

INTRODUCTION

In some germanic languages, including Swedish, there is a complementary relationship between the durations of [V] and [C] in [VC]-sequences. In positions carrying word-accent, Swedish phonotax allows only /V:/, /V:C/, and /VC:/ (Elert, 1964). In search for an invariant description of the durational properties of these sequences, invariant with respect to variations in speech rate and context, it has been proposed to consider the ratio of vowel duration to consonant duration (V/C), or the ratio of vowel to [VC]-sequence duration (V/(V+C)), (Bannert, 1976). Within the scope of the present investigation, the validity of these ratios as descriptors of the quantity distinction was tested, also considering the perception of the durations of any kind of speech segments in general.

METHOD

A female speaker of standard Swedish produced each word among the minimally distinctive pairs [œ:n:a] - [œ:n:a], [de:sa] - [des:a], and [mɛ:ta] - [mɛ:t:a] in a carrier phrase [de va ju ____ ja sɔ:], that was uttered at five different rates of speech, distributed over the whole range of variation occurring in natural speech.

The utterances were recorded and a computer program allowing the elimination or repetition of single glottal periods, or sections of similar duration in unvoiced speech segments, was used to manipulate the durations of the [V]- and [C]-segments in such a way as to obtain a number of utterances in which these segments had durations interpolated between those occurring in the natural utterances with /V:C/ and /VC:/ sequences. The actually obtained durations of vowel, consonant, and utterance are listed in Table 1 for one example. The final vowel of the carrier phrase has been excluded from what we here refer to as "utterance duration". An extra fast speech rate was obtained by shortening the whole utterance spoken with the fastest speech rate our speaker could achieve by uniformly distributed removal of glottal periods.

Table 1: Segment durations (in ms) in one series of stimuli (1 to 5): Manipulated versions of [de va ju mɛ:ta ja sɔ:].

	Orig. [ɛ:t]	1	2	3	4	5	Orig. [ɛ:t]
"Utterance"	1144	1136	1132	1128	1141	1129	1165
Vowel	220	185	155	129	107	92	101
Consonant	154	181	207	229	264	267	266

The utterances were presented via headphones in order of increasing speech rate, and in randomized order within each rate, to 25 adult subjects, all natives of Central Sweden, using a computerized system of response collection. The subjects had to identify the word containing the manipulated segments. The response alternatives were 1 (/CV:Ca/), 2 (/CVC:a/), and X (undecided).

RESULTS AND DISCUSSION

Figure 1 shows the results obtained for two series of stimuli based on original utterances opposed in quantity only. There were 18 such pairs of stimulus series, but the following presentation is limited to one crucial point in each of these: the mean position of the quantity

boundary, i. e., the mean of the two crossover points of the dashed and whole-drawn curves labeled 1 and 2 in Figure 1.

The durations of the vowel and the consonant for stimuli at the /V:C/-/VC:/-boundary are shown in Figure 2 as a function of total utterance duration. Figure 3 shows the ratio $V/(V+C)$ as a function of utterance duration.

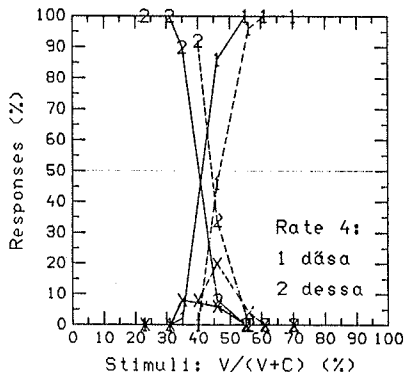


Figure 1: Identification results (see text).

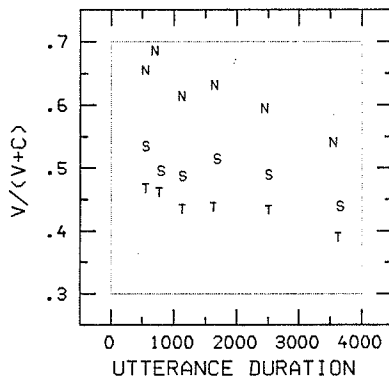


Figure 3: The durational ratio $V/(V+C)$ at the /V:C/-/VC:/-boundary, as a function of utterance duration in ms.

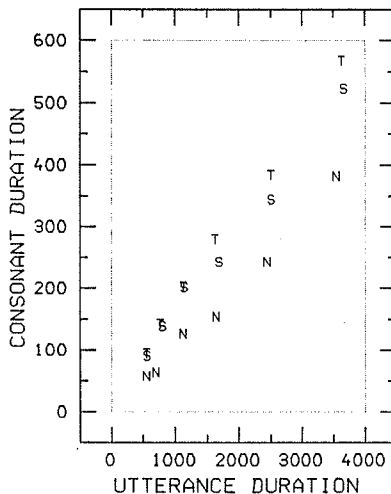
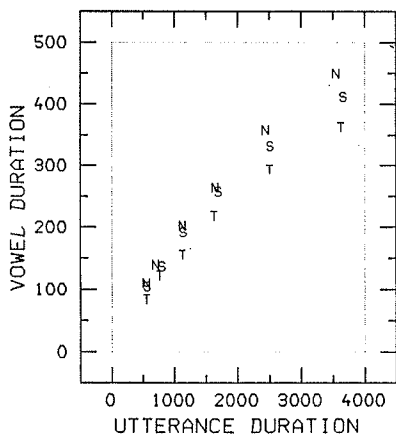


Figure 2: Vowel (left) and consonant durations (right) at the /V:C/-/VC:/-boundary, as a function of utterance duration. Durations in ms.

Footnote: There was an [h]-like segment (preaspiration) at the end of the vowel when followed by [t] or [s]. In Figure 1 this is included in the consonant duration, while it is included in the vowel duration in all the following figures.

If all segmental durations would vary proportionally with changing speech rate, implying invariant ratios V/C and $V/(V+C)$, the data points in Figure 2 would fall on straight lines going through origo. In Figure 3 they would fall on lines in parallel with the x-axis. While this may be said to be the case in rough approximation, we observe regular deviations from such a course.

Non-proportional variations in segment durations have been observed and taken into consideration previously (Klatt, 1979; Lindblom et al, 1981). According to the approaches taken by these authors, our data points in Figure 2 should still fall on straight lines, but these need not go through origo. While an improved fit would be obtained in this way, inspection of our figures led us to the conjecture that a certain class of curved lines which do go through origo might describe our data still better and more adequately.

Figure 2 (right) shows substantial differences in duration between the consonants [n], [s], and [t]. In addition, the duration of the vowel is dependent on the following consonant, as can be seen in Figure 2 (left). Both observations agree qualitatively with what has been observed previously (Elert, 1964; Lindblom et al, 1981).

Figure 4 is equivalent to Figure 2, except for the scaling of the durations, which now is logarithmic. The linear regression lines shown in Figure 4 can be seen to fit the data very well. According to these lines, the relation between the durations D_y (y-axis) and D_x (x-axis) is described by $\log(D_y) = p \log(D_x) + c$, where p is the slope of the line and c its displacement. The slope of the line indicates the compliance, with respect to changes in speech rate, of the segment whose duration is plotted along the y-axis, as compared with the one plotted along the x-axis. If these segments vary proportionally in duration, then $p = 1.0$. Figure 4 shows the compliance of the vowel segment to be distinctly lower (0.76) than that of the whole utterance (1.00). This figure also shows the influence of the following consonant on the duration of the vowel to be roughly constant, expressed in logarithmic units, i. e., also in terms of a percentage.

All this means that a power-law with exponent p holds between D_y and D_x . We expect this kind of law to hold, in satisfactory approximation, for any kinds of segments, including also pauses, whose compliance p will be higher than that of any real speech segments. The power law also holds satisfactorily for the duration of the [VC] segment, as produced by our speaker, with $p = 0.86$, relative to the duration of the whole utterance.

In this connection, it should also be noted that a power law with $p = 2.0$ has been found to hold between speech rate measured physically (syllables per minute) and psychophysically (Grosjean and Lane, 1981).

The power-law for segmental durations, in the following form, can be used, i. a., to vary the speech rate of synthetic speech:

$$D = k^D D_0 ,$$

where D is segment duration, D_0 its initial value, k the factor by which an utterance is to be stretched in duration, and p the compliance of the segment relative to that of the whole utterance. However, before this law can be used practically, the segment specific values of p must be known, in addition to the initial durations.

Figure 5 is equivalent to Figure 3, but with utterance duration and the ratio $V/(V+C)$ both scaled logarithmically. The slopes of the regression lines shown in this figure reflect the difference in compliance p of the vowel [e] and the consonants [n], [s], and [t].

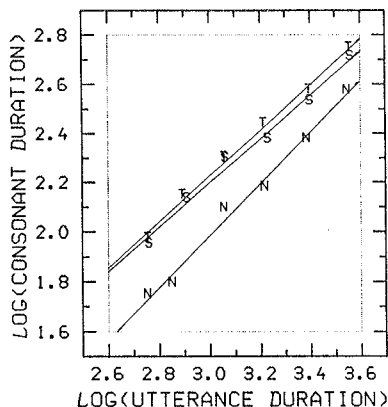
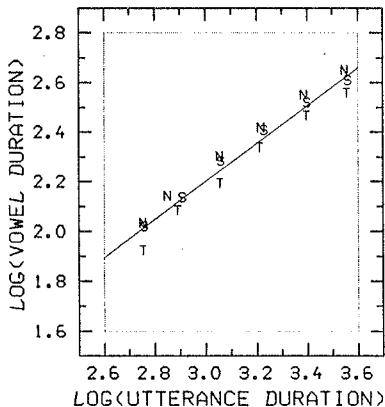
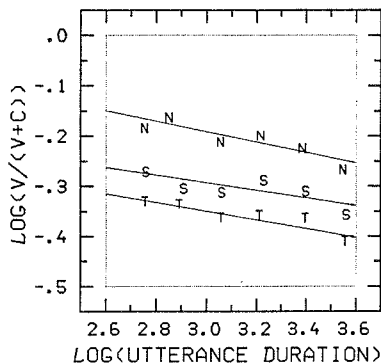


Figure 4: Vowel (left) and consonant durations (right) at the /V:C/-/VC:/-boundary, as a function of utterance duration. Durations logarithmically scaled. Same data as in Figure 2.

Figure 5: The durational ratio $V/(V+C)$ at the /V:C/-/VC:/-boundary, as a function of utterance duration. Quantities logarithmically scaled. Same data as in Figure 3.



ACKNOWLEDGMENT

This research is supported by a grant from HSFR, the Council for Research in the Humanities and Social Sciences.

REFERENCES

R. Bannert (1976): Mittelbairische Phonologie auf akustischer und perceptorischer Grundlage, Travaux de l'institut de phonetique de Lund, vol X, Gleerups, Lund.

C.C. Elert (1964): Phonologic Studies of Quantity in Swedish, Almqvist & Wiksell, Uppsala.

F. Grosjean, H. Lane (1981): Temporal variables in the perception and production of spoken and sign languages. In P.D. Eimas, J.L. Miller (eds.) Perspectives on the Study of Speech, Lawrence Erlbaum Ass., Hillsdale, New Jersey, pp. 297-237.

D.H. Klatt (1976): Linguistic uses of segmental duration in English: Acoustic and perceptual evidence, J. Acoust. Soc. Am., 59: 1208-1221.

B. Lindblom, B. Lyberg, K. Holmgren (1981): Durational Patterns of Swedish Phonology: Do they reflect short-term motor memory processes?, Indiana Univ. Linguistics Club.