Constructing Asian English Speech Corpus for Universal Purposes

Mariko Kondo\textsuperscript{1,2} & Yoshinori Sagisaka\textsuperscript{1,3}
\textsuperscript{1}Language and Speech Science Laboratory
\textsuperscript{2}SILS & \textsuperscript{3}GITI
Waseda University
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
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**.
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 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]**?
(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
## (1) Consonants

Japanese consonants

<table>
<thead>
<tr>
<th></th>
<th>bilabial</th>
<th>labio-dental</th>
<th>dental</th>
<th>alveolar</th>
<th>post-alveolar</th>
<th>alveolo-palatal</th>
<th>palatal</th>
<th>velar</th>
<th>glottal</th>
</tr>
</thead>
<tbody>
<tr>
<td>plosives</td>
<td>p</td>
<td>b</td>
<td>t</td>
<td>d</td>
<td></td>
<td></td>
<td></td>
<td>k</td>
<td>g</td>
</tr>
<tr>
<td>nasals</td>
<td>m</td>
<td>n</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>η</td>
<td>η</td>
</tr>
<tr>
<td>fricatives</td>
<td>ɸ</td>
<td>s</td>
<td>z</td>
<td>ɕ</td>
<td>ɕ</td>
<td></td>
<td>h</td>
<td></td>
<td></td>
</tr>
<tr>
<td>affricates</td>
<td>ts</td>
<td>dz</td>
<td></td>
<td>tc</td>
<td>dz</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>lateral</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>approximants</td>
<td>(w)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>j</td>
<td>(w)</td>
</tr>
<tr>
<td>Tap/flap</td>
<td></td>
<td></td>
<td></td>
<td>r</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
**Language Specific Features (L1 Japanese) (2/4)**

**English consonants**

<table>
<thead>
<tr>
<th></th>
<th>bilabial</th>
<th>Labiodental</th>
<th>dental</th>
<th>alveolar</th>
<th>post-alveolar</th>
<th>palato-alveolar</th>
<th>palatal</th>
<th>velar</th>
<th>glottal</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>plosives</strong></td>
<td>p</td>
<td>b</td>
<td>t</td>
<td>d</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>k</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>g</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(?)</td>
</tr>
<tr>
<td><strong>nasals</strong></td>
<td>m</td>
<td></td>
<td>n</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>η</td>
</tr>
<tr>
<td><strong>fricatives</strong></td>
<td>f</td>
<td>v</td>
<td>θ</td>
<td>ð</td>
<td>s</td>
<td>z</td>
<td>j</td>
<td></td>
<td>h</td>
</tr>
<tr>
<td><strong>affricates</strong></td>
<td>ts</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>tʃ</td>
</tr>
<tr>
<td><strong>lateral</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>dʒ</td>
</tr>
<tr>
<td><strong>approximants</strong></td>
<td>(ʍ)</td>
<td>(w)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>j</td>
<td></td>
<td>(ʍ)</td>
</tr>
<tr>
<td></td>
<td>(w)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(w)</td>
</tr>
</tbody>
</table>

**Non-phonemic in Japanese**

**No phonemic contrast in Japanese**

**Contextual problems for Japanese speakers**
(2) Vowel system

<table>
<thead>
<tr>
<th>Japanese</th>
<th>vs.</th>
<th>English</th>
</tr>
</thead>
<tbody>
<tr>
<td>5 vowels</td>
<td></td>
<td>GenAm: 15 vowels</td>
</tr>
<tr>
<td></td>
<td></td>
<td>RP: 19 vowels</td>
</tr>
<tr>
<td></td>
<td></td>
<td>AU: 18-19 vowels</td>
</tr>
<tr>
<td></td>
<td></td>
<td>NZ: 17-19 vowels</td>
</tr>
<tr>
<td></td>
<td></td>
<td>CAN: 14 vowels</td>
</tr>
</tbody>
</table>

- **Timing:** vowel sequences vs. diphthongs, vowel weakening (stress & rhythm)
(3) Syllable structures

<table>
<thead>
<tr>
<th>Japanese</th>
<th>vs.</th>
<th>English</th>
</tr>
</thead>
<tbody>
<tr>
<td>(C)(j)(V)(V)(C)(C)</td>
<td></td>
<td>(C)(C)(C)V(C)(C)(C)</td>
</tr>
</tbody>
</table>

\*CC

No deletion rule  \rightarrow  vowel epenthesis

stranged  /streɪndʒd/  \rightarrow  |sutore(i)Nzido|

\[\text{u o i o}\

strengths  /streŋθs/  \rightarrow  |sutoreNgusu|

\[\text{u o u u}]

Japanese vs. English (C)(j)(V)(V)(C)(C) vs. (C)(C)(C)V(C)(C)(C)
### Target words

<table>
<thead>
<tr>
<th></th>
<th>2-1</th>
<th>3-1</th>
<th>3-2</th>
<th>3-3</th>
<th>4-1</th>
<th>4-2</th>
<th>4-3</th>
<th>4-4</th>
<th>LH</th>
<th>RH</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Y-N</strong> (rise)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>money</td>
<td>wonderful</td>
<td><strong>apartment</strong></td>
<td>overnight</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>white wine</td>
<td></td>
</tr>
<tr>
<td><strong>WH</strong> (fall)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>elevator</td>
<td>available</td>
<td>information</td>
<td>misunderstand</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>supermarket</td>
<td></td>
</tr>
<tr>
<td><strong>Cont.</strong> (rise)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>January</td>
<td>experience</td>
<td>California</td>
<td>Vietnamese</td>
<td>department store</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Decl.</strong> (fall)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>morning</td>
<td>video</td>
<td><strong>tomorrow</strong></td>
<td><strong>Japanese</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>afternoon</td>
<td></td>
</tr>
<tr>
<td>Narrow focus</td>
<td>money</td>
<td>morning</td>
<td>wonderful video</td>
<td>apartment</td>
<td><strong>tomorrow</strong></td>
<td><strong>Japanese</strong></td>
<td>overnight</td>
<td>elevator</td>
<td>January</td>
<td>available</td>
</tr>
</tbody>
</table>

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** /'eləvətə(r)/
- apartment /ə'pɑː(r)tmənt/
- **tomorrow** /tə'mɔːrəʊ/tə'mɔːrəʊ/  
- **Japanese** /ˌdʒɑːpə'niːz/  
- (white wine /'waɪt 'waɪn/)
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

I said apartment five times.

/ə/ /ɑː(r)/
Example of Task 4

If the birthday party wasn't for Mary, then who was it **for**?

/ʃə/  /fɔː/
Example of Task 7

You are a reservation agent for EVA Airlines. Help this customer reserve a flight from Taipei to New York.

Customer: Good morning. I'd like to reserve a ticket from Taipei to JFK airport in New York.

Play

When would you like to travel?
Task 8

What items are on the man's shopping list?
Picture of Task 8

Shopping List:
- Watermelon
- Orange juice
- Red wine
- Noodles
- Strawberries
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

- 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
“… should be considered …”  ➞  …shouldu be considered …”

Automatic Annotation by HTK

Manual segmentation
Segmental changes

/l/ \rightarrow /r/

“...wind blew as hard...”  \rightarrow  “…wind brew as hard...”

Automatic annotation by HTK

Manual segmentation
“... they agreed ...”

Automatic annotation by HTK

Manual segmentation
You are a reservation agent for EVA Airlines. Help this customer reserve a flight from Taipei to New York.

Customer: Good morning. I'd like to reserve a ticket from Taipei to JFK airport in New York.

Play

When would you like to travel?

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 cashier where you can find the items on the list.
Testing pragmatic skills in English.
L2 speech timing evaluation modeling

Learner’s 2nd language speech

Native’s speech (reference)

Automatic feature extraction for objective L2 timing evaluation

Native’s evaluation

Subjective score (MOS)

Correlation analysis

Objective timing evaluation model
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

T: Training set
C: Close-speaker open set
O: Open-speaker open set
L1: Learner open set 1: 11 ≤ Year ≤ 15
L2: Learner open set 2: 6 ≤ Year ≤ 10
L3: Learner open set 3: 0 < Year ≤ 5
L4: Learner open set 4: Year = 0
Loudness weighted perceptual measure

Duration difference $\Delta t$ weighted by
Loudness of the current segment
Loudness jumps from the adjacent segments

Learner

Loudness

th ae ng k y uw

Native

th ae ng k y uw
Higher correlation using a loudness-weighted perceptual measure.

The chart shows the correlation between subjective scores with and without loudness weighting for different sentence lengths: All sentences, long sentences, and short sentences. The correlation is higher with loudness weighting compared to without it.