Measuring Rhythmic Deviation in Second Language Speech

Felix Schaeffler
IPSK - Department of Phonetics and Speech Communication
University of Munich, Germany
felix@phonetik.uni-muenchen.de

Abstract
This study deals with the question of whether recently provided methods to determine the rhythm class of languages can be transferred to foreign-accented speech. Therefore read German speech of Venezuelan Spanish native speakers was compared with read speech of a native German control group by means of four different measurements. Three of the four applied measurements showed significant differences between the two groups, with one of the differences contradicting earlier expectations. The study has shown that the measurements can be successfully transferred to foreign-accented speech, but slightly modified measurements are suggested.

1. Introduction
Speech Rhythm is still a controversial topic within the speech sciences. This is partly due to the fact that it has never been possible to provide clear empirical evidence for the 'stress-timed' vs. 'syllable-timed' distinction on the basis of the classic isochrony hypothesis, although the impression of rhythmic differences between 'stress-timed', 'syllable-timed' and 'mora-timed' languages is quite striking. Consequently, the isochrony hypothesis has repeatedly been attacked. Newer accounts (cf. e.g. [1], [2], [3]) have made the claim that the impression of syllable-timed, stress-timed or mora-timed rhythm in speech is rather the outcome of the phonological properties of the respective language than a phonological primitive. On the basis of these assumptions, there have been some methods recently introduced which seem to provide durational correlates for rhythmic impression.

The impression of a 'wrong' or deviant speech rhythm is often very strong with foreign-accented speech. This paper thus addresses the question of whether this deviation can be measured by the methods proposed. For this purpose, read German speech of Venezuelan Spanish students of German has been compared with read speech of German native speakers.

2. Measuring Rhythm
The methods applied in the present study have been presented by Grabe et al. [4], (cf. also [5]) and Ramus et al. [3]). Grabe et al. proposed a 'Pairwise Variability Index' (PVI) based on durational differences of vowels in sequence within an intonational phrase. They were able to show that the PVI differentiates between the more 'syllable-timed' Singapore English and the more 'stress-timed' Standard British English, furthermore they demonstrated that the index is sensitive to differences in rhythm of child speech and adult speech in both French and English. PVI values were always lower for syllable-timed languages than for 'stress-timed' languages. Ramus et al. calculated the percentual proportion of vocalic intervals within a sentence (%V), the standard deviation of the duration of vocalic parts (ΔV) and the standard deviation of consonantal parts (ΔC). At least two of these indices (ΔC and %V) were more difficult to interpret, this value also showed a tendency to be higher for classical 'stress-timed' languages.

3. Rhythm in Second Language Acquisition
There is evidence that rhythmic patterns of a native language interfere with those of a second language (cf. e.g. [6]) and that this interference could lead to reduced intelligibility and the impression of 'foreign accent'. Tajima et al. [7], for example, corrected the timing of foreign-accented English by manipulation of duration and intensity and found a strong increase in intelligibility. Missaglia [8] showed that rhythmic training enhanced pronunciation of a foreign language. A deeper understanding of the role of rhythm in second language acquisition could provide valuable insights for a rhythm typology theory and at the same time lead to alternative methods of pronunciation teaching.

4. Hypotheses
If speech rhythm is an outcome of syllabic structure, then difficulties with the syllabic structure of the target language should result in deviant speech rhythm. German syllable structure differs from Spanish syllable structure at least in the following aspects:

- contrastive vowel length
- vowel reduction
- more complex consonant clusters

If difficulties with these three features are assumed, we could expect that the four measurements (PVI, %V, ΔV and ΔC) show different values for the Venezuelan subjects and the German control group. It is tempting to presume that the values of the measurements should shift towards the direction of so-called 'syllable-timed' languages for the data from the Venezuelans. In the case of vocalic segments this could happen if the subjects weakened the durational difference between long and short vowels and did not reduce unstressed vowels appropriately. The mean values for PVI and the standard deviation of the duration of vocalic parts could therefore be lower for the IPs of the Venezuelan subjects.

Consonant clusters may be accordingly modified by substitution or elision of consonants or by the insertion of vowels (cf.
e.g. Magen [9] who reports similar effects for the English pronunciation of South-American Spanish natives). Most of these modifications should also lead to a reduction and assimilation of the duration of the contoid parts of the speech signal, thus reducing ΔC and increasing %V.

On the other hand, it is quite possible that cases of hyper-correction or prolonging of segments due to pronunciation difficulties shift some values in unexpected directions. Moreover, as to our knowledge the measurements have never been applied to foreign-accented speech before, and as the segmentation procedure has been slightly modified compared to previous examinations (see below), we decided, for statistical reasoning, to only assume that the values should be different for the Venezuelan group.

5. Method

5.1. Material, Subjects and Recordings

The text read in the task was the German version of ‘The Northwind and the Sun’. The Venezuelan group was recorded in a classroom of the ‘Asociation Cultural Humboldt’ in Caracas, Venezuela. Eight speakers participated in these recordings. From these eight speakers, the data of five speakers (3 female, 2 male) was analyzed in the present experiment as these speakers had studied German for approximately three years. Three of the speakers were born and grown up in Caracas, Venezuela, one speaker came from Valencia, Venezuela and one speaker came from Ciudad Bolivar, Venezuela. All speakers had studied in Caracas and also been living there during their studies.

The data of the German group was taken from the ‘Strange Corpus 1’ (SC1) of the ‘Bavarian Archive of Speech Signals’ (cf. [10]). This corpus includes 16 German versions of ‘The Northwind and the Sun’, read by German natives. All of it has been segmented by hand. From these 16 versions, five were chosen. These five matched the sex of the Venezuelan speakers and showed as few dialectal colouring as possible.

All recordings were made with DAT recorders in a noise-protected environment, although the recordings of the Venezuelan speakers did not reach studio quality.

5.2. Segmentation of sounds

It seems crucial to us to describe the segmentation techniques for the sound classes used in the present study in order to make comparisons with other studies possible. Therefore they will be presented here in some detail.

As the speakers reduced unstressed vowels to different degrees, it was necessary to find solutions for cases of elision. The word /mant@l/ (coat), for example, was often pronounced as [mantil]. Because of that it was decided not to distinguish between vocoids and contoids but between syllabic nucleus and syllabic edge (syllabic onset and coda). This means that if the schwa in /mant@l/ was realized, it was judged as the nucleus, when it was elided, the [l] got the label ‘nucleus’.

A second problem concerned the realization of /l/ in the coda. This sound is often realized in German as a reduced vowel, forming a falling diphthong with the preceding vowel (cf. [ve:G] ‘who’ or [de:G] ‘the’). A separation of these two sounds would almost always be quite arbitrary and it was decided to treat these two sounds as one segment. When the /l/ was realized as a trill, fricative, or approximant, it was still counted as belonging to the nucleus.

Stop consonants after pauses are another problem for segmentation, as their onsets normally can’t be specified. In case of their occurrence, these sounds have therefore been treated as part of the preceding pause.

The rest of the segmentation procedure followed the conventions provided by Geumann et al. [12], which should be quite similar to rules provided by e.g. Peterson et al. [13].

For the German data, the segmentation was taken from the SC1 corpus and modified with the help of the Speech Analysis Software ‘PRAAT’ (cf. [14]). The Venezuelan data was pre-segmented with the ‘Munich Automatic Segmentation System’ (MAUS, cf. [15]) and corrected by hand with ‘PRAAT’ as well.

5.3. Segmentation of intonational phrases

The reading of the same text does not guarantee that subjects split it in the same intonational phrases (IPs). Thus, the borders of the IPs had to be specified for each subject separately. In most cases this was rather straightforward, as there were often pauses in the speech signal at the end of IPs (this might be partly a side effect of the reading task). Every pause was taken as an IP border. In some cases an IP border was inserted, when pitch changes, anacrusis or final lengthening suggested it.

5.4. Calculation of the indices

The precise formula of the PVI is given by Grabe et al. [4] as

\[
PVI = 100 \times \frac{\sum_{k=1}^{m} \left( \frac{d_k - d_{k+1}}{d_k + d_{k+1}} \right)}{(m-1)}
\]

(m=number of vowels in utterance, d=duration of the kth vowel)

This value was calculated for every intonational phrase containing at least four nuclei. The last nucleus of each IP was excluded from the calculation, according to the procedure described in Grabe et al.

Where the calculation of the PVI was possible, the three values suggested by Ramus et al. [3] were calculated as well. They are:

- the percentual proportion of nucleus intervals within an IP (noted as %N).
- the standard deviation of the duration of nucleus intervals, multiplied by hundred (noted as ΔN).
- the standard deviation of the duration of syllabic edge intervals, multiplied by hundred (noted as ΔE).

6. Results

6.1. Results concerning the hypotheses

Table 1 shows the mean results for the PVI measurements and intonational phrase duration (IPD), table 2 shows the results for the other three measurements, averaged by native language of the speakers. Values in brackets give the respective standard deviations.

The differences in PVI, ΔN and %N turned out to be significant: Two-tailed t-tests yielded significance (p < .001) for these three differences. The difference in ΔE was not significant. The values of PVI and ΔN are lower for the Venezuelan group than for the German group, pointing towards a more syllable-timed rhythm. The situation is different for %N. The similar %V measurements in the study of Ramus et al. were higher for syllable-timed languages.

### 6.2. Further results

Since the duration of intonational phrases was not controlled in the task, it has to be taken into account that this could be a disturbing factor. Table 1 shows that the means of intonational phrase duration were almost equal for both groups, suggesting that this should not have influenced the results to a great extent.

The present study was primarily designed for comparisons of means over intonational phrases. The results for the individual speakers are nevertheless interesting, as they could give some information about the strength of the effects and the predictive power of the measurements. Table 3 shows the average PVI value for each speaker. Table 4 shows the average values for the three other measurements. As the tables show, only %N would divide German and Venezuelan subjects into the correct groups. This value is lower for all Venezuelan subjects.

Additionally, correlations between all four measurements and intonational phrase duration were calculated to get clues about possible interactions. Table 5 shows the Pearson correlation coefficient for all measurements, separated by language. For the sake of clarity, very weak correlation values between -0.2 and 0.2 were not included in the table. Significance has not been tested for these results, which is due to the explorative character of this part of the study.

Intonational phrase duration shows a positive correlation with ΔE for both groups and a negative correlation with %N and PVI, at least for the Venezuelan group (the values for the German group, which are not displayed in the table, were also negative but very small). %N and ΔE showed a relatively strong negative correlation for both groups, corresponding to results reported by Ramus et al. [3]. These authors observed a negative correlation for %N and ΔC across their different language samples. It is interesting to note that this relation also consists within the same language.

The relatively strong positive correlations for %N and ΔN are not very surprising (variation of nuclear intervals may increase if the relative proportion of nuclear intervals increases), but it should be noted that a much weaker relationship appeared for %N and PVI (correlation was positive but below 0.2 for both groups).

ΔN and PVI correlated positively for both groups, which is also not very surprising as these measurements are based on similar causes.

### 7. Discussion

If one accepts the notion that the obtained measurements measure ‘rhythm’, then the results for PVI and ΔN point in the expected direction: Both means were significantly lower for the

### 8. Conclusion

There’s no denying the fact that PVI, ΔN an %N are valuable indices for rhythm classes, even in the case of foreign accented speech. In order to support this assumption, one should examine how well these measurements agree with rhythmic impressions in the case of foreign accented speech. Perceptual tests with foreign accented speech could be an appropriate way to do this.

PVI and ΔN seemed to behave rather similar. Thus, it is not
Table 5: Pearson correlations for all measurements (except values between -0.2 and 0.2), separated by language

<table>
<thead>
<tr>
<th></th>
<th>IPD</th>
<th>%N</th>
<th>ΔN</th>
<th>ΔΕ</th>
<th>PVI</th>
</tr>
</thead>
<tbody>
<tr>
<td>IPD</td>
<td>German</td>
<td>X</td>
<td>-313</td>
<td>.299</td>
<td>.463</td>
</tr>
<tr>
<td></td>
<td>Spanish</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>%N</td>
<td>German</td>
<td>X</td>
<td>X</td>
<td>.476</td>
<td>.434</td>
</tr>
<tr>
<td></td>
<td>Spanish</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>ΔN</td>
<td>German</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td></td>
<td>Spanish</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>ΔΕ</td>
<td>German</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td></td>
<td>Spanish</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
</tbody>
</table>

possible to decide on the basis of the present data whether one of the two measurements is better than the other. The compensation of speech rate, which is only provided by PVI is maybe not very important in a reading task, but could be crucial for spontaneous speech.

The values for %N seem to be the most striking results of this study. Although the result may be opposite to the expectations one has concerning the speech of Spanish speakers learning German, the difference was significant and showed a tendency towards a possible discrimination between German and Venezuelan speakers.

Based on the present study it is an open question whether non-vocalic nuclei should be included in the measurements or not. In our opinion, the present results encourage to choose the solution of including them. The individual speakers’ differences for this measurement showed that it may, under certain circumstances, provide a stable index when combined with appropriate segmentation criteria. Furthermore, the segmentation of speech is a labour-intensive and difficult task, even for a trained phonetician. Less difficult segmentation would be without doubt an advantage, especially in the context of automatic rhythm testing. The measurement %N would point into this direction and it should be tested whether even easier segmentation criteria like voiced rhyme segmentation or a simple voiced vs. voiceless distinction could lead to comparable results.

The absolute values of the measurements in the present study have not been compared with values from other studies, as such a comparison seemed to be too problematic due to rather different experimental procedures and our modified calculation of the measurements in the case of vocalic segments. It has to be emphasized, though, that language-specific stability across experimental settings seems to us an important prerequisite for a rhythm index.

One possible deficiency of our experimental design has to be mentioned, too. Experiments with read speech are maybe not the best way to elicit natural speech rhythm, as even native speakers of a language sometimes show rather strange rhythmic patterns under these circumstances. On the other hand, read speech provides us with highly controlled data and is therefore at least appropriate as a starting point. The present results have to be checked against data from spontaneous speech, though.

Finally, it would be interesting to see what happens the other way round, that is in the case of German native speakers learning Spanish. Spanish syllabic structure seems to be easier to some extent than German syllabic structure. If German learners thus have less problems with Spanish syllabic structure, they should also show less rhythmic deviation in Spanish - a hypothesis which seems a little bit surprising at first sight.

9. References

[10] http://www.phonetik.uni-muenchen.de/Bas/BasHome.html