Fluency in non-native read and spontaneous speech

Catia Cucchiarini, Joost van Doremalen, Helmer Strik

Department of Linguistics, Radboud University Nijmegen, The Netherlands
{c.cucchiarini, j.vandoremalen, h.strik} @let.ru.nl

Abstract

Various studies have investigated the temporal aspects of non-native speech and their relation to perceived fluency, because fluency constitutes an important aspect of second language proficiency. For this purpose it is important to determine which measures are most strongly correlated with perceived fluency and how these measures vary. In the present study objective measures related to perceived fluency were calculated for read and spontaneous speech of non-native speakers of Dutch. The results indicate that the objective measures vary as a function of different variables. Suggestions are made for future investigations so as to facilitate comparisons between studies and meta-analyses.

Index Terms: fluency, non-native speech, temporal measures

1. Introduction

Fluency is viewed as an important aspect of second language speech production and as such it is often included in tests of second language skills [14]. A review of the fluency literature reveals, however, that the term fluency has been used to refer to different constructs (for a brief review, see [1]), varying from overall language performance to more restricted definitions that concern the temporal aspects of L2 speech. This latter definition, which is found in [12; 11; 14; 15; 16], makes reference to ‘‘native-like rapidity’’ [11: 390]. According to this interpretation, the goal in second language learning would be to produce ‘‘speech at the tempo of native speakers, unimpeded by silent pauses and hesitations, filled pauses, self-corrections, repetitions false starts and the like’’ [11: 390]. owever, HH However, various studies have indicated that even native speech is not always smooth and continuous and exhibits many hesitations and repairs [11; 14].

The interest in studying and evaluating temporal aspects of L2 learners’ speech production to establish its degree of fluency has different motivations, varying from gaining insight into the factors that affect L2 acquisition to developing automatic tests of oral proficiency [18]. For this latter purpose it is very important to find out which temporal measures are most strongly correlated with perceived fluency and how these measures vary depending on task and type of speech.

In this paper we present the results of a study on temporal measures of fluency in non-native read and spontaneous speech. In section 2 we present and discuss related work on this topic and explain the rationale behind the present study. In sections 3, 4 and 5 we describe the method adopted in our study, the results obtained and relate them to those of previous studies, with the aim of increasing our understanding of fluency in non-native speech.

2. Related research

In previous studies we addressed the question of whether and how objective, temporal measures of speech production could be employed to develop automatic fluency tests [1; 2]. In these studies traditional temporal measures as proposed by Grosjean [9] were employed. The precise definitions of these measures were slightly adapted to make them suitable for automatic calculation and to relate them to sentence length. This made it possible to use these measures in comparative studies that employ different speech material [2]. These studies revealed that some objective, temporal measures of speech are good indicators of perceived fluency, that is fluency as assessed by human raters. However, the magnitude of the correlation between objective measures and human ratings of fluency turned out to vary depending on the type of speech and the type of task the speaker is engaging on.

Other authors addressed the question of how objective measures of fluency are related to ratings assigned by expert judges [10; 6; 17]. The results of the various studies reveal similarities and differences. In general, measures such as phonation time ratio and speech rate appear to be good indicators of fluency, although the magnitude of the correlations varies in the different studies. This is not only related to differences between the tasks investigated, but could also be connected to the precise definitions of the objective measures employed. For instance, while [1; 2] employed phones as the unit to calculate measures such as articulation rate and speech rate, [10; 6; 17] used the syllable as the unit of computation. Although the relative advantages of using either phones or syllables might be related to the specific language under investigation and its syllabic structure, it is worth mentioning that the predicting power of the phone-based measures employed by [1; 2] appeared to be higher than that of the syllable-based measures.

Another possible influencing factor is the rubric adopted for the human ratings. In the studies by [1; 2; 3]. the judges were asked to distinguish different dimensions, such as fluency, speech rate and pronunciation quality. The raters managed to keep these dimensions distinct, which is borne out by the correlations between the various objective measures and the rating scales. Other studies adopted less strictly defined dimensions such as accentedness [17].

Finally, the differences between the results obtained in previous studies may be related to the definition of pauses that was adopted. In [1; 2] a silent pause was defined as: a stretch of silence with a duration of no less than 0.2 s, just like in [10]. In [17] pauses were 0.1 s, while in [6] a threshold of 0.4 s was adopted. The two latter studies found no strong correlation between number of pauses and perceived fluency. [10] found no strong correlation either, but number of pauses appeared not to vary much between low and advanced speakers, while in [2] considerable differences were observed between the three proficiency groups beginners, intermediate and advanced.

To summarize, although the various studies on the relationship between objective measures and human ratings of fluency reveal some differences, there are indications that temporal objective measures of speech production are related to fluency ratings to various degrees and are in any case worth studying.

3. The present study

In [2] we presented data from two experiments that employed read and spontaneous speech and studied the relationship between objective measures of speech and fluency ratings assigned by human judges. These studies provided useful insights into how objective measures can vary between read and spontaneous speech. However, since the data were collected from different speakers, they were not optimally suited for studying the differences between read and spontaneous speech. To make a fair comparison it would be better to compare read and spontaneous speech data that stem from the same speakers. Such data were collected within the framework of the JASMIN project [4], which was aimed at compiling a corpus of speech of children, non-natives and elderly people.

For the purpose of getting a better understanding of how objective measures of fluency vary between read and spontaneous speech we calculated such measures for a group of 44 speakers for each of which samples of read and spontaneous speech were available, as will be explained below. So, the data presented in the rest of this paper concern objective measures of fluency and not perceived fluency, since the latter would imply collecting ratings by human judges.

4. Method

4.1. Speech material

The speech material for the present experiments was taken from the non-native component of the JASMIN speech corpus [4]. Recording of read and spontaneous speech were made of speakers with different mother tongues and relatively low proficiency levels (A1, A2 and B1 of the Common European Framework).

The read speech material consists of utterances produced by 44 speakers while reading aloud short texts from the screen and sets of phonetically rich sentences. The spontaneous speech material was derived from human-machine dialogues which were collected through a Wizard-of-Oz-based platform [4]. In these dialogues the speakers reply 39 questions about a journey. The dialogues reflect typical situations in human-machine interaction where speakers produce phenomena such as hyperarticulation, syllable lengthening, shouting, stress shift, restarts, filled pauses, silent pauses, self talk, talking to the machine, repetitions, prompt/question repeating and paraphrasing.

4.2. Automatic speech analysis

To calculate the quantitative measures, a segmentation of the speech signal into phonemes is needed. These segmentations were created by doing a Viterbi alignment with the SPRAAK toolkit [5] on the basis of an orthographic transcription, a pronunciation lexicon and acoustic models.

Orthographic transcriptions were created manually. The pronunciation lexicon was based on the CGN database lexicon [13]. Some pronunciation phenomena, such as broken words and pauses in a word, were annotated manually in the lexicon.

To obtain reliable acoustic models initial acoustic models trained on Dutch native speech were adapted by performing a single training pass with the non-native data in the JASMIN corpus.

Acoustic preprocessing was done using a 32 ms Hamming window, with a 10 ms step size. Acoustic feature vectors consisted of 13 mel-based cepstral coefficients, including C0, plus their first and second order derivatives.

The accuracy of the Viterbi alignment was checked manually and the segmentation appeared to be of good quality and was then used to calculate the quantitative measures which are described in detail in the following section.

4.3. Objective temporal measures

Previous studies of temporal phenomena in native and non-native speech have identified a number of quantitative variables that appear to be related to perceived fluency [8; 9; 12; 11; 14; 15; 16]. The clearest taxonomy is provided by Grosjean [9: 40], who distinguishes between primary and secondary variables. Primary variables are ‘‘variables that are always present in language output’’ [9: 40]. Secondary variables are related to hesitation phenomena such as filled pauses, repetitions, repairs, and restarts. These variables are not necessarily present in speech and seem to be infrequent in read speech [9: 42].

Before introducing the variables used, we first give some definitions (for more details, see [2]):

- *silence*: every frame of silence detected by the Automatic Speech Recognizer (ASR)
- *silent pause*: a stretch of silence with a duration of no less than 0.2 s,
- *nph*: number of phonemes,
- *dur1*: duration of speech without utterance internal silences,
- *dur2*: duration of speech including utterance internal silences.
- *Broken words*: initial parts of words.

The following measures were calculated

1. Articulation rate: *nph*/*dur1*
2. Rate of speech: *nph*/*dur2*
3. Phonation/time ratio: 100% x *dur1*/*dur2*
4. Mean length of runs: Mean number of phonemes between silent pauses
5. Mean length of silent pauses
6. Mean length of all silent pauses
7. Duration of silent pauses per minute: Total duration of all silent pauses/(*dur2*/60)
8. Number of silent pauses per minute: Number of silent pauses/(*dur2*/60)
9. Number of filled pauses per minute: Number of filled pauses/(*dur2*/60)
10. Number of broken words per minute: Number of broken words/(*dur2*/60)

5. Results

Table 1 shows the values of the objective measures for the two types of speech, read and spontaneous. In the fourth column the significance of the differences is indicated (paired t test). For nine of the ten measures significant differences are observed between read and spontaneous speech, but they are not all in the same direction.

In general, the measures indicate that the speakers in this experiment are more fluent in read speech than in spontaneous speech, but there are exceptions. Specifically, Rate of speech, Phonation/time ratio, Mean length of silent pauses, Duration of silent pauses per minute, Number of filled pauses per minute and Duration of filled pauses per minute indicate that read speech is more fluent than spontaneous speech. On the other hand, articulation rate is higher in spontaneous speech than in read speech and there are more silent pauses and broken words in read speech than in spontaneous speech. Our previous study on the impact of objective measures on fluency ratings ([2]) showed that articulation rate is a good predictor of perceived fluency in read speech, but not in spontaneous speech.

Table 1. Objective measures for read and spontaneous speech.

	Read	Spontaneous	p value
Articulation rate	9.44	10.17	0.001
Rate of speech	5.60	4.90	0.001
Phonation/time ratio	59.18	48.32	0.000
Mean length of runs	14.54	14.52	0.351
Mean length of silent pauses	0.91	1.31	0.000
Duration of silent pauses per min.	22.10	26.28	0.001
Number of silent pauses per min.	24.78	20.87	0.000
Number of filled pauses per min.	1.97	9.34	0.000
Number of broken words per min.	2.28	1.39	0.001
Duration of filled pauses per min.	0.18	0.38	0.000

As to the number of silent pauses, this finding should probably be related to the one that indicates that filled pauses are much more frequent in spontaneous speech than in read speech. In other words, in spontaneous speech speakers tend to produce more filled pauses than silent pauses, maybe to signal that they are still speaking, somehow, and to keep their turn.

The finding that broken words are more frequent in read speech than in spontaneous speech is in line with previous results. In [2] we found that restarts, repetitions of initial parts of words, were more frequent in read speech than in spontaneous speech, probably because in read speech speakers are forced, as it were, to produce words they see on paper some of which they might not be familiar with. Articulating such words is evidently more difficult than producing words speakers planned in their minds, as is the case in spontaneous speech. It is therefore not surprising that speakers stumble more in pronouncing these words, than when they have to pronounce words which they have chosen themselves.

At this point it seems interesting to compare the differences between read and spontaneous speech observed in this study with those found in [2]. For the sake of comparison, we present the values of the measures investigated in both studies in Table 2.

To make the comparison as fair as possible we present the results for read and spontaneous speech for the various proficiency levels. This is particularly important because the non-native speakers in the current study are, in general, low proficient compared to those in [2].

Although the tendencies are largely the same in the two experiments, the differences between read and spontaneous speech are much larger in [2] than in the present study. There could be several explanations for this. First of all, recall that while in the present study we analyze speech data of the same speakers, the speech data in [2] stem from different speakers. Second, in [2] the read speech seem to be less disfluent than in the current study, which might be related to the generally lower proficiency level of the speakers involved in the present study. Furthermore, the protocol used to collect read speech in the two studies differed in various respects. In [2] a printed version of the utterances had previously been sent to the speakers who had the opportunity of reading them beforehand, rehearsing them and then read them aloud over the phone. In the current study, on the other hand, the participants performed a sort of cold reading because they had to read aloud a text from a computer screen without rehearsing or practicing in advance. In both studies the subjects had to read out phonetically rich sentences; in addition, in the present study the participants had to read also short stories properly chosen for their proficiency level.

Table 2. Values for the objective measures in the current study and in the previous study [2].

	Current study		Previous study				
	Read	Spontaneous	Read Beginners	Read Intermediate	Read Advanced	Spontaneous Beginners	Spontaneous Intermediate
Articulation rate	9.44	10.17	10.87	11.15	12.47	12.25	11.85
Rate of speech	5.60	4.90	8.54	8.95	11.03	5.99	5.31
Phonation/time ratio	59.18	48.32	77.97	79.62	88.28	49.33	44.92
Mean length of runs	14.54	14.52	16.51	18.10	27.73	9.50	9.33
Mean length of silent pauses	0.91	1.31	0.40	0.40	0.34	0.92	1.02
Duration of silent pauses per min.	22.10	26.28	9.29	8.67	3.97	27.90	31.02
Number of silent pauses per minute	24.78	20.87	22.33	20.11	10.18	31.00	31.41
Number of filled pauses per minute	1.97	9.34	0.31	0.35	0.32	10.83	10.55

The finding that filled pauses are particularly frequent in spontaneous speech and infrequent in read speech emerges from both studies and is in line with the definitions of primary and secondary variables formulated by Grosjean [9]. The frequency of filled pauses in spontaneous speech is similar in the two studies. As to silent pauses, these appear to be more frequent in the read speech in the present study than in [2], probably because the speakers in [2] had the possibility of reading the text beforehand.

6. Discussion

In the previous section we have presented the values of several objective, temporal measures of fluency that have been calculated for fragments of read and spontaneous speech produced by the same non-native speaker in a group of 44 learners of Dutch as L2. Not all these objective measures appear to vary to the same extent between read and spontaneous speech. The most striking difference concerns the number of filled pauses, whose frequency in spontaneous speech is at least five times higher than in read speech, thus suggesting that spontaneous speech is more challenging, which is in line with previous findings.

On the other hand, read speech may be challenging in a different way. In the present study subjects tended to produce more broken words in read speech than in spontaneous speech. This may be related to the use of cold reading in this study, i.e. they had to read the words in the prompts without being able to prepare them beforehand, while in spontaneous speech subjects can select their own words and prepare them in their minds. Cold reading is particularly challenging if the prompts contain words that are not (well-)known to the subjects, the more so for L2 learners.

Comparison of these results with those of previous studies in which similar objective measures of fluency had been calculated for read and spontaneous speech reveals many similarities, but also some differences. In particular, it appears that the values of the objective measures investigated can vary as a function of different variables such as the specific nature and details of the task the speakers have to carry out, the protocol and instructions used for data collection, and of course the proficiency level of the speakers.

7. Conclusions

Our results suggest that a number of factors such as task, speaker's proficiency level, and instructions received can influence measures of fluency, and therefore should accurately be varied or kept under control when collecting data for fluency investigations, and at least reported in publications, as this would make it easier to make comparisons between studies. For this purpose it would also be useful to choose appropriate and comparable measures in the different studies, as this would also facilitate meta-analysis and lead to a better understanding of L2 fluency and the factors that influence it.

8. Acknowledgements

This research was carried out within the DISCO project and made use of the results of the DISCO and JASMIN projects. Both projects are part of the STEVIN programme, which is funded through grants by the Dutch and Flemish Governments (http://taalunieversum.org/taal/technologie/stevin/).

9. References

[1] Cucchiarini, C., Strik, H. and Boves, L. "Quantitative assessment of second language learners' fluency". Journal of the Acoustical Society of America, 107 (2), 989-999, 2000a.

[2] Cucchiarini, C., Strik, H. and Boves, L. "Quantitative assessment of second language learners' fluency: Comparisons between read and spontaneous speech. Journal of the Acoustical Society of America, 111(6), 2862-2873, 2002.

[3] Cucchiarini, C., Strik, H. and Boves, L. "Different aspects of expert pronunciation quality ratings and their relation to scores produced by speech recognition algorithms". Speech Communication, 30 (2-3), 109-119, 2000b.

[4] Cucchiarini, C., Driesen, J., Van hamme, H. and Sanders, E., "Recording speech of children, non-natives and elderly people for HLT applications: the JASMIN-CGN corpus", Proc. LREC, 2008.

[5] Demuynck, K., Roelens, J., Compernolle, D. Van Wambacq, P. "SPRAAK: an open source Speech Recognition and Automatic Annotation Kit," In Proc. ICSLP, p. 495, 2008.

[6] Derwing, T. M., Rossiter, M. J., Munro, M. J. and Thomson, R. I. "Second language fluency: Judgments on different tasks", Language Learning, 54, 655-679, 2004.

[7] Derwing, T, Munro, M., Thomson, R., and Rossiter, M. "The relationship between L1 fluency and L2 fluency development". *Studies in Second Language Acquisition*, 533-557, 2009.

[8] Goldman-Eisler, F. "Psycholinguistics: Experiments in Spontaneous Speech", Academic Press, New York, 1968.

[9] Grosjean, F.. "Temporal variables within and between languages,'' in Towards a Cross-Linguistic Assessment of Speech Production, edited by H. Dechert and M. Raupach, Lang, Frankfurt, 39–53, 1980.

[10] Kormos, J. and D´enes, M., "Exploring Measures and Perceptions of Fluency in the Speech of Second Language Learners", System: An International Journal of Educational Technology and Applied Linguistics, 32(2):145-164, 2004.

[11] Lennon, P. "Investigating fluency in EFL: A quantitative approach", Language Learning 3, 387–417, 1990.

[12] Nation, P. "Improving speaking fluency", System, 377–384, 1989.

[13] Oostdijk, N. "The design of the spoken Dutch corpus," in New Frontiers of Corpus Research, P. Peters, P. Collins, and A. Smith, Eds. Rodopi, pp. 105-112, 2002.

[14] Riggenbach, H. "Toward an understanding of fluency: A microanalysis of non-native speaker conversations", Discourse Process. 14, 1991.

[15] Schmidt, R. "Psychological mechanisms underlying second language fluency", Studies in Second Language Acquisition 14, 357–385, 1992.

[16] Towell, R., Hawkins, R., and Bazergui, N. "The development of fluency in advanced learners of French", Applied Linguistics 1, 84–119, 1996.

[17] Trofimovich, P., and Baker, W. ."Learning second language suprasegmentals: Effect of L2 experience on prosody and fluency characteristics of L2 speech". Studies in Second Language Acquisition, 28, 1-30, 2006.

[18] Zechner, K., Higgins, D., Xi, X. and Williamson, D. "Automatic scoring of non-native spontaneous speech in tests of spoken English", Speech Communication, 883-895, 2009.