The concurrent validity of several speech rate determination methodologies

Klaas Bakker
Maureen Darby
Missouri State University
Springfield, MO

Abstract: Speech rates of 12 monologues (syllables per minute [SPM], # stressed vowels per minute [SVPM], and perceptual ratings) were analyzed. SPM and SVPM inter-correlated strongly but were weakly related to the perceptual ratings. SVPM shows promise for automated speech rate determination applications. Perceptual ratings contribute unique clinical information otherwise not available.

INTRODUCTION

Speech rate measurement is an intricate component in many clinical procedures regarding fluency disorders; speech rate may be (1) a problem by itself (e.g., tachylalia), (2) a symptom of another clinical manifestation (e.g., cluttering), or (3) used as a target in treatment or research. Speech rate related procedures call for practical procedures with demonstrated levels of validity and reliability. This study emphasized aspects of concurrent validity comparing 3 different speech rate assessment options in preparation for the development of an automated speech rate assessment procedure.

Bakker, Brutten & McQuain, (1995), assessed the concurrent and construct validity of several automated syllable counting procedures with promise for automated speech rate applications. Their automatic stressed syllable counter most closely mirrored the conventional method of determining # of syllables by hand (Bakker, et al, 1995). Two studies addressed automated procedures for talking time (articulation time) estimation (Bakker, 2010; Bakker, et al, 2007). A strong correlation was established between a workstation based manual procedure and a computer-automated procedure, thereby
indicating that the faster, easier computer-automated procedure, is a desirable alternative for determining talking time in clinical applications. A very recent approach based on a Script written for the PRAAT software (De Jong and Wempe, 2009) produced similar results but did not point to an obvious real time application unless the analysis can somehow be developed independently from PRAAT.

This study forms the last step in evaluating the potential for automated speech rate determination based on automated stressed vowel detection and simultaneous determination of talking time. It compared several available speech rate measurement options head to head, including determination of (1) the total number of syllables per minute (SPM), the most traditional and frequently followed approach, (2) the number of stressed vowels per minute (SVPM), and (3) clinician speech rate ratings. Furthermore, it distinguished use of automated “talking time” versus “total sample duration” to determine if elimination of pauses would improve the quality of the measures.

METHOD

Twelve normal adult speakers, 6 men and 6 women, ranging in age from 22 to 52 years of age, native English speakers, produced two-minute monologues. Following listening to the recordings, perceptual judgments about speech rate were first determined by one of the experimenters. Then, syllables as well as stressed vowels were counted from orthographic transcription using the stress definitions available from a dictionary, and subsequently verified through listening to the actual productions.

Speech analysis software (PRAAT, Note 1), was used to determine total sample duration (time lapse between the first and last acoustic evidence of the speech sample) as well as determination of
*total talking time* (by eliminating pauses greater than 200ms and non-speech related sounds for example because of breathing). Selections were made from the sound intensity display.

**RESULTS**

Table 1 displays the inter-correlations between the measures.

Table 1. Inter-correlations between the different speech rate measures (based on syllables or stressed vowels per duration as measured in sample duration or talking time) on the one hand, and the correlations of each of these measures with a perceptual speech rate rating.

<table>
<thead>
<tr>
<th>Measure</th>
<th>Correlation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Syllables per minute (SPM) duration and SPM talking time</td>
<td>0.85*</td>
</tr>
<tr>
<td>Stressed vowels per minute (VPM) duration and VPM talking time</td>
<td>0.92*</td>
</tr>
<tr>
<td>SPM duration and VPM duration</td>
<td>0.94*</td>
</tr>
<tr>
<td>SPM talking time and VPM talking time</td>
<td>0.90*</td>
</tr>
<tr>
<td>Perceptual ratings and SPM duration</td>
<td>0.69</td>
</tr>
<tr>
<td>Perceptual ratings and SPM talking time</td>
<td>0.51</td>
</tr>
<tr>
<td>Perceptual ratings and VPM duration</td>
<td>0.61</td>
</tr>
<tr>
<td>Perceptual ratings and VPM talking time</td>
<td>0.48</td>
</tr>
</tbody>
</table>

*). The critical value for a two tailed significance level of $r_{xy}$ (df=10; $p<=.01$) = .708.

The correlation between SPM (sample duration) and SPM (talking time) was strong and statistically significant ($r_{xy} = 0.85$, df=10, $p<.01$). While this suggests that these alternatives are strongly related, it also shows that there is some difference. This urges empirical evaluation through research to determine if this difference matters for clinical usage of speech rate measures. As syllables per minute talking time appears to measure the amount of speech produced per unit time, one may wonder if in situations where highly accurate speech rate estimates are needed this measure would be preferable.
Stressed vowels per minute (SVPM) for sample duration and talking time (Table 2) also revealed a strong correlation, 0.92. Similar strong correlations were obtained when comparing SPM (sample duration) to SVPM (sample duration) and when comparing SPM (talking time) to SVPM (talking time). Overall, the quantified “physical” speech rate measures were strongly interrelated with each other, suggesting that in large degree they reflect the same characteristic.

The perception of speech rate was not correlated strongly to any of the measured speech rates as can be seen in Table 2. This despite a sizable degree of internal reliability ($r_{xy} = 0.90$, df=10, p<=.01)

**DISCUSSION**

Syllables per minute total sample duration is the most frequently chosen method for speech rate measurement. Therefore, in the present study, it was used as a standard for evaluating a few new speech rate assessment options (that is a procedure that conducts stressed vowel counts, and a subjective severity rating). This study revealed that determining number of stressed vowels per minute correlates strongly with the traditional SPM measure. Also there was not much difference using total sample duration or articulation time. As stressed vowels have the potential to be easily and reliably detected through automated procedures, an automated procedure for the estimation of speech rate based on detecting stressed vowel productions therefore is a feasible option. Use of this automated alternative for estimating speech rate would reduce the burden of many clinicians who work with speech rate or related clients and among others would provide a mechanism for online feedback about speech rate in clinical applications. Since these data were determined from normal speakers it is important that a similar study evaluates if number of stressed vowels per minute is also valid when determined from speech that is not normally fluent. Particularly, it will be interesting to determine if
speech rate can be estimated from speech that contains stuttering (or reflects the use of prolonged speech for therapy) or cluttering.

Surprisingly, perceptual decisions about speech rate were not strongly correlated with any of the measured speech rates. Future research will hopefully address the relationship between speech rate perception and measures based on the physical manifestation of speech rate. The lack of relationship in the present study may point to a weakness of the particular perceptual approach that was followed, but it could also be interpreted as a sign that speech rate perception is a complex variable and not related to physical speech rate in a simple or straightforward way. In other words, neither of the measures may fully cover the rate at which speech is produced, and both measures may need to be used in situations where speech rate plays a crucial role.

To the extent that number of stressed vowels per unit time contributes to this important assessment, clinical technologies that allow the measure to be determined automatically are anticipated to greatly facilitate assessments regarding the speech production rate.

Note 1 – Praat is a free software program available on the Internet. Paul Boersma and David Weenink developed it for speech analysis and synthesis. http://www.praat.org/

REFERENCES

