

Speech Hesitation Patterns in English by Native Speakers of Japanese

Ralph L. Rose
<rose@waseda.jp>

Center for English Language Education (CELESE)
Waseda University Faculty of Science and Engineering



Acknowledgments

Hiroaki Suzuki, Junichi Inagaki, Masayuki Motoori,
Yukikatsu Fukuda, & Tatsuhiro Nomaguchi

Introduction

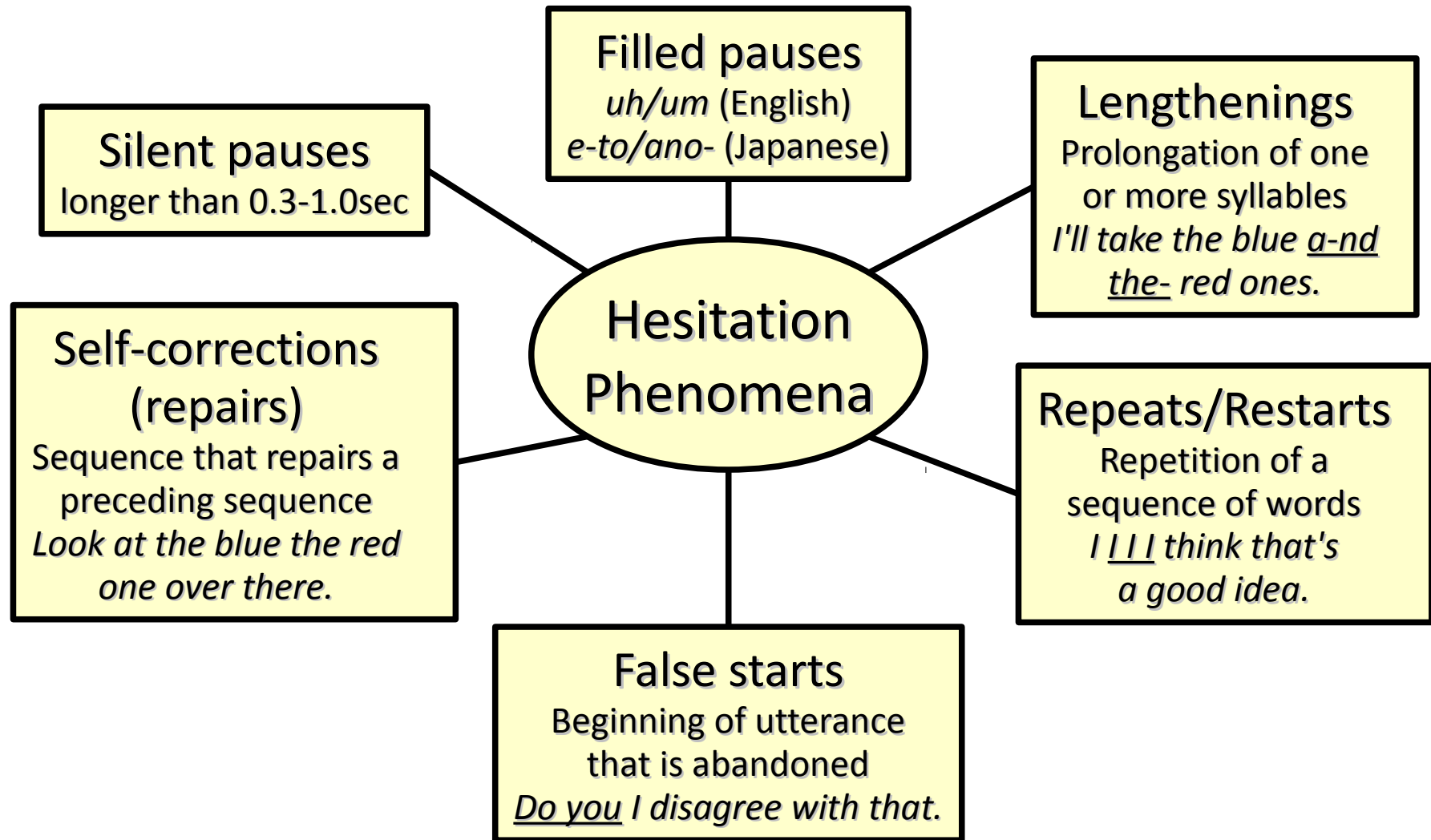
Well, **i- i- i-** in my own life **I'd-** I'd break it up in stages, when **uh** I had a difficult youth. **Uh** my father wasn't in the house, **uh** I've written about this, **uh there- uh uh** you know there were times where **uh** I've experimented with drugs, and I drank, **uh yeah** in my teenage years, **a-nd wh-** 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

Overview

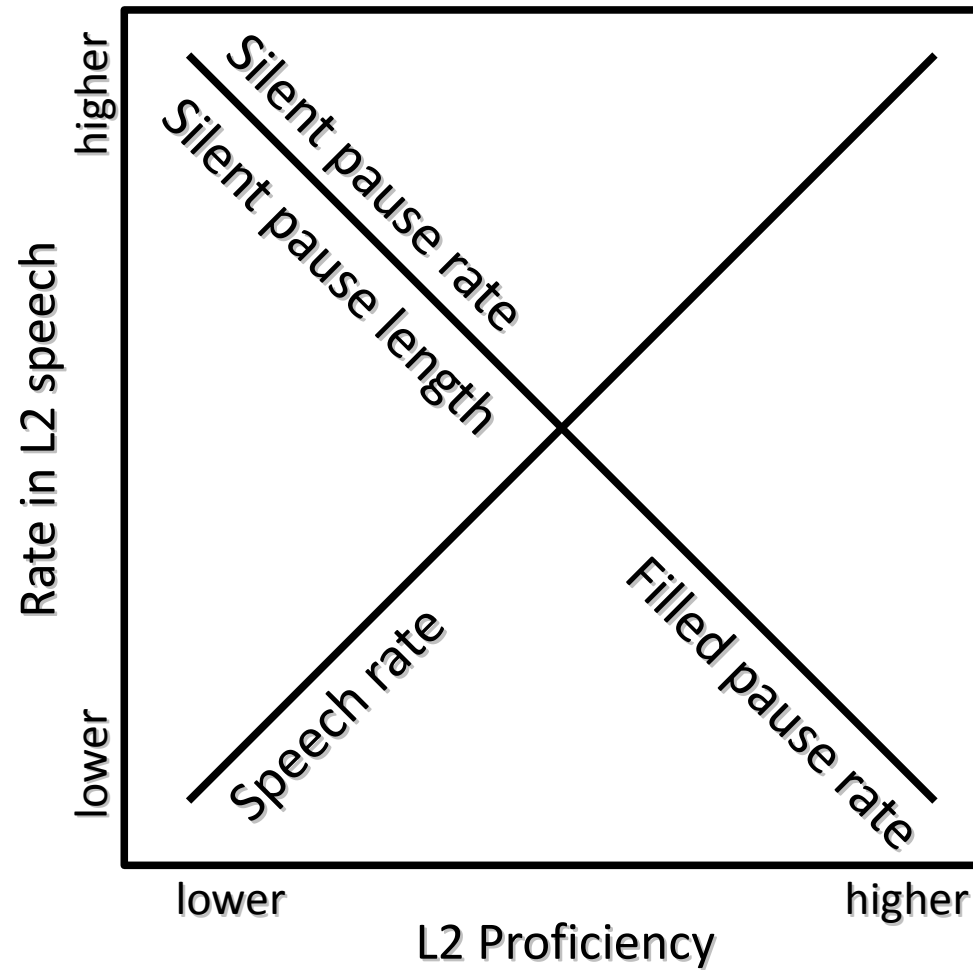
- Hesitation phenomena
 - Overview
 - HP in L2 speech
- Crosslinguistic Corpus of Hesitation Phenomena
 - Description
 - Results
- HP Developmental Trajectory
- Accessing the CCHP

Hesitation Phenomena



(Goldman-Eisler 1961, Levelt 1983, 1989, Maclay and Osgood 1959, Rochester 1973, inter alia)

HP in L2 Production



(Cucchiari et al 2010, Kormos and Dénes 2004, Riazantseva 2001, Rieger 2003, Tavakoli 2011, Trofimovich and Baker 2006, 2007, Wu 2008)

HP in L2 production

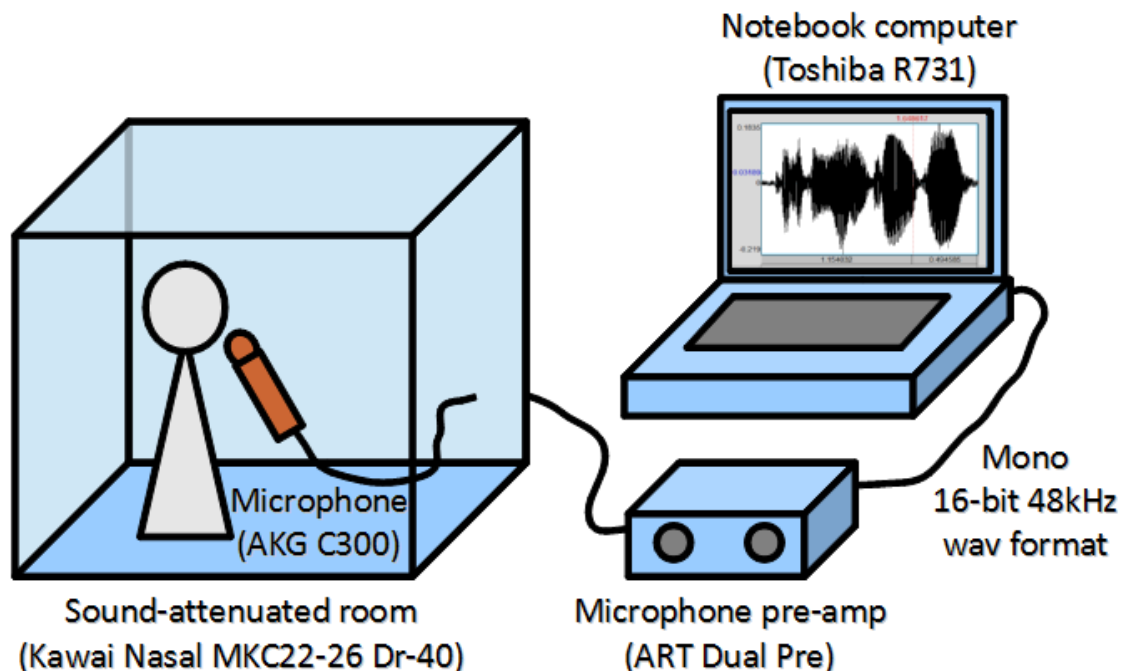
- As a whole, work has been quite comprehensive.
- However, individual works are limited in that many do not take individual variation into account (cf., de Leeuw 2007).
- Gradually, more studies are including L1 observations.
 - Derwing et al (2009) and Cox and Baker-Smemoe (2012) observed that both speech rate and pause rate in L1 and L2 production are correlated.
- The current research is designed to contribute to greater understanding of the influence of L1 hesitation on L2 hesitation.

Research Questions

- What is the relationship between hesitation patterns in L1 and L2 speech?
- What is the developmental trajectory of HP use in L2?

Crosslinguistic Corpus of Hesitation Phenomena (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 Hesitation Phenomena (CCHP)

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

```
<UTTERANCE>
<T>in</T>
<T>America</T>
<T FILLED-PAUSE="yes">uh</T>
<T>there's</T>
<T>a</T>
<T FILLED-PAUSE="yes">uh</T>
<T>very</T>
<T>famous</T>
<T FILLED-PAUSE="yes">uh</T>
<T>and</T>
<T>loved</T>
<T FILLED-PAUSE="yes">uh</T>
<T>basketball</T>
<RP>
<O>
  <T>cl#</T>
</O>
<T FILLED-PAUSE="yes">uh</T>
<E>
  <T>association</T>
</E>
</RP>
<T>which</T>
<T>is</T>
<T>called</T>
<T>NBA</T>
<T>National</T>
<T>Basketball</T>
<T>Association</T>
<T>I</T>
<T>think</T>
</UTTERANCE>
```

CCHP Results: Basic Statistics

- Participants: 15 Japanese L1, English L2 speakers
- Full corpus
 - 24,201 words
 - 5 hrs, 4 min
- Spontaneous speech
 - 15,058 words
 - 3 hrs, 56 min
- Read speech
 - 9,143 words
 - 1 hr, 10 min
- 6,947 silent pauses
- 1,277 filled pauses
- 579 self-corrections
- 177 repeats

Sample recordings

Speaker 1

(low L2 proficiency)

Topic Narrative
Japanese
“Explain TV”

Topic Narrative
English
“Explain basketball”

Speaker 2

(high L2 proficiency)

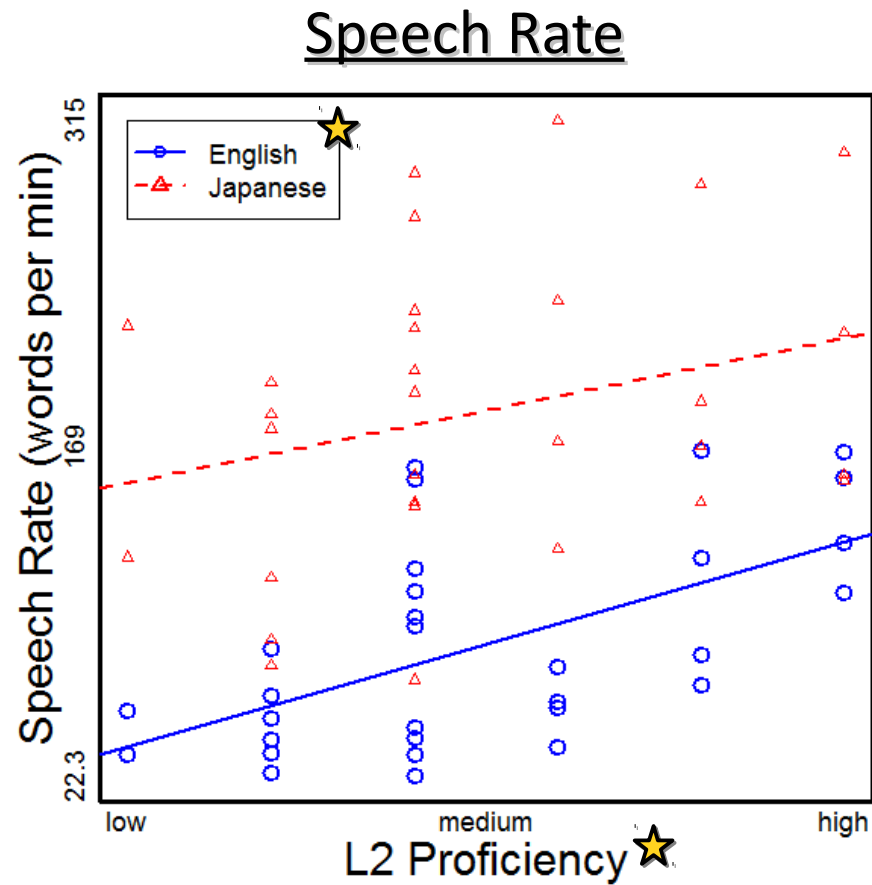
Topic Narrative
Japanese
“Explain TV”

Topic Narrative
English
“Explain basketball”

CCHP Results: Analysis

- Used spontaneous speech data only.
- Computed rates of speech, silent pauses, filled pauses, repeats, and self-corrections for each recording.
- Performed repeated measures ANOVA
 - (between) L2 Proficiency as numerical variable, estimated from test scores, experience abroad, self-assessment
 - (within) Language as categorical variable: Japanese, English
- Used $\alpha = 0.05$ for significance testing (marked with ★).

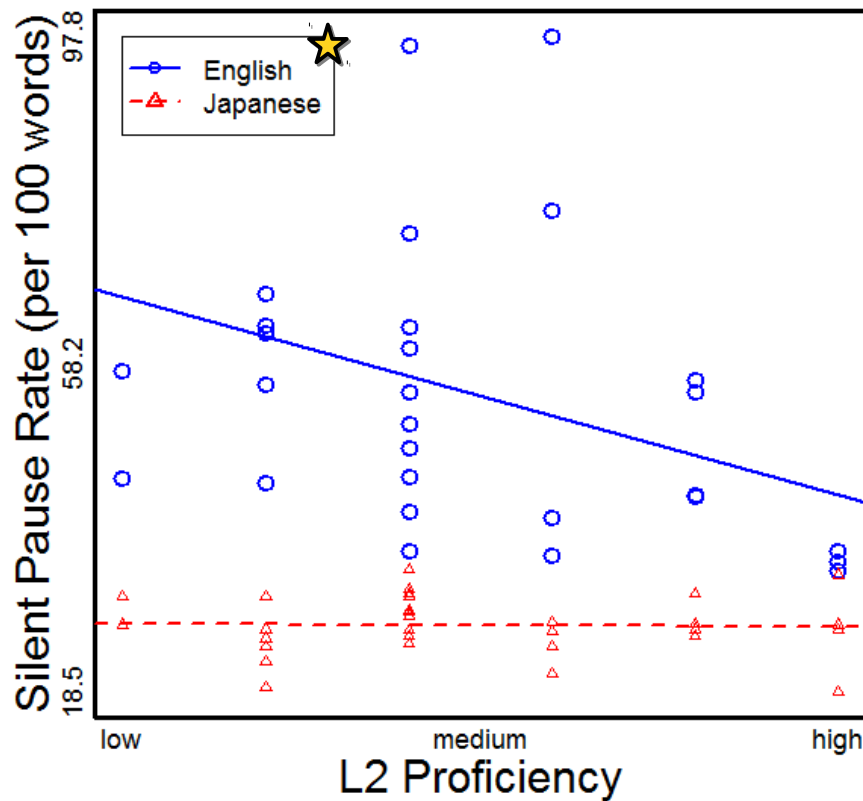
CCHP Results: Speech Rate



Consistent with Derwing et al (2009) and Cox and Baker-Smemoe (2012)

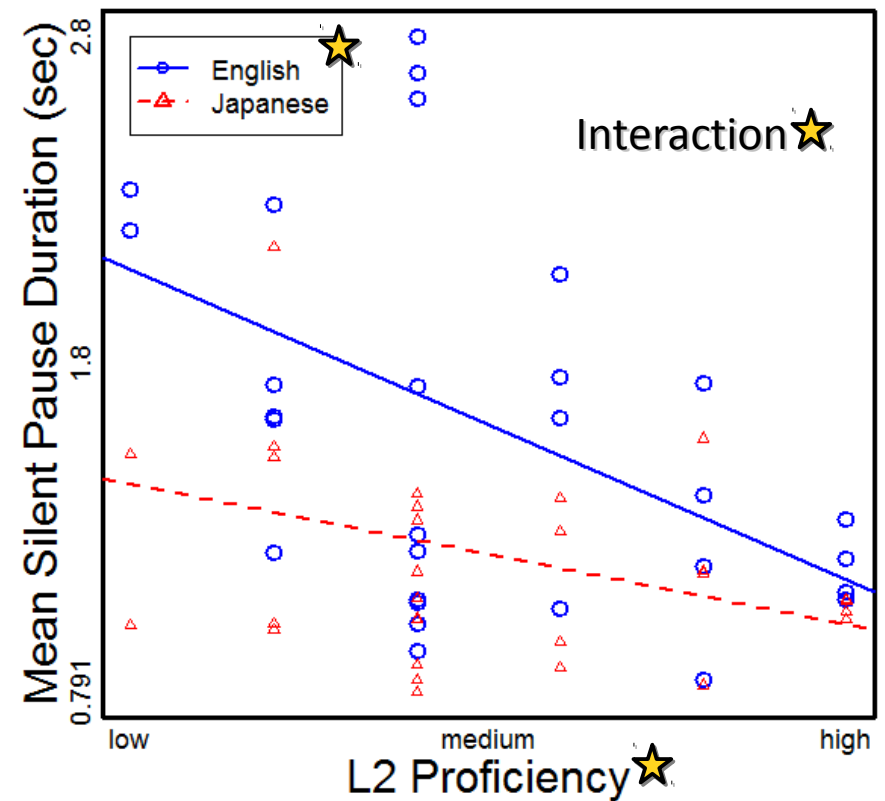
CCHP Results: Silent Pauses

Silent Pause Rate



Consistent with Derwing et al (2009)
and Cox and Baker-Smemoe (2012)

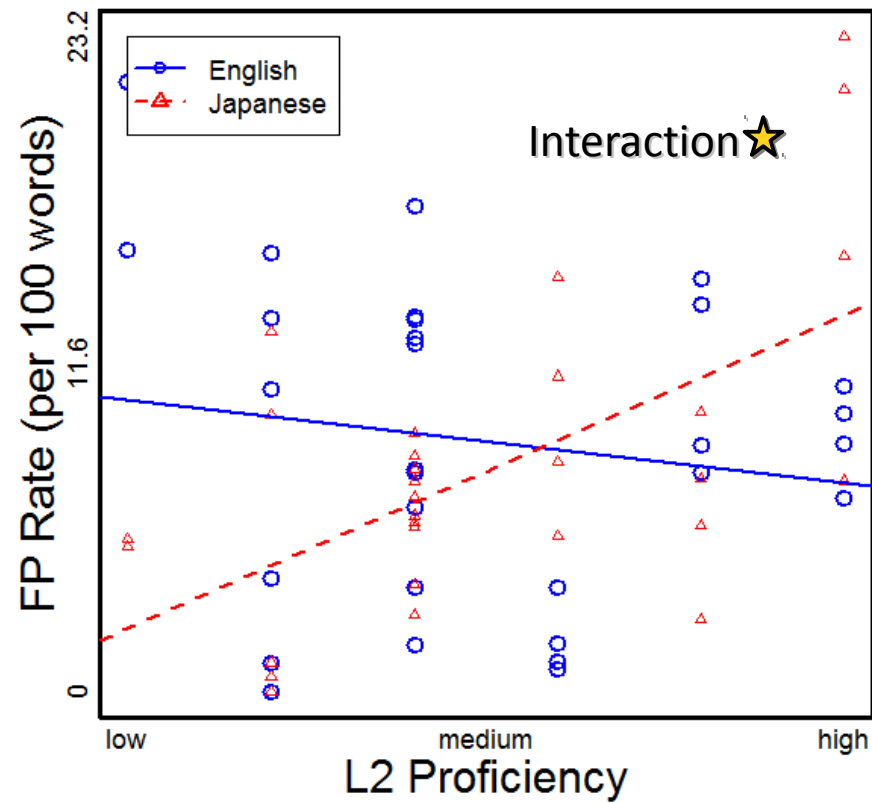
Silent Pause Duration



Contra Derwing et al (2009) and
Cox and Baker-Smemoe (2012)

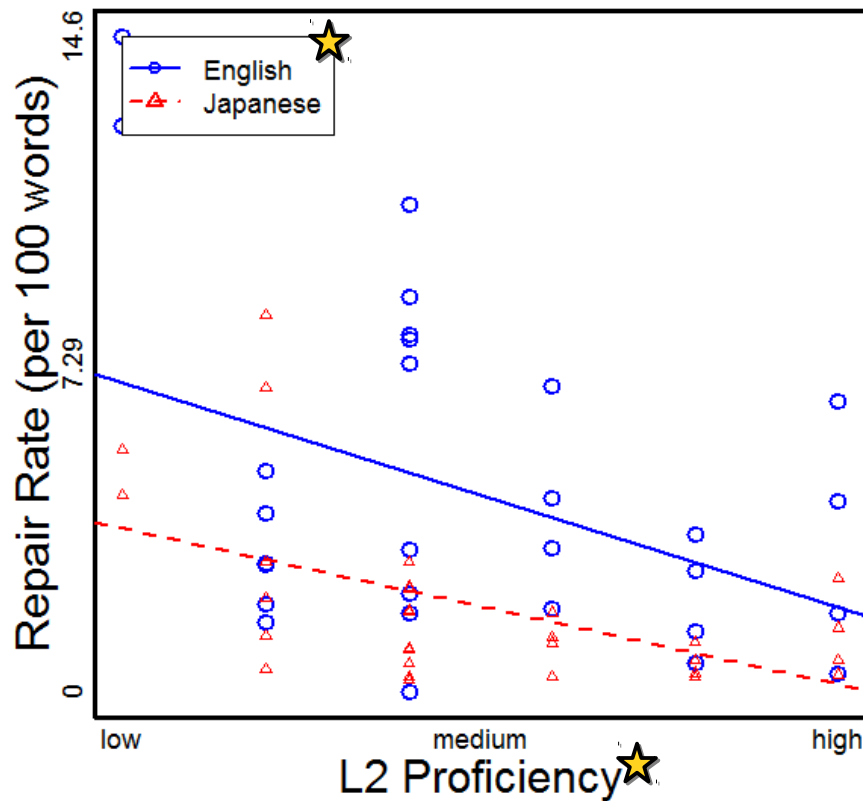
CCHP Results: Filled Pauses

Filled Pause Rate

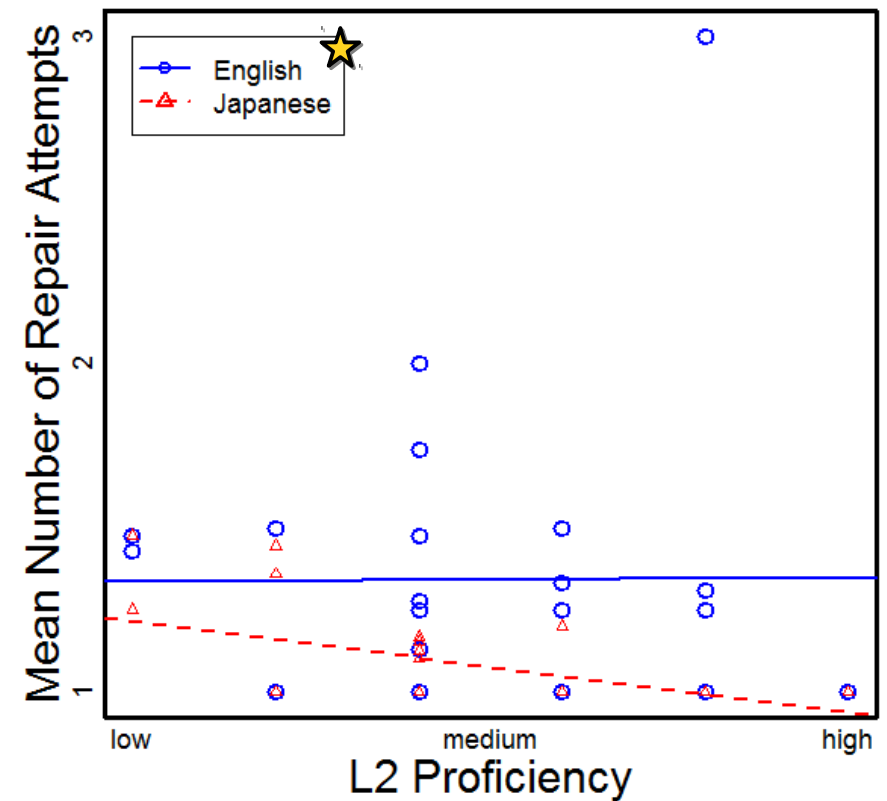


CCHP Results: Self-corrections

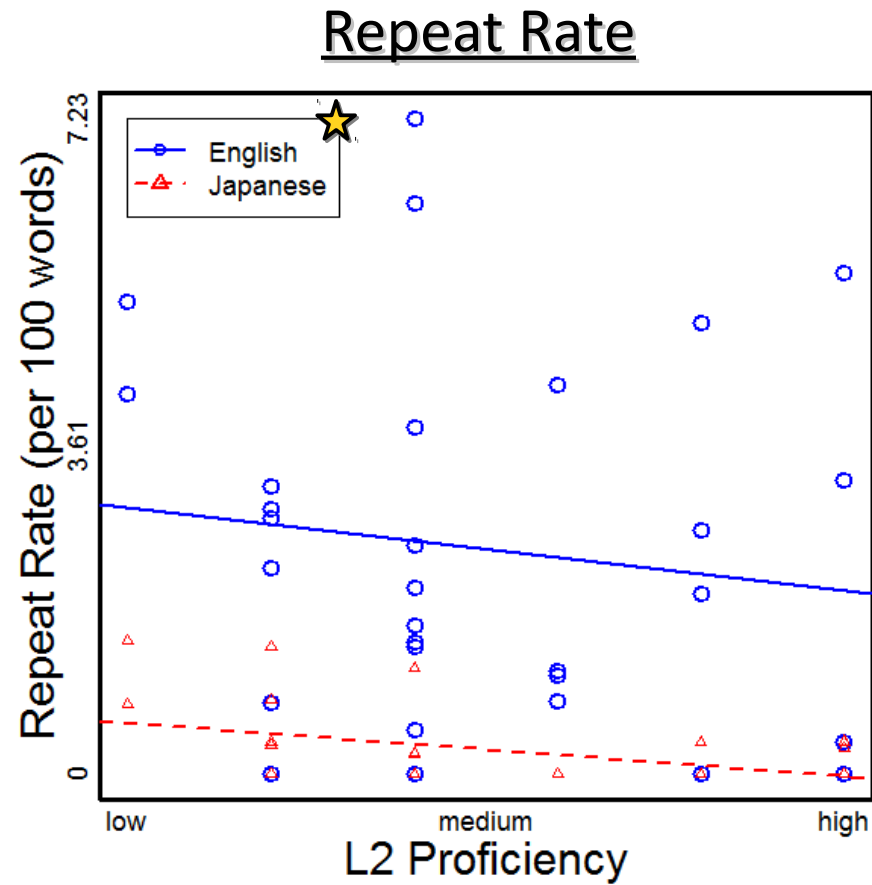
Self-correction Rate



Self-correction Attempts

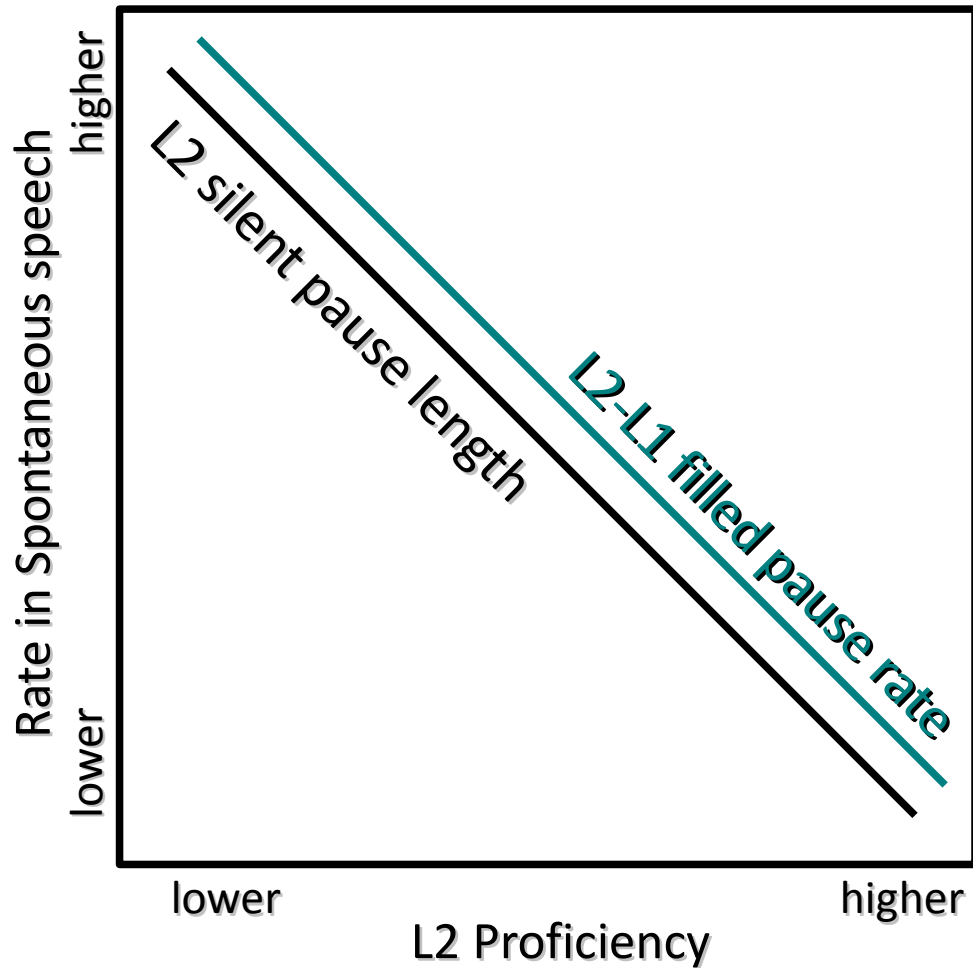


CCHP Results: Repeats



Repeats are uncommon in Japanese (Fox et al 1996)

HP Developmental Trajectory



To-do:

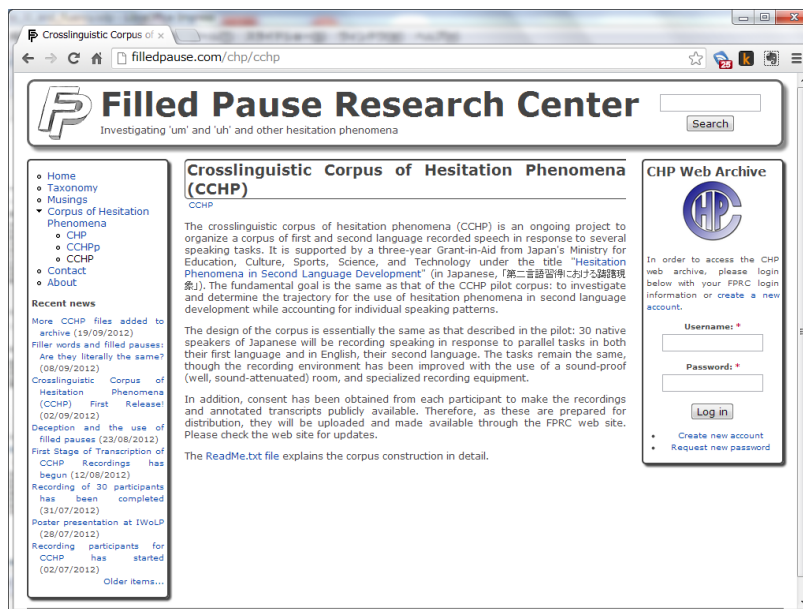
- Filled pause duration
- Lengthenings
- Repair type distribution
- Structural distribution
- Syllable counts

Summary

- Recent studies of L2 speech performance are taking L1 speech performance more and more into account.
- The Crosslinguistic Corpus of Hesitation Phenomena allows us to account for L1 factors in the study of L2 hesitation patterns.
- Results show that silent and filled pause features indicate learners' L2 proficiency.
- Results show that speakers at all proficiency levels use more repeats.
- Results suggest that other aspects of L2 hesitation use correlate with that of L1.

CCHP Public Corpus

- Assembling a larger (n=35), 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>



The screenshot shows the homepage of the Filled Pause Research Center. The header includes the logo and the text "Investigating 'um' and 'uh' and other hesitation phenomena". A search bar is located in the top right. The main content area is titled "Crosslinguistic Corpus of Hesitation Phenomena (CCHP)" and contains a paragraph describing the project. Below this, there is a "CHP Web Archive" section with a login form. The left sidebar contains a navigation menu and a "Recent news" section with several bullet points.

Filled Pause Research Center
Investigating 'um' and 'uh' and other hesitation phenomena

Crosslinguistic Corpus of Hesitation Phenomena (CCHP)

The crosslinguistic corpus of hesitation phenomena (CCHP) is an ongoing project to organize a corpus of first and second language recorded speech in response to several speaking tasks. It is supported by a three-year Grant-in-Aid from Japan's Ministry for Education, Culture, Sports, Science, and Technology under the title "Hesitation Phenomena in Second Language Development" (in Japanese, 第二言語習得中心における躊躇現象). The fundamental goal is the same as that of the CCHP pilot corpus: to investigate and determine the trajectory for the use of hesitation phenomena in second language development while accounting for individual speaking patterns.

The design of the corpus is essentially the same as that described in the pilot: 30 native speakers of Japanese will be recording speaking in response to parallel tasks in both their first language and in English, their second language. The tasks remain the same, though the recording environment has been improved with the use of a sound-proof (well, sound-attenuated) room, and specialized recording equipment.

In addition, consent has been obtained from each participant to make the recordings and annotated transcripts publicly available. Therefore, as these are prepared for distribution, they will be uploaded and made available through the FRPC web site. Please check the web site for updates.

The ReadMe.txt file explains the corpus construction in detail.

CHP Web Archive

In order to access the CHP web archive, please login below with your FRPC login information or create a new account.

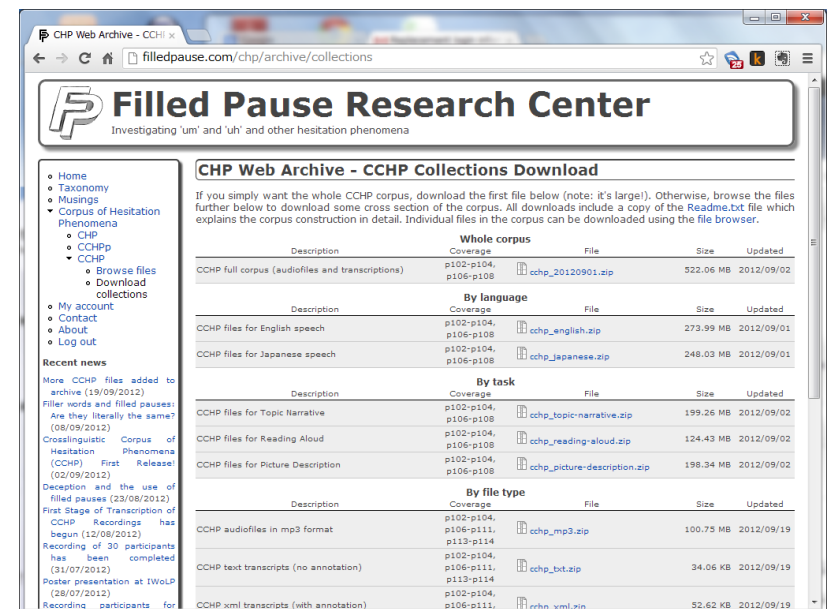
Username: *
Password: *

Log in

• Create new account
• Request new password

Recent news

- More CCHP files added to archive (19/09/2012)
- Filler words and filled pauses: Are they literally the same? (08/09/2012)
- Crosslinguistic Corpus of Hesitation Phenomena (CCHP) First Release! (02/09/2012)
- Deception and the use of filled pauses (23/08/2012)
- First Stage of Transcription of CCHP Recordings has begun (12/08/2012)
- Recording of 30 participants has been completed (31/07/2012)
- Poster presentation at IWOLP (28/07/2012)
- Recording participants for CCHP has started (02/07/2012)
- Older items...



The screenshot shows the "CCHP Collections Download" page. It features a navigation menu on the left and a main content area with a table of download links. The table is organized into sections: "Whole corpus", "By language", "By task", and "By file type". Each section contains a table with columns for Description, Coverage, File, Size, and Updated.

Filled Pause Research Center
Investigating 'um' and 'uh' and other hesitation phenomena

CHP Web Archive - CCHP Collections Download

If you simply want the whole CCHP corpus, download the first file below (note: it's large!). Otherwise, browse the files further below to download some cross section of the corpus. All downloads include a copy of the ReadMe.txt file which explains the corpus construction in detail. Individual files in the corpus can be downloaded using the file browser.

Description	Coverage	File	Size	Updated
CCHP Full corpus (audiofiles and transcriptions)	p102-p104, p106-p108	cchp_20120901.zip	522.06 MB	2012/09/02

By language

Description	Coverage	File	Size	Updated
CCHP files for English speech	p102-p104, p106-p108	cchp_english.zip	273.99 MB	2012/09/01
CCHP files for Japanese speech	p102-p104, p106-p108	cchp_japanese.zip	248.03 MB	2012/09/01

By task

Description	Coverage	File	Size	Updated
CCHP files for Topic Narrative	p102-p104, p106-p108	cchp_topic-narrative.zip	199.26 MB	2012/09/02
CCHP files for Reading Aloud	p102-p104, p106-p108	cchp_reading-aloud.zip	124.43 MB	2012/09/02
CCHP files for Picture Description	p102-p104, p106-p108	cchp_picture-description.zip	198.34 MB	2012/09/02

By file type

Description	Coverage	File	Size	Updated
CCHP audiofiles in mp3 format	p102-p104, p106-p111, p113-p114	cchp_mp3.zip	100.75 MB	2012/09/19
CCHP text transcripts (no annotation)	p102-p104, p106-p111, p113-p114	cchp_txt.zip	34.06 KB	2012/09/19
CCHP xml transcripts (with annotation)	p102-p104, p106-p111	rxhn_xml.zip	52.62 KB	2012/09/19

Recent news

- More CCHP files added to archive (19/09/2012)
- Filler words and filled pauses: Are they literally the same? (08/09/2012)
- Crosslinguistic Corpus of Hesitation Phenomena (CCHP) First Release! (02/09/2012)
- Deception and the use of filled pauses (23/08/2012)
- First Stage of Transcription of CCHP Recordings has begun (12/08/2012)
- Recording of 30 participants has been completed (31/07/2012)
- Poster presentation at IWOLP (28/07/2012)
- Recording participants for CCHP has started (02/07/2012)
- Older items...

References

- Cox, T. and Baker-Smemoe, W. (2012). The relationship between L1 fluency and L2 fluency across different proficiency levels and L1s. Presentation at Workshop Fluent Speech (Utrecht University, The Netherlands).
- Cucchiari, C., van Doremalen, J., & Strik, H. (2010). *Fluency in non-native read and spontaneous speech*. Paper presented at Proceedings of Disfluency in Spontaneous Speech (DiSS) and Linguistic Patterns in Spontaneous Speech (LPSS) Joint Workshop.
- de Leeuw, E. (2007). Hesitation markers in English, German, and Dutch. *Journal of Germanic Linguistics*, 19(2), 85-114.
- Derwing, T. M., Munro, M. J., Thomson, R. I., & Rossiter, M. J. (2009). The relationship between L1 fluency and L2 fluency development. *Studies in Second Language Acquisition*, 31(4), 533-557.
- Fox, B., Hayashi, M., and Jasperson, R. (1996). Resources and repair: A cross-linguistic study of syntax and repair. In Ochs, E., Schegloff, E., and Thompson, S. (eds). *Interaction and Grammar*. Cambridge University Press, Studies in Interactional Sociolinguistics, volume 13: pages 185-237.
- Goldman-Eisler, F. (1961). A comparative study of two hesitation phenomena. *Language and Speech*, 4(1), 18-26.
- Kormos, J. & Dénes, M. (2004). Exploring measures and perceptions of fluency in the speech of second language learners. *System*, 32(2), 145-164.
- Levelt, W. J. M. (1983). Monitoring and self-repair in speech. *Cognition*, 14(1), 41-104.
- Levelt, W. J. M. (1989). *Speaking: from intention to articulation*, MIT Press, ACL-MIT Press series in natural-language processing.
- Maclay, H., & Osgood, C. (1959). Hesitation phenomena in spontaneous English speech. *Word*, 15, 19-44.
- Riazantseva, A. (2001). Second language proficiency and pausing a study of Russian speakers of English. *Studies in Second Language Acquisition*, 23(4), 497-526.
- Rieger, C. L. (2003). *Disfluencies and hesitation strategies in oral L2 tests*. Paper presented at Proceedings of DiSS '03, Disfluency in Spontaneous Speech Workshop.
- Rochester, S. (1973). The significance of pauses in spontaneous speech. *Journal of Psycholinguistic Research*, 2(1), 51-81.
- Tavakoli, P. (2011). Pausing patterns: differences between L2 learners and native speakers. *ELT Journal*, 65(1), 71-79.
- Trofimovich, P., & Baker, W. (2006). Learning second language suprasegmentals: Effect of L2 experience on prosody and fluency characteristics of L2 speech. *Studies in Second Language Acquisition*, 28, 1-30.
- Trofimovich, P., & Baker, W. (2007). Learning prosody and fluency characteristics of second language speech: The effect of experience on child learners' acquisition of five suprasegmentals. *Applied Psycholinguistics*, 28(2), 251-276.
- Wu, C.-H. (2008). *Filled pauses in L2 Chinese: A comparison of native and non-native speakers*. Paper presented at Proceedings of the 20th North American Conference on Chinese Linguistics (NACCL-20), Columbus, Ohio.