Behavior Research Methods
2009, 41 (2), 385-390
doi:10.3758/BRM.41.2.385

Praat script to detect syllable nuclei and measure speech rate automatically

Nivja H. de Jong
University of Amsterdam, Amsterdam, The Netherlands
and Utrecht University, Utrecht, The Netherlands

and

Ton Wempe
University of Amsterdam, Amsterdam, The Netherlands

In this article, we describe a method for automatically detecting syllable nuclei in order to measure speech rate without the need for a transcription. A script written in the software program Praat (Boersma & Weenink, 2007) detects syllables in running speech. Peaks in intensity (dB) that are preceded and followed by dips in intensity are considered to be potential syllable nuclei. The script subsequently discards peaks that are not voiced. Testing the resulting syllable counts of this script on two corpora of spoken Dutch, we obtained high correlations between speech rate calculated from human syllable counts and speech rate calculated from automatically determined syllable counts. We conclude that a syllable count measured in this automatic fashion suffices to reliably assess and compare speech rates between participants and tasks.

Speech rate is used as a measure of fluency in second language acquisition and bilingualism research (e.g., O'Brien, Segalowitz, Freed, & Collentine, 2007; Riggenbach, 1991; Towell, Hawkins, & Bazergui, 1996); in research diagnosing speech and language disorders (e.g., Feyereisen, Pillon, & de Partz, 1991; Redmond, 2004; Shenker, 2006); and in research on speech of pathological populations (e.g., Cannizzaro, Harel, Reilly, Chappell, & Snyder, 2004; Covington et al., 2005). However, calculating speech rate by hand is a tedious task, which is therefore often not carried out. In this article, we present a script written in the software program Praat (Boersma & Weenink, 2007) that automatically detects syllable nuclei in order to calculate speech rate. The nucleus of a syllable, also called the peak, is the central part of a syllable (most commonly, the vowel in the syllable). Locating these syllable nuclei allows for a computation of number of syllables, which can be used to calculate speech rate.

According to Tavakoli and Skehan (2005), fluency is multifaceted in nature. They distinguish three different facets of fluency: breakdown fluency (number and length of pauses), speed and density per time unit (speech rate), and repair fluency (false starts and repetitions). In second language testing practice, fluency is usually a score awarded by human judges who use several aspects of fluency in their judgment. However, Cucchiarini, Strik, and Boves (2002) have shown that of several objectively measured aspects of fluency, speech rate (as measured by phonemes per time unit) is the best predictor of subjective fluency. Kormos and Dénes (2004) likewise have shown that speech rate (in terms of number of syllables per time unit) is a good predictor of subjective fluency. We conclude that, for researchers wanting to include a measure of fluency, speech rate is an important factor to take into account. However, because of time constraints, this measure is often impossible to carry out. For instance, the script presented in this article was written in order to be able to measure the speech rate of 250 participants in a corpus of over 45 h of speech, a task that would take at least 8 months of full-time work for one person to measure by hand. In the context of a large-scale research project on the correlates of speaking proficiency carried out at the University of Amsterdam (What Is Speaking Proficiency: www.hum.uva.nl/wisp), we developed two tools to measure fluency automatically. For the purpose of measuring pauses in running speech, we wrote a script in the software program Praat to automatically detect silence in speech [a simplified version of which is now incorporated in the button *To TextGrid* (*silences*) in the Praat software]. For the purpose of estimating the speech rate of speech performances, we wrote a script in Praat that automatically detects syllable nuclei to compute speech rate in terms of syllables per time unit. In this article, we will present and validate the script for detecting syllable nuclei.

For automatic speech recognition, speech rate is an important factor as well. Human listeners are able to understand both fast and slow speech. Speech recognizers implemented in computers, however, perform relatively poorly when

N. H. de Jong, nivja.dejong@let.uu.nl

speech rate is very fast or very slow. In order to improve computer performance, several researchers have proposed that measuring speech rate prior to speech recognition will result in higher success rates of automatic speech recognizers (Pfau, Faltlhauser, & Ruske, 2000), and several ways to automatically measure speech rate in terms of phones and/or syllables per time unit have been put forward.

Mermelstein (1975) developed an algorithm with which to segment speech into syllables by finding minima in loudness that serve as possible syllable boundaries. Verhasselt and Martens (1996) presented an automatic speech detector that measures phone boundaries and thus calculates rate of speech as phone rate. The phone boundaries are provided by a multilayer perceptron that is trained on a subset of the data that must be hand-segmented at the phone level. Pfitzinger (1999) used a combination of syllable rate and phone rate for correlations with perceptual speech rate. Syllable rate was calculated by counting peaks in the energy contour, whereas phone rate was calculated by use of transcription. The syllable, phone, and perceptual speech rates were measured over (very) short stimuli (625 msec). Hunt (1993) used recurrent neural networks to detect syllables. Pfau and Ruske (1998) determined syllable nuclei by detecting vowels on smoothed modified loudness, and then calculated speech rate. Pellegrino and Andre-Obrecht (2000) used a vowel-detection algorithm based on the spectral analysis to detect formant–antiformant structure that is specific for vowels (and some nasals). In their study, the vowel-detection algorithm was used for automatic language identification. In the study by Pellegrino, Farinas, and Rouas (2004), the same vowel-detection algorithm was used to estimate speech rate. They compared hand-measured and automatically measured speech rates, calculated for speech performances of around 45 sec with speakers of several languages. Finally, Wang and Narayanan (2007) also developed a method for speech-rate estimation; they made use of spectral subband correlation including temporal correlation, prominent spectral subbands, and pitch information to estimate syllable nuclei.

All of these different automatic ways to measure syllable and/or phone rate are quite successful. It is difficult, however, to compare the success of these automatic measurers, because they were all used on different corpora, and their success was reported in different ways. Some researchers have reported correlations between human and automatic speech rate; others, a percentage of syllables (or phones) undetected and falsely detected as compared with human-measured syllables (or phones); and still others, a correlation between the number of manually measured and automatically measured syllables (or phones). The difference in corpora used should also be noted, because some studies have used large corpora with many different speakers and others have used quite small corpora with few speakers; some have used corpora of speech read aloud, whereas others have used (semi-) spontaneous speech corpora. Finally, a noteworthy difference between studies concerns the length of the spurt on which speech rate was calculated. Some studies have used extremely short time windows to calculate speech rate, and others have used much longer windows. Perhaps the most obvious reason why we cannot compare success of these different automatic speech-rate measurers is that the length of the time window (or spurt) as well as the variance in spurt length will strongly influence calculations of speech rate.

Many of the proposed speech-rate measurers need to be trained on a subset of the data that is transcribed or preprocessed by hand (Hunt, 1993; Pfau & Ruske, 1998; Pfitzinger, 1999; Verhasselt & Martens, 1996). In this article, we will present an easy way to automatically measure speech rate without the use of preprocessing or the need for transcriptions and test it on two different corpora of spontaneously spoken Dutch. To be able to compare the success of the script over these two different corpora, spurt length was controlled. We wrote a script in Praat using a combination of intensity (similar to Pfitzinger, 1999) and voicedness (similar to Pfau & Ruske, 1998) to find syllable nuclei.

The Praat Script

In what follows, we describe the sequence of actions that the script completes to find syllable nuclei using intensity (dB) and voicedness. We use intensity first to be able to find peaks in the energy contour, since a vowel within a syllable (the syllable nucleus) has higher energy than do surrounding sounds (described in Steps 1 and 2, below). We then use the intensity contour to make sure that the intensity between the current peak and the preceding peak is sufficiently low. With this procedure, we delete multiple peaks within one syllable (described in Step 3, below). Finally, we use voicedness to exclude peaks that are unvoiced, which is required to delete surrounding voiceless consonants that have high intensity (described in Step 4, below). Before the script is run, sound files that are quite noisy should be filtered so that the frequency range is speech-band limited.

Step 1. We extract the intensity, with the parameter "minimum pitch" set to 50 Hz, using autocorrelation. With this parameter setting, we extract intensity smoothed over a time window of 64 msec, with 16-msec time steps.

Step 2. We consider all peaks above a certain threshold in intensity to be potential syllables. We set the threshold to 0 or 2 dB above the median intensity measured over the total sound file (0 dB if the sound is not filtered; 2 dB if the sound is filtered). We use the median, rather than the mean, to calculate the threshold in order to avoid including extreme peaks in the calculation of the threshold.

Step 3. We inspect the preceding dip in intensity and consider only a peak with a preceding dip of at least 2 or 4 dB with respect to the current peak as a potential syllable (2 dB if the sound is not filtered; 4 dB if the sound is filtered).

Step 4. We extract the pitch contour, this time using a window size of 100 msec and 20-msec time steps, and exclude all peaks that are unvoiced.

Step 5. The remaining peaks are considered syllable nuclei and are saved in a TextGrid (point tier).

Figure 1 shows a part of a sound file together with the output as TextGrid made by the script. The speech utterance depicted here is *dat uh was wel goed bevallen toen* ("that uhm was quite well liked then"), totaling nine syllables, including "uh." The script is available at sites.google

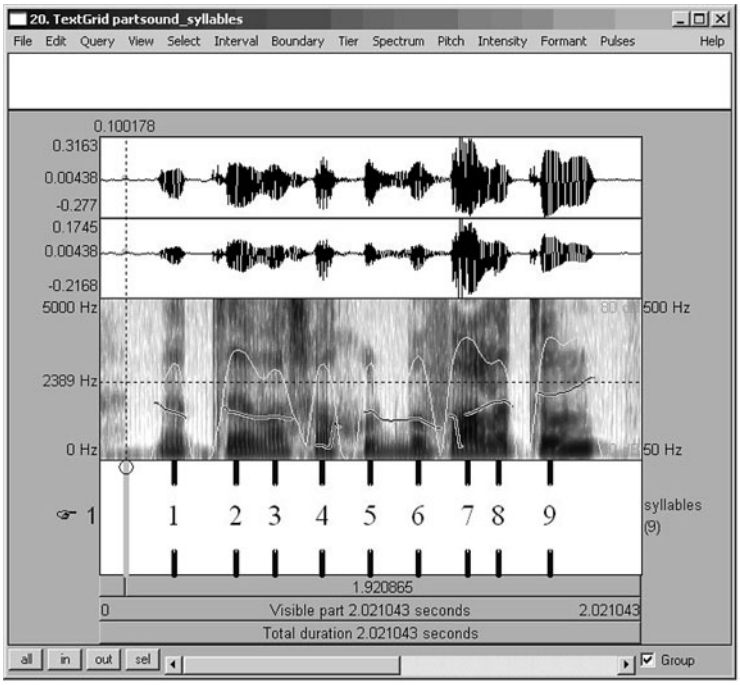

Figure 1. Part of a speech file in Praat with intensity (white line in spectrum) and pitch (dark line in spectrum) shown. The points in the tier are the syllable nuclei as detected by the script.

.com/site/speechrate under the terms of the GNU General Public License (de Jong & Wempe, 2008).

Performance

As a part of the project "What Is Speaking Proficiency" (WISP), conducted at the Amsterdam Center for Language and Communication at the University of Amsterdam, we have collected speech data from 258 participants—200 nonnative speakers of Dutch (with various first languages) and 58 native speakers of Dutch. Each participant performed 8 speaking tasks, resulting in a total of approximately 46 h of speech. See de Jong, Steinel, Florijn, Schoonen, and Hulstijn (2007) for a description of the speaking tasks and an application of the fluency measures. In order to be able to include measures of fluency in our research, we made two scripts written in Praat. The first script automatically detects pauses [a modified version is now incorporated in Praat in the *TextGrid* (*to silences*) button], and the second script automatically detects syllables. The second script is described in this article. In what follows, we report a validation of the computation of syllables per time unit as generated by the script in two different corpora. First, we randomly selected 50 out of the total of 258 * 8 speaking tasks and measured syllables by hand. This corpus comprised 75 min of speech. Second, we tested the script on a subset of the IFA corpus that was comparable to the speaking tasks in the WISP study (van Son, Binnenpoorte, van den Heuvel, & Pols, 2001). This part of the corpus comprised 125 min of speech summed over 8 participants.

Speech Data for the WISP Study

We counted the syllables of 50 speech files. Pauses longer than 0.4 sec were considered possible spurt boundaries. We used all spurts of 5 sec or more to calculate speech rate, and combined consecutive shorter spurts to get to 5 sec (excluding pauses). We thus avoided calculating speech rate over very short periods of time. We then automatically detected syllables using the Praat script. Many sound files in this corpus were moderately noisy; therefore, we filtered all sounds prior to the syllable measuring, using 100 Hz as the lower edge of the pass band, 5000 Hz as the upper edge of the pass band, and 50 Hz as the width of the smoothing region. With these settings, we attenuate nonspeech frequency components and keep all possible voice-related information about intensity and voicedness, across all formants. Measuring peaks in intensity (dB), we used 2 dB above the median intensity per sound file as threshold, and 4 dB as minimum dip between peaks, excluding peaks that are unvoiced.

We computed speech rate by hand and from the script per spurt by dividing the number of syllables per spurt for both measures by the spurt length in seconds. Figure 2 shows the scatterplot of the human and automatic speech-rate calculations per spurt; the correlation was .71. For our purposes, of comparing speakers and/or tasks, however, we needed a less refined calculation of speech rate. Figure 3 shows the scatterplot when we calculated speech rate over the total speech file: total number of syllables per speech file divided by total speaking time (correlation .88). In other words, the automatically measured speech

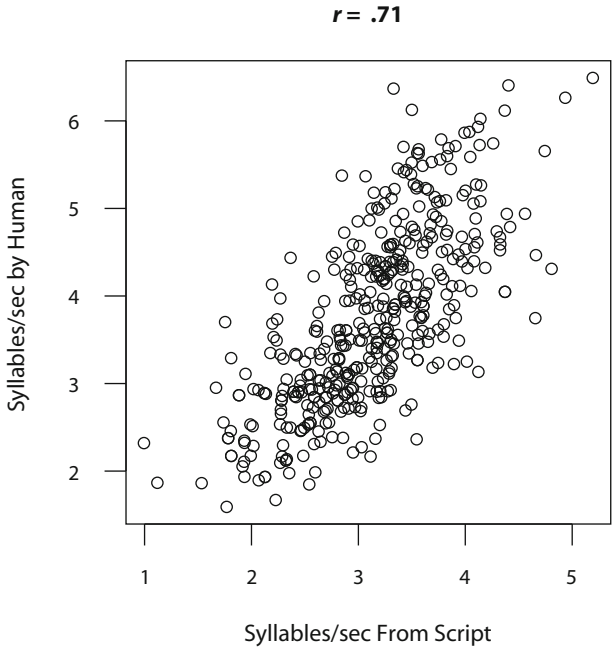

Figure 2. Scatterplot of WISP speech data of 50 participants; 441 spurts. The number of syllables per second is calculated per spurt by hand and automatically.

rate correlates well with the human-measured speech rate. However, with these data and these parameters, it seems to be the case that the script tended to miss syllables that were actually present. Upon inspection of the TextGrids made by the script, we concluded that the script missed mostly unstressed syllables that were detected by hand.

Speech Data for the IFA Corpus

The IFA corpus is an open source database of hand-segmented Dutch speech. Eight participants (4 female, 4 male) performed several speech tasks, ranging from reading aloud lists of syllables to informal storytelling. To validate the script on another corpus of Dutch, we selected the three tasks that were similar to the tasks used in the WISP study, eliciting (semi-) spontaneous speech. The three tasks were informal storytelling, face to face, to an "interviewer," retelling of the story previously told, and retelling of a story previously read (van Son et al., 2001).

For this corpus, we decided not to use a filter because the speech data of this corpus were not as noisy as the above-described speech data. Furthermore, filtering long sounds takes a lot of time and uses up a lot of computer memory (too much for the computer this script was run on at the time). As a result, we decided to lower the threshold and minimum preceding dip in intensity. We used the median intensity per sound file as threshold and 2 dB as the minimum preceding dip in intensity. In this corpus, sentences are defined on the basis of pauses as well as syntax, and the number of hand-measured syllables could therefore be counted per sentence. Because sentences were also defined on syntax, many sentences were very short. Such sentences comprised a single word, such as *uh* or *en* ("uh" or "and"), mostly as beginnings of unfinished sentences.

To test the automatic measures against these existing human-made measures, and to be able to compare success of speech-rate measures across the two corpora, we redefined spurts in this corpus as stretches of speech (including pauses) of at least 5 sec (except when the end of the speech file was reached, in which case the remaining shorter spurt was selected). We then counted the number of syllables using the human transcripts and counted the number of syllables measured automatically for the same time period. For this corpus, we had 8 participants for whom human-measured information was available in speech tasks quite comparable to those in the WISP study. In Figure 4, we show, for all 8 participants, the scatterplot of human-measured speech rate per spurt with automatically measured speech rate for that same spurt.

Again, for the purpose of comparing tasks and speakers, we need a calculation of speech rate computed per task. Figure 5 shows the correlation of the 8 speakers in three different tasks ($r = .8$). As with the speech data for the WISP study, the script missed syllables that were detected by hand. An inspection of the TextGrids produced by the script revealed that most of the undetected syllables were unstressed syllables. We think that many of these unstressed syllables might be phonological syllables and therefore can be detected when measured by hand, but probably not all are also phonetic syllables in the sense that they are present in the signal in any detectable way. For example, Ernestus (2000, pp. 129–132) explains that unstressed vowels and even adjacent consonants may be absent—in particular, for casual Dutch. Therefore, we may conclude that the algorithm picks up on prominent syllables. As shown by the correlations between human

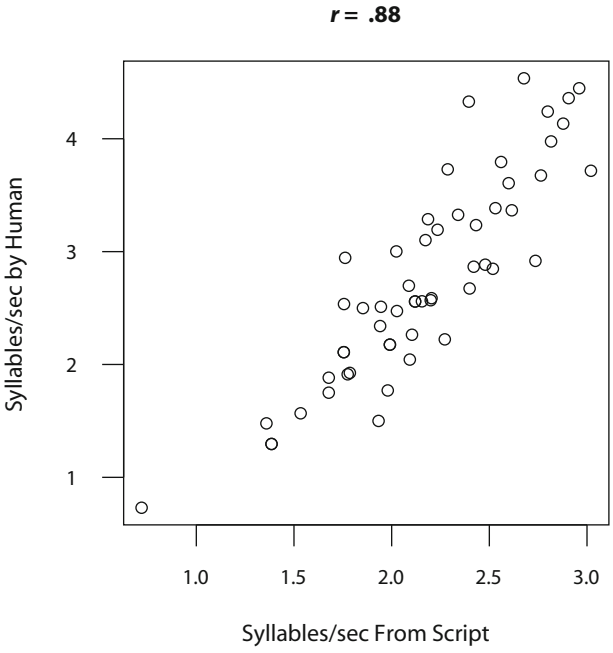

Figure 3. Scatterplot of WISP speech data of 50 participants. The number of syllables per second is calculated per task (participant) by hand and automatically.

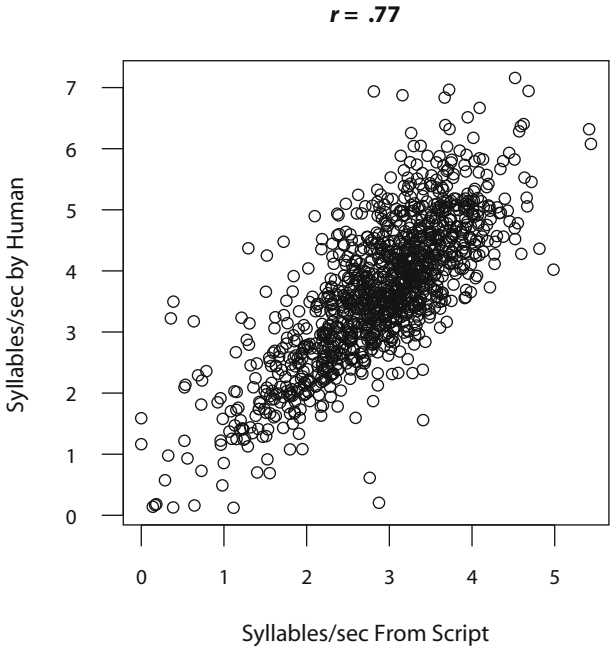

Figure 4. Scatterplot of the IFA corpus; 8 participants, 1,171 spurts. The speech rate and number of syllables per second are counted per spurt by hand and automatically.

measures and automatic measures, missing such unprominent syllables is not problematic for researchers who want to compare speech rate between speakers and tasks, since the underestimation of speech rate is consistent. The script will miss syllable nuclei that are very soft in comparison with most syllables in the speech signal (since they do not exceed the set threshold of intensity) and syllables that are not present in the speech signal in any detectable way, as explained above. To determine how the speech rate as detected by hand relates to the speech rate as computed from the script, we fitted linear regression models. When we fix the intercept of these models to zero, we can compare the slopes of the two regression models of the two corpora (using the speech rates calculated over tasks).[1] For the speech rates in the WISP corpus, we found a slope of 1.27, and in the IFA corpus, the slope in the model was 1.29. In other words, to predict hand-measured speech rate using the speech rate as derived from the script, you should multiply by approximately 1.28.

Research by Kormos and Dénes (2004) suggests that in fact it is the number of stressed syllables that correlates best with subjective fluency. Perhaps it is the case that number of prominent syllables better reflects speech rate in the sense that it measures density of content per time unit. Future research is needed to further explore this thesis.

Discussion

In this article, we have described a script written in Praat that automatically detects syllables in sound files of speech. No transcription of the speech data is necessary to run this script. The script takes sound files as input and writes a TextGrid file with syllable nuclei marked in a point tier. In two validation studies, we found high correlations between human-measured speech rate and automatically measured speech rate. Although the script misses (mostly unstressed) syllables that are detected by human judges, the correlations suggest that the algorithm works well in predicting the actual number of syllables. We conclude that for the purpose of measuring speech rate as number of syllables per time unit when comparing speakers and tasks, this script suffices.

In second language testing (see, e.g., the speaking rubrics of the TOEFL test as reported on the ETS Web site [Educational Testing Service, 2004]) and second language research (e.g., Kormos & Dénes, 2004), as well as in the diagnosis of different language and speech disorders (Feyereisen et al., 1991; Redmond, 2004; Shenker, 2006), fluency is an important factor to take into account. The script described and validated in this article may be useful to easily and objectively measure speech rate in terms of syllables per second without the need to transcribe speech beforehand. For research investigating differences in speech rate for pathological populations, this script may also be useful. For instance, Cannizzaro et al. (2004) investigated the relation between ratings of major depression and speech rate (measured by hand) and found strong correlations. Depressed and schizophrenic patients may speak monotonously (see, e.g., Covington et al., 2005, for an overview). Even though the script uses pitch (along with intensity) to detect syllable nuclei, monotonous speech will not pose a problem since the pitch contour is used only to ascertain voicedness of syllable nuclei, and not to measure any changes in pitch. In any case, the recorded speech should not be too soft, since for soft speech the signal/noise ratio is too low for syllable nuclei to be reliably detected as a function of intensity (dB).

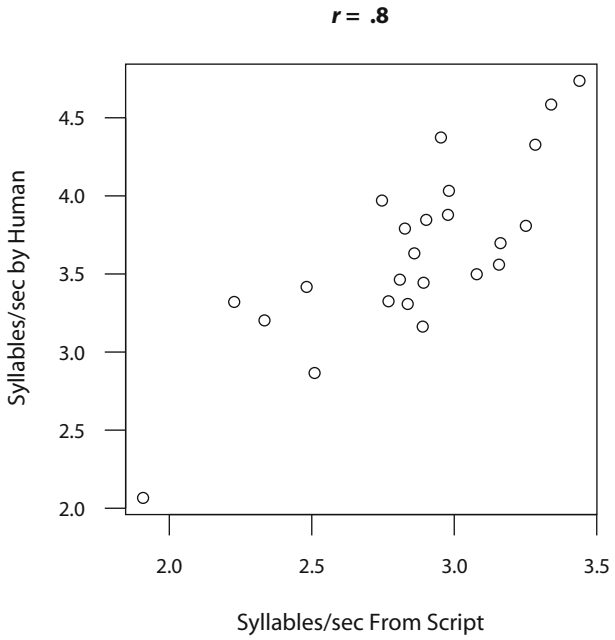

Figure 5. Scatterplot of the IFA corpus; 8 participants in three tasks. The speech rate and number of syllables per second are calculated per task by hand and automatically; $N = 24$.

As yet, it is impossible to directly compare the amount of success of the different syllable measurers described in the introduction. First of all, other syllable measurers have been developed to detect syllables in spoken English or German, which might be different from detecting syllables in Dutch. For instance, Pellegrino et al. (2004) compared success of speech-rate estimation for several languages and found that the correlations between hand-measured speech rate and automatically measured speech rate (calculated over approximately 45 sec) differed between languages. For German speakers, the Pearson correlation was $r = .73$, but for English it was $r = .82$. Furthermore, the different corpora on which the existing syllable measurers have been tested have been transcribed by different criteria. Finally, most researchers report Pearson correlations for speech rate or for number of syllables per spurt. However, comparisons are confounded if spurt length is uncontrolled. For longer spurts, a count of one or two extra or fewer syllables will not result in a large deviation of the calculated speech rate. For short spurts, a count of a single extra or fewer syllable will result in an enormous difference in the calculated speech rate. Future research should take these mathematical issues into account when comparing different methods that automatically measure speech rate. In the present article, we opted for choosing at least 5 sec as a constant spurt length. In this way, we were able to compare success in syllable detection across corpora.

AUTHOR NOTE

Research was funded by the Netherlands Organization of Scientific Research (NWO) under Grant 254-70-030, with the project title, "Unraveling Second Language Proficiency." Project leaders were Jan H. Hulstijn and Rob Schoonen. We thank Renske Berns for her help in counting syllables in speech. We also thank Jan Hulstijn, Rob Schoonen, and two anonymous reviewers for comments on earlier versions of this article. Correspondence relating to this article should be addressed to N. H. de Jong, Department of Dutch Language and Culture, Utrecht University, Trans 10, 3512 JK Utrecht, The Netherlands (e-mail: nivja.dejong@let.uu.nl).

REFERENCES

Boersma, P., & Weenink, D. (2007). Praat (Version 4.5.25) [Software]. Latest version available for download from www.praat.org.

Cannizzaro, M., Harel, B., Reilly, N., Chappell, P., & Snyder, P. J. (2004). Voice acoustical measurement of the severity of major depression. *Brain & Cognition*, **56**, 30-35. doi:10.1016/j.bandc.2004.05.003

Covington, M. A., He, C., Brown, C., Naçi, L., McClain, J. T., Fjordbak, B. S., et al. (2005). Schizophrenia and the structure of language: The linguist's view. *Schizophrenia Research*, **77**, 85-98. doi:10.1016/j.schres.2005.01.016

Cucchiarini, C., Strik, H., & Boves, L. (2002). Quantitative assessment of second language learners' fluency: Comparisons between read and spontaneous speech. *Journal of the Acoustical Society of America*, **111**, 2862-2873. doi:10.1121/1.1471894

de Jong, N. H., Steinel, M. P., Florijn, A. F., Schoonen, R., & Hulstijn, J. H. (2007). The effect of task complexity on fluency and functional adequacy of speaking performance. In S. Van Daele, A. Housen, M. Pierrard, F. Kuiken, & I. Vedder (Eds.), *Complexity, accuracy and fluency in second language use, learning and teaching* (pp. 53-63). Brussels: Koninklijke Vlaamse Academie van België voor Wetenschappen en Kunsten.

de Jong, N. H., & Wempe, T. (2008). *Praat script speech rate*. Retrieved October 14, 2008, from sites.google.com/site/speechrate/.

Educational Testing Service (2004). *iBT/Next Generation TOEFL Test: Independent Speaking Rubrics*. Retrieved December 10, 2007, from www.ets.org/Media/Tests/TOEFL/pdf/Speaking_Rubrics.pdf.

Ernestus, M. T. C. (2000). *Voice assimilation and segment reduction in casual Dutch: A corpus-based study of the phonology–phonetics interface*. Ph.D. dissertation, Vrije Universiteit, Amsterdam (LOT Series 36).

Feyereisen, P., Pillon, A., & de Partz, M.-P. (1991). On the measures of fluency in the assessment of spontaneous speech production by aphasic subjects. *Aphasiology*, **5**, 1-21. doi:10.1080/02687039108248516

Hunt, A. (1993). Recurrent neural networks for syllabification. *Speech Communication*, **13**, 323-332. doi:10.1016/0167-6393(93)90031-F

Kormos, J., & Dénes, M. (2004). Exploring measures and perceptions of fluency in the speech of second language learners. *System*, **32**, 145-164. doi:10.1016/j.system.2004.01.001

Mermelstein, P. (1975). Automatic segmentation of speech into syllabic units. *Journal of the Acoustical Society of America*, **58**, 880-883. doi:10.1121/1.380738

O'Brien, I., Segalowitz, N., Freed, B., & Collentine, J. (2007). Phonological memory predicts second language oral fluency gains in adults. *Studies in Second Language Acquisition*, **29**, 557-582. doi:10.1017/S027226310707043X

Pellegrino, F., & Andre-Obrecht, R. (2000). Automatic language identification: An alternative approach to phonetic modelling. *Signal Processing*, **80**, 1231-1244. doi:10.1016/S0165-1684(00)00032-3

Pellegrino, F., Farinas, J., & Rouas, J.-L. (2004). Automatic estimation of speaking rate in multilingual spontaneous speech. *Proceedings of Speech Prosody 2004*, *Nara, Japan* (pp. 517-520).

Pfau, T., Faltlhauser, R., & Ruske, G. (2000). A combination of speaker normalization and speech rate normalization for automatic speech recognition. *Proceedings of ICSLP 2000, Peking, China,* **4**, 362-365.

Pfau, T., & Ruske, G. (1998). Estimating the speaking rate by vowel detection. *Acoustics, Speech, and Signal Processing* (*ICASSP 2005 Proceedings)*, **2**, 945-948. doi:10.1109/ICASSP.1998.675422

Pfitzinger, H. R. (1999). Local speech rate perception in German speech. *Proceedings of the XIVth International Congress of Phonetic Sciences*, **2**, 893-896.

Redmond, S. (2004). Conversational profiles of children with ADHD, SLI and typical development. *Clinical Linguistics & Phonetics*, **18**, 107-125. doi:10.1080/02699200310001611612

Riggenbach, H. (1991). Toward an understanding of fluency: A microanalysis of nonnative speaker conversations. *Discourse Processes*, **14**, 423-441.

Shenker, R. C. (2006). Connecting stuttering management and measurement: I. Core speech measures of clinical process and outcome. *International Journal of Language & Communication Disorders*, **41**, 355-364. doi:10.1080/13682820600623861

Tavakoli, P., & Skehan, P. (2005). Strategic planning, task structure, and performance testing. In R. Ellis (Ed.), *Planning and task performance in a second language* (pp. 239-276). Amsterdam: John Benjamins.

Towell, R., Hawkins, R., & Bazergui, N. (1996). The development of fluency in advanced learners of French. *Applied Linguistics*, **17**, 84-119. doi:10.1093/applin/17.1.84

van Son, R. J. J. H., Binnenpoorte, D., van den Heuvel, H., & Pols, L. C. W. (2001). The IFA corpus: A phonemically segmented Dutch "open source" speech database. *EUROSPEECH 2001*, 2051-2054.

Verhasselt, J. P., & Martens, J. P. (1996). A fast and reliable rate of speech detector. *Spoken Language (ICSLP 96 Proceedings)*, **4**, 2258-2261. doi:10.1109/ICSLP.1996.607256

Wang, D., & Narayanan, S. (2007). Robust speech rate estimation for spontaneous speech. *IEEE Transactions on Speech, Audio and Language Processing*, **15**, 2190-2201. doi:10.1109/TASL.2007.905178

NOTE

1. For the IFA corpus, this simplification does not lead to loss of fit (R^2 for both models is .645). For the WISP corpus, R^2 changes from .77 (intercept is −.9) to .73 (intercept fixed to 0). However, for the present goal of comparing slopes for the two models, we can assume the intercept to be 0. Theoretically, this simplification is warranted because, if nothing is said, both automatically calculated speech rate and speech rate measured by hand should be 0.

(Manuscript received October 14, 2008;
revision accepted for publication January 6, 2009.)