Pausing patterns in the first and second language speech of native Japanese speakers

Ralph ROSE <rose@waseda.jp>
Center for English Language Education (CELESE)
Faculty of Science and Engineering
Waseda University; Tokyo, Japan

10 Maio 2017
Faculdade de Letras
Universidade de Lisboa
Acknowledgments

● Grants
  ○ Japan Society for the Promotion of Sciences (JSPS) Grants-in-aid #24520661, #24520661
  ○ Waseda University Special Research Grant #2011B-512

● Research support staff: Richard Varela, Wataru Okuzumi, Yutaka Shirasugi, Hiroaki Suzuki, Junichi Inagaki, Masayuki Motoori, Yukikatsu Fukuda, Tatsuhiro Nomaguchi, Aiko Ooe, Maiko Serizawa

● Colleagues: Chris Sheppard, Hinako Masuda, Helena Moniz, Michiko Watanabe

● Spoken Language Systems Lab (L²F)

● Universidade de Lisboa
Features of fluency: Case of Kei Nishikori

- **Japanese speech**
  - Filled pauses: “e-”
  - Very few silent pauses
  - Some fixed expressions

- **English speech**
  - Filled pauses: “uh” / “um”
  - Very few silent pauses
  - Discourse markers: “you know”, “kind of”

https://youtu.be/CQuAZGyEsu0
Fluency

● Scope of fluency
  ○ Broad: speak a language proficiently
  ○ Narrow: speak smoothly with minimal but natural hesitation

● Segalowitz (2010) taxonomy of fluency types
  ○ Cognitive fluency (in speech planning)
  ○ Utterance fluency (in speech production/articulation)
  ○ Perceived fluency (from listener's perspective)

● Numerous investigations of second language fluency development (De Jong et al 2012, inter alia)

● Numerous investigations of classroom methods for developing L2 fluency (Nation 1989, De Jong and Perfetti 2011, inter alia)
Temporal variables in utterance fluency

Silent pauses longer than 0.3-1.0sec

Self-corrections (repairs) Sequence that repairs a preceding sequence

Filled pauses
uh/um (English)
e-to/ano- (Japanese)

Lengthenings Prolongation of one or more syllables
I'll take the blue and the red ones.

Repeats/Restarts Repetition of a sequence of words
I I I I think that's a good idea.

False starts Beginning of utterance that is abandoned
Do you I disagree with that.

Speech rate by word, by syllable, with/without pauses

Crosslinguistic Corpus of Hesitation Phenomena

- CCHP (Rose 2013)
- Participants: L2 learners of varying proficiency levels
- Elicitation tasks (both L1 and L2)
  - Spontaneous speech: picture description, topic narrative
  - Reading aloud
- Annotation
  - Transcript with FPs, repairs, etc.
  - Two annotators, one checker
  - Temporal measurements: auto (Quéné et al 2011) & manual
CCHP: Basic statistics

- Participants: 35 Japanese L1 / English L2 speakers

<table>
<thead>
<tr>
<th></th>
<th>Word count</th>
<th>Time</th>
</tr>
</thead>
<tbody>
<tr>
<td>Read speech</td>
<td>21,406</td>
<td>2 hr, 41 min</td>
</tr>
<tr>
<td>Picture description</td>
<td>19,732</td>
<td>4 hr, 39 min</td>
</tr>
<tr>
<td>Topic narrative</td>
<td>21,138</td>
<td>4 hr, 35 min</td>
</tr>
<tr>
<td>Total</td>
<td>62,276</td>
<td>11 hr, 56 min</td>
</tr>
</tbody>
</table>

Hesitation phenomena

- 15,480 silent pauses
- 3,741 filled pauses
- 1,635 repairs
- 566 repeats

Transcriber agreement = 91.5%
L1-L2 temporal pattern correspondence

- **Articulation rate (syll/sec)**
  - \( r = 0.41 \)
  - \( p < 0.001 \)

- **Silent pause rate (per min)**
  - \( r = 0.34 \)
  - \( p < 0.005 \)

- **Mean pause duration (lgs)**
  - \( r = 0.64 \)
  - \( p < 0.001 \)

- **Filled pause rate (per min)**
  - \( r = 0.58 \)
  - \( p < 0.001 \)

- **Repair rate (per min)**
  - \( r = 0.41 \)
  - \( p < 0.001 \)

**High L1-L2 correlation:**
- L2 speech explained by L1 speech behavior

**Low L1-L2 correlation:**
- More likely predictors of L2 fluency development
Crosslinguistic communication

- Native speaker (blue) talking to nonnative listener (pink)
- Nonnative speaker (pink) talking to native listener (blue)
- Nonshared native language
- Native language
- Shared native language
Crosslinguistic speech perception

- Perception of nonnative speech modulated by listener's language background
  - Wester and Mayo 2014 – nonnative listeners judge accentedness more harshly than native listeners
  - Bent and Bradlow 2003 – nonnative listeners comprehend better than native listeners

- Crosslinguistic perceptions of fluency, too
  - Rossiter 2009
    - Fluency ratings: nonnative listeners < native listeners
    - Both native/nonnative listeners' ratings correlated with articulation rate and pause frequency
  - Foote and Trofimovich 2016 – native listeners attend to pause frequency; nonnative listeners to speech rate
Fluency ratings

- Extracted 7 30-second clips of English speech per speaker
  - Reading aloud x 1
  - Picture description x 3
  - Topic narrative x 3

- Rater instructions
  - Rate fluency on 9-point scale (1 – low … 9 – high)
  - Rate “smoothness” of the speech
  - Ignore pronunciation, grammar, word-choice, etc.
Fluency ratings by native listeners

- Obtained via Amazon Mechanical Turk
- Workers limited to native English speakers (self-reported)
- Used attention checks and monitoring of audio player activity to check that instructions were followed.
Fluency ratings by nonnative listeners

- Obtained via local web pages in computer lab
- Participants limited to native Japanese speakers (recruited)
- Used attention checks and monitoring of audio player activity to check that instructions were followed.
Fluency rating results

- Participants: 34 native English raters; 20 native Japanese raters
- Nonnative (Japanese) raters judge fluency lower than native (English) raters (similar to previous work)
- Reading aloud judged more fluent than other tasks

<table>
<thead>
<tr>
<th>Effect</th>
<th>DFn</th>
<th>DFd</th>
<th>F</th>
<th>p</th>
<th>p&lt;.05</th>
</tr>
</thead>
<tbody>
<tr>
<td>raterlang</td>
<td>1</td>
<td>52</td>
<td>7.581507</td>
<td>8.104075e-03</td>
<td>*</td>
</tr>
<tr>
<td>task</td>
<td>2</td>
<td>104</td>
<td>485.145647</td>
<td>1.850916e-53</td>
<td>*</td>
</tr>
<tr>
<td>raterlang:task</td>
<td>2</td>
<td>104</td>
<td>2.312539</td>
<td>1.040802e-01</td>
<td></td>
</tr>
</tbody>
</table>
Fluency rating results

- Linear regression modeling (using \texttt{lm} in \texttt{R})
  - Dep. var: fluency rating
  - Ind. vars: articulation rate, pause rate, pause length, filled pause rate, repair rate, rater’s listener status, speech task
- With full model, all variables significant except task
- But by rater group, relevant temporal features are different

<table>
<thead>
<tr>
<th>native raters</th>
<th>nonnative raters</th>
</tr>
</thead>
<tbody>
<tr>
<td>✓ higher articulation rate</td>
<td>✓</td>
</tr>
<tr>
<td>✓ lower pause rate</td>
<td>✓</td>
</tr>
<tr>
<td>✓ shorter pause length</td>
<td>✓</td>
</tr>
<tr>
<td>✓ higher filled pause rate</td>
<td>✓</td>
</tr>
<tr>
<td>✓ lower repair rate</td>
<td>✓</td>
</tr>
</tbody>
</table>

Higher fluency associated with...

\[ F(7, 406) = 64.9, \ p < 0.001; \ \text{adjusted } R^2 = 0.52 \] \quad \text{[R}^2\text{=0.45] [R}^2\text{=0.57]}
Summary, so far

- **Pause duration** and filled pause rate are highly correlated between first and second language speech.
- Hence, articulation rate, silent pause rate and repair rate are more likely to be predictive of fluency development.
- Articulation rate and **pause duration** are most reliable predictive factors of perceptual fluency for both native and nonnative listeners.

Conundrum: pause duration!

So, be like Kei Nishikori!
Automated assessment of L2 speech

- Pronunciation (with visual feedback*)
  - Segmental: Cucchiarini et al 2009; Patten and Edmonds 2013*

- Fluency
  - ETS SpeechRater (Zechner et al 2009)
  - Versant (Pearson, Ordinate; Bernstein 1999)
  - CASEC (Hayashi et al 2004)

Useful overviews: Eskenazi 2009; Gamper and Knapp 2010
Feedback to learner

- Eskenazi (1999) - “Learners must receive pertinent corrective feedback”
- Most systems provide rapid feedback.
- In human-human communication, some feedback is in real-time
  - Back-channeling (uh-huh)
  - Head movements (nodding, shaking)
  - Facial expressions
- Is it possible to provide real-time feedback on fluency-related matters in human-computer interaction?
Fluidity: fundamental aims

- Measure various utterance fluency characteristics and update them in real-time.
- Provide real-time feedback to learner on utterance fluency measures.
- Provide opportunity for learner to review their production together with visual representation of fluency measures.
- Provide feedback in a manner that emulates human-human communication.

A work in progress!
Fluidity: fluency measures

- Phonation time
- Silence time
- Syllable count: energy peaks
  (cf., Bhat et al 2010)
- Silent pause count: silence > 300ms
  (cf., De Jong and Bosker 2013)
- Filled pause count: stable formants and pitch
Fluidity: main window

- Requires Java SE 6 or greater
- Detection settings
- Audio input settings
- Audio level meter
- Fluency measure indicators

Uses TarsosDSP (Joren Six) and AudioInfo.java (Jonathan Simon) libraries

- Transition threshold
- Pause threshold
- Silence threshold
- Filled pause sensitivity
- Smoothing factor
- Elapsed time: 00:00:00
- Speech time: 00:00:00
- Pause time: 00:00:00
- Syllable count: 0
- Silent pause count: 0
- Filled pause count: 0

- 8000 Hz
- 11025 Hz
- 16000 Hz
- 22050 Hz
- 44100 Hz
- 48000 Hz
- 8-bit
- 16-bit
Fluidity: playback window

- Playback controls
- Waveform representation
- Fluency visualization

- Filled pauses
- Speech
- Silent pauses
Fluidity: usability testing

- Participants (n=21)
- Procedure
  - Practice speaking with Fluidity.
  - Adjust settings to fit their production.
  - Respond to survey questions about the experience.
Fluidity: user response

- Fluidity was easy to use.
- I enjoyed using Fluidity.
- The indicators gave accurate measurements without any adjustment.

- I could adjust my speech based on feedback from the indicators.
- Fluidity made me think about English fluency in a new or different way.
- Fluidity would make a good commercial application.
Fluidity: desirable features

- **Capability to save recordings**
- **Capability to import sample recordings of native speakers**
- **Background noise through headphones to simulate different environments**

- **Animated face/head that blinks its eyes in a realistic manner**
- **Animated face/head that nods in a realistic manner**
- **Animated face/head that interrupts me if I've been silent too long**
Fluidity: user comments

● 「語学を専攻していましたが、発音や文法にとらわれることが多く、流暢さを考えることがあまりなかったので、勉強になりました。」
  ○ Although I majored in languages, I have mostly studied about pronunciation and grammar and have not studied much about fluency. So, this was very educational.

● 「具体的にどうすれば良いかは分かりませんが、この「Fluidity」を基板としたゲーム形式のアプリを使えば、すごく楽しく使えるかと思います。」
  ○ I wasn't really sure how to make use of Fluidity objectively, but if I could use it like a game application, I think it would be very enjoyable to use.
Putting things together...

● Different audiences have different perceptions of fluency
  ○ Native vs. nonnative
  ○ Shared native language vs. non-shared native language

● Audience description is critical aspect of language program design (cf., CEFR guidelines)
  ○ Emphasize articulation rate?
  ○ Emphasize minimal silent pauses?
  ○ Emphasize productive use of filled pauses?

● Various utterance fluency profiles are needed …

● To guide learners with possibly different needs
Goals during this year

- Implement profiles in Fluidity
- Test them with Portuguese (and other) students
- Return to Japan to use with students there.
- (And distribute application widely, of course.)

Muito Obrigado!
References


Universidade de Lisboa

Pausing Patterns in L1/L2 Speech

R. Rose