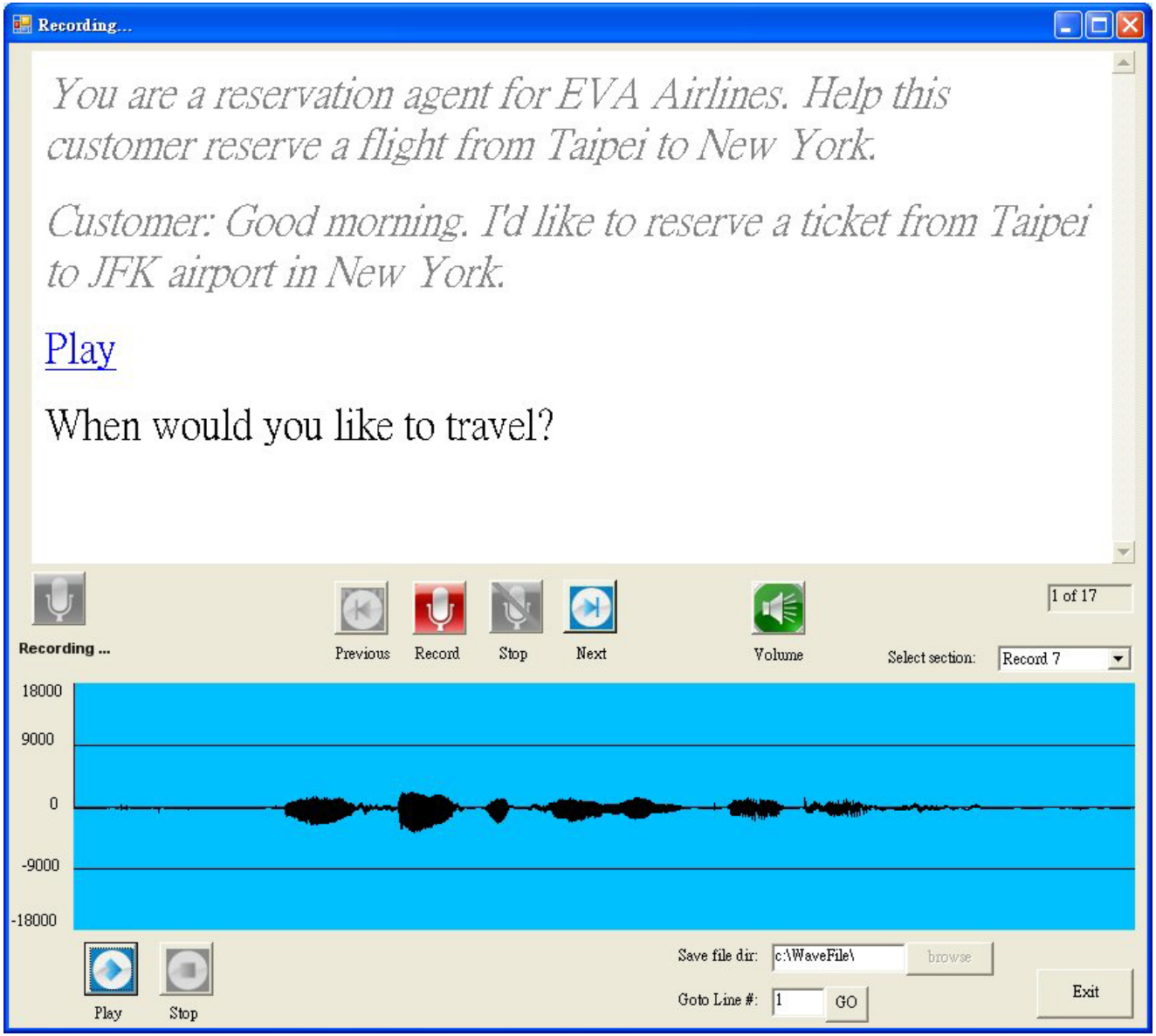
Example of Task 1



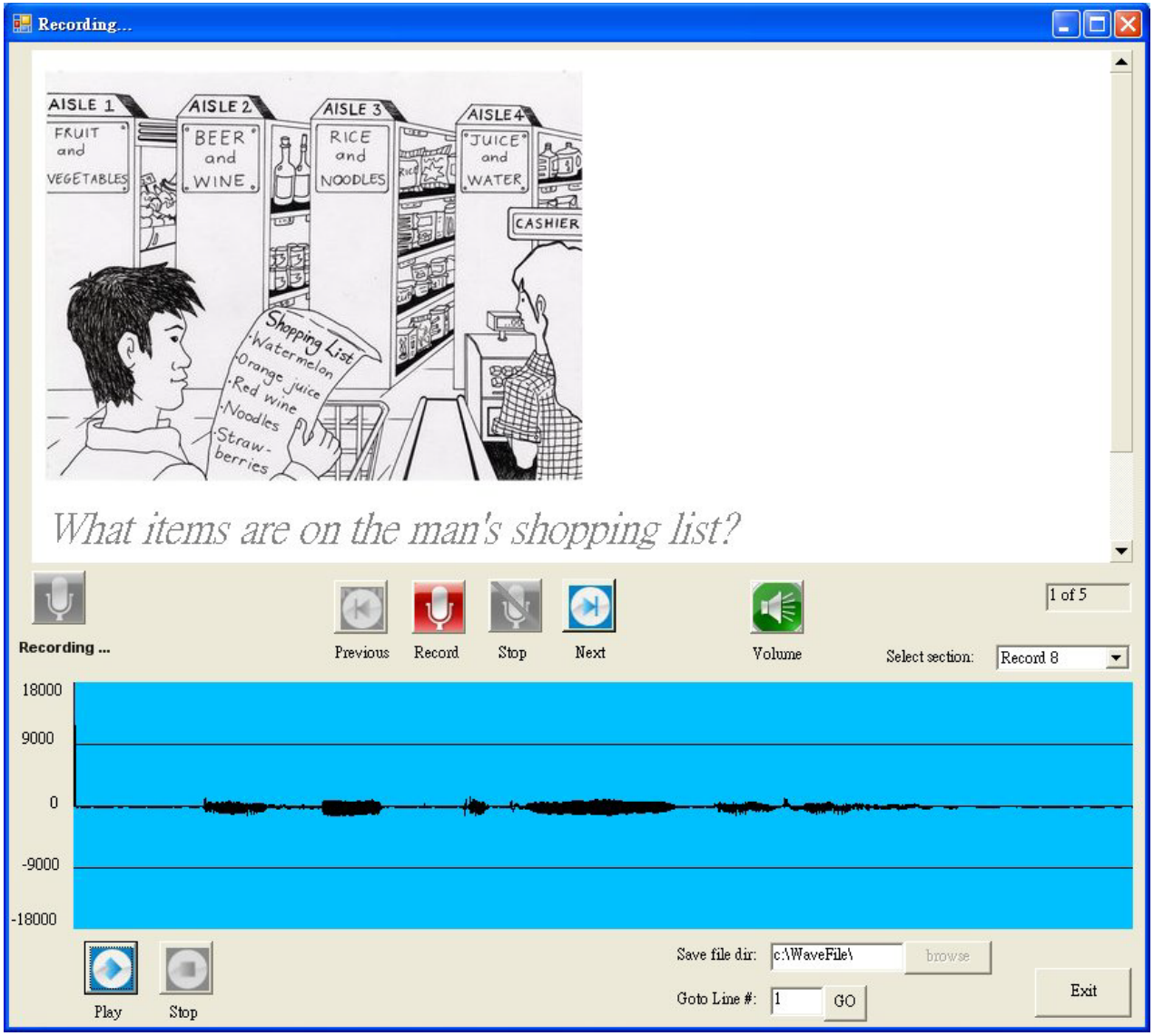
Example of Task 4



Example of Task 7



Task 8


Recording...
AISLE 1
FRUIT and VEGETABLES
AISLE 2
BEER and WINE
AISLE 3
RICE and NOODLES
AISLE 4
JUICE and WATER
CASHIER
Shopping List
·Watermelon
·Orange juice
·Red wine
·Noodles
·Straw-berries
What items are on the man's shopping list?
Recording ...
Previous
Record
Stop
Next
Volume
1 of 5
Select section:
Record 8
18000
9000
0
-9000
-18000
Play
Stop
Save file dir:
c:\WaveFile\
browse
Goto Line #:
1
GO
Exit

Picture of Task 8

AESOP Data: further issues (1/2)

- Current Problems
  - Mostly read speech
  - Not catering for other research purposes
- Other research purposes
  - Speech technology
    - Large speech samples
    - Main users=Asian English speakers → more L2 English data
    - L2 English-L2 English interaction:
      e.g. Japanese English vs. Chinese English
  - Education
    - Data of different level speakers
    - Grammar & Pragmatic skill is important in communication

AESOP Data: further issues (2/2)

- Phrase accent and intonation
  - Important in communication
  - Convey meaning
- Syntax, pragmatics, vocabulary
  - Crucial in communication
  - Important in L2 assessment

⇒ spontaneous speech data

- Problems of data analysis
  - Automatic phoneme alignment labelling using HTK format

(by Shigeki Matsuda, NICT)

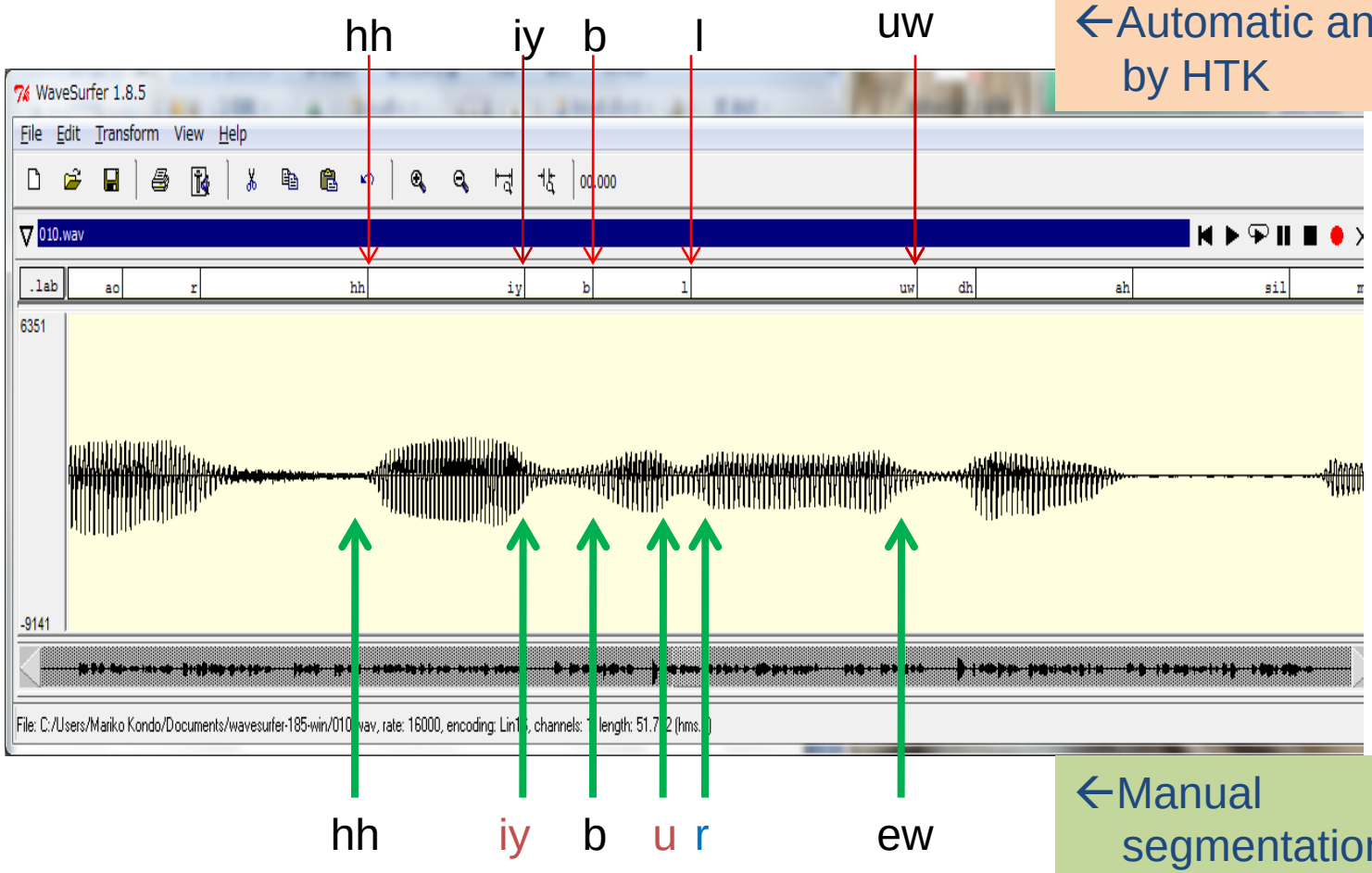
→ Not adjusted to L2 specific phenomena

Waseda AESOP Data (1/4)

- Vowel insertion

/bl/ → /bur/

"…he blew the …" → "… he burew the …"



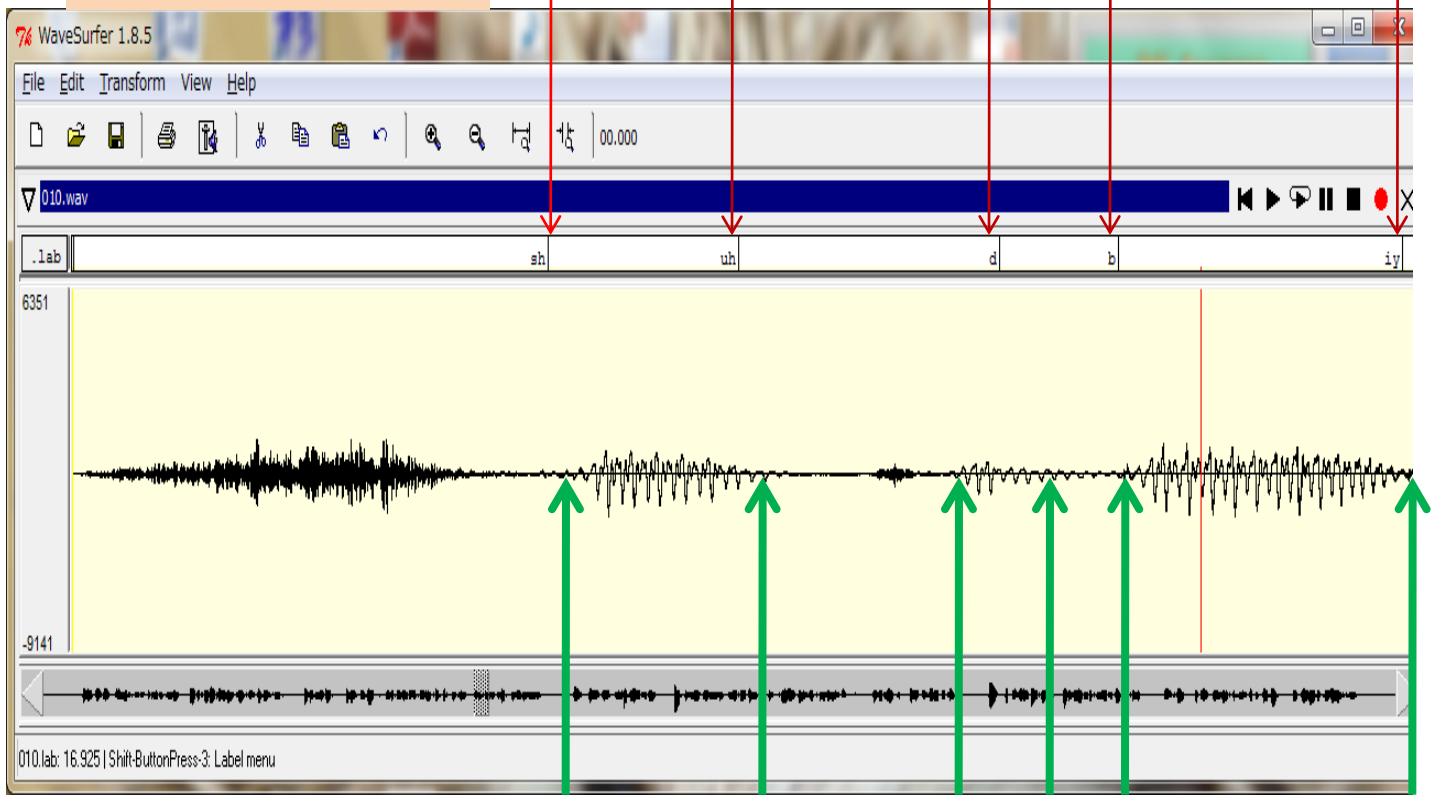
←Automatic annotation by HTK

←Manual segmentation

Waseda AESOP Data (2/4)

"… should be considered …" → …shouldu be considered …"

Automatic Annotation by HTK →



Manual segmentation →

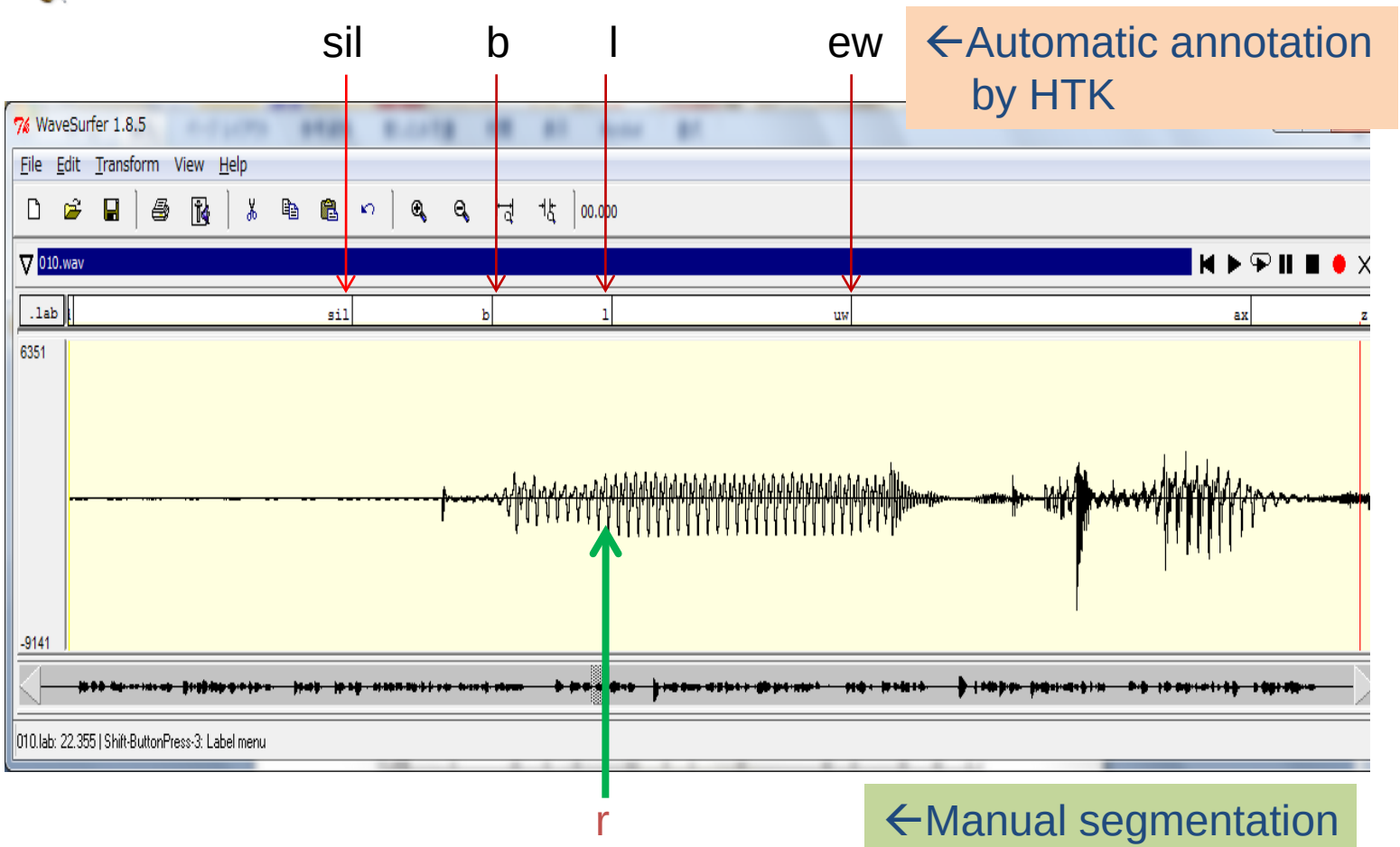
sh uh d u b iy

Waseda AESOP Data (3/4)

- Segmental changes

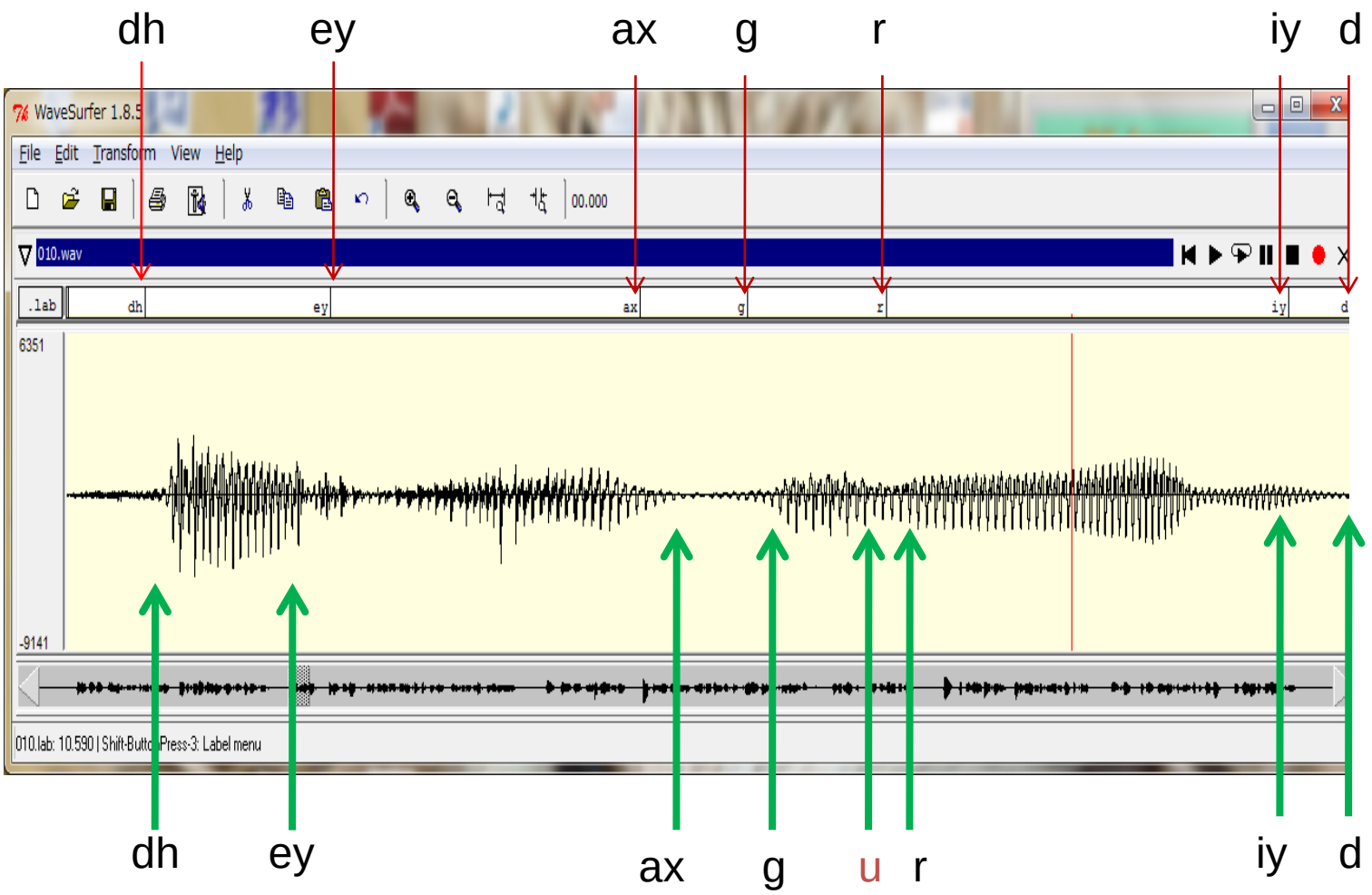
/l/ → /r/

"…wind blew as hard…" → "…wind brew as hard…"



Waseda AESOP Data (4/4)

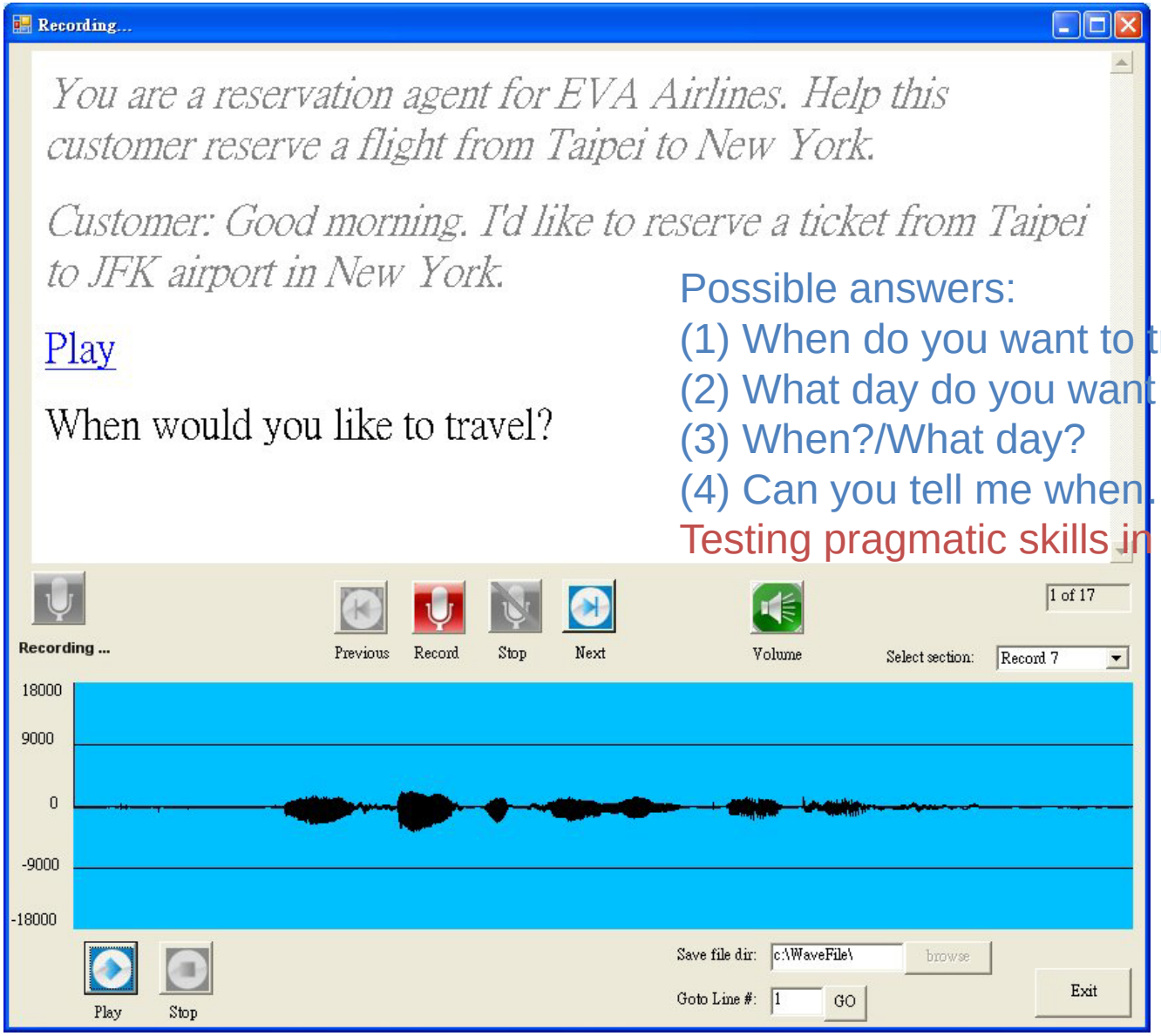
"… they agreed …" → "… they agureed …"



← Automatic annotation by HTK

←Manual segmen-tation

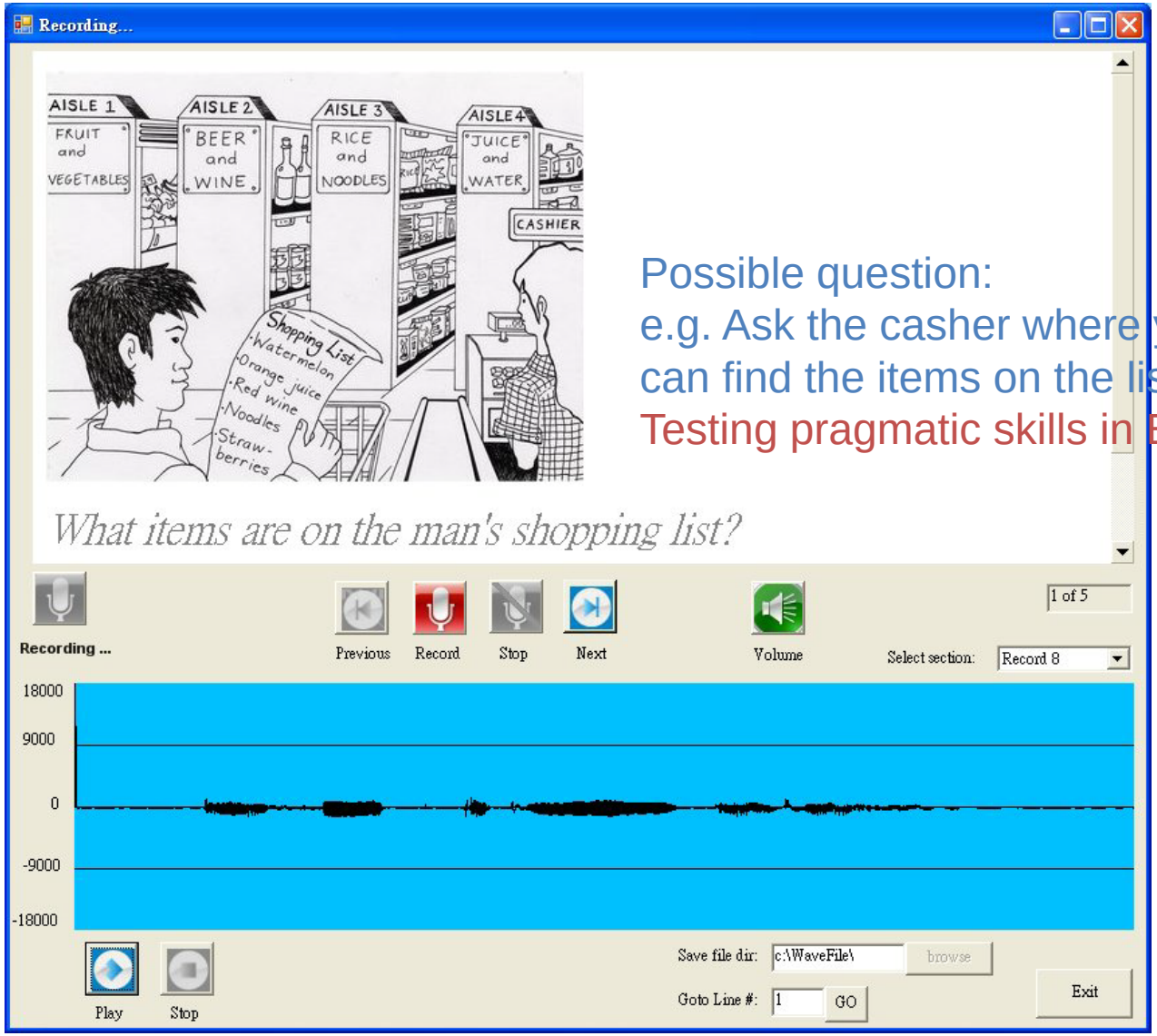
e.g. Task 7 -> change to free answer question (?)



Possible answers:
(1) When do you want to travel?
(2) What day do you want to fly?
(3) When?/What day?
(4) Can you tell me when….? etc.
Testing pragmatic skills in English.

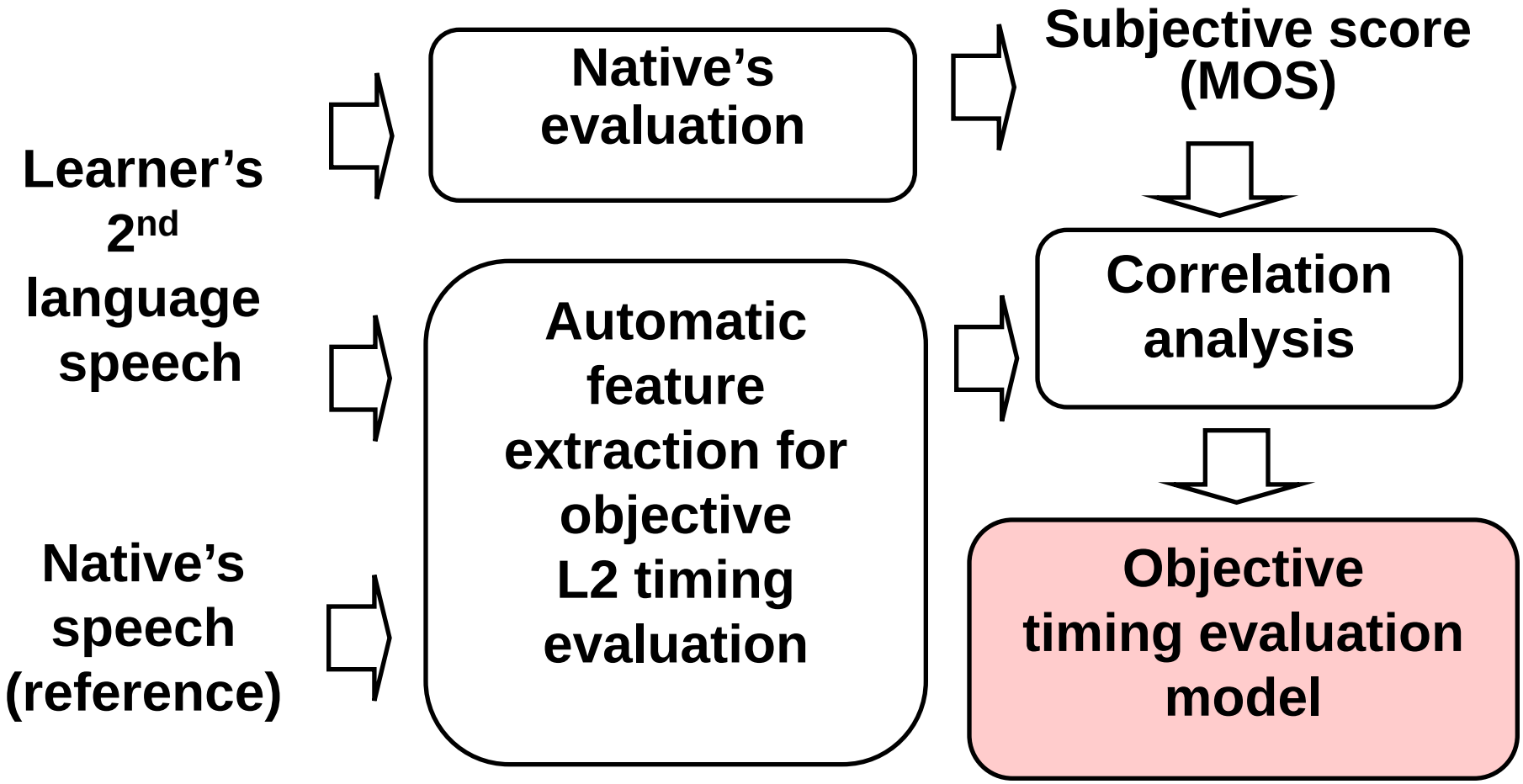
e.g. Task 8 →

Change questions to more spontaneous responses



Possible question:
e.g. Ask the casher where you can find the items on the list.
Testing pragmatic skills in English.

L2 speech timing evaluation modeling



L2 speech timing characteristics

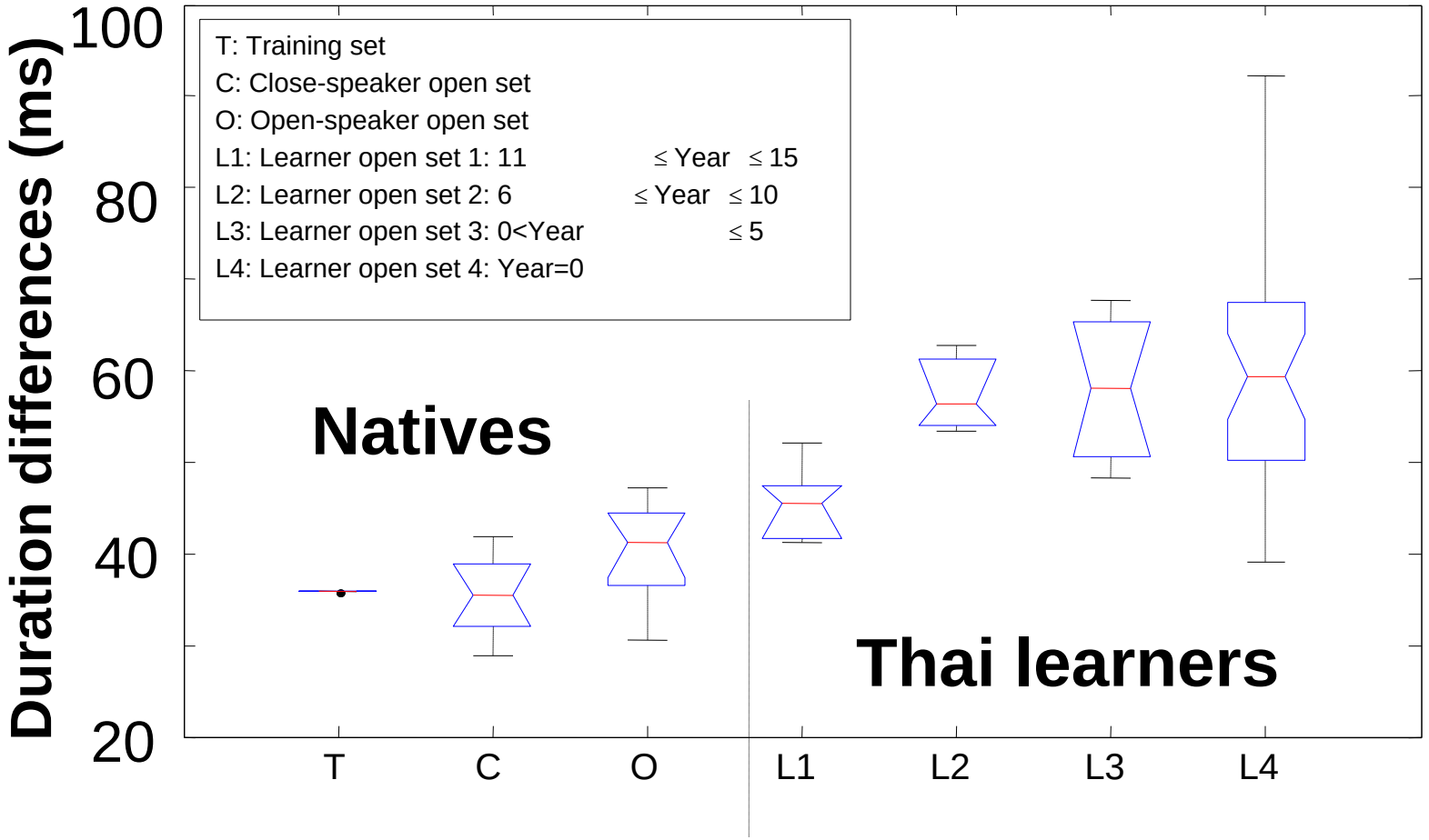
Interference of L1 (Japanese)

- Open syllable only (Vowel insertion)
- Mora timing (Function word lengthening)

Beginner's characteristics

- Slower tempo for longer sentences
- Pause insertion (Frequent & long pause)

Thai learners' duration differences from predicted English durations

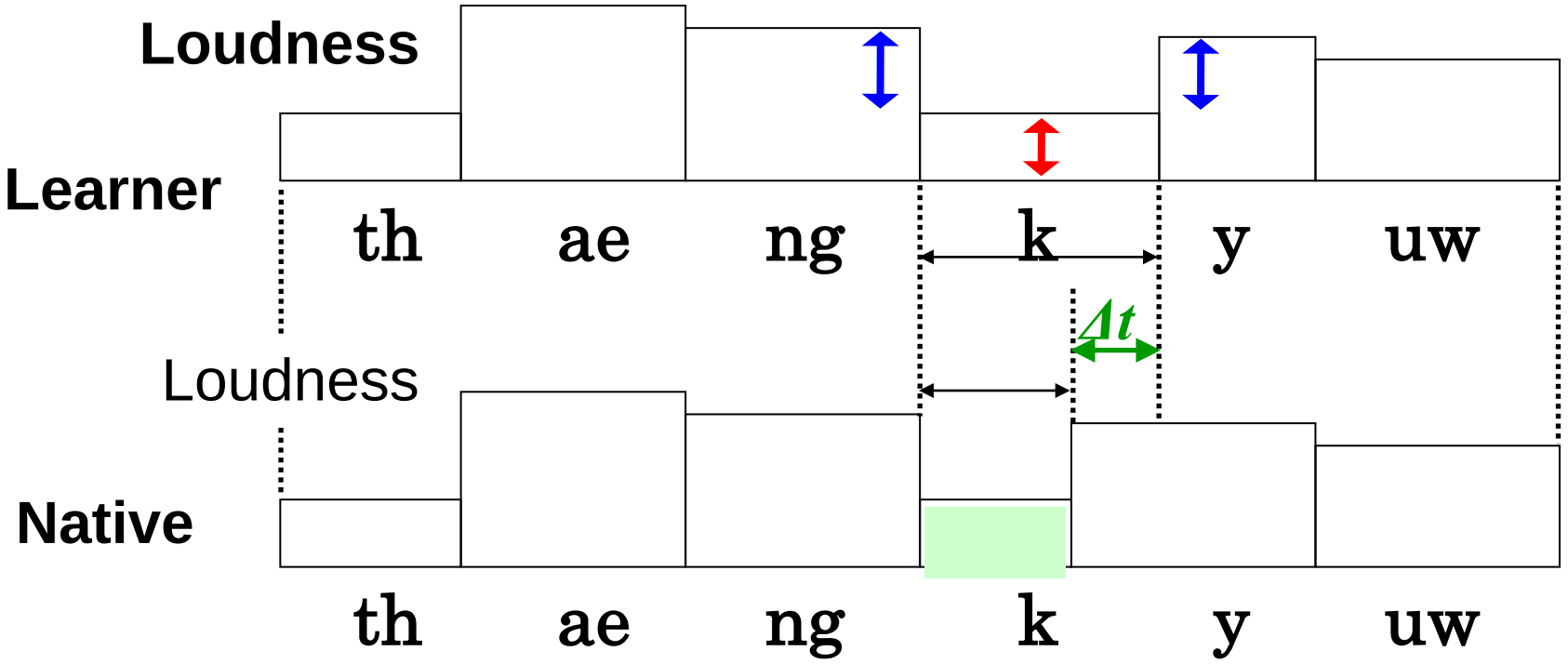


Loudness weighted perceptual measure

Duration difference Δt weighted by

Loudness of the current segment

Loudness jumps from the adjacent segments



Higher correlation using a loudness-weighted perceptual measure