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

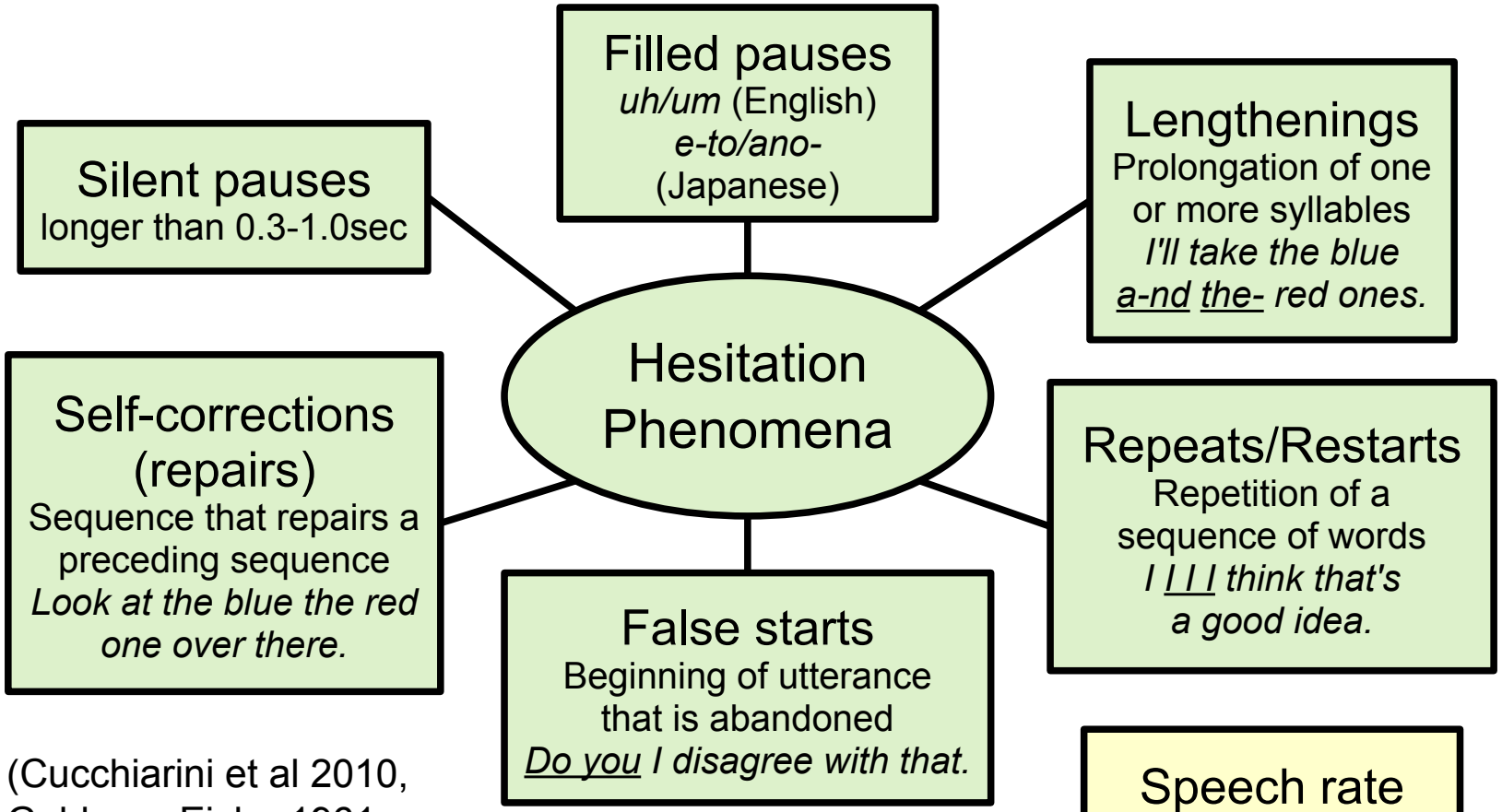
- Japanese speech
  - Filled pauses: “e-”
  - Very few silent pauses
  - Some fixed expressions
- English speech
  - Filled pauses: “uh” / “um”
  - Very few silent pauses
  - Discourse markers: “you know”, “kind of”

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



(Cucchiarini et al 2010, Goldman-Eisler 1961, Levelt 1983, 1989, Maclay and Osgood 1959, Rochester 1973, inter alia)

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 (Quené et al 2011) & manual

```
<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: Basic statistics

- Participants: 35 Japanese L1 / English L2 speakers

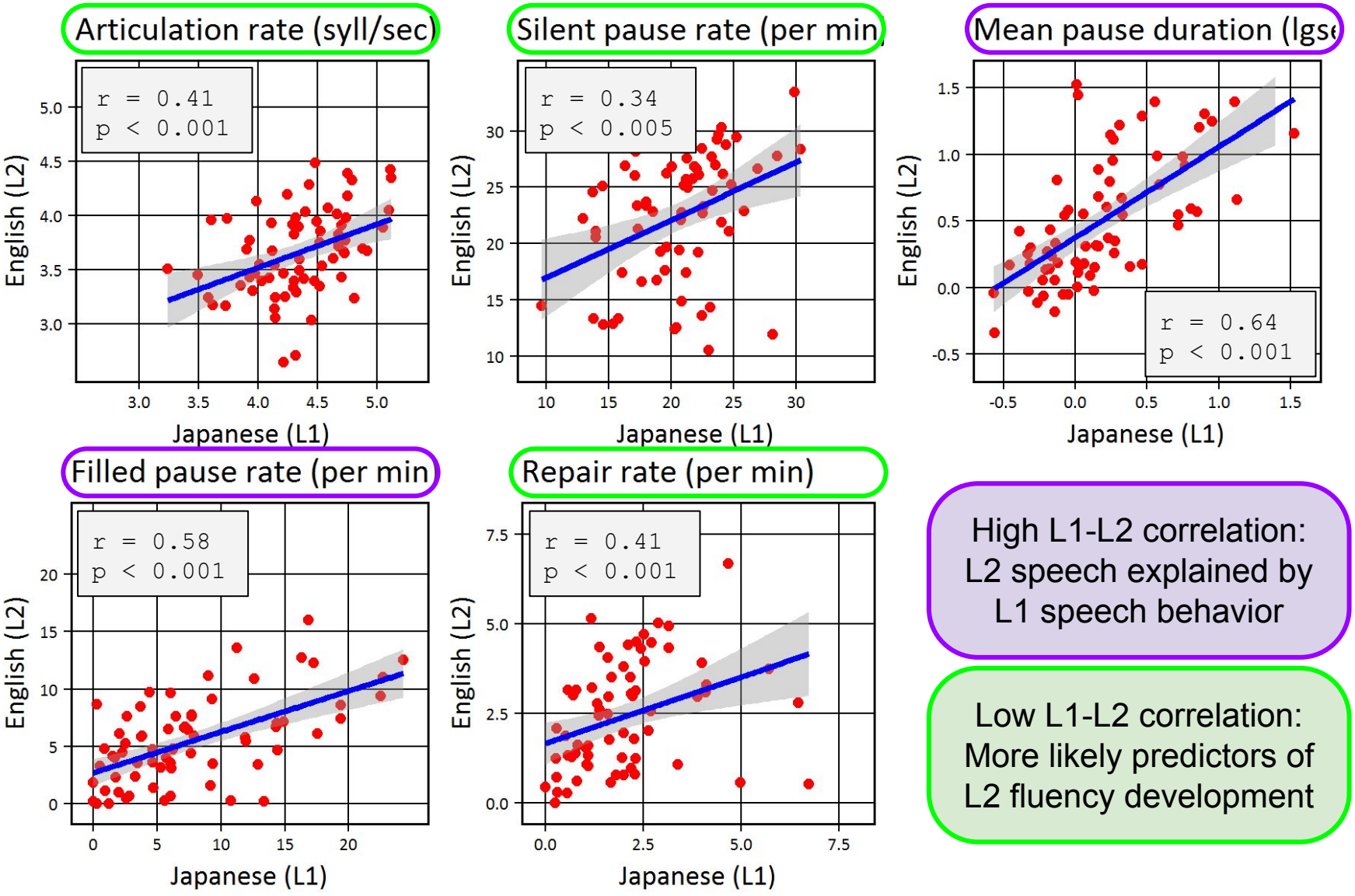
	Word count	Time
Read speech	21,406	2 hr, 41 min
Picture description	19,732	4 hr, 39 min
Topic narrative	21,138	4 hr, 35 min
Total	62,276	11 hr, 56 min

Hesitation phenomena

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

Transcriber agreement = 91.5%

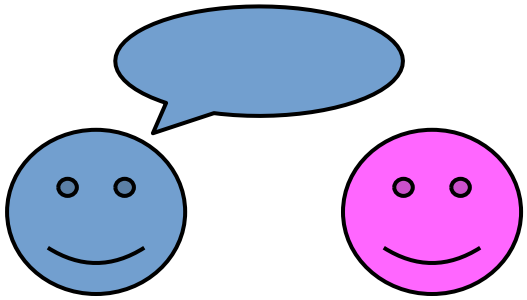
L1-L2 temporal pattern correspondence



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

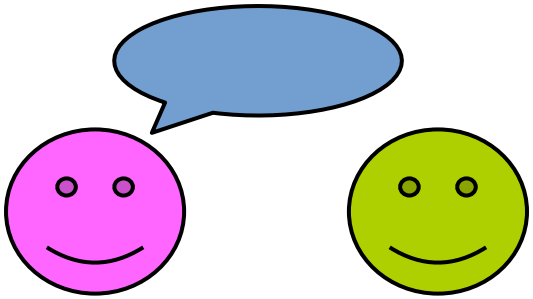
Low L1-L2 correlation: More likely predictors of L2 fluency development

Crosslinguistic communication



native speaker

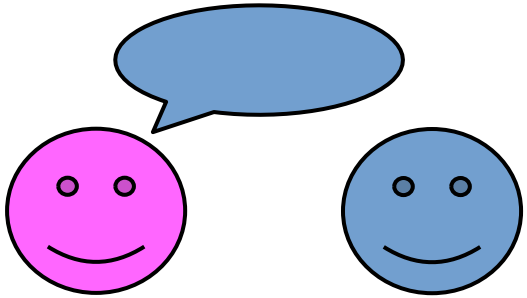
nonnative listener



nonnative speaker

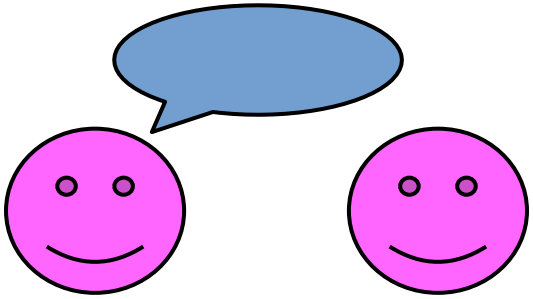
nonnative listener

non-shared native language



nonnative speaker

native listener



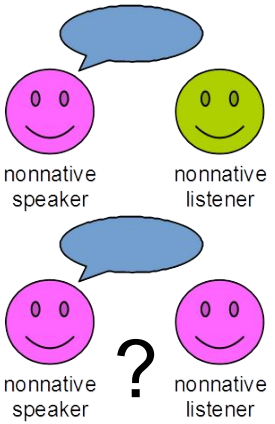
nonnative speaker

nonnative listener

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

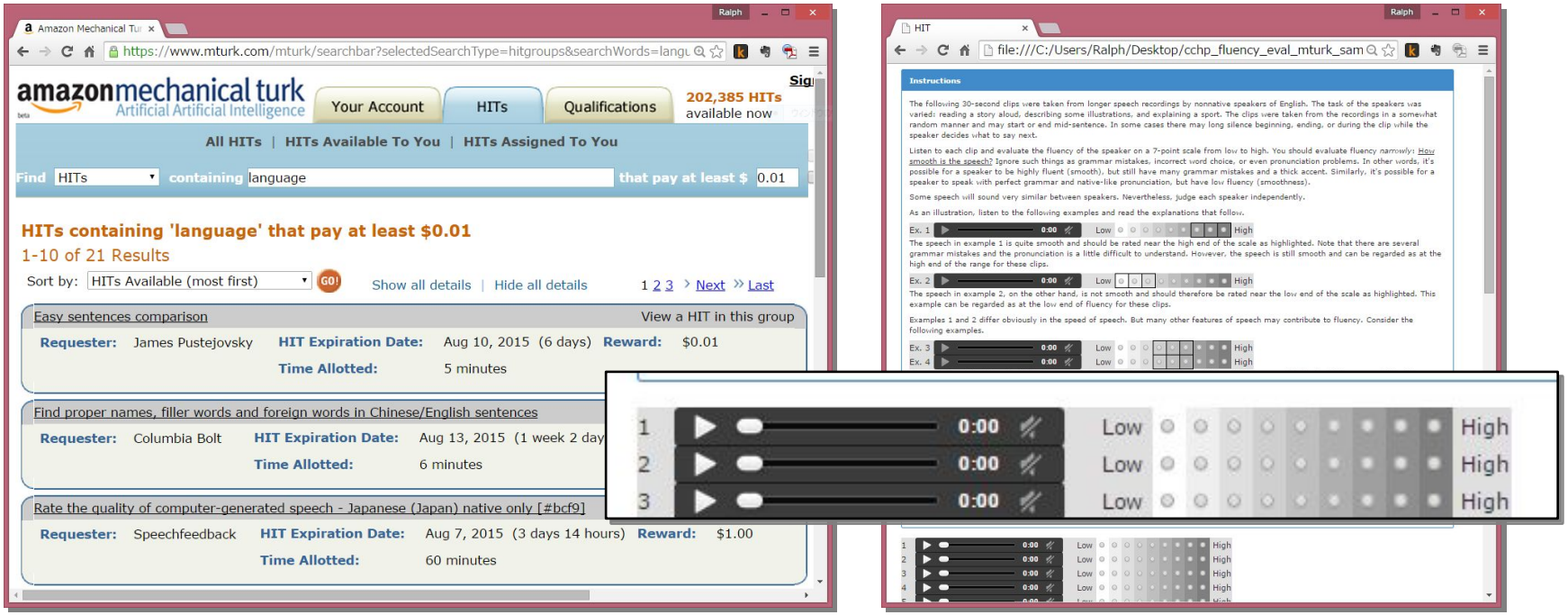


Fluency ratings

- 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.

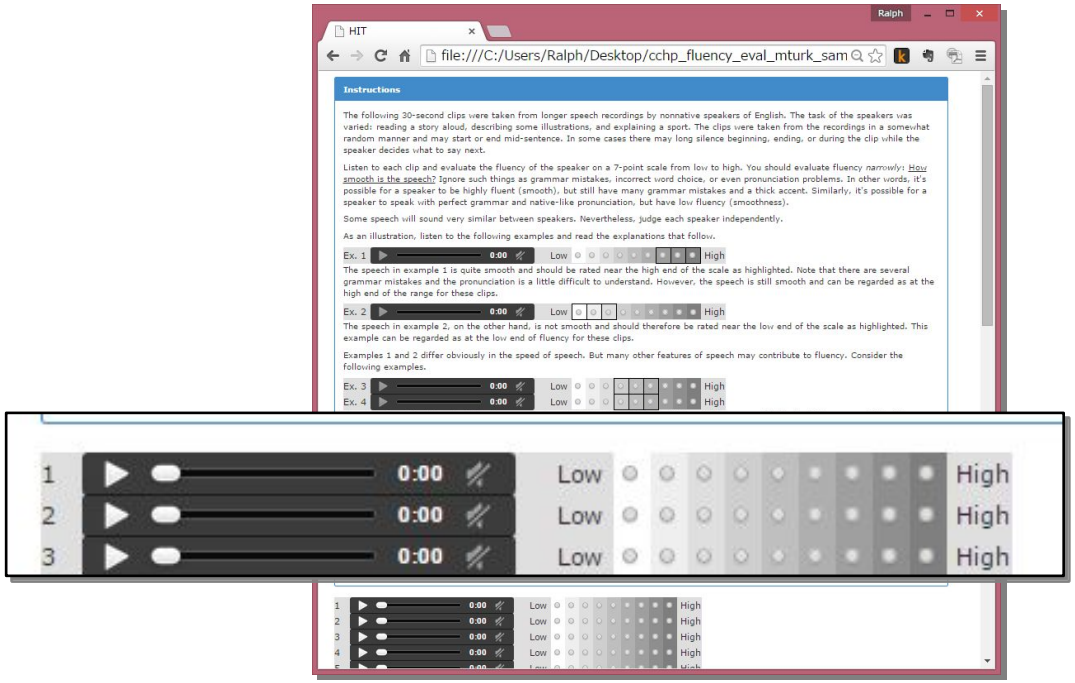
Fluency ratings by native listeners

- 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.



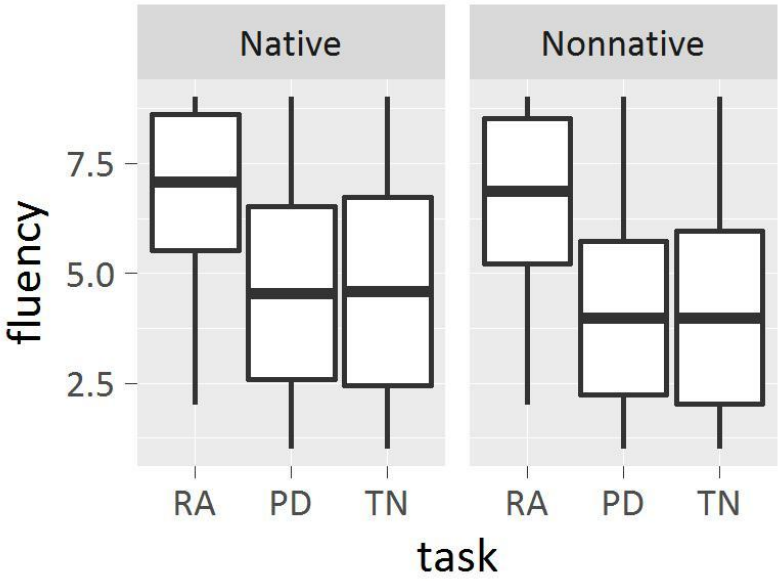
Fluency ratings by nonnative listeners

- 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



Effect	DFn	DFd	F	p	p<.05
raterlang	1	52	7.581507	8.104075e-03	*
task	2	104	485.145647	1.850916e-53	*
raterlang:task	2	104	2.312539	1.040802e-01	

Fluency rating results

- Linear regression modeling (using `lm` in `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

Higher fluency associated with...	native raters	nonnative raters
higher articulation rate	✓	✓
lower pause rate		✓
shorter pause length	✓	✓
higher filled pause rate		✓
lower repair rate	✓	
	[R^2=0.45]	[R^2=0.57]

[F(7,406)=64.9, p<0.001; adjusted R^2=0.52]

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*
  - Supra-segmental: Anderson-Hsieh 1992*; Taniguchi and Abberton 1999*; de Wet et al 2009
- 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

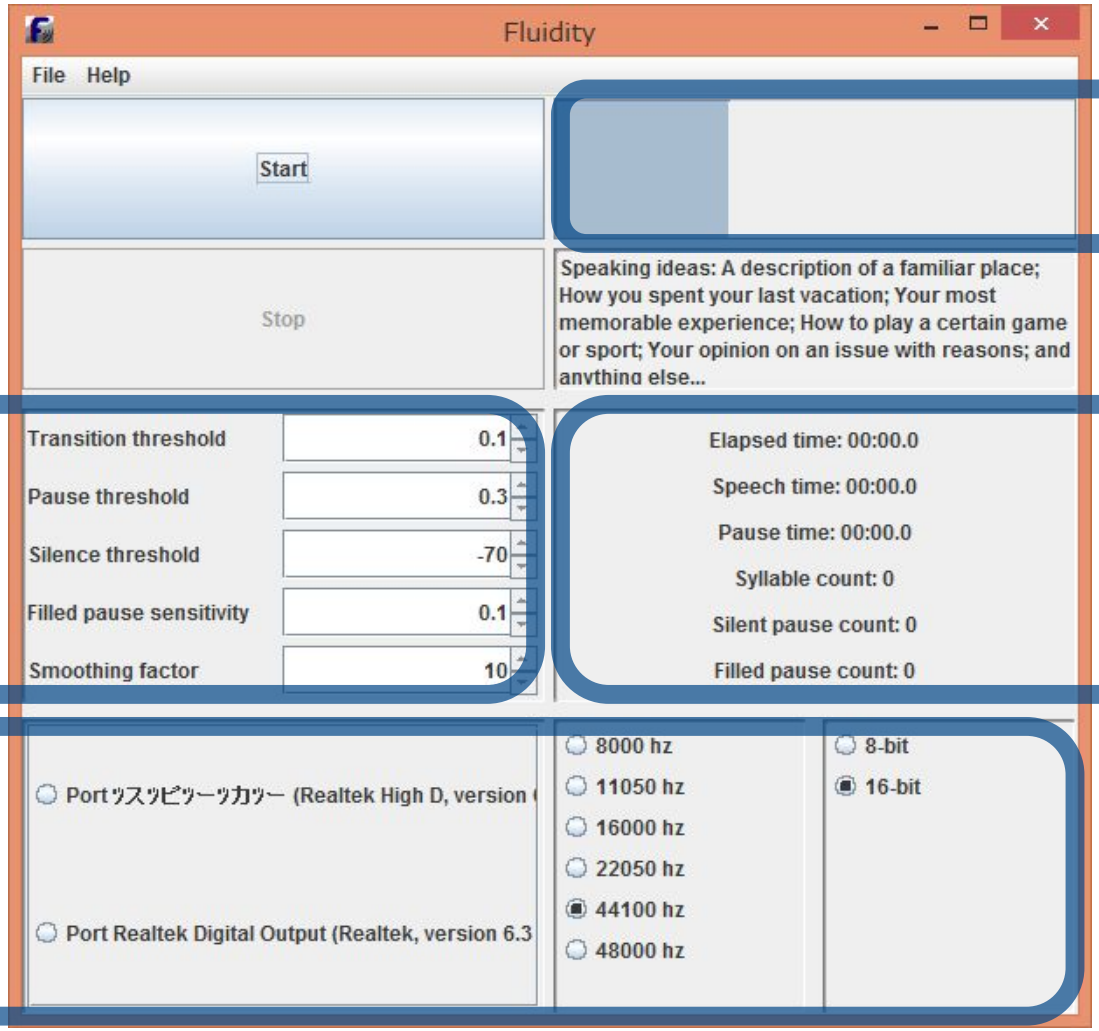
- Measure various utterance fluency characteristics and update them in real-time.
- Provide real-time feedback to learner on utterance fluency measures.
- Provide opportunity for learner to review their production together with visual representation of fluency measures.
- Provide feedback in a manner 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
  (cf., Shriberg and Lickley 1993, Tseng 1999, Audhkhasi et al 2009)

Fluidity: main window



Requires Java SE 6 or greater

audio level meter

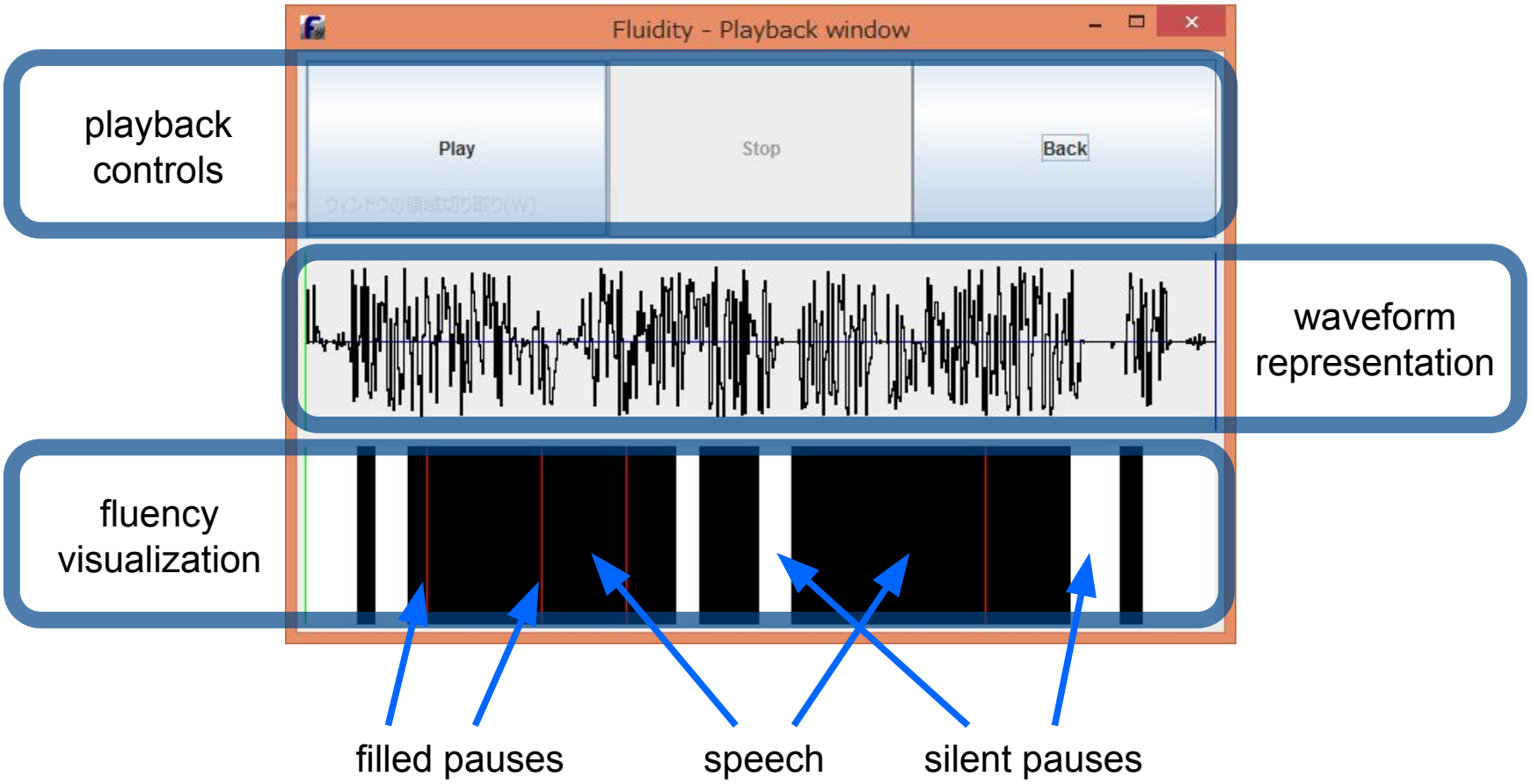
detection settings

fluency measure indicators

audio input settings

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

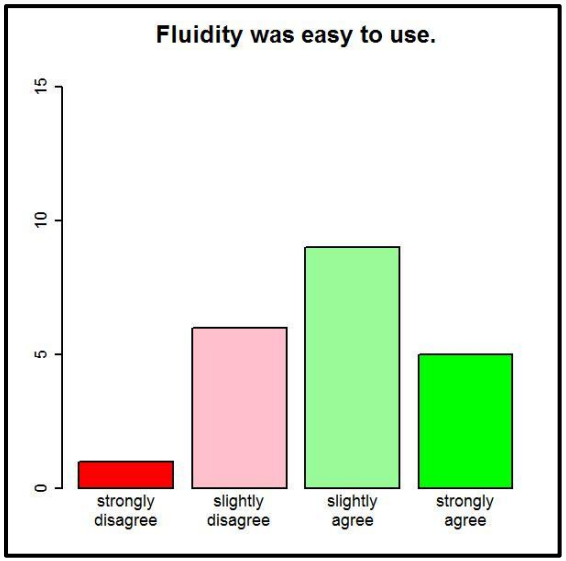
Fluidity: playback window

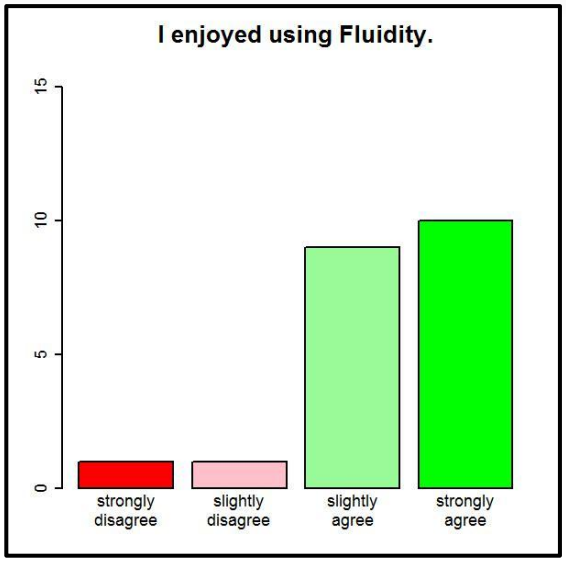


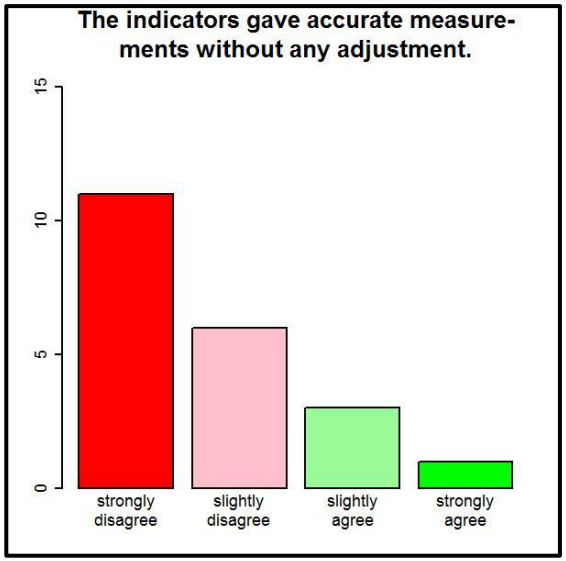
Fluidity: usability testing

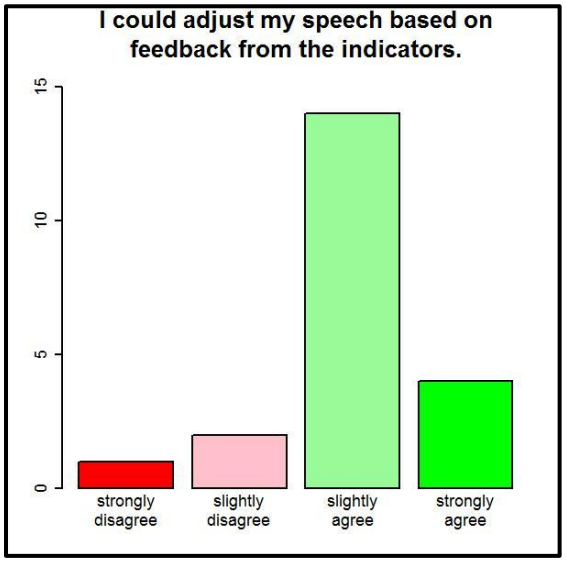
- Participants (n=21)
- Procedure
  - Practice speaking with Fluidity.
  - Adjust settings to fit their production.
  - Respond to survey questions about the experience.

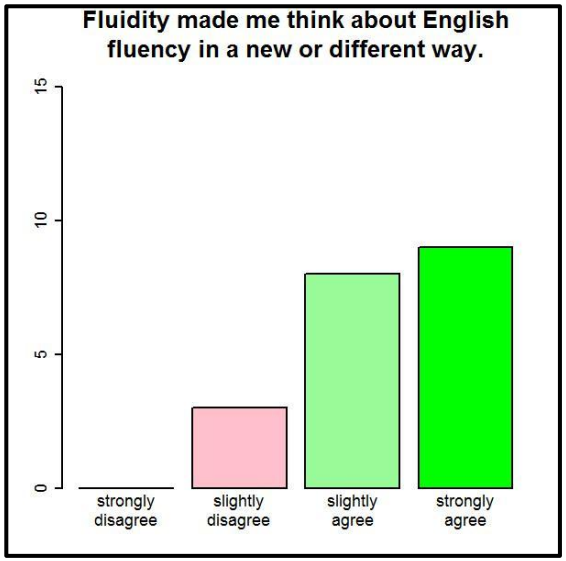
Fluidity: user response

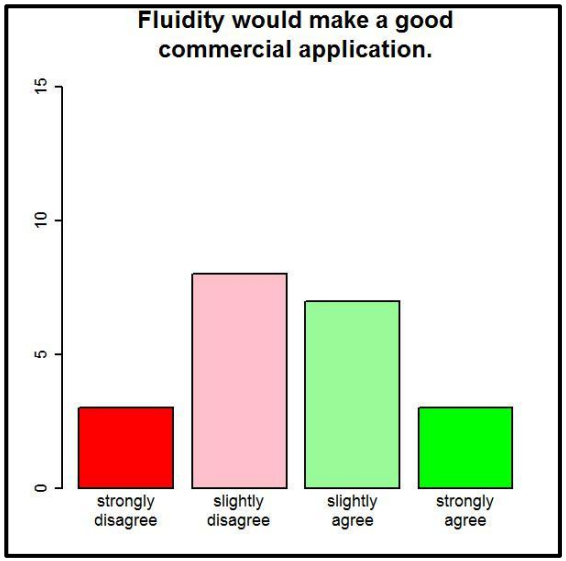




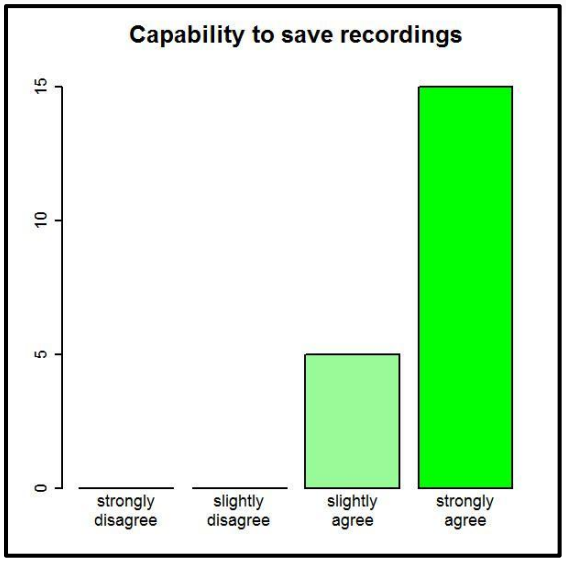


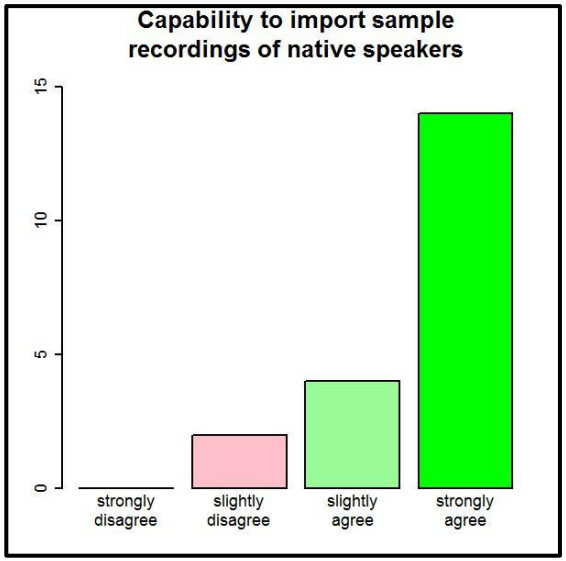


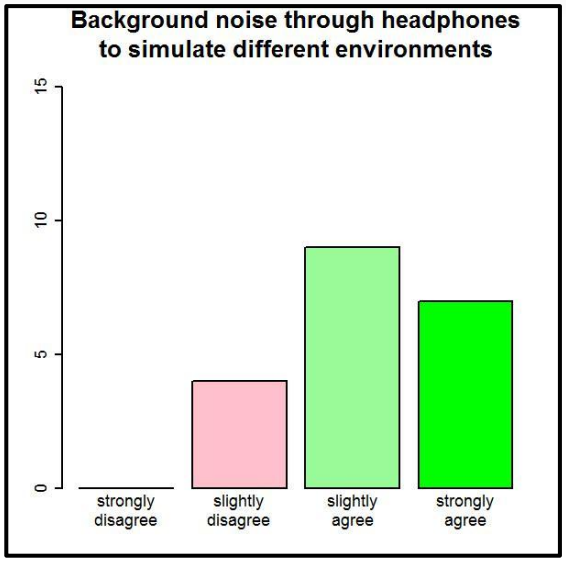


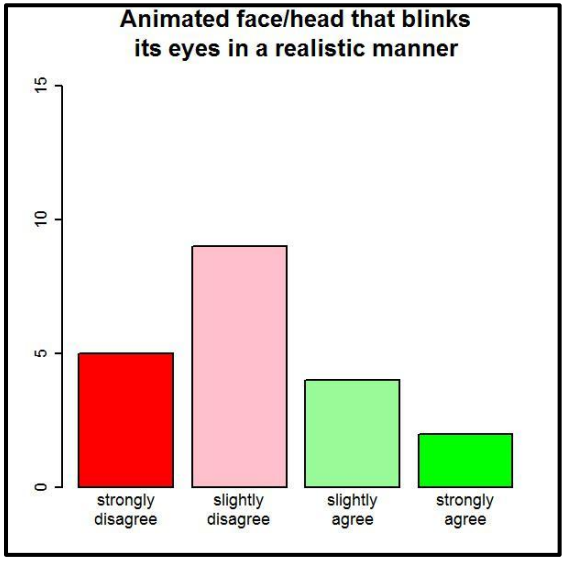


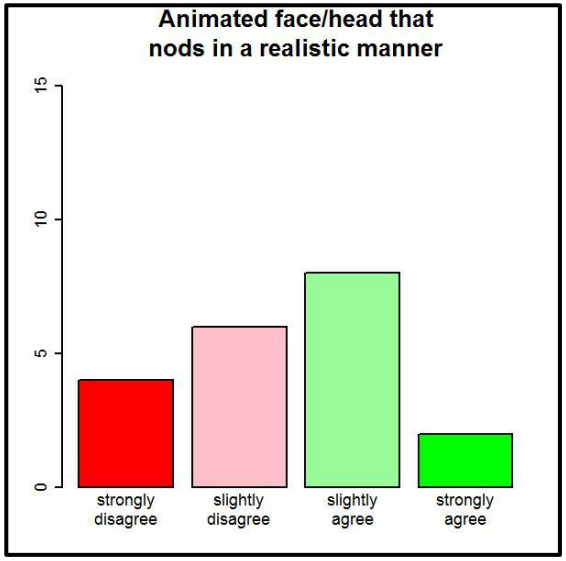
Fluidity: desirable features

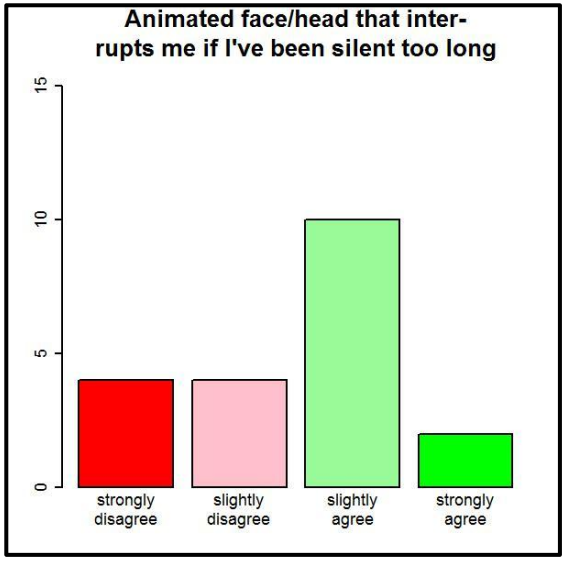












Fluidity: user comments

- 「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れば，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.*

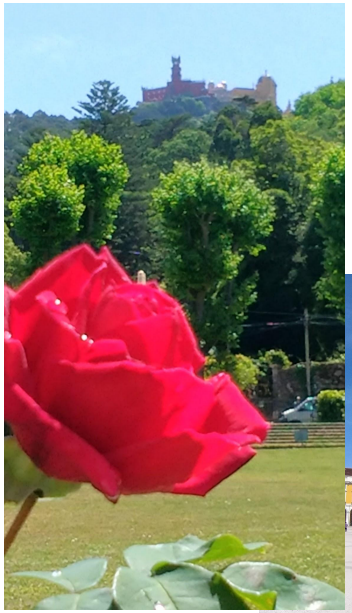
Putting things together...

- 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

Anderson-Hsieh, J. (1992) Using electronic visual feedback to teach supra-segmentals. System, Vol. 20, pp. 51-62.

Audhkhasi, K.; Kandhway, K.; Deshmukh, O.D.; and Verma, A., (2009) Formant-based technique for automatic filled-pause detection in spontaneous spoken English. IEEE International Conference on Acoustics, Speech and Signal Processing, 2009, pp. 19-24 April 2009.

Belz, M., and Reichel U. (2015). Pitch Characteristics of Filled Pauses. Proceedings of The 7th Workshop on Disfluency in Spontaneous Speech (DiSS 2015)

Bent, T. and Bradlow, AR. (2003) The interlanguage speech intelligibility benefit. In Journal of the Acoustical Society of America, 114(3): 1600-1610.

Bernstein J (1999). PhonePassTM testing: Structure and construct. Menlo Park, CA: Ordinate.

Bhat, S., Hasegawa-Johnson, M., and Sproat, R. (2010) Automatic Fluency Assessment by Signal-Level Measurement of Spontaneous Speech. 2010 INTERSPEECH Satellite Workshop on Second Language Studies: Acquisition, Learning, Education and Technology, Makuhari, Japan.

Cucchiarini, C., Neri, A., and Strik, H. (2009) Oral proficiency training in Dutch L2: The contribution of ASR-based corrective feedback, Speech Communication, 51(10):853-863.

Cucchiarini, 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 Jong, N., & Perfetti C. A. (2011). Fluency Training in the ESL Classroom: An Experimental Study of Fluency Development and Proceduralization. Language Learning. 61(2), p. 533-568.

De Jong, N., Steinel M. P., Florijn A., Schoonen R., & Hulstijn J. H. (2012). Facets of Speaking Proficiency. Studies in Second Language Acquisition. 34(1), p. 5-34.

Eskénazi, M. (1999) Using Automatic Speech Processing for Foreign Language Pronunciation Tutoring: Some Issues and a Prototype. Language Learning & Technology 2(2): 62-76.

Eskenazi M. (2009) An overview of spoken language technology for education, Speech Communication, 51(10): 832-844

Foote, J. A., & Trofimovich P. (2016). A Multidimensional Scaling Study of Native and Non-Native Listeners' Perception of Second Language Speech. Perceptual and Motor Skills. 122(2), p. 470-489.

Gamper, J. and Knapp, J. (2002) A Review of Intelligent CALL Systems, Computer Assisted Language Learning, 15(4): 329-342

Goldman-Eisler, F. (1961). A comparative study of two hesitation phenomena. Language and Speech, 4(1), 18-26.

Goldman-Eisler, F. (1961). A comparative study of two hesitation phenomena. Language and Speech, 4(1), 18-26.

Goto, M. Itou, K. and Hayamizu, S. (1999) A real-time filled pause detection system for spontaneous speech recognition, Proc. of Eurospeech, September, pp. 227–230.

Hayashi, N., Nogami, Y., Maeda, K., and Ikeda, H. (2004) Practical use of computerized adaptive testing in Japan -Development and operation of "CASEC". ICP-2004 at Beijing, China. August 13th.

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.

Nation, P. (1989). Improving speaking fluency. System, 17(3), 377–384.

Patten, I. and Edmonds, L.A. (2013) Effect of training Japanese L1 speakers in the production of American English /r/ using spectrographic visual feedback. Computer Assisted Language Learning.

Rochester, S. (1973). The significance of pauses in spontaneous speech. Journal of Psycholinguistic Research, 2(1), 51-81.

Rose, R. (2013) Crosslinguistic Corpus of Hesitation Phenomena: A corpus for investigating first and second language speech performance. In Proceedings of the 14th Annual Conference of the International Speech Communication Association (Interspeech 2013), Lyon, France, pp. 992-996.

Rossiter, MJ. (2009) Perceptions of L2 Fluency by Native and Non-native Speakers of English. In Canadian Modern Language Review, 65(3):395.

Segalowitz, N. (2010). Cognitive bases of second language fluency. Routledge.

Shriberg, E., and Lickley R. J. (1993) Intonation of clause-internal filled pauses. Phonetica. 50(3), p. 172-179.

Simon, J. AudioInfo.java, audio info library. Retrieved from http:// www.javafaq.nu/java-example-code-479.html on 25 December 2014.

Six, J. TarsosDSP, Java library. Retrieved from https://github.com/JorenSix/ TarsosDSP on 25 December 2014.

Taniguchi, M. and Abberton, E. (1999) Effect of Interactive Visual Feedback on the Improvement of English Intonation of Japanese EFL Learners. Speech, Hearing and Language: Work in Progress No. 11 (pp.76-89) Department of Phonetics and Linguistics, University College London.

Tseng, S. - C. (1999). Grammar, prosody and speech disfluencies in spoken dialogues. Bielefeld: Bielefeld University.

Wester, M. and Mayo, C. (2014) Accent rating by native and non-native listeners. In Proceedings of International Conference on Acoustics, Speech, and Signal Processing, 7749-7753.

de Wet, F., Van der Walt, C., and Niesler, T.R. (2009) Automatic assessment of oral language proficiency and listening comprehension, Speech Communication, 51(10):864-874.

Zechner, K., Higgins, D., Xi, X., and Williamson, D.M. (2009) Automatic scoring of non-native spontaneous speech in tests of spoken English, Speech Communication, 51(10)883-895.