A note on segment duration in Swedish polysyllables

Lindblom, B.

journal: STL-QPSR
volume: 5
number: 1
year: 1964
pages: 001-003

http://www.speech.kth.se/qpsr
I. SPEECH ANALYSIS AND SPEECH PRODUCTION

A. A NOTE ON SEGMENT DURATION IN SWEDISH POLYSyllABLES

A speech material has been constructed to enable the study of the acoustic correlates of stress contours within Swedish polysyllables. Words are \( n \) syllables long \((n = 1, 2, 3, 4, 5)\). For a given word length a \( CV_{1}V_{2} \) sequence occurs in all possible positions. The rest of the word is formed by prefixing or suffixing the number of \( CV \) and \( VC \) syllables necessary to make it \( n \) syllables long. Each word has one main stress that occurs in all possible positions. In this way \( n \frac{(n-1)(n-2)}{2} \) words are generated.

Below a few preliminary results will be reported. Measurements were made of segment durations within words in which the \( CV_{1}V_{2} \) syllable was \([\text{bal}]\) and the added syllable \([-\text{al}]\) and/or \([\text{lal}-]\). Thus words like \(['\text{balal}'], ['\text{lalbal}'], ['\text{lala\text{-}bal}']\) etc. were obtained. Each word had the tonal accent I and was preceded by a carrier phrase "så är det ----". A male Swedish talker read this material which comprised three repetitions of each utterance. Since no major changes could be observed in the length of the carrier phrase it was assumed that the talker had managed to keep the overall speaking rate fairly constant.

Informal listening to the recorded material revealed that the resulting stress contours of the nonsense words were similar to those characteristic of some normal Swedish words as they are uttered in lexical pronunciation. These stress patterns are usually denoted:

\[
\begin{array}{c|c}
  n=1 & 4 \\
  n=2 & 40 04 \\
  n=3 & 400 040 104
\end{array}
\]

using a scale of prominence increasing from 0 to 4.\(^{(1)}\)

Within the syllable \([\text{bal}]\) measurements were made from spectrograms of the duration of the \([b]\)-occlusion, the vowel segment of the interval between the \([b]\)-losion and the moment of contact for \([l]\) (visible in spectrograms as a sudden lowering of \(F_{1}\) and a decrease in level) and the \([l]\)-segment
defined as the interval between the moments of contact and release. Except for a few cases of [l] in final position which have been excluded from the presentation below the events demarcating the segments were in general easy to determine. The following results emerge at this preliminary stage:

1. There is a tendency for the average segment duration in accented syllables to decrease as the number of syllables in the word increases. This effect is particularly marked in the final [l] and the vowel can be observed to some extent also for the [b]-segment. Defining syllable duration for the moment simply as the sum of constituent segment lengths the duration of syllables to which stress level 4 can be assigned apparently varies by approximately 250 msec. (Fig.1-1)

2. The duration of segments within an accented syllable depends not only on the number of syllables within a word but also on its position. This is shown in Fig.1-2 in which points to the extreme left and right pertain to initial and final positions respectively. In final syllables sequences of [al] tend to be considerably longer than in initial positions. In the [b]-segment this difference was not noticeable to any marked extent.

3. Comparing the duration of segments in accented and unaccented syllables for a given word length and when they occur in identical positions we find a fairly lawful relationship indicating that segments in unaccented syllables may also vary with position and syllable number. In Fig.1-3 the length of segments in unstressed syllables appear to be about 65% of that of segments occurring under stressed conditions and in corresponding positions.

4. As shown in Fig.1-1 changes in vowel segment duration are accompanied by similar variations in the length of the final consonant. Included in Fig.1-4 are data for unstressed as well as accented pairs of [al]. Each point represents the mean of three cases.

It should be remembered that the present investigation was undertaken as a pilot study. More data are being collected
The mean of 3, 6, 9, 12, 15 measurements from left to right each point represents a word. The number of syllables in a stressed syllable is shown against the average segment duration in msec.
Fig. 1-2. The figure demonstrates the effect of position.

The number of syllables.

When stressed, the numbers in the plot refer to
within the word on the duration of an [a]-segment

Position

Number of Syllables

Duration

Deviation of [a] Duration

From Average [a] Duration

(accented syllables)

 msec

Initial position  Final position

2/2 3/3  4/4  5/5  6/6  7/7  8/8  9/9  10/10
SEGMENT LENGTH IN CORRESPONDING UNACCENTED SYLLABLES

SEGMENT LENGTH IN ACCENTED SYLLABLE

msec

Fig. 1-3. For a given position and length of a word the average duration of a segment in unstressed syllables has been plotted against its duration in identical position under stressed conditions.
Figure 4. The duration of [n] is shown plotted against the mean of three cases. Each point is based on the mean of symbols. For stressed as well as unstressed that of [n] is shown plotted against [n].
along similar lines from a more comprehensive material in order to gain a broader understanding of stress contours and temporal organization in speech.

B. Lindblom

Reference:


B. ON THE VOWEL SOURCE SPECTRUM

Two different methods have been developed for investigating vowel source spectra:

a) Analysis by synthesis

b) Inverse filtering

The analysis by synthesis method was based on spectrum sampling with our 51-channel spectrum analyzer and synthesis by means of a digital computer. This frequency domain method is described in more detail by Tappert, Mártony, and Fant (1).

The inverse filter is described by Fant (2) and Lindqvist (3). It operated in the time-domain providing an output similar to the glottal wave form. The transform of the output to a spectrum was also made with the 51-channel analyzer.

The speech material used in the present study are different steady state vowels produced by a single speaker with normal voice effort.

F.M. recordings were made for the inverse filtering analysis in order to avoid phase distortion when studying wave forms. The F.M. system had a high frequency cutoff of 2.5 kc/s.

Fig. 1-5 shows the deviations of the source spectra from a reference 12 dB/oct falling source. The solid lines