Syllable durations obtained from the KTH speech database

Strangert, E.

journal: STL-QPSR
volume: 29
number: 4
year: 1988
pages: 051-057

http://www.speech.kth.se/qpsr
III. SPEECH PRODUCTION

A. SYLLABLE DURATIONS OBTAINED FROM THE KTH SPEECH DATA BASE

Eva Strangert*

Abstract

The work reported is a data base analysis of the temporal differences between stressed and unstressed syllables in Swedish. In addition to the stressed/unstressed distinction, effects of syllable structure, position and boundaries were studied. The data revealed relatively stable unstressed/stressed duration ratios with minor variations between .50 and .60 in the different conditions. The analysis generally confirmed previous results based on a more restricted material.

1. Introduction

Speech data are normally obtained from very restricted material. Prosodic analysis, for example, is usually based on a relatively small number of sentences, that are systematically varied so as to reveal the effects of the prosodic aspect under study. There can be heavy restrictions on their segmental composition, and reiterant or hummed speech may be used.

An analysis of data from a real-speech data base offers a valuable complement to this approach, provided that the material is wide enough in scope. A large number of different sentences may be used to test the generality of results revealed in experimental studies. The work reported here includes both a test and an elaboration on previous data on syllable durations.

In Swedish, the mean stressed and unstressed syllable durations are clearly different, at least when the syllables contain 1 or 2 consonants (Strangert, 1985, pp. 73-80; Strangert, 1987). A stable temporal relation between the (primary) stressed and the following unstressed syllables in a stress group was found in that study when a set of sentences was systematically varied. The ratio between the mean unstressed and stressed syllable duration remained close to 50-60% irrespective of the number of unstressed syllables (0-4) in the stress group. This was equally true for speech at normal speed as well as for fast and slow speech. The three speakers in that study all had similar ratios. The stability was seen to reflect the importance of the stressed/unstressed distinction in Swedish. Also, the low percentages were seen as a consequence of the greater complexity of stressed as compared to unstressed syllables. Two characteristics, in particular, contribute to this differentiation: there are fewer restrictions on the segmental composition of stressed syllables and the quantity distinction is only found in stressed syllables (cf., Strangert, 1985, pp. 29-31).

* Department of Phonetics, Umeå University, S-901 87 Umeå, Sweden. This study was carried out while the author was a guest researcher at the Department of Speech Communication and Music Acoustics.
That analysis was part of a study of Swedish speech rhythm. In the present work, data from that study were compared to the stressed and unstressed syllable durations found in the data base. In addition to the stressed/unstressed distinction, the following factors were systematically studied:

a. syllable structure (the number and type of consonants in the syllable)
b. position (before a stressed or unstressed syllable)
c. boundaries (presence or absence of an intervening word or formative boundary**)

2. General characteristics of the data base and its operation

The data base consists of 150 Swedish sentences containing from 2 to 13 words read by a male speaker.

Each sentence is stored together with its respective phonetic transcription that is semi-automatically generated by the KTH text-to-speech system. It is further segmented into phoneme-sized units. Thus, the data base contains both a segmented phonetic transcription and durational information about the segments. For details on the transcription and segmentation procedure, see Carlson & Granström (1985; 1986).

The data base is searched by means of rules that are similar in structure to those used in the KTH text-to-speech system. These rules permit all instances of a certain phoneme or phoneme string to be identified and their durations measured.

3. Accessing stressed and unstressed syllable durations

At the outset a few general rules were formulated:

a. A syllable was defined as one vowel and any following consonant(s) up to the next vowel irrespective of any intervening word or formative boundary.***

b. The category "stressed syllable" included syllables with the primary stress of accent I and accent II words and the secondary stress of compound accent II words. Other types of syllables were referred to the "unstressed" category.

c. Syllables immediately before or after pauses were excluded from further analysis as they may be especially influenced by onset and offset phenomena.

General characteristics of the data

The distribution of unstressed and stressed syllable durations in the entire material (N=1733 syllables) is shown in Fig. 1a. The dispersion is partly a consequence of the lack of consonant restrictions, both categories including syllables containing from no to several consonants. There are, however, two quite distinct peaks. The mean durations of the unstressed and stressed syllables are 150 and 234 msec, respectively (see also Table I). The unstressed-to-stressed ratio of the means is .64. It is somewhat higher than the previously reported ratios that ranged between .50 and .60 (Strangert, 1985, p. 79).

** "formative boundary" here refers to the boundary between the parts of a compound word.

*** Vowel onsets had been used as segmentation points for syllables in the previous studies reported (Strangert, 1985; 1987). To get comparable data the same segmentation points were used here.
However, still higher ratios may be demonstrated in individual stress groups. In groups of two syllables, there are instances of extreme ratios like .08 and 2.0. These ratios occurred in the following sentences:

Små grytor har också öron.  
(Little pitchers have big ears.)

|  /óks(oː)ː/ | 260 20 | .08 |

Man måste betala skatt på lönen.  
(You have to pay tax on your earnings.)

|  /ˈaː l a s k ā/ | 120 240 | 2.0 |

Though both instances are extreme, the first is less so than the second. The first exhibits an exaggerated "normal" pattern: a structurally complex stressed syllable and a structurally simple, often reduced unstressed syllable (cf., Strangert, 1985, pp. 29-31). The second example shows the opposite "unmoral" pattern: a complex unstressed syllable and a simple stressed syllable. This is the principal reason for the very high ratio, though other factors may have contributed, too. Thus, the presence of a phrase boundary between 'betala' and 'skatt' and extra prominence (due to focal accent) on 'skatt' may well have induced temporal effects that exaggerated the differences between the syllables.

**Effects of syllable structure**

The two examples above illustrate the durational effects of differences in syllable structure. Fig. 1b gives the distribution of durations of syllables containing just one consonant. Figs. 1c and 1d refer to the same one-consonant data broken down into specific subsets, syllables with an obstruent or a non-obstruent consonant respectively. Table I gives the number of occurrences, means and standard deviations for these categories. It also includes the ratios between the mean unstressed and stressed syllable durations of each category.

<table>
<thead>
<tr>
<th></th>
<th>unstressed</th>
<th>stressed</th>
<th>U/S ratio</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>M  SD  N</td>
<td>M  SD  N</td>
<td></td>
</tr>
<tr>
<td>no restrictions</td>
<td>151 68 1007</td>
<td>234 60 726</td>
<td>.65</td>
</tr>
<tr>
<td>cons (1,1)</td>
<td>110 34 507</td>
<td>210 49 392</td>
<td>.52</td>
</tr>
<tr>
<td>cons, obst (1,1)</td>
<td>117 33 273</td>
<td>222 51 230</td>
<td>.53</td>
</tr>
<tr>
<td>cons, -obst (1,1)</td>
<td>102 34 234</td>
<td>194 38 162</td>
<td>.53</td>
</tr>
</tbody>
</table>

| cons = consonant; obst = obstruent; -obst = nonobstruent; (1,1) = one, no more, no less; U/S ratio = unstressed-to-stressed duration ratio

**Table I.** Durations (mean and SD in msec.) and mean ratios of unstressed and stressed syllables with additional variation of syllable structure.

The restriction to one consonant affects the range of durations, especially in the unstressed category. There is less dispersion and the peaks are farther apart. This lowers the unstressed-to-stressed ratios to .52 or .53 for the three categories.
Position effects

Table II shows the duration of unstressed and stressed syllables in different syllable contexts. Each category has been divided into two subsets, syllables occurring either before unstressed or stressed syllables. The number of consonants in the syllables is systematically studied in the same way as before.

Fig. 1. Duration of unstressed and stressed syllables a) with no consonant restrictions, b) with one consonant of any type, c) with one obstruent, d) with one non-obstruent.
The position clearly affects the duration of syllables in these data. The unstressed syllables are always longer before a stressed than before an unstressed syllable, and this is true irrespective of the number and type of consonants.

The position differences amount to 14 msec for the one-consonant data and 48 msec. for syllables without consonant restrictions. The 14 msec difference may be explained with reference to a previous search in this data base. Carlson & Granström (1987) found single consonants immediately preceding a stressed vowel to be 16 msec longer than consonants before an unstressed vowel. As consonants are grouped with the preceding vowel by the definition of a syllable adopted here, the position difference may be a consequence of this specific definition. The difference of 48 msec may be similarly explained. In this case differences in syllable complexity between stressed and unstressed syllables may play a major role.

The position differences amount to 14 msec for the one-consonant data and 48 msec. for syllables without consonant restrictions. The 14 msec difference may be explained with reference to a previous search in this data base. Carlson & Granström (1987) found single consonants immediately preceding a stressed vowel to be 16 msec longer than consonants before an unstressed vowel. As consonants are grouped with the preceding vowel by the definition of a syllable adopted here, the position difference may be a consequence of this specific definition. The difference of 48 msec may be similarly explained. In this case differences in syllable complexity between stressed and unstressed syllables may play a major role.

The stressed syllables before unstressed and stressed syllables respectively present a more complex pattern. When there are no restrictions on the number of consonants, the mean stressed syllable is longer in prestress position, just like the mean unstressed syllable. The position differences may be similarly explained, too. The one-consonant data, however, show the opposite pattern. Unfortunately, the number of occurrences are too small to permit any conclusions. One might, however, speculate on the reason for this converse pattern. If this difference is real, why should stressed syllables with just one consonant be shorter before another stressed syllable? A possible explanation might be that it is a way to adjust to "stress clash" (two stressed syllables immediately following each other). According to Bruce (1987) one of the two clashing syllables may be weakened. Such a weakening could lead to shortening, and perhaps it is facilitated if the syllable structure is not too complex.

The contrast between stressed and unstressed syllables is stronger before an unstressed than before a stressed syllable in this material. The mean difference is 10% in the categories with no restrictions on the number of consonants.

Table II. Durations (mean and SD in msec.) and mean ratios of unstressed and stressed syllables with additional variation of the following syllable (unstressed/stressed) and variation of the syllable structure. Unreliable estimations within parentheses.

<table>
<thead>
<tr>
<th></th>
<th>unstressed</th>
<th>stressed</th>
<th>U/S ratio</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>M</td>
<td>SD</td>
<td>N</td>
</tr>
<tr>
<td>before unstressed</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>no restrictions</td>
<td>123</td>
<td>56</td>
<td>427</td>
</tr>
<tr>
<td>cons (1,1)</td>
<td>104</td>
<td>32</td>
<td>283</td>
</tr>
<tr>
<td>cons, obst (1,1)</td>
<td>110</td>
<td>30</td>
<td>135</td>
</tr>
<tr>
<td>cons, -obst (1,1)</td>
<td>99</td>
<td>33</td>
<td>148</td>
</tr>
<tr>
<td>before stressed</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>no restrictions</td>
<td>171</td>
<td>69</td>
<td>580</td>
</tr>
<tr>
<td>cons (1,1)</td>
<td>118</td>
<td>35</td>
<td>224</td>
</tr>
<tr>
<td>cons, obst (1,1)</td>
<td>124</td>
<td>33</td>
<td>138</td>
</tr>
<tr>
<td>cons, -obst (1,1)</td>
<td>108</td>
<td>36</td>
<td>86</td>
</tr>
</tbody>
</table>
Effects of boundaries

Only the one-consonant data, cons (1,1), have been analyzed. The result is shown in Table III.

<table>
<thead>
<tr>
<th></th>
<th>unstressed</th>
<th>stressed</th>
<th>U/S ratio</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>M</td>
<td>SD</td>
<td>N</td>
</tr>
<tr>
<td>without a boundary</td>
<td>99</td>
<td>29</td>
<td>173</td>
</tr>
<tr>
<td>with a boundary</td>
<td>116</td>
<td>35</td>
<td>334</td>
</tr>
</tbody>
</table>

*Table III. Durations (mean and SD in msec.) and mean ratios of unstressed and stressed syllables with additional variation of word and formative boundaries (absence/presence). Analysis restricted to syllables with one consonant. unreliable estimations within parentheses.*

The unstressed syllables including a word or formative boundary are somewhat longer than those without a boundary. In the stressed category the pattern is the opposite. Syllables with a word or formative boundary are somewhat shorter than syllables without a boundary. However, to decide if an intervening boundary in effect influences stressed and unstressed syllables differently, more data need to be analyzed. What seems to be true is that syllable durations are only marginally affected by the presence or absence of boundaries. This supports some of the earlier data on word boundaries. The boundary effects, however, differed considerably between speakers (Strangert, 1985, p. 64-68).

The unstressed-to-stressed ratio is small, indicating a sharp distinction for syllables without an intervening boundary. For syllables with a boundary the estimated ratio is based on insufficient data on stressed syllables. It should be noted that to a great extent these syllables also appeared in the problematic category "stressed syllable before another stressed syllable" above.

4. Concluding remarks

The unstressed-to-stressed ratios come close to the previously reported range of .50 - .60 in all conditions. However, the ratio is larger in some cases and smaller in others.

The temporal distinction is weaker, that is, the ratio approaches or exceeds unity, when the number of consonants is not severely restricted. Extremely high ratios result when the stressed syllable has fewer consonants than the unstressed. Further, the distinction is weaker before a stressed than before an unstressed syllable, and it seems to be weaker when there is a boundary within the syllable than when there is not. Such weakening in prestress position as well as in the presence of a boundary could in both cases be traced back to the stressed syllable. The differences compared to unstressed syllables might either be "real" replaceable effects or they may be due to the very small number of instances of stressed syllables in the conditions in question. More data are needed to shed light on this problem.

Acknowledgments

I am very grateful to Björn Granström and Rolf Carlson for permission to use the KTH data base and for all their valuable help with the search of the material. I am also indebted to Björn Granström for helpful comments on the manuscript.
References