Duration of syllable-sized units in casual and elaborated speech: Cross-language observations on Swedish and Spanish

Engstrand, O.

journal: Proceedings of Fonetik, TMH-QPSR
volume: 44
number: 1
year: 2002
pages: 069-072

http://www.speech.kth.se/qpsr
Duration of syllable-sized units in casual and elaborated speech: cross-language observations on Swedish and Spanish

Olle Engstrand and Diana Krull
Department of Linguistics
Stockholm University

Abstract

Recordings of careful readings and unscripted monologues in Swedish and Argentinean Spanish were auditorily and acoustically analyzed to test the hypothesis that temporal equalization of syllable-sized contoid-vocoid (CV) sequences typically occurs in casual as opposed to elaborated speaking styles. Results indicated 1) that CV durations and standard deviations appearing in unscripted Swedish tended to resemble those in Argentinean Spanish, 2) that durational differences between unstressed and stressed CV units in unscripted Swedish were smaller than those observed in Swedish careful reading, and 3) that, in both languages and speaking conditions, timing of syllabic units was partly determined by their size. Timing of CV units was, thus, significantly affected by alterations in speaking style. This supports the view that rhythmic patterns associated with stress- and syllable-timing may arise as consequences of more primary phonetic intentions rather than themselves representing such intentions.

Introduction

Swedish has a complex phonotax permitting heavy consonant clusters in all positions (Sigurd, 1965). However, Engstrand & Krull (2001) found that sound sequences produced in casual speech tended to consist of alternating contoid and vocoid articulations relating to more complex consonant and vowel structures at the phonological level. This observation invited the hypothesis that durations of syllable-sized contoid-vocoid sequences might be less variable than durations of syllabic units found in more formal speaking styles. One reason was that, in casual speech, effects such as cluster simplification and syllable opening appeared to reduce the amount of time spent on each syllable-sized unit. Another reason was that many unstressed syllables were observed to be dropped altogether. Thus, it appeared as if durational equalization would result from avoidance of extremes, a possible consequence being that casual Swedish approximates the rhythmic structure of ‘syllable-timed’ languages such as Spanish. It is the purpose of this paper to present a preliminary test of this hypothesis, using read and unscripted speech samples from Standard Swedish and Argentinean Spanish.

Methods

The unscripted speech material consisted of lively monologues produced by one male native speaker of Central Standard Swedish and one male native speaker of Argentinean Spanish (henceforth ‘Spanish’). The monologues were supported by brief comments and questions by one of the experimenters during the Swedish recording, and by a Castilian interlocutor during the Spanish recording. The Spanish speaker also read a text in an elaborated speaking style. A Swedish text was read, also in an elaborated style, by a female actress, who was also a speaker of Central Standard Swedish. All recordings were made in sound-treated recording studios using high quality professional equipment.

The recordings were digitized at 16 kHz. Parts of the digitized material were labeled and analyzed using the Soundswell Signal Workstation. Utterances were segmented into syllable-sized units. However, since casual speech is typically characterized by a wealth of coarticulation and reduction phenomena, a conventional morpho-phonologically based syllabification (cf., e.g., Gårding, 1967) was not possible. For example, a word such as behandla ‘treat’ may well be pronounced as [bɛala] with...
1) complete /h/ reduction, 2) partial /e/ reduction leaving just a slight diphthongization and 3) superimposed nasalization as the sole noticeable trace of the nasal consonant. For another example, för att ‘because’ may be realized as [frat] or [fɔat] in casual speech without any audible or visible trace of the /ø/ vowel. Identifying sequences that can be reliably and consistently segmented is, thus, problematical. For the present purpose, contoid-vocoid(-contoid) sequences reflecting opening-closing movements of the vocal tract turned out to be the most convenient units. These were identified auditorily and visually using spectrograms. Largely disregarding grammatical word and morpheme boundaries, single contoids were counted as unit onsets. Contoid clusters were counted as onsets as long as they constituted well-formed initial clusters according to the so-called sonority hierarchy (Jespersen 1926, Engstrand & Krull 2001). For instance, a word such as valurna ‘ballot-box’, which is morphophonologically parsed as /val+urna/ would be phonetically parsed [va.łu.ɾə] (where the dots represent boundaries between contoid-vocoid sequences). For convenience, these sequences will be referred to as ‘CV units’. It should be remembered, however, that they are units only in a strictly phonetic sense and may well contain traces of ‘underlying’ segments.

Results

Figure 1 summarizes the durational data in graphical form. The histograms depict the distribution of CV durations in Swedish (left column) and Spanish (right column). For each language, the top row represents careful reading and the bottom row represents unscripted speech. CV units in prepausal position have been removed throughout the data presentation. The number of observations for each case is given in table 1.

Looking first at the Swedish data, it can be seen that the frequency distribution for the read condition is broader and flatter than that for the unscripted condition. The same tendency is apparent also in the Spanish distributions. It is also clear that the Spanish distributions are narrower than the Swedish distributions in both conditions. However, the distribution for the Swedish unscripted condition bears a closer distributional similarity to the Spanish read condition than does that for the Swedish read condition. This tendency is also evident from the numerical data in table 1.

Table 1 shows that the mean CV duration for the Swedish unscripted condition is 178 ms, i.e., 22 ms smaller than that for the Swedish read condition. For Spanish, the mean values are about 155 ms for both conditions. Thus, durations pertaining to the Swedish unscripted condition tend to fall inbetween those of the Swedish read condition and those of the Spanish read and unscripted conditions. The effect of speaking condition on the dispersion about the means is even more evident. The standard deviation for the Swedish unscripted condition is 62 ms; it is thus 24 ms smaller than that for the Swedish read condition but practically identical to the 59 ms standard deviation found for the Spanish read condition. The standard deviation for the Spanish unscripted condition is slightly less than that for the read condition.

Table 2 presents mean durations for stressed vs. unstressed CV’s in the two languages and speaking conditions. The general picture is similar to that observed in the above data (where stress conditions were not separated). It is evident, however, that the difference between the read and unscripted conditions is strongly linked to stress. Thus, the stressed CV units in the Swedish read and unscripted conditions are 261 and 226 ms, respectively; the corresponding figures for the unstressed CV’s is 154 and 150 ms. In Spanish, the pattern is similar but less extreme. Thus, the stressed CV units in the Spanish read and unscripted conditions are 198 and 170 ms, respectively, whereas the corresponding figures for the unstressed CV’s is 134 and 144. It can be noted that, in Spanish, the unstressed CV’s in the unscripted condition tend to be longer than those pertaining to the read condition.

The differences between stressed and unstressed syllable durations in the respective languages and speaking conditions are summarized in the rightmost column of table 2. The 107 ms stress effect in the Swedish read condition agrees with the expected durational effect of stress in stress-timed languages (Eriksson 1991), whereas the durational quantum added by stress in the remaining conditions (unscripted Swedish as well as read and unscripted Spanish) would be more typical of syllable-timed languages.

Effects on duration of CV unit size is shown in figure 2 depicting mean durations of stressed
Table 1. Mean durations and standard deviations (ms) for CV units in read and unscripted Swedish and Spanish.

<table>
<thead>
<tr>
<th>Language</th>
<th>Condition</th>
<th>Mean</th>
<th>Std</th>
<th>N</th>
</tr>
</thead>
<tbody>
<tr>
<td>Swedish</td>
<td>Read</td>
<td>200</td>
<td>86</td>
<td>350</td>
</tr>
<tr>
<td></td>
<td>Unscripted</td>
<td>178</td>
<td>62</td>
<td>306</td>
</tr>
<tr>
<td>Spanish</td>
<td>Read</td>
<td>156</td>
<td>59</td>
<td>167</td>
</tr>
<tr>
<td></td>
<td>Unscripted</td>
<td>155</td>
<td>51</td>
<td>287</td>
</tr>
</tbody>
</table>

and unstressed CV's as a function of the number of contoid or vocoid segments. For both languages and speaking conditions (top: Swedish, bottom: Spanish, left: read, right: unscripted), a linear function fits the means reasonably well except for the sometimes negligible difference between units consisting of one segment (a single vocoid) and those consisting of two segments (a contoid-vocoid sequence). This suggests that CV size is a strong determinant of duration. The Swedish data also display a relatively constant difference between stressed and unstressed durations whereas, in the Spanish data, there is an interaction between CV size and the amount of time added by stress. However, the smallest and the largest CV sizes are represented by very few tokens which makes these effects somewhat unreliable.

Summary and conclusions

Previous observations have suggested that phonotactic patterns at the phonological level...
tend to undergo considerable simplification in unscripted Swedish speech. This has led to the hypothesis that temporal equalization of CV-sized units takes place in casual speaking styles. Segment and syllable reduction might, thus, entail timing effects usually considered typical of syllable-timed languages such as Spanish. Preliminary data presented here have supported this expectation in that CV durations and standard deviations appearing in unscripted Swedish tended to resemble those appearing in Spanish. In addition, the durational difference found in unscripted Swedish between unstressed and stressed CV units was much smaller than that found in careful reading. In that respect, too, the unscripted speaking condition was less typically stress-based than the read condition. Finally, a clear relationship was observed between number of segments and duration of CV units. This suggested that timing of syllabic units is crucially determined by their size. It should be noted, however, that Spanish CV’s were consistently shorter than the corresponding Swedish CV’s. This may reflect traces of the reduced gestures remaining in the Swedish CV’s. In Spanish, which is essentially CV-based, timing is less affected by such remnants.

In a much cited article, Dauer (1983) wrote: “The difference between stress-timed and syllable-timed languages has to do with differences in syllable structure, vowel reduction, and the phonetic realization of stress and its influence on the linguistic system. Languages, language varieties, or historical stages of a language can be considered more or less stress-based depending on differences in these characteristics” (p. 51). Subsequent studies (e.g., Strangert 1985, Engstrand 1987, Eriksson 1991) have reached similar conclusions. This study has also demonstrated that aspects of stress- and syllable-timing may be significantly affected by speaking style. This lends additional credence to the claim that, in their temporal aspects, speech rhythms arise as consequences of more primary phonetic intentions rather than themselves representing such intentions.

References
