How can we determine, detect, and develop fluency in second language speech?

Ralph L. Rose
<rose@waseda.jp>
Center for English Language Education (CELESE)
Waseda University Faculty of Science and Engineering

Acknowledgments
Hinako Masuda, Hiroaki Suzuki, Junichi Inagaki, Masayuki Motoori, Yukikatsu Fukuda, Tatsuhiro Nomaguchi, Aiko Ooe, & Maiko Serizawa

Spoken Language Systems Lab (L²F)
INESC-ID (Lisbon, Portugal)
11 March 2016

This research is partially supported by Japan Society for the Promotion of Sciences (JSPS) Grants-in-Aid (#24520661, #15K02765)
Introduction

Well, in my own life I'd break it up in stages, when I had a difficult youth. My father wasn't in the house. I've written about this. You know there were times where I've experimented with drugs, and I drank in my teenage years. And what I trace this to is a certain selfishness on my part. I was so obsessed with me, and the reasons that I might be dissatisfied, that I couldn't focus on other people. And I think the process for me of growing up was to recognize that it's not about me, it's about ...

Absolutely. But look, you know, when I find myself taking the wrong step, I think a lot of the times it's because I'm trying to protect myself, instead of trying to do God's work. And so that I think is my own failure

Barack Obama (August, 2008); Saddleback Presidential Forum
Introduction

Well, i- i- i- in my own life I'd- I'd break it up in stages, when uh I had a difficult youth. Uh my father wasn't in the house, uh I've written about this, uh there- uh uh you know there were times where uh I've experimented with drugs, and I drank, uh yeah in my teenage years, and what I trace this to is uh a certain selfishness on my part, I- I was so obsessed with me, and you know the- the reasons that I might be dissatisfied, that I- I- w- I couldn't focus on other people. And uh y- you know I think the process for me of growing up was to recognize that it's not about me, it's about ...

it's about- absolutely, so- so- but- but look, you know, th- the uh wh- when I uh wh- when I find myself um taking the wrong step, I think a lot of the times it's because I'm trying to protect myself, instead of trying to do God's work. And- and- an- and so that I think is- is my own failure

Barack Obama (August, 2008); Saddleback Presidential Forum
Overview

- Determining L2 fluency
  - Hesitation phenomena in speech
  - Crosslinguistic Corpus of Hesitation Phenomena
  - Hearer perception of fluency

- Detecting and developing L2 fluency
  - Feedback in L2 development
  - Fluidity application
  - ...and complexity, too
Fluency

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

• Segalowitz (2010): levels of fluency
  – Cognitive fluency: ease of mental preparation
  – Utterance fluency: smoothness of articulation
  – Perceptual fluency: hearer's view of smoothness

• Many researchers looking at cognitive-utterance dimension
Observations of fluency

Silent pauses
longer than 0.3-1.0 sec

Self-corrections (repairs)
Sequence that repairs a preceding sequence
*Look at the blue the red one over there.*

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

Lengthenings
Prolongation of one or more syllables
*I'll take the blue a-nd the- red ones.*

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

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

Hesitation Phenomena

Speech rate
by word, by syllable, with/without pauses
*(Cucchiarini et al 2010)*

Hesitation phenomena in L2 production

Hesitation phenomena in L2 production

• Gradually, more studies are including L1 observations.
  – Derwing et al (2009) and Cox and Baker-Smemoe (2012) observed that both speech rate and pause rate in L1 and L2 production are correlated.
  – De Jong et al (2015) found measures of L2 articulation rate were more meaningful when “corrected” for L1 speech patterns.

• The current research is part of a project designed to contribute to greater understanding of the relationship between L1 hesitation patterns and L2 hesitation patterns.
Research questions

- What is the relationship between hesitation patterns in L1 and L2 speech?
- What is the developmental trajectory of the use of hesitation phenomena in L2?
- What is the relationship between hesitation patterns in L2 speech and hearer's perception of fluency?
Crosslinguistic Corpus of HP (CCHP)

- Participants: L2 learners of varying proficiency levels
- Elicitation tasks
  - Spontaneous speech: picture description, topic narrative
  - Reading aloud
  - Performed in both L1 and L2
Crosslinguistic Corpus of HP (CCHP)

- Demographic information: age, gender, L2 proficiency (standardized test scores, experience abroad, self-assessment)
- Annotation
  - Transcripts, HP, word & pause intervals
  - Two annotators, one checker
CCHP Results: Basic Statistics

- Participants: 36 Japanese L1 / English L2 speakers
- Full corpus
  - 62,632 words
  - 11 hrs, 31 min
- Spontaneous speech
  - 40,296 words
  - 8 hrs, 43 min
- Read speech
  - 22,336 words
  - 2 hr, 48 min
- Transcriber agreement
  - 91.5%

- 15,837 silent pauses
- 3,516 filled pauses
- 1,689 self-corrections
- 518 repeats
L1-L2 Utterance Flu. vs. Perceptual Flu.

• Aim:
  − Compare L1 vs. L2 utterance fluency characteristics.
  − Examine which utterance fluency characteristics correlate with perceptions of fluency by hearers.
• L1-L2 utterance fluency factors measured with praat script (Quené et al 2011)
• Perceptual fluency
  − Fluency ratings (1=low ... 9=high) obtained via Amazon Mechanical Turk
  − Obtained fluency ratings on 7 30-second clips of L2 speech from all corpus participants.
  − Used attention checks and background monitoring of audio player activity to check that instructions were followed.
L1-L2 Utterance Fluency

Articulation rate (sylls/sec)

- English (L2) vs. Japanese (L1)
  - $t(83) = 15.4$
  - $p < 0.001$
  - $R^2 = 0.48$

Mean pause duration (lgsec)

- English (L2) vs. Japanese (L1)
  - $t(83) = 6.0$
  - $p < 0.001$
  - $R^2 = 0.11$

Silent pause rate (per 100 tokens)

- English (L2) vs. Japanese (L1)
  - $t(83) = 18.5$
  - $p < 0.001$
  - $R^2 = 0.64$

Filled pause rate (per 100 tokens)

- English (L2) vs. Japanese (L1)
  - $t(83) = 2.1$
  - $p < 0.05$
  - $R^2 = 0.02$
Utterance Fluency vs. Perceptual Fluency

- Articulation rate (syll/sec)
  - $t(5310) = 9.2$
  - $p < 0.001$

- Mean pause duration (lgsec)
  - $t(5310) = 39.2$
  - $p < 0.001$

- Silent pause rate (per 100 tokens)
  - $t(5310) = 26.2$
  - $p < 0.001$

- Filled pause rate (per 100 tokens)
  - $t(5310) = 4.9$
  - $p < 0.001$

$R^2 = 0.21$
CCHP results: Summary

- Results show that for utterance fluency, silent pause duration and filled pause rate indicate learners' L2 proficiency.
  - Other L2 hesitation phenomena correlate with those of L1.
- Fluency raters, however, seem to rely on silent pause duration and rate, and less so on articulation rate and filled pause rate.
Future Work with CCHP

• Deeper annotation
  – Syntactic structure
  – Part-of-speech information
  – Syllable and phoneme intervals
  – (F1,F2) measurements

• More speakers

• More L1-L2 combinations
  – Taiwan Chinese L1 – English L2
  – English L1 – French L2
  – English L1 – Spanish L2
  – Korean L1 – English L2
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
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?
Fluency in L2 speech

• Scope
  – Broad: speak a language proficiently
  – Narrow: speak smoothly with minimal but natural hesitation

• Segalowitz (2010): levels of fluency
  – Cognitive fluency: ease of mental preparation
  – Utterance fluency: smoothness of articulation
  – Perceptual fluency: hearer's view of smoothness

Focus of the present work
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
  (cf., Audhkhasi et al 2009)
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

Speaking ideas: A description of a familiar place; How you spent your last vacation; Your most memorable experience; How to play a certain game or sport; Your opinion on an issue with reasons; and anything else...

Transition threshold: 0.1
Pause threshold: 0.3
Silence threshold: -70
Filled pause sensitivity: 0.1
Smoothing factor: 10

Audio input settings:
- Port Realtek Digital Output (Realtek, version 6.3)
- Port Realtek High Definition Audio Controller (Realtek High Definition Audio Controller, version 6.3)

Audio level meter:
- 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
Fluidity: playback window

- **Playback controls**
  - Play
  - Stop
  - Back

- **Waveform representation**

- **Fluency visualization**
  - Filled pauses
  - Speech
  - Silent pauses
Fluidity: usability testing

- Participants (n=14)
- Procedure
  - Practice speaking with Fluidity.
  - Adjust settings to fit their production.
  - Respond to survey questions about the experience.
Fluidity: user response

1. Fluidity was easy to use.
2. I enjoyed using Fluidity.
3. The indicators gave accurate measurements without any adjustment.
4. I could adjust my speech based on feedback from the indicators.
5. Fluidity made me think about how I can improve my English fluency.
6. 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: technical considerations

- Silence threshold often needs to be adjusted.
  - Solution: Do a better job of auto-detecting the sound configuration and microphone settings.

- Filled pause detection is difficult. Even after adjusting sensitivity, many participants did not see their filled pauses detected accurately (or at all).
  - Solution: Try other algorithms for detecting stable formants and pitch.
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.
Fluidity: Summary

- Fluidity is designed to give real-time feedback to L2 English speakers on their utterance fluency.
- The application is still under development, but is capable of providing real-time feedback on most fluency measures, plus visualizations for review.
- Users find the application interesting, fun, and they are motivated to think about how to improve fluency.
- Users noted that the interface is difficult to use, and filled pauses are not accurately detected.
- Future development will include improving accuracy of fluency measurements and creating a user interface that better matches human-human interaction.
Distribution of silent and filled pauses

• Campione and Véronis 2005: Examined SPs and FPs in the Spoken French Reference Corpus
  – SPs and FPs overlap to mark syntactic structure
  – FPs (with following SPs) mark non-syntactic interruptions
  – SP duration is longer at stronger boundaries and before longer clauses
  – FP duration is not affected
• In short, SP and FP are not simply free alternatives when hesitation is needed.
Utterance length in L1 and L2

Utterance length and SP rate

<table>
<thead>
<tr>
<th>Language</th>
<th>t(54)</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Japanese</td>
<td>15.1</td>
<td>&lt; 0.001</td>
</tr>
<tr>
<td>English</td>
<td>2.2</td>
<td>&lt; 0.05</td>
</tr>
</tbody>
</table>

Utterance length (words) vs. SP rate (per 100 words)

$R^2 = 0.65$

Utterance length and FP position

<table>
<thead>
<tr>
<th>Language</th>
<th>t(54)</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Japanese</td>
<td>4.5</td>
<td>&lt; 0.001</td>
</tr>
<tr>
<td>English</td>
<td>2.6</td>
<td>&lt; 0.05</td>
</tr>
</tbody>
</table>

Proportional position of FP vs. Utterance length (words)

$R^2 = 0.01$
**Fluency ... and complexity**

- In pedagogical models of L2 development, complexity, accuracy, and fluency (CAF) are often treated as independent variables.
- Some aspects of utterance fluency may give us insight on the complexity of L2 speakers’ utterances.
Wrapping up

• Determining L2 fluency
  – While many temporal features of L2 utterances are observable, not all of them correlate to a perception of fluency on the part of listeners: Silent pause rate and duration > articulation rate and filled pause rate

• Detecting L2 fluency
  – Fluidity shows that certain aspects of utterance fluency can be detected in (nearly) real-time.

• Developing L2 fluency
  – Fluidity users feel that the real-time feedback on their utterance-level fluency features raises their awareness of their own fluency and provides a basis to improve it.
References

Amazon Mechanical Turk web site. www.mturk.com


