Correlation of perceived fluency with phonetic measures of speech rate and pausing

Peter Kleman
Department of English and American studies
Faculty of Philosophy
Constantine the Philosopher University
Štefánikova trieda 38/67, Nitra, 949 10, Slovakia
peter.kleman@ukf.sk

Štefan Beňuš
Department of English and American studies
Faculty of Philosophy
Constantine the Philosopher University
Štefánikova trieda 38/67, Nitra, 949 10, Slovakia

Institute of Informatics of the Slovak Academy of Sciences
Dúbravská cesta 9, 841 04 Bratislava, Slovakia
sbenus@ukf.sk

Abstract. *The paper studies the relationship between perceived fluency of L2 semi-spontaneous utterances and phonetic measures such as speech rate and the number of pauses. The data for the correlation analysis comes from a word guessing experiment conducted with Slovaks speaking English. Subjects provided cues for target words intended to facilitate the correct guessing of those words. In the second phase, speakers were asked to guess the words to which the interlocutors were providing cues. The guessers were also asked to evaluate the fluency of the interlocutors for each of the words that the speakers were guessing. The data from the recordings is analysed through a correlation analysis of the phonetic measures extracted from the acoustic signal and the level of perceived fluency that was elicited for each target word. The study found that phonetic measures do correlate with the levels of perceived fluency. The findings may be used for improvements in automated computer assisted fluency assessment.*

Introduction

The study of the relationship of fluency and phonetic measures is an endeavour that will prove to be useful when it comes to fully understanding how humans perceive fluency of their peers and will aid in the pursuit of creating of automatic fluency measuring algorithms and programs. Such technological advances will be useful in the coming age of intelligent self-learning computer that will be able to understand, evaluate, and perhaps even study human languages.

De Jong and Wempe conducted a study in 2009 [1] using PRAAT to automatically detect syllable nuclei in order to measure speech rate. The data used in the study came from experiments performed by 8 participants with tasks such as reading aloud syllable lists and informal storytelling. They conducted a correlation analysis on the predicted data obtained from the analysis in relation to human syllable counts done on the data from the experiments. This study concluded that automatic syllable count could reliably assess and compare speech rates.

Kallio, Suni, Virkkunen, and Šimko conducted a study in 2018 [2] on whether prosodic prominence levels of syllables could be used to predict the prosodic competence of L2 speakers of Swedish. They used a continuous wavelet transformation analysis of syllable prominence with combinations of f_0, energy, and duration features. The data for the test was gathered from a larger corpus created during a computer-aided oral test. They manually annotated the data to syllable-level and measured f_0 using PRAAT. This data was assessed using wavelet transformation analysis. The second set of assessments was gathered from expert raters. The results showed that the assessments correlated to the assessments of expert raters. This data provided strong support for future use of wavelet-based prominence estimation in automatic assessment of L2 proficiency.

Ramanarayanan, Lange, and Evanini studied the human and automated scoring of fluency, pronunciation, and intonation [3]. They collected interactions of L2 speakers of English and used both human and machine learning for creation of scores for each of the aspects. The study showed that trained scoring models were generally on par with human raters' scores.

Therefore, for such automated assessments we need two separate sets of data. The first set consists of subjective data gathered from evaluation of fluency provided by subjects [4]. The second set of data consists of phonetic measures that were previously studied and had their importance assessed [5]. Such approach to data gathering was also used in the following study. With the increased volume of such data available, the algorithms can be improved to incorporate more measures that aid the computer in better assessing various aspects of human speech.

The aim of the study was to search for a statistically significant correlation between perceived fluency and phonetic measures. This was firstly studied across the data from all the speakers in one group. Secondly, they were also divided into groups, which consisted of assessors of the same proficiency level. We expected that the correlation should be better with all subjects taken into account as assessors.as opposed to only same proficiency group assessors. The rationale behind this statement is that the more varied points of view we have on assessment, the better the correlation results will be. This was also meant to avoid the extremes that were predicted to come up in the analyses.

1.1 Definitions

For a number of L2 speakers of English, fluency seems to be an elusive language feature that they can never quite master. Various disfluencies can have an impact on the speech of a person, both natives and non-natives, as previously demonstrated in research [4]. Previous research in fluency provides several definitions of what fluency actually is [3, 7, 8, 9, 10, 11, 12], but there does not seem to be an agreed upon definition that is accepted by all. In general, fluency is considered to be the overall proficiency of a speaker that uses a language at a high level [13, 14, 15]. The same general definition can be used for L2 Fluency as well. Fluency was also used as an umbrella term, when it was divided into a broad sense and a narrow sense of fluency [16]. The broad sense shares a similar definition to the previously mentioned, while the narrow sense of fluency is referring only to the speed and smoothness of delivery.

Perceived fluency is defined as "inferences listeners make about a speaker's cognitive fluency based on their perception of utterance fluency" [17]. This aspect of fluency was important for the creation of the experiment, since it provided understanding of how subjective fluency is perceived and what constitutes as fluent speech in the narrow sense that can be used for analysis. The analysis of perceived fluency and phonetic measures is a new direction for the automated assessment of fluency.

2 Methodology

Two previously mentioned ideas [16, 17] were joined in the creation of the current study. Smoothness was be represented by the frequency and length of pauses and the speed with words per second and the overall wordcount. Perceived fluency [17] was used as a subjective measure that was collected from subjects in the experiment.

The basis for the study was a semi-spontaneous word guessing experiment conducted on 13 L2 speakers of English with proficiency levels of C1, B2, and B1. The experiment was divided into two phases, where in the first phase the subjects were tasked with creating cues for a set of provided words. These words were randomly chosen from the British National Corpus with the criteria of being at most three syllables long and were either a noun, verb, or an adjective. Each speaker was given a set of ten words and they were asked to create two cues for each word. They were asked not to use the words that they were hinting at. The cues that they provided were recorded and concatenated into a single recording for each of the speakers. These recordings always consisted of the first cue for the word, three second pause provided for the guessers as thinking space, then the second cue for the word, followed by another three second pause.

The recordings processed in this way were used in phase two, where the subjects were asked to try and guess the words to which the interlocutors were providing cues. Each subject listened to the recordings of all other subjects. They were asked to listen to the cues and try to guess the word that the interlocutor was providing the cues for. The success of guesses was recorded for future use. After the subjects listened to both of the cues for each word, they were asked to evaluate the fluency level of the interlocutors on a scale of 1 to 7. Since all subjects were naïve assessors, they were mainly asked to focus on guessing the words from cues. They were asked to provide spontaneous assessments of fluency. The experimenter marked the perceived fluency assessments for each of the words. Each of the subjects provided 10 assessments for all of the 12 speakers resulting in a data set of 120 assessments for each subject.

2.1 Data processing

In the data processing, the recordings from phase one were labelled using PRAAT speech analysis software. Each recording was annotated in three tiers. The first was the cue tier in which the cues were labelled from their beginning to their end. The second was the word tier, where each of the words was labelled from its beginning to its end. And the third was the pause tier, where each of the pauses was labelled from its beginning to its end.

A Praat script was then used to extract the number of words in each cue and their length, and also the number of inside cue pauses and their length from these annotations. The data was transferred into an Excel sheet where the words per second were counted as the sum of words in both cues divided by the sum of word durations in both cues and the inside cue pause duration in both cues. The overall wordcount was calculated as the sum of words in both cues. The overall pause count was calculated as the sum of inside cue pauses in both cues. Lastly, the overall duration of pauses was calculated as the sum of inside cue pause duration in both cues. The levels of perceived fluency were also added to each word as evaluated by each of the subjects.

The first data set was created from the evaluations of fluency that were provided by subjects during the word guessing experiment. The second set of data consisted of four different phonetic measures that were chosen for the correlation analysis in relation with the evaluated levels of fluency. These measures are words per second, wordcount, length of pauses, and the number of pauses. Such pair of data is referred to as an objective-subjective pair or subjective-objective approach [18]. The measures were used as an objective means of assessing fluency in relation to the subjective evaluation of perceived fluency that were provided by the participants while listening to cues from their peers.

The research examined the correlation of perceived fluency and phonetic measures analysed in the recording data from phase one. The average level of perceived fluency was calculated for each of the words from the normalised fluency evaluations in the following way. Since the data was displayed as a chart, we had the perceived fluency evaluations from each speaker as columns. Each of the cue pairs had an original evaluation value of one to seven and was represented as a row. In order to normalise the data, we took each of the evaluations and subtracted from it the minimum score that the speaker provided in their entire column. This number was divided by the difference between the maximum per column and

minimum per column. The result was a number between 0 and 1, where 0 represented the lowest score provided by the speaker and 1 the highest score.

2.2 Data analysis

The correlation of data was studied in four cases calculating the Pearson correlation coefficient and also multiple linear regression. Each pair for the calculation of Pearson correlation coefficient consisted of perceived proficiency evaluation, and a phonetic measure. The first pair used words per second as the independent variable, the second used wordcount, the third used the number of pause, and the fourth used the total duration of inside cue pauses as its independent variable.

3 Results

3.1 Results for all speakers

As mentioned before, four pairs of data sets were created for the calculation of Pearson correlation coefficient. In the first pair of data sets, which consisted of words per second and perceived fluency, a Pearson r was computed to assess the relationship between perceived fluency and words per second. We found positive significant relationship ($r = 0.574$, $p < 0.001$). The relationship between the two variables is visualised in a scatterplot shown in Fig. 1.

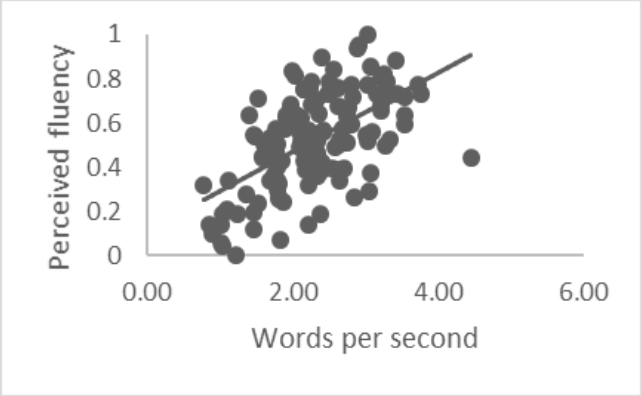

Fig. 1. Correlation data for words per minute and perceived fluency

In the second pair of data sets, which consisted of the wordcount in both cues per word and perceived fluency, a Pearson r was computed to assess the relationship between perceived fluency and wordcount. We found positive significant relationship ($r = 0.316$, $p < 0.001$). The data sets were visualised in a scatterplot graph as shown in Fig. 2.

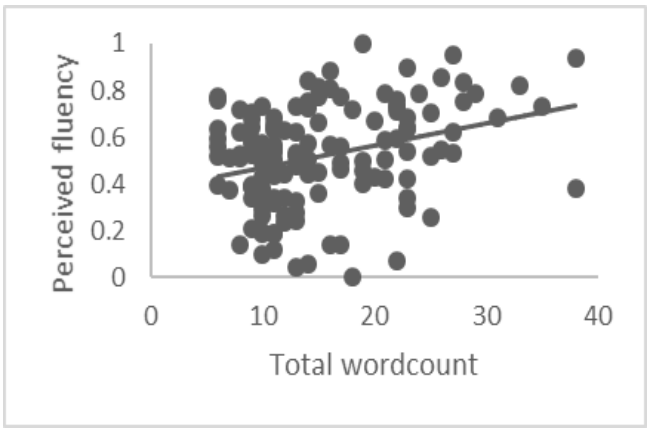

Fig. 2. Correlation data for wordcount and perceived fluency

In the third pair of data sets, which consisted of the sum of the number of inside cue pauses and perceived fluency, a Pearson r was computed to assess the relationship between perceived fluency and total pause count. We have not found a significant relationship suggesting that the pair does not correlate ($r = -0.098$, $p < 0.579$). The data visualisation is available in a scatterplot graph as shown in Fig. 3.

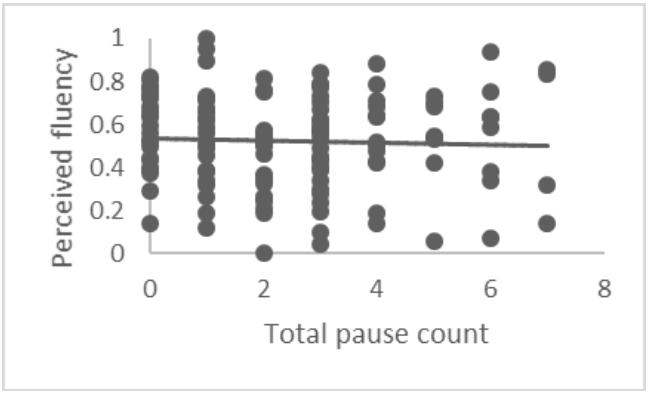

Fig. 3. Correlation data for the number of pauses and perceived fluency

In the fourth pair of data sets, which consisted of the total duration of pauses inside both cues per word and perceived fluency, a Pearson r was computed to assess the relationship between perceived fluency and total pause duration. We found negative significant relationship ($r = -0.479$, $p < 0.001$). The data visualisation is visible in Figure 4.

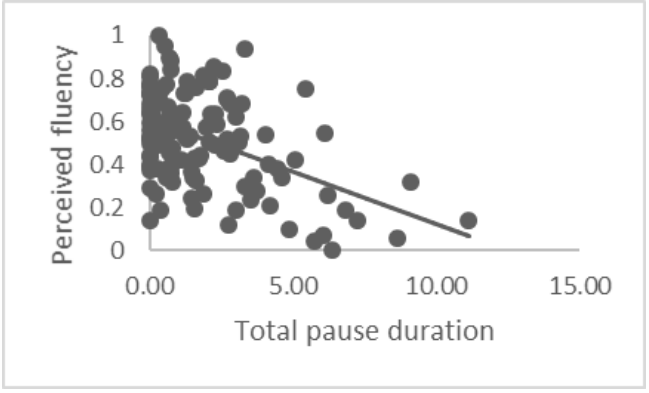

Fig. 4. Correlation data for the total inside cue pause duration and perceived fluency.

A multiple linear regression was calculated to predict perceived fluency based on the words per second, wordcount, and pause duration. Pause count was omitted, as it did not seem to have an effect on perceived fluency based on the correlation result above. A significant regression model was found ($F(3,126) = 39.333$, $p < 0.001$), with an R^2 of 0.484. Subject's predicted perceived fluency is shown in Table 1. Subject's perceived fluency increased by 0.068 for each word per second, by 0.013 for each word, and decreased by -0.046 for each second in total pause duration. The coefficients in the table represent each of the phonetic measure that were used. The Intercept represents the perceived fluency. All three measures were significant predictors of perceived fluency.

Table 1. Results of multiple linear regression calculations

R Square	0.484		
	Coef	*t Stat*	*P-value*
Intercept	0.243	3.076	0.003
wps	0.068	2.195	0.030
wordcount	0.013	5.921	0.000
icp_dur	-0.046	-4.322	0.000

3.2 Results for each proficiency group

The data was then divided into three proficiency groups and was again analysed using the Pearson correlation coefficient and multiple linear regression. This was done in order to study which phonetic measures influence the relationship between produced and perceived fluency in each of the proficiency groups. Three groups were created, each consisting of either only C1 level speakers, B2 level speakers, or B1 level speakers. All the assessments made by these speakers were taken into account and a new value for perceived fluency was calculated from their evaluations.

3.2.1 Level C1

Firstly, we will talk about the results for the group of C1 assessors. Four Pearson r values were computed to assess the relationship between the four data pairs. In this group, only the perceived fluency values of the C1 subjects were taken into account. The Pearson r values were also measured for their statistical significance with a p-value. This data is visible in Table 2.

Table 2. Pearson r results for group C1

R_wps_pf	0.500
p-value	$p < 0.001$
R_wc_pf	0.339
p-value	$p < 0.001$
R_icpc_pf	-0.069
p-value	$p < 0.437$
R_icpd_pf	-0.410
p-value	$p < 0.001$

In their first pair of data sets, which consisted of words per second and perceived fluency, the Pearson r suggests positive significant relationship ($r = 0.500$, $p < 0{,}001$).

In their second pair of data sets, which consisted of the wordcount in both cues per word and perceived fluency, the Pearson r suggests positive significant relationship ($r = 0.339$, $p < 0{,}001$).

In their third pair of data sets, which consisted of the total number of inside cue pauses and perceived fluency, the Pearson r suggests no significant relationship ($r = -0.069$, $p < 0{,}437$).

In their fourth pair of data sets, which consisted of the total duration of pauses inside both cues per word and perceived fluency, the Pearson r suggests negative significant relationship ($r = -0.410$, $p < 0{,}001$).

A multiple linear regression was calculated to predict perceived fluency based on the words per second, wordcount, and pause duration. Pause count was omitted, as it did not seem to have an effect on perceived fluency based on the correlation result above. A significant regression model was found ($F(3,126) = 29.793$, $p < 0.001$), with an R^2 of 0.415. Subject's predicted perceived fluency is shown in Table 3. Subject's perceived fluency increased by 0.049 for each word per second, by 0.014 for each word, and decreased by -0.046 for each second in total pause duration. The coefficients in the table represent each of the phonetic measure that were used. The Intercept represents the perceived fluency All three measures were significant predictors of perceived fluency.

Table 3. Results of multiple linear regression calculation in group C1

R Square	0.415		
	Coef	*t Stat*	*P-value*
Intercept	0.238	2.772	0.006
wps	0.049	1.441	0.152
wordcount	0.014	5.856	0.000
icp_dur	-0.046	-3.936	0.000

3.2.2 Level B2

The second set of analyses was conducted on the B2 group. The results for the group are shown below in the tables and they consist of four Pearson r values, which were computed to asses the relationship between the data pairs. In this group, only the perceived fluency values of the B2 subjects were taken into account. The p-values were also

measured for their statistical significance. This data is visible in Table 4.

Table 4. Pearson r results for group B2

R_wps_pf	0.487
p-value	p < 0.001
R_wc_pf	0.257
p-value	p < 0.003
R_icpc_pf	0.019
p-value	p < 0.828
R_icpd_pf	-0.408
p-value	p < 0.001

In their first pair of data sets, which consisted of words per second and perceived fluency, the Pearson r suggests positive significant relationship (r = 0.487, p < 0,001).

In their second pair of data sets, which consisted of the wordcount in both cues per word and perceived fluency, the Pearson r suggests positive significant relationship (r = 0.257, p < 0,003).

In their third pair of data sets, which consisted of the total number of inside cue pauses and perceived fluency, the Pearson r suggests no significant relationship (r = 0.019, p < 0, 828).

In their fourth pair of data sets, which consisted of the total duration of pauses inside both cues per word and perceived fluency, the Pearson r suggests negative significant relationship (r = -0.408, p < 0,001).

A multiple linear regression was calculated to predict perceived fluency based on the words per second, wordcount, and pause duration. Pause count was omitted, as it did not seem to have an effect on perceived fluency based on the correlation result above. A significant regression model was found (F (3,126) = 21.742, p < 0.001), with an R^2 of 0.341. Subject's predicted perceived fluency is shown in Table 5. Subject's perceived fluency increased by 0.071 for each word per second, by 0.013 for each word, and decreased by -0.046 for each second in total pause duration. The coefficients in the table represent each of the phonetic measure that were used. The Intercept represents the perceived fluency. All three measures were significant predictors of perceived fluency.

Table 5. Results of multiple linear regression calculation in group B2

R Square	0.341		
	Coef	*t Stat*	*P-value*
Intercept	0.276	2.598	0.011
wps	0.071	1.697	0.092
wordcount	0.013	4.284	0.000
icp_dur	-0.046	-3.192	0.002

3.2.3 Level B1

The final group of assessors that we will talk about is the B1 group. The relationship between the four data pairs was assessed with the help of four Pearson r values, which were computed. These values were also measure for their statistical significance with a p-value. All the data belonging to B1 group can be seen in Table 6.

Table 6. Person r results for group B1

R_wps_pf	0.579
p-value	p < 0.001
R_wc_pf	0.246
p-value	p < 0.003
R_icpc_pf	-0.104
p-value	p < 0.309
R_icpd_pf	-0.505
p-value	p < 0.001

In their first pair of data sets, which consisted of words per second and perceived fluency, the Pearson r suggests positive significant relationship (r = 0.579, p < 0,001).

In their second pair of data sets, which consisted of the wordcount in both cues per word and perceived fluency, the Pearson r suggests positive significant relationship (r = 0.246, p < 0,003).

In their third pair of data sets, which consisted of the total number of inside cue pauses and perceived fluency, the Pearson r suggests no significant relationship (r = -0.104, p < 0, 309).

In their fourth pair of data sets, which consisted of the total duration of pauses inside both cues per word and perceived fluency, the Pearson r suggests negative significant relationship (r = -0.505, p < 0,001).

A multiple linear regression was calculated to predict perceived fluency based on the words per second, wordcount, and pause duration. Pause count was omitted, as it did not seem to have an effect on perceived fluency based on the correlation result above. A significant regression model was found (F (3,126) = 35.438, p < 0.001), with an R^2 of 0.458. Subject's predicted perceived fluency is shown in Table 7. Subject's perceived fluency increased by 0.079 for each word per second, by 0.013 for each word, and decreased by -0.050 for each second in total pause duration. All three measures were significant predictors of perceived fluency.

Table 7. Results of multiple linear regression calculation in group B1

R Square	0.458		
	Coef	*t Stat*	*P-value*
Intercept	0.239	2.697	0.008
wps	0.079	2.257	0.026
wordcount	0.013	5.032	0.000
icp_dur	-0.050	-4.152	0.000

4 Discussion

In this study we set out to search for a statistically significant correlation between perceived fluency and phonetic measures that would be observable across the data from all the speakers and also in groups, which consist of assessors of the same proficiency level. We expected that the correlation should be better with all subjects taken into account as assessors.as opposed to only using assessors of certain proficiency groups. The rationale behind this statement is that the more varied points of view we have on assessment, the more accurate the results will be.

The study found some of the phonetic measures seemed to correlate with perceived fluency much more in simple pair tests. One such measure is words per second. If we look purely at its relationship to perceived fluency, we see a moderately high positive correlation. However, this did not seem right, since such analysis did not take into account the relation with the other measures. The pause count showed no significant relationship. This could probably be caused, because the subjects were mainly tasked with guessing a word from the cues. Since they were probably more focused on the message, the number of pauses did not seem to play a role. They started noticing the pauses only when their duration was too long.

Even though the Pearson r showed a lesser correlation in the initial analyses, this changed after a linear regression analysis was used. This analysis took into account all the data necessary for the correlation analysis. This means that it measured the significance of all the measures in relation to perceived fluency at the same time and not only in individual pairs. The results of this analysis showed a different picture of the measure significance. The most prominent became the wordcount with its positive relationship, the second was the duration of pauses with a negative relationship, and words per second were third with a positive relationship.

The same ordering of measures was also observed in the group phase of analyses. The speakers were divided into groups based on their proficiency levels. In these groups only their fluency assessments were taken into account. We saw a change in the strength of correlation of all the pairs in all the groups. This means that pair one, which is the words per second and perceived fluency pair, had a completely different value in all the pairs. This difference is easily observed between the B2 pair one r = 0.487 and B1 pair one r = 0.579. Such differences were observed across all the pairs and suggest that each different proficiency level evaluates fluency based on different criteria.

The study showed that the best correlating data was observed, when all speaker were used as assessors. This suggests that the before mentioned differences in pair correlations are equalized. This offers a better correlation analysis partially also because of the higher number of assessors.

Acknowledgment

This work was funded by the Slovak Scientific Grant Agency VEGA “Automatic assessment of acute stress from speech”, grant No. 2/0161/18 and also by University Grant Agency UGA “Manipulation of acoustic signal of speech for improvement of fluency in a foreign language and targeted reduction of mother tongue interference”, grant No. I-19-208-02.

References

[1] N.H. De Jong and T. Wempe, “Praat script to detect syllable nuclei and measure speech rate automatically,” *Behavior Research Methods* 41, pp. 385-390, 2009.

[2] H. Kallio, A. Suni, P. Virkkunen, and J. Simko, “Prominence-based evaluation of L2 prosody,” *Interspeech* 2018, pp. 1838-1842, 2018.

[3] V. Ramanarayanan, P. Lange, K. Evanini, H. Molloy, and D. Suendermann-Oeft, “Human and automated scoring of fluency, pronunciation and intonation during human–machine spoken dialog interactions,” *Interspeech* 2017, pp. 1711–1715, 2017.

[4] H. R. Bosker, H. Quené, T. Sanders, and N. H. Jong, . “The perception of fluency in native and non-native speech,” *Language Learning*, 64 (3), pp. 579-614, 2014.

[5] J. Kormos and M. Dénes, “Exploring measures and perceptions of fluency in the speech of second language learners,” *System* 32 (2), pp. 145-164, 2004.

[6] T. Rasinski, “The Fluent Reader: Oral Reading Strategies for Building Word Recognition, Fluency, and Comprehension,” *Scholastic Inc*., 2003.

[7] A. Hasselgreen, “Testing the Spoken English of Young Norwegians: A Study of Testing Validity and the Role of Smallwords in Contributing to Pupils' Fluency,” *Cambridge University Press*, 2005.

[8] Z. Breznitz, “Fluency in Reading: Synchronization of Processes,” Routledge, 2006.

[9] J. B. Gilquin and S. De Cock, “Errors and Disfluencies in Spoken Corpora,” *John Benjamins Publishing*, 2013.

[10] A. Khateb and I. Bar-Kochva, “Reading Fluency: Current Insights from Neurocognitive Research and Intervention Studies,” *Springer*, 2016.

[11] P. Kendale, “WORKBOOK for Spoken English Fluency Development – 4,” *Independently Published*, 2017.

[12] L. Wang, J. Zhang, F. Pan, B. Dong, and Y. Yan, “Automatic Fluency Assessment of Non/native English Reading,” *Journal of Convergence Information Technology* 7, pp. 636-642, 2012.

[13] P. Lennon, “Investigating fluency in EFL: A quantitative approach,” *Language Learning*, vol. 40, pp. 387-417, 1990.

[14] H. Riggenbach, “Toward an understanding of fluency: A microanalysis of non-native speaker conversations,” *Discourse Processes*, vol. 14, pp. 423-441, 1991.

[15] J. Kormos, “Speech production and second language acquisition,” *Lawrence Erlbaum Associates*, 2006.

[16] P. Lennon, “The lexical element in spoken second language fluency,” *In H. Riggenbach (Ed.), Perspectives on fluency Ann Arbor, University of Michigan Press*, pp. 25-42, 2000.

[17] N. Segalowitz, “Cognitive bases of second language fluency,” *New York: Routledge*, 2010.

[18] N. H. De Jong, et. al. “Facets of Speaking Proficiency,” *Studies in Second Language Acquisition*, vol. 34 (1), pp. 5-34, 2010.