Differences in second language speech fluency ratings: Native versus nonnative listeners

Ralph L. ROSE

e-mail: rose@waseda.jp  Twitter: @fildpauz

Center for English Language Education (CELESE)
Faculty of Science and Engineering
Waseda University; Tokyo, Japan

Acknowledgments: Chris Sheppard, Wataru Okuzumi, Yutaka Shirasugi, Hinako Masuda, Hiroaki Suzuki, Junichi Inagaki, Masayuki Motoori, Yukikatsu Fukuda, Tatsuhiro Nomaguchi, & Aiko Ooe

Supported by Japan Society for the Promotion of Sciences (JSPS)
Grants-in-Aid #15K02765 (PI, Ralph Rose)

Fluency & Disfluency Across Languages and Language Varieties
Louvain-la-Neuve, Belgium; 15-17 February 2017
Crosslinguistic communication

- Native speaker
- Non-native listener
- Non-native speaker
- Native listener
- Non-native speaker
- Non-native listener
- Non-shared native language
- Shared native language
Overview

• Background
  – Crosslinguistic speech perception

• Method
  – Crosslinguistic Corpus of Hesitation Phenomena
  – Fluency ratings
    • by native listeners
    • by nonnative listeners

• Results

• Discussion

• Accessing the CCHP
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
Research questions

• Do native listeners and nonnative listeners rate the fluency of nonnative speech differently?
  – Particularly when the nonnative listeners share the same native language as the speakers?

• What temporal features of nonnative speech do native and nonnative listeners each attend to in their judgments of fluency?
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 (Quené et al 2011)
## CCHP: Basic Statistics

- **Participants:** 35 Japanese L1 / English L2 speakers

<table>
<thead>
<tr>
<th>Activity</th>
<th>Word count</th>
<th>Time</th>
<th>Hesitation phenomena</th>
</tr>
</thead>
<tbody>
<tr>
<td>Read speech</td>
<td>21,406</td>
<td>2 hr, 41 min</td>
<td>• 15,480 silent pauses</td>
</tr>
<tr>
<td>Picture description</td>
<td>19,732</td>
<td>4 hr, 39 min</td>
<td>• 3,741 filled pauses</td>
</tr>
<tr>
<td>Topic narrative</td>
<td>21,138</td>
<td>4 hr, 35 min</td>
<td>• 1,635 repairs</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td>62,276</td>
<td><strong>11 hr, 56 min</strong></td>
<td>• 566 repeats</td>
</tr>
</tbody>
</table>

Transcriber agreement = 91.5%
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.

(Previously reported in part in Rose 2015)
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.
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

![Box plot showing fluency differences across tasks](image)

<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>
Results

- Linear regression modeling (using `lm` in R)
  - Dep. variable: fluency rating
  - Ind. variables: 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>Feature</th>
<th>Native Raters</th>
<th>Nonnative Raters</th>
</tr>
</thead>
<tbody>
<tr>
<td>Higher articulation rate</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Lower pause rate</td>
<td></td>
<td>✓</td>
</tr>
<tr>
<td>Shorter pause length</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Higher filled pause rate</td>
<td></td>
<td>✓</td>
</tr>
<tr>
<td>Lower repair rate</td>
<td></td>
<td>✓</td>
</tr>
</tbody>
</table>

Higher fluency associated with:

- [F(7, 406) = 64.9, p < 0.001; adjusted R² = 0.52]
- [R² = 0.45] [R² = 0.57]
Discussion

- Native and nonnative raters both attend to articulation rate and silent pause length.
- Native raters do not attend to silent pause rate
  - Native listeners have comprehension disadvantage (Bent & Bradlow 2003)
  - Increased silent pauses may aid native listeners' comprehension, ...
  - But they may not be perceptually salient
- Nonnative raters do not attend to repair rate
  - Recognition of repairs requires higher syntactic competence.
  - Some participants may lack this competence.
  - Hence, they do not notice repairs.
Discussion

● Native raters do not attend to filled pause rate
● Nonnative raters evaluate filled pauses more highly.
● Compare to other advantages for filled pauses in speech
  − Cues to greater complexity (Watanabe et al 2008)
  − Feeling of another's knowing (Brennan and Williams 1995)
● Filled pauses are more frequent in L1 Japanese than in L1 English (Watanabe and Toyama 2016)
● Perhaps raters regarded a higher filled pause rate as more natural, and hence “smoother”.
Summary

- Nonnative listeners who share native language with speaker judge fluency more harshly than do native listeners.
- Nonnative and native listeners both judge fluency based on articulation rate and silent pause length.
  - Native listeners also attend to repair rate.
  - Nonnative listeners attend to silent pause rate, and
  - Judge a higher filled pause rate more positively.
- Differences between listener groups may stem from different processing difficulties and language backgrounds.
- Yet undetermined is whether the non-native-language-sharing listeners would perform differently. → Future work
CCHP Public Corpus

- Assembling a public version of the Crosslinguistic Corpus of Hesitation Phenomena is ongoing.
- When complete, audio files and annotated transcripts will be available for free download.
- Some files are already available for download: http://www.filledpause.com/chp/cchp
References

Amazon Mechanical Turk web site. www.mturk.com


