SLaTE 2019: 8th ISCA Workshop on Speech and Language Technology in Education
20-21 September 2019, Graz, Austria

Fluidity: Developing second language fluency with real-time feedback during speech practice

Ralph L. Rose[1]

[1]Waseda University Faculty of Science and Engineering
rose@waseda.jp

Abstract

Fluidity is a JavaFX framework application that is designed to give second language learners instruction in how to develop their speech fluency. While learners are practicing giving a monologic speech, the application measures certain features of their speech fluency (e.g., phonation vs. silence time, syllable count, pause count) and updates these measures in real-time. The application also gives real-time feedback to learners through an on-screen avatar nicknamed "Fludie" that changes its facial expression according to the fluency measures: for example, pleased when fluent, but not satisfied when disfluent. After each speech practice, learners may review their practice through playback and while viewing multiple visualizations of their speech fluency. Future development plans of Fluidity include increased capability for gamification, more sophisticated feedback to learners, customization of desired fluency profiles, and more speech practice methods.

Index Terms: speed fluency, breakdown fluency, automated feedback, second language speech, avatar

1. Introduction

Language teaching pedagogy is often built on a componential model of language proficiency such as that of Skehan [1] which consists of three components: complexity, accuracy, and fluency. The present work concerns the last of these, fluency, understood here to refer somewhat narrowly to the smoothness of speech. Here, fluency may yet be seen in three aspects: speed fluency, comprising the rate of speech production; breakdown fluency, including various ways that speech 'breaks down' by such interruptions as silent or filled pauses (e.g., in English, *uh*/*um*); and repair fluency, comprising the ways that speakers repair their ongoing speech production, such as through self-corrections. Methods for developing fluency involve such things as repeated practice on speaking tasks [2-3], among other techniques.

When learners receive feedback on their fluency, it often takes the form of a holistic or composite measure, derived algorithmically from a variety of objective measures, such as speech rate, pause rate and duration, or length of runs (a stretch of uninterrupted speech). Applications exist which can produce such feedback automatically including both commercial (e.g., ETS SpeechRater [4], Versant [5]) and non-commercial [6-7] applications. These systems may provide feedback with relatively low latency, taking perhaps minutes. However, none of these systems provide real-time feedback comparable to that received during human-human communication (e.g., facial expressions, head and eye movements). The present work describes a system that is designed to provide such real-time feedback to learners through an on-screen avatar.

2. Fluidity application

2.1. Objectives

There are five fundamental objectives on which the Fluidity application is designed, as follows.

- To provide second language learners with various means to practice speaking to a simulated interlocutor.
- To measure fluency-related features of learners' speech and update these measures in real-time.
- To provide feedback on learners' fluency in real-time in a manner that emulates human-human communication
- To provide learners an opportunity to review their practice and raise their own fluency awareness.
- To provide learners an opportunity to practice and develop their fluency in a game-like environment.

2.2. User experience

When users launch Fluidity, they initially access the dashboard (see Figure 1), which provides links to different sorts of speech practice options, their speech practice history, a record of achievements obtained so far, as well as links to help and system settings. Also featured is an on-screen face avatar, nicknamed "Fludie", which is animated and appears on nearly every Fluidity screen.

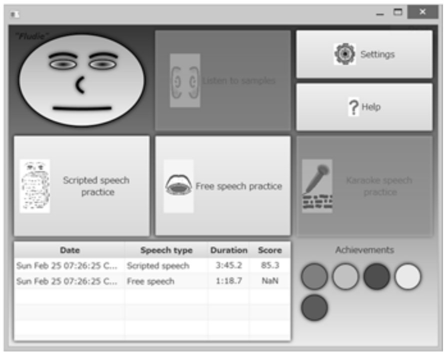

Figure 1: *Fluidity application dashboard*

Users can engage in different sorts of practice. In each, they record themselves speaking in response to different task assignments (Figure 2, left and center). While they speak, Fludie's facial expressions change in response to the ongoing fluency of their speech. That is, if a learner's speech is well-paced, then Fludie will continue to show a pleased face and show interest in the speech. But if a learner becomes too hesitant or their speech is too slow, then Fludie may show a

Figure 2: *Fluidity scripted speech (left) and free speech (center) practice screens and post-practice visualization screen (right)*

more unsatisfied expression, or possibly even interrupt. The triggers for these expression changes are dependent on certain thresholds related to the measured fluency features.

After any practice, learners may also see a visualization of their production (Figure 2, right), showing such things as a waveform as well as a silence/speech interval map. On this screen, they may playback and review their speech.

2.3. Application architecture

Fluidity is programmed in Java using the JavaFX framework and is therefore a cross-platform application. It uses the TarsosDSP library [8] for signal processing. The system is designed to detect and keep track of several basic speech parameters: phonation time, silence time, syllable count (estimated by energy peak detection; cf., [7]), silent pause count (using a default minimum of 300 ms for pause length; cf., [9]), and filled pause count (estimated by detecting stable formants and pitch; cf., [10-11]). Fluidity does not utilize any speech recognition engine. In short, it tracks several features of the speaker's speed and breakdown fluency, but does not perform any higher level linguistic analysis.

3. Testing

To date, Fluidity has undergone some testing. Usability evaluations [12] have been performed in Japan, Portugal, France, and USA. Users report that the application is easy to use and that it gave them a different understanding and awareness of their speech fluency than before.

Other testing of Fluidity has included validation of the detection mechanism [13]. Results show that the mechanism outperforms a widely used script [14] in fluency studies on most measures and is robust against increased background noise. However, it remains weak in filled pause detection with both false positive and false negative cases.

4. Conclusions

Future development of Fluidity is expected to focus on the creation of further practice methods for learners, validation of and testing of the effectiveness of Fludie as a vehicle for feedback, and the development of more sophisticated algorithms for detecting fluency features. For instance, although real-time measurements will likely be constrained to the current set, more sophisticated off-line analysis might be feasible, based on speech recognition, parsing, alignment, and then syntactic, semantic, and prosodic analysis. Finally, an on-line component to the application is planned, which might allow teachers to monitor their students' practice or allow learners to use the application in a more socially productive manner in coordination with other learners.

5. Acknowledgements

This work was partially funded by a Japan Society for the Promotion of Sciences Grant-in-aid (Scientific Research (C), #24520661) and a Waseda University Grant for Special Research Projects (#2017K-220). I am grateful to Laurence Anthony, Fernando Batista, Sylvie De Cock, Lorenzo García-Amaya, Gaëtanelle Gilquin, Sylviane Granger, Nicholas Henriksen, Helena Moniz, Rubén Solera Ureña, and Jürgen Trouvain for insightful conversations about this work.

6. References

[1] P. Skehan, *A Cognitive Approach to Language Learning*. Oxford: Oxford University Press, 1998.

[2] M. Bygate Effects of task repetition: Appraising the developing language of learners. In J. Willis & D. Willis (Eds.), *Challenge and change in language teaching*. Oxford: Heinemann, 1996, pp. 136–146.

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

[4] K. Zechner, D. Higgins, X. Xi, and D. M. Williamson, "Automatic scoring of non-native spontaneous speech in tests of spoken English," *Speech Communication*, vol. 51, no. 10, pp. 883–895, 2009.

[5] J. Bernstein, *PhonePassTM testing: Structure and construct*. Menlo Park, CA: Ordinate, 1999.

[6] C. Cucchiarini, A. Neri, and H. Strik, "Oral proficiency training in Dutch L2: The contribution of ASR-based corrective feedback," *Speech Communication*, vol. 51, no. 10, pp. 853–863, 2009.

[7] S. Bhat, M. Hasegawa-Johnson, and R. Sproat, "Automatic Fluency Assessment by Signal-Level Measurement of Spontaneous Speech," *Proceedings of Second Language Studies: Acquisition, Learning, Education and Technology*, paper O2-1, 2010.

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

[9] N. H. De Jong and H. R. Bosker, "Choosing a threshold for silent pauses to measure second language fluency," in *Proceedings of The 6th Workshop on Disfluency in Spontaneous Speech, August 21-23, Stockholm, Sweden*, pp. 17–20, 2013.

[10] K. Audhkhasi, K. Kandhway, O. D. Deshmukh, and A. Verma, "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.

[11] S.-C. Tseng, *Grammar, prosody and speech disfluencies in spoken dialogues*. Dissertation, Bielefeld University, 1999.

[12] R. Rose, "Real-time feedback on speed fluency with Fluidity," Japan Association for Language Teaching (JALT) International Conference, November 23–26, 2018.

[13] R. Rose, "Fluidity: Real-time Feedback on Acoustic Measures of Second Language Speech Fluency," Unpublished study.

[14] H. Quené, I. Persoon, and N. de Jong, Syllable Nuclei v2 [Praat Script]. Version 28 Feb 2011. https://sites.google.com/site/speechrate/Home/praatscript-syllable-nuclei-v2 (accessed on December 26, 2014)