INTERNATIONAL JOURNAL
OF LINGUISTICS

Published by The Linguistic Circle
of Copenhagen

Acta Linguistica Hafniensia

Linguistic Studies in Honour of Jørgen Rischel

Edited by
Jens Elmegård Rasmussen
in collaboration with
Birgitte Jacobsen and Peter Molbæk Hansen

VOLUME 27, part one

C. A. REITZEL · COPENHAGEN 1994

Wolfram, Walt: *A Sociolinguistic description of Detroit Negro Speech,*
1969 Center for Applied Linguistics, Washington D.C.
Wolfson, Nessa: Speech Events and Natural Speech: Some Implications for Sociolinguistics Methodology, *Language in Society,* 5, 189-209
1976
Wolfson, Nessa: *The Conversational Historical Present Tense in American English Narratives,* Foris Publications, Dordrecht
1982
Ømålsordbogen, C.A.Reitzel, København
1992

RHYTHM, DURATION AND PITCH IN REGIONAL VARIANTS OF STANDARD DANISH

by

NINA GRØNNUM
Institute af General and Applied Linguistics
University of Copenhagen

Abstract

The acoustic analysis is an attempt to establish the durational correlates of differences in perceived rhythm in six regional variants of Standard Danish. Durational relations between vowels and consonants in disyllabic sequences vary considerably across the six regions, but the correlation with differences in perceived rhythm is nearly nil. Fo patterning – over and above its influence on the perception of duration – seems to be the main determinant of perceived rhythm.

Introduction

The analysis is based on a material recorded for intonation investigation purposes (cf. Thorsen 1988a, 1988b, Grønnum 1989, 1991, 1992), and is motivated by my clear auditory impression from the recordings that the six regions concerned differ not only in stress group patterning and intonational parameters, but also with respect to the internal rhythmical organization of the stress group (or the foot, i.e. the succession of a stressed and all succeeding unstressed syllables). The six are: Copenhagen, Næstved (South Zealand), Aalborg (North Jutland), Tønder (West South Jutland), Sønderborg (East South Jutland) and Bornholm (an island south of Sweden).

Because of the nature of the material and the question I am posing, there are a number of rhythmical phenomena which are *not* considered here: (a) inter-stress intervals – which have received a great deal of attention in the literature, cf. e.g. Allen (1975), Dauer (1983, 1987), Hoequist (1983), Kohler (1983a), Lehiste (1977), Strangert (1985), Uldall (1971); (b) variations in tempo within the phrase or utterance which might also be subsumed under the 'rhythm' heading, cf. Fant and Kruckenberg (1989); (c) the more global rhythmical variation which may be induced by varying the degree of prominence on successive stressed syllables; (d) the role of timing in disambiguating syntactic constructions, i.e. its role in the grouping of successive stimuli, cf. e.g. Lehiste et al. (1976); (e) rhythmical alternation within sequences of unstressed syllables, cf. Bruce (1984), Strangert (1985); (f) segment duration as a function of phonological structure or position, cf. Fischer-Jørgensen (1982), Klatt and Cooper (1975), Lehiste (1971); (g) perceptual centres, i.e. the psychological moment of occurrence of an event, for instance a stress beat, cf. Marcus (1979), Morton et al. (1976), Fowler (1979), Kohler (1981). In brief, I am not studying the effect of *systematically* varying parameters that may influence the perception of rhythm (duration, intensity, pitch, phonological structure). I am rationalizing after the fact, so to speak: looking for acoustic differences in recordings that appear auditorily to be different rhythmically at the level of the prosodic stress group, the foot.

Insofar as the stress-timed versus syllable-timed dichotomy is operative at all (cf. e.g. Dauer 1983, Eriksson 1991, Hoequist 1983, Miller 1984), all of the varieties under investigation here are stress-timed, and the rhythmic groups all fall under the heading 'falling' or 'trochaic', 'dactylic', respectively, i.e. their durational, pitch and other characteristics are such that the accent unmistakably falls on the first syllable in the sequence, cf. Woodrow (1911, 1951). However, I am not so certain that the present data support the general assumption that syllables (rather than segments) are the relevant units for the establishment of rhythm, cf. Huggins (1972), or if so their boundaries may not be altogether catholic, cf. below.

For general overviews of the role of rhythm in speech and its acoustic correlates, of factors that influence segment duration, and the production and perception of rhythm and timing, see Allen (1975), Eriksson (1991), Huggins (1972), Klatt (1976), Lehiste (1970, 1977), Ohala (1975) and Strangert (1985). For a general introduction to Danish intonation, see Grønnum (1992, in print), and for a complete account of the regional variants investigated here, see Thorsen (1988a, 1988b) and Grønnum (1990), all summarized in Grønnum (1991, 1992).

Material

The material analyzed was to a large extent homogenous across regions, so measurements and comparisons could be carried out in a meaningful man-

ner. The underlined passages in the ten utterances below constitute the stress groups analyzed. Words in final position were excluded, because 'final lengthening' is not equally extensive in all six varieties, cf. Grønnum (1989). The questions were succeeded by answers in the reading task, the declaratives were isolated. The utterances are thus all void of any (default or focal) sentence accents, cf. Grønnum (1992).

Kamma stammer fra Svaneke/Næstved/Salling/Padborg.
([ˈg̊ʰma] [ˈsd̥ɑmɐ fʁ̥ɑ])
Anders og Kamma skal til Sandvig/Fakse/Strandby/Ballum.
([ˈɑnɐs ʌ] [ˈg̊ʰɑma sg̊a d̥ʰe])
Torbens søster hedder Kamma.
([ˈd̥ʰɒ:b̥əns] [ˈsøsd̥ɐ heðɐ])
Hvad hedder Torbens søster?
([ˈd̥ʰɒ:b̥əns])
Hvem har en søster der hedder Kamma?
([ˈsøsd̥ɐ d̥ɑ heðɐ])
De fik kanerne frem til nytår.
([ˈg̊ʰæ:nɐnə] [ˈfʁ̥amˀ d̥ʰe])
Hun fik kanderne fyldt til kanten.
([ˈg̊ʰanɐnə] [ˈfylˀd̥ d̥ʰe])
Hun fik kanden fyldt med mælk.
([ˈg̊ʰann̩]) [ˈfylˀd̥ mɛ])
Hvem skal til Sandvig/Fakse/Strandby/Ballum foruden Anders?
([ˈvɛmˀ sg̊a d̥ʰe])
Hvor skal de unge holde ferie?
([ˈvɒ:ˀ sg̊a d̥i])

Thus there are 16 stress groups in this analysis at the outset, recorded six times by two to four speakers from each region. Two speakers (the most fluent readers) from each area were selected for the present purpose.

Recordings, registration, segmentation and measurement

Speakers from Tønder, Sønderborg and Aalborg were recorded with high quality portable equipment in their homes or at work, whereas speakers from Copenhagen, Næstved and Bornholm were recorded in professional recording studios. The procedures otherwise are standard (if old-fashioned) in that the analysis was performed on analog hardware instrumentation (pitch and intensity analyzers connected to a mingograph) and measuring done by hand. For a complete technical account, see Grønnum (1989, p. 30-31). Note only here that C-V boundaries were set to the point in time where a high-pass filtered (500Hz) intensity curve begins to drop sharply, and V-C boundaries where the same curve begins to rise sharply. Consonant clusters were measured as a whole. Initial consonants are excluded from calculation, because stressed vowel onset seems to determine the onset of the unit within which temporal compensations may take place, cf. e.g. Allen (1975), Fant and Kruckenberg (1989), Fischer-Jørgensen (1982), Huggins (1972), Strangert (1985). (In fact, Hoequist 1983 is the only recent reference I am aware of who includes initial consonants in his durational measurements of stressed syllables.) Stressed vowel onset also constitutes the boundary between successive Fo patterns, cf. Thorsen (1984a, 1984b). The importance of Fo patterning for the perceptual identification of stress group onsets is the subject of Kohler (1983b). On this line of reasoning initial consonants in the succeeding stress group should be included in the stress group under scrutiny. And so they were in the initial analyses, to the extent that they were identical across regions; but see the next section.

Results

It is reasonable to assume that it will be *stressed* vowel (or syllable) duration relative to one or more immediately succeeding vowels (or syllables) that determines our perception of foot internal rhythm. And perhaps we should expect only temporal relations within the stressed and the first post-tonic syllable to have rhythmical relevance, cf. Ohala (1975, p. 452) who – in a discussion of the temporal *control* of speech – observes: "Typically it is ... at most two adjacent syllables whose timing characteristics have *perceptual* import." [emphasis mine] He takes this as support of a contention that at least a short-term temporal precision is required in speech production, "... whereas there is no corresponding evidence that a speaker needs to maintain long-term temporal precision in his speech ...". Such evidence is also lacking in this material. Thus, e.g., when the first ˈV(:) and C(C) durations in the stress group were expressed in percent of the *total* stress group duration, including all post-tonic syllables, differences across regions in relative stressed vowel and post-vocalic consonant durations were indistinct and did not follow the same pattern as when only the stressed and first post-tonic syllables were considered as the basis against which relative ˈV(:) and succeeding C(C) durations were calculated. It also appeared that distinctions blurred when the consonant(s)

succeeding the first *post-tonic* vowel were included in the calculations and graphs. Thus, the search for durational parameters that would successfully and quasi-consistently differentiate the speech samples from the six regions was successful only for the segments within the ˈV(:)C(C)$_{\circ}$V sequence, i.e. the ˈV(:)/C(C) and ˈV(:)/$_{\circ}$V ratios in the 16 stress groups. In the remaining part of the paper, "V' will stand for both long and short stressed vowels, and 'C' will cover post-vocalic C as well as CC, unless otherwise specified by the text.

There appeared a further limitation, however. In the stress groups '... frem til ..., ... fyldt til ..., ... fyldt med ..., Hvem skal til ..., Hvor skal de ...', once more, distinctions between regions in vowel and consonant duration ratios, which are rather clear otherwise, do not appear. And, furthermore, a pair of speakers from the same region (e.g. Aalborg), who otherwise have near identical relative short stressed vowel durations for a given type, diverge in some of these problematic cases. There may be several reasons for this. Some regions/speakers have stød (in casu in 'frem, fyldt, hvem, hvor') (Copenhagen, Næstved, Aalborg – and it is decidedly strong phonetically in Næstved but weak in Aalborg) others do not (Bornholm, Sønderborg), and one Tønder speaker did (although erratically) and the other did not produce stød. Stød influences segment duration, but perhaps differently across regions (this is a matter which has not been investigated previously). Thus, in Standard Copenhagen Danish, a (phonologically) long vowel with stød will have a duration about half-way between a short vowel and a long stød-less vowel (Fischer-Jørgensen 1955), and consonants are likewise affected. (In the remaining 11 stress groups no stød occurs anywhere.) But the stød or lack of it may not be the only reason. The stressed syllables in the five problematic stress groups are all monosyllabic words, and it is possible that the two syllables to form a rhythmical unit have not only to be adjacent but must belong to the same word. If this is so, then it has interesting, if complicated, implications for the prosody/syntax interface. Because, on the one hand, where Fo patterning is concerned, the prosodic stress group *is* an autonomous unit, i.e. Fo patterns pertaining to a stressed and succeeding unstressed syllables, cut across any sentence internal syntactic boundary, and *initial* consonants in the stressed syllable team up with the preceding stress group, both as far as Fo *and* duration goes, cf. above. But, on the other hand, we may have to establish that (perceptually relevant) durational *relations* between successions of syllables within the stress group are sensitive to word boundaries. This is very clearly a subject matter to be further pursued.

Having to leave out the five monosyllables, 11 ˈVC$_{\circ}$V structures remain to be further examined. Five of these ('Kamma(2x), Ander(s), stammer, kander(ne)') could be averaged without obscuring significant differences, as could the two times 'Torben(s)', and the two times 'søster', whereas 'kaner(ne)' is presented separately. 'kanden' had to be left out, because of its rendering with a nasal diphthong and deleted /n/ in Bornholm, and because the pre-positioned definite article in Tønder and Sønderborg ('a kande' [ə ˈg̊ʰan̩]) produces a shorter final nasal than does 'kanden' ([ˈg̊ʰann̩]).

Average durations of segments in percent of the total ˈVC̥V sequence (expressed in msec rounded off to the nearest integer) in each of these four stress group types are given in Tables 1-4. Not every subject is represented in every table/figure, partly due to unfelicitous renderings (hesitations and pauses), but mainly due to problems of segmentation in individual cases, and subjects are only represented if they produced at least 3 repetitions for averaging of a given type.

Table I

Relative durations (in percent of the total) of the short stressed vowel, the intervocalic nasal and the post-tonic vowel, averaged over 3-6 recordings of 2-5 (indicated in raised numerals after the speaker identification) ˈVN̥V sequences. The average of the total duration in msec is given to the right, and average percentages at the bottom. N = Næstved, B = Bornholm, C = Copenhagen, A = Aalborg, T = Tønder, S = Sønderborg. Starred speakers are 'outliers' and are excluded from the calculation of averages and the correlation coefficients shown in the figures.

Kamma/Anders/stammer/kander(ne)				
	ɑ	m/n	a/ɐ	total
N1[5]	49.8	19.8	30.4	251
N2[2]	43.2	20.4	36.4	224
B1[3]	42.7	23.4	33.9	319
B2[2]	46.3	24.5	29.2	286
C1[4]	44.6	21.9	33.5	224
C2[5]	41.5	23.1	35.4	222
A1[5]	42.4	23.9	33.7	304
A2[4]	41.2	26.3	32.5	216
T1[5]	39.3	29.3	31.4	236
T2[3]*	39.8	33.5	26.7	225
S1[5]	34.1	29.0	36.9	213
S2[4]	34.5	29.4	36.1	210
average	41.8	24.6	33.6	

Table II

Relative durations (in percent of the total) of the short stressed vowel, the intervocalic obstruent cluster and the post-tonic vowel, averaged over 3-6 recordings of 1 or 2 (indicated in raised numerals after the speaker identification) 'VCC$_\circ$V sequences. See further the caption to Table 1.

søster				
	ø	sd̥	ɐ	total
N1^{2}	43.3	41.4	15.3	289
N2^{2}	42.4	37.7	19.9	264
B1^{1}	36.2	41.3	22.5	298
B2^{2}	34.6	46.0	19.4	289
C1^{2}	37.5	43.8	18.7	240
C2^{2}	30.2	48.3	21.5	232
A1^{2}	37.2	42.6	20.2	323
A2^{2}*	35.6	52.2	12.2	246
T1^{1}	38.1	45.5	16.4	257
T2^{2}*	28.4	59.4	12.2	247
S1^{2}	29.3	46.6	24.1	285
S2^{2}	29.8	50.2	20.0	275
average	35.9	44.3	19.8	

Table III

Relative durations (in percent of the total) of the long stressed vowel, the intervocalic nasal and the post-tonic vowel in a 'V:C$_\circ$V sequence, averaged over 3-6 recordings. See further the caption to Table 1.

kaner(ne)				
	æ:	n	ɐ	total
N1	54.9	16.7	28.4	288
B1	60.5	15.9	23.6	347
B2	63.8	15.7	20.5	332
C1	58.2	15.2	26.6	263
C2*	50.8	13.2	36.0	250
A1	59.1	17.6	23.3	330
A2	55.6	19.4	25.0	248
T1	62.7	16.4	20.9	287
T2*	63.9	20.1	16.0	313
S1	53.8	18.9	27.3	264
average	58.6	17.0	24.5	

Table IV

Relative durations (in percent of the total) of the long stressed vowel, the intervocalic stop and the post-tonic syllabic nasal, averaged over 3-6 recordings of 1 or 2 (indicated in raised numerals after the speaker identification) 'V:C$_o$ (V)N sequences. See further the caption to Table 1.

Torben(s)				
	ɒ:	b̥	ŋ̍/ən	total
N1^{2}*	55.1	10.1	34.8	288
N2^{2}	56.3	12.8	30.9	258
B1^{2}	47.7	19.7	32.6	408
B2^{2}	44.7	22.5	32.8	362
C1^{2}	47.1	23.6	29.3	290
C2^{2}	48.5	19.2	32.3	268
A1^{2}*	56.1	24.1	19.8	342
A2^{2}	51.7	19.5	28.8	266
T1^{2}	54.6	18.9	26.5	302
T2^{2}	54.0	15.8	30.2	285
S1^{2}	45.6	21.2	33.2	264
S2^{2}	52.4	14.7	32.9	289
average	50.3	18.8	31.0	

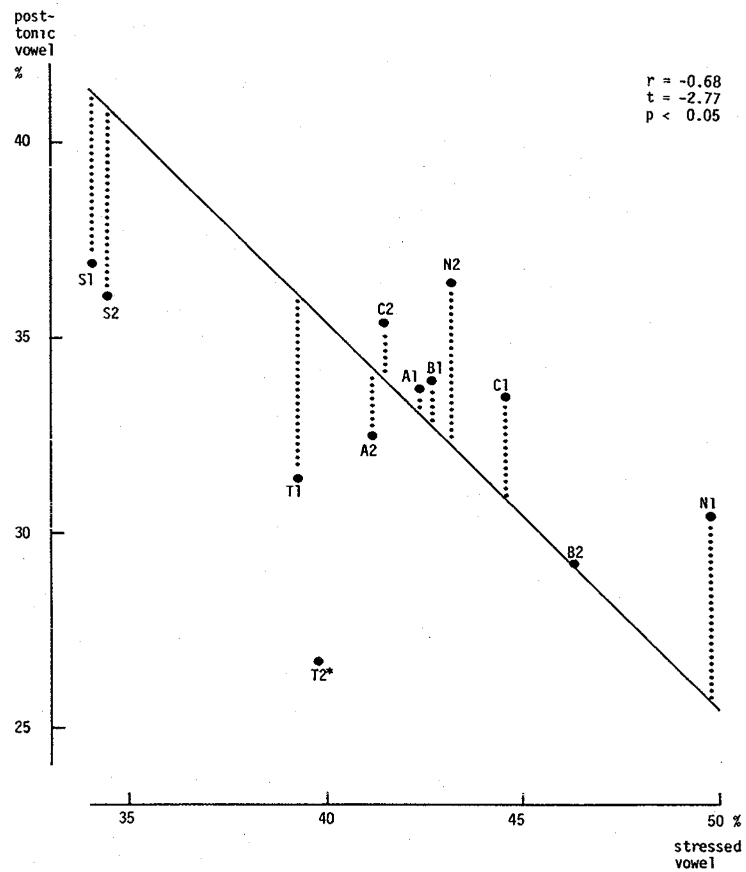

Figure 1a. *Plot of the short stressed and post-tonic vowel durations listed in Table 1. The slant is drawn with a slope of -1 through the average 'V and V̥ values. The vertical distance from each data point to the slant equals its difference (to be read off the y axis) from the mean intervocalic consonant duration. Points above and below the diagonal represent negative and positive differences, respectively. Starred data points are 'outliers'. The Pearson product moment correlation coefficient, with associated t-value and probability of significance – if below 0.05 – are given in the upper right of the figure.*

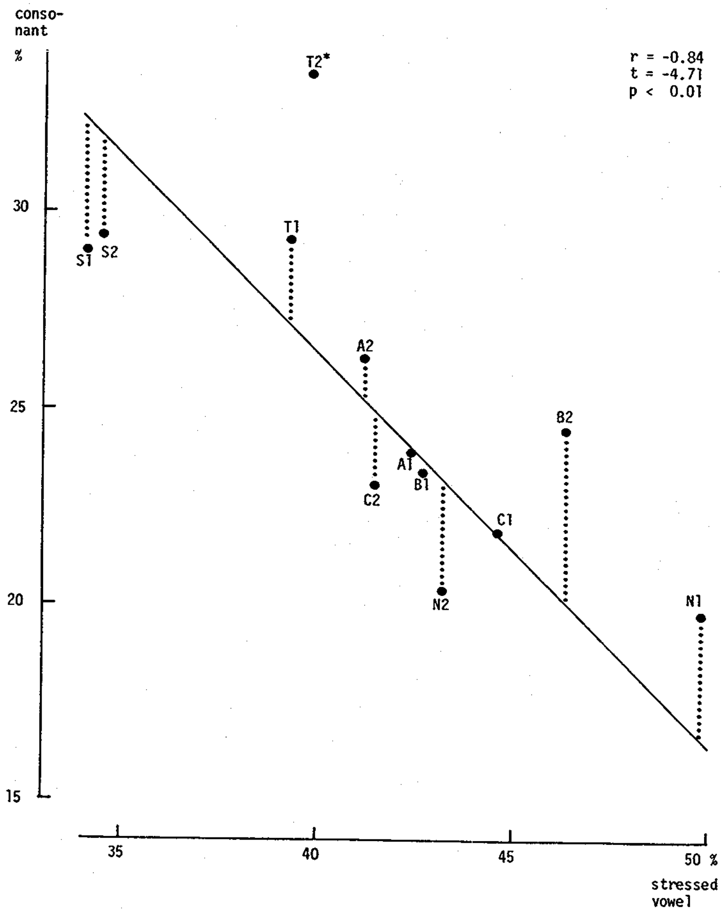

Figure 1b. *Plot of the short stressed vowel and intervocalic nasal durations listed in Table 1. The diagonal is drawn with a slope of -1 through the average V and C values. The vertical distance from each data point to the slant equals its difference (to be read off the y axis) from the mean post-tonic vowel duration. See further the legend to Fig. 1a.*

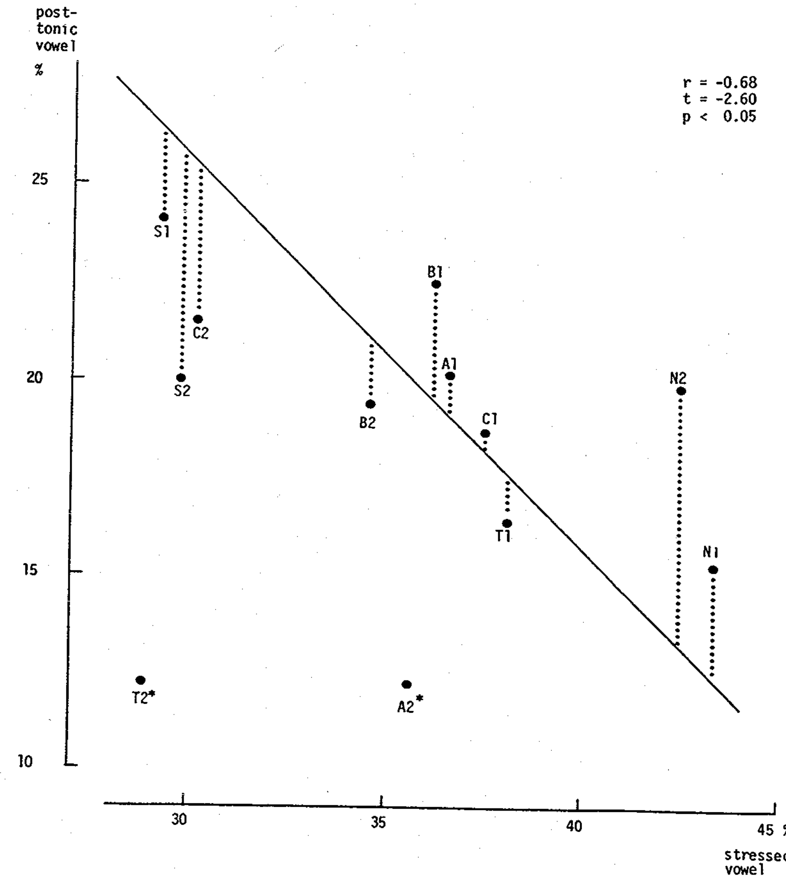

Figure 2a. *Plot of the short stressed and post-tonic vowel durations listed in Table II. See further the legend to Fig. 1a.*

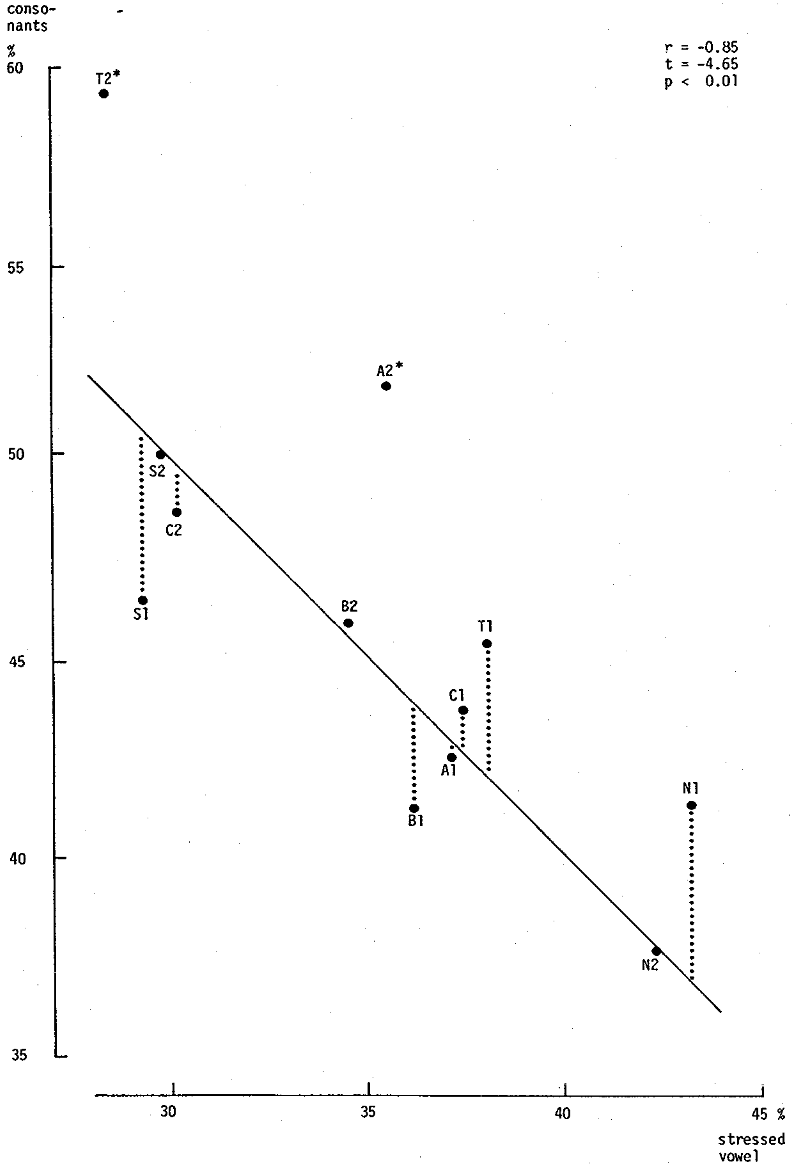

Figure 2b. *Plot of the short stressed vowel and intervocalic obstruent cluster durations listed in Table II. Se further the legends to Figs. 1a and 1b.*

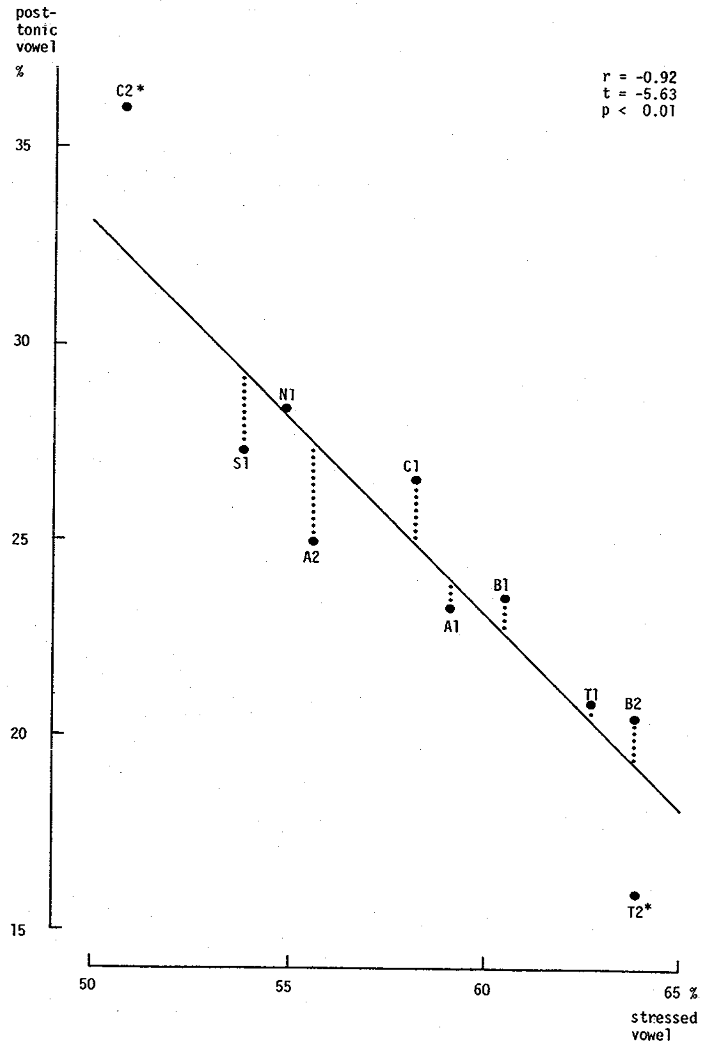

Figure 3a. *Plot of the long stressed and post-tonic vowel durations listed in Table III. See further the legend to Fig. 1a.*

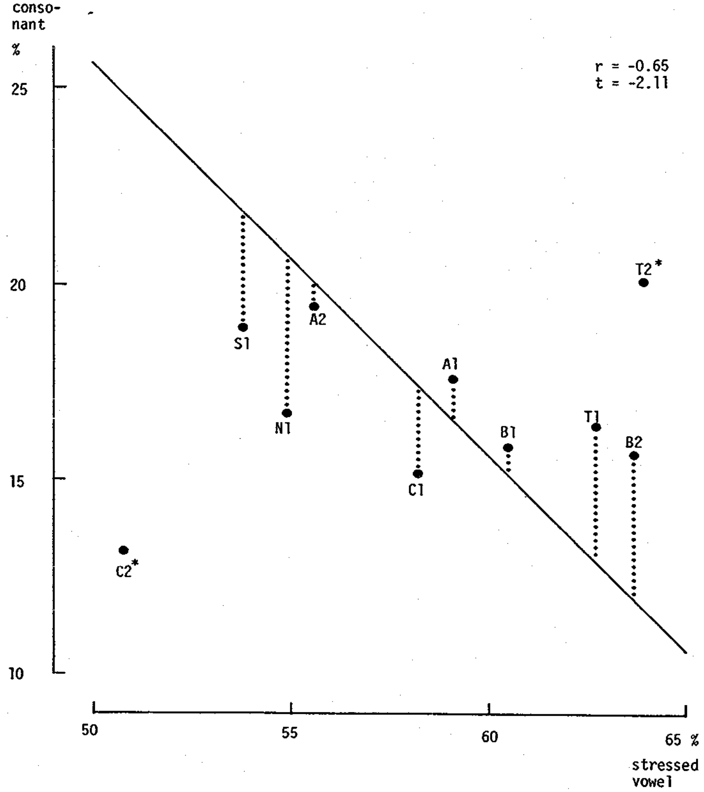

Figure 3b. *Plot of the long stressed vowel and intervocalic nasal durations listed in Table III. See further the legends to Figs. 1a and 1b.*

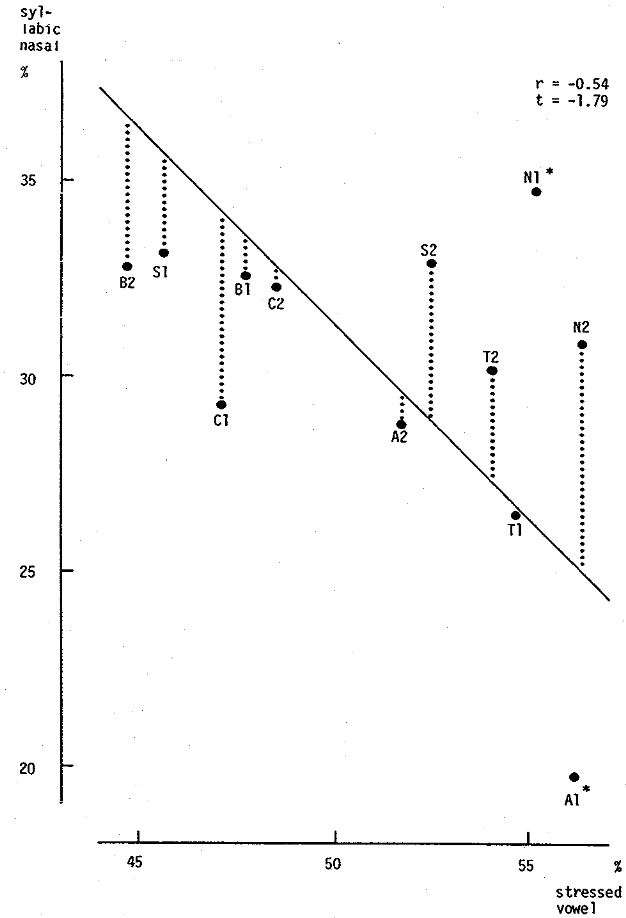

Figure 4a. *Plot of the long stressed vowel and post-tonic syllabic nasal durations listed in Table IV. See further the legend to Fig. 1a.*

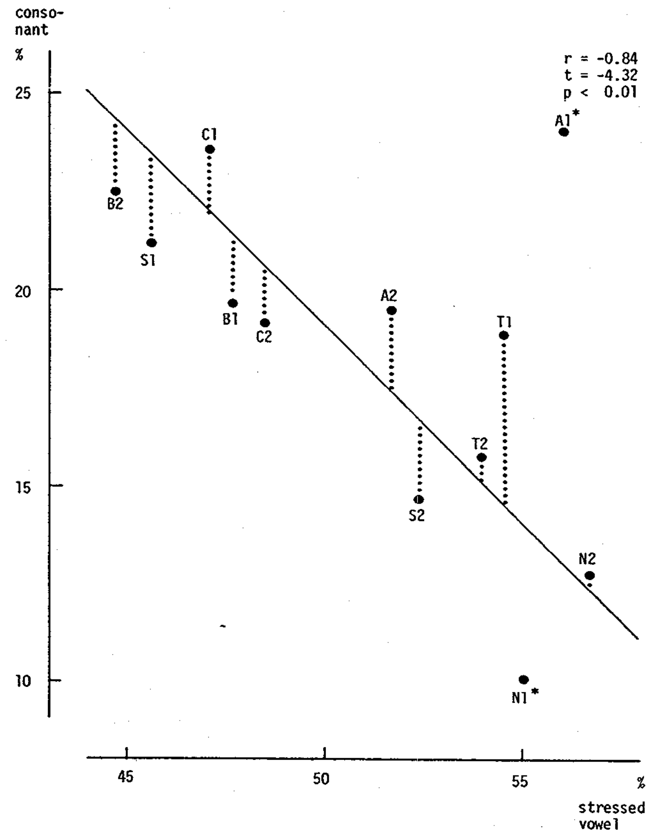

Figure 4b. *Plot of the long stressed vowel and intervocalic stop consonant durations listed in Table IV. See further the legends to Figs. 1a and 1b.*

The results are presented graphically in Figs. 1-4. They display stressed vowel duration against post-tonic vowel duration (Figs. 1a-4a) and consonant duration (Figs. 1b-4b), respectively. The slant through the graphs is drawn with a slope of -1 through the point which represents the average percentage of the x and y variables, excluding the starred 'outliers', cf. the tables. It is the fictional line along which data points would disperse if only x and y varied, which is of course not the case: Thus, points above the slant represent (a) consonant or (b) post-tonic vowel durations below the average, and points below the slant represent corresponding durations above the average. The vertical distance of each point to the diagonal indicates the deviation (to be read off the y axis) from the mean intervocalic consonant or post-tonic vowel duration, respectively. To take an example from Fig. 1a: N1 represents a long stressed vowel (49.8% – the average is 41.8%), a below average post-tonic vowel (30.4% – the average is 33.6%) and an intervocalic nasal which is shorter than the average (24.6%) by 4.8%. At the other extreme S1 has a short stressed vowel (34.1%), a long post-tonic (36.9%) and an intervocalic nasal which is longer than the average by 4.4%. Thus the longer stressed vowel with N1 is balanced off by a shorter consonant and post-tonic vowel; and the shorter stressed vowel with S1 is compensated for by a longer C and $_{\circ}$V. The consonant seems to carry slightly more of the compensation than does the post-tonic vowel. Note also the difference between N1 and T1: they differ in stressed vowel duration (by 10.5%), not much in post-tonic vowel duration (1.0%), and thus the difference in 'V is compensated for mainly by the intervocalic nasal duration which is considerably longer in T1 (by 9.5%). The (a) and (b) versions of the figures present the same information, of course, but from two different perspectives, and they are both reproduced here, because it becomes so immediately apparent which correlations in the data are the tightest. Cf. also the Pearson product moment correlation coefficients given in the upper right of each frame.

There is some resemblance between Figs. 1 and 2 (the short stressed vowels), but not much of a repetition of the pattern of plots in Figs. 3 and 4 (the long stressed vowels), and it will be expedient to treat long and short stressed vowels separately. Note that the rather confusing durational data presented will be treated somewhat cursorily, due to the final conclusion that duration cannot be made primarily responsible for listeners' perception of rhythm in the six varieties anyway.

Short stressed vowel (Figs. 1-2). 'V duration below the average is associated (though not quite monotonously so) with C duration above the average, and vice versa (1a-2a), whereas 'V and $_{\circ}$V durations are not as straightforwardly correlated (cf. the data points *above* the slant to the left of the mean 'V duration in 1b-2b). Calculation of the 'V~C, 'V~$_{\circ}$V, and C~$_{\circ}$V correlations for each speaker separately across the values in the four tables also comes out with unanimously the highest values for 'V~C, occasionally high also for C~$_{\circ}$V (N1, N2, C1, C2, A2, S2), but nil for 'V~$_{\circ}$V, except for one speaker (C2). The closer correlation between 'V and C would indicate that the primary trade-off between segments in the 'VC$_{\circ}$V sequence is between the stressed vowel and the succeeding consonant(s), and that variation in the post-tonic vowel falls out roughly from that. This could be taken to indicate that the relevant boundary in a 'V(:)C$_{\circ}$V/'V(:)CC$_{\circ}$V sequence is 'V(:)C-$_{\circ}$V/'V(:)CC-$_{\circ}$V rather than 'V(:)-C$_{\circ}$V/'V(:)-CC$_{\circ}$V/'V(:)C-C$_{\circ}$V. I.e. the relevant boundary prosodically in the post-tonic syllable is immediately before the vowel, by analogy with the case for the stressed syllable, cf. above. That is what I alluded to at the end of the introduction: a boundary thus located may coincide, but need not do so, with phonological syllable boundaries, and it is certainly at variance with what is traditionally conceived of as phonetic syllable boundaries.

With a tri-partition of the range of variation in the three parameters in Figs. 1 and 2, excluding 'outliers', and corresponding coarse denominations in terms of 'long', 'short' and intermediate ("–"), we achieve the classification below, where designations to the left and right of the slash are from Figs. 1 and 2, respectively. Stars denote 'outliers' in the figures:

	'V	C	$_{\circ}$V
N1	long/long	short/short	short/short
N2	–/long	short/short	long/–
B1	–/–	–/short	–/long
B2	long/–	–/–	short/–
C1	long/–	short/–	–/–
C2	–/short	–/long	long/long
A1	–/–	–/–	–/–
A2	–/–*	long/long*	–/short*
T1	–/–	long/–	short/short
T2	–*/short*	long*/long*	short*/short*
S1	short/short	long/long	long/long
S2	short/short	long/long	long/–

With a certain leniency in the interpretation – necessitated by the lack of strict uniformity in the data – we can conclude that only three regions are really distinct with respect to relative 'V, C and $_{\circ}$V durations, i.e. Næstved, Tønder and Sønderborg, as summarized below:

	'V	C	$_{\circ}$V
N	long	short	–
T	–	long	short
S	short	long	long

I.e. Næstved is distinct from Tønder and Sønderborg by the long stressed vowel and the short succeeding consonant(s). Tønder and Sønderborg are distinguished by relative stressed vowel and most pronouncedly by post-tonic vowel durations.

Long stressed vowel (Figs. 3-4). The long stressed vowels yield a somewhat different result: First of all, the distribution along the x axis, the stressed vowel duration, is much more even than with the short vowels. Secondly, the range of variation in the dental nasal consonant (Fig. 3) is smaller than in the other consonants, and its correlation with stressed vowel duration (still disregarding the 'outliers', i.e. C2 and T2), is poor (r = -0.65), whereas the 'V:~$_{\circ}$V correlation is high (r = -0.92) in 'kaner(ne)'. Thirdly, there are speakers (N1, B1, B2, C1, T2) who take opposite values for the 'same' segment in the two figures.

	'V	C	$_{\circ}$V
N1	short/long*	–/short*	long/long*
N2	/long	/short	/long
B1	long/short	short/–	–/long
B2	long/short	short/long	short/long
C1	–/short	short/long	long/–
C2	short*/short	short*/–	long*/long
A1	–/long*	–/long*	–/short*
A2	short/–	long/–	–/–
T1	long/long	short/–	short/short
T2	long*/long	long*/short	short*/–
S1	short/short	long/long	long/long
S2	/–	/short	/long

Relative to the short stressed vowels in Figs. 1-2, Copenhagen seems to acquire some distinction, i.e. to deviate from the mean values, and in Næstved, Tønder and Sønderborg durational realtions between vowels and consonants have changed, as summarized below:

	ˈV	C	$_\circ$V
C	short	–	long
N	–	short	long
T	long	short	short
S	short	–	long

The distinctly long stressed vowels are found now with Tønder, and Sønderborg and Tønder are different here also in relative consonant duration.

If short and long stressed vowels do not yield the same pattern of relations between regions, it may be because different regions have different manifestations of the phonological long:short vowel difference, i.e. different regions lengthen long vowels to different degrees relative to the short vowels, ceteris paribus. This, again, is a matter for further inquiry, but the present data seem to indicate that Tønder would have the highest and Copenhagen and Næstved the smallest long:short vowel ratio. It is perhaps also debatable whether the (phonetically) long vowels in 'kaner(ne)' and 'Torben' should be treated on a par with each other. 'kanerne' has a phonologically long vowel, /a:/, but [ɒ:] in 'Torben' derives from /ɔʁ/, i.e. a phonologically short vowel which merges phonetically with the succeeding pharyngeal /ʁ/, and it seems as if Bornholm, and perhaps also Næstved, speakers treat them differently.

It is hard to draw a unified conclusion from the durational data, partly because speakers from the same region do not invariably cluster in the figures, partly because some seem to behave erratically (the 'outliers' in the figures, notably T2), and partly because short and long stressed vowels do not pattern alike. Nevertheless, it does not stretch the data too far to say that Næstved, Tønder and Sønderborg do seem to establish different types, diverging from the average in their specific ways, in the ˈV~C~$_\circ$V relations, whereas Bornholm, Copenhagen and Aalborg generally cluster around the average durational relations. If durational relations in the stressed and first post-tonic syllable have any role in the perception of foot-internal rhythm, then we should expect Næstved, Tønder and Sønderborg to be judged differently among themselves and from the rest.

Auditory evaluation

Four colleagues in the department were supplied with tapes which contained 18 different complete utterances by each of two speakers from each region, rendered speaker by speaker, region by region. (They were given complete utterances on the assumption that perceived rhythmical differences between

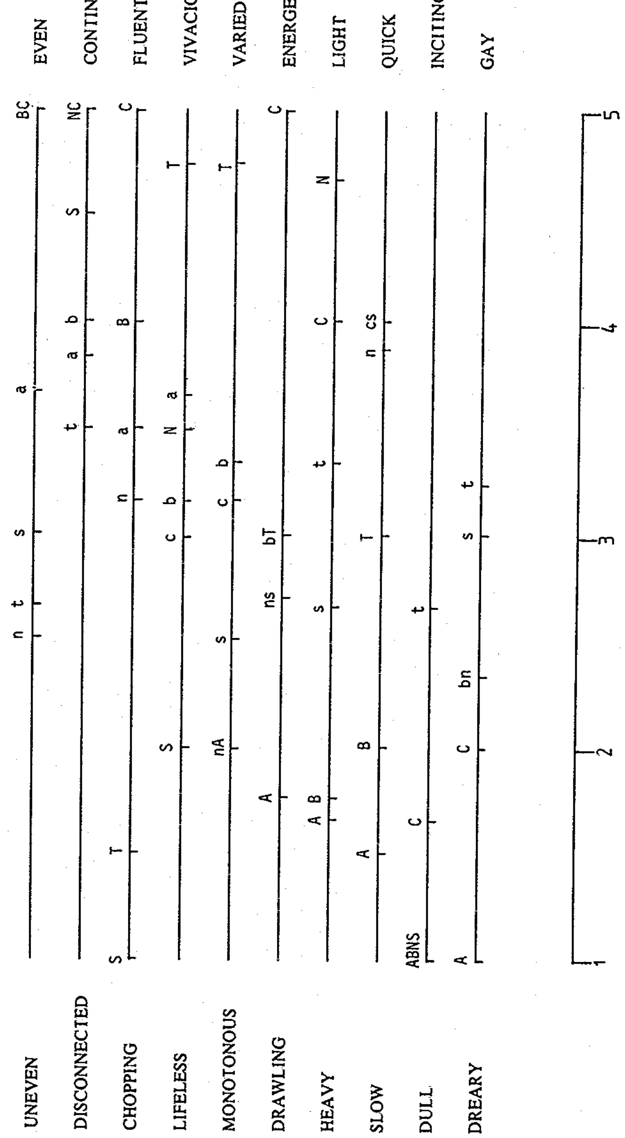

Figure 5 Auditory evaluation of foot-internal rhythm, in terms of 10 antonymous adjective pairs, in six Danish varieties: Aalborg (A, a), Bornholm (B, b), Copenhagen (C, c), Næstved (N, n) Sønderborg (S, s), and Tønder (T, t) Capital letters represent averages on the five-point scale over near-unanimous scores by four listeners, lower case letters represent averages of non-unanimous scores See further the text

different regions will cut across any type of stress group in any – except perhaps final – position. And, further, that measurable differences in any type of directly comparable structure will be representative of differences under other segmental conditions.) The listeners were asked to evaluate the foot internal rhythm of each region in terms of 11 pairs of antonymous adjectives: even-uneven, continuous-disconnected, fluent-chopping, vivacious-lifeless, varied-monotonous, energetic-drawling, light-heavy, quick-slow, inciting-dull, gay-dreary, syncopated-staccato, the last one of which nobody seemed able to apply. Both poles of each pair were listed, and subjects could mark one or the other (or both) with "+" or "–" for each region. More or less of a given quality could be indicated with question marks. No marking would indicate neutrality. Subjects complained afterwards that the task had been extremely difficult, which is reflected in a very poor inter-subject agreement in a large number of cases.

"+", "+?", –, "-?" and "-" for a given 'positive' adjective (or "-", "-?", –, "+?" and "+" for its 'negative' antonym) was translated into a five-point scale (+ = 5, +? = 4, – = 3, -? = 2, - = 1) and averaged over the four listeners. Provided that the four listeners did not diverge by more than two points on a given adjective, averages are presented in capital letters in Fig. 5. Lower case letters represent the remaining non-unanimous scores.

Considering only the near unanimous scores (capital letters) and collapsing scores in the intervals 1-2, 2-4, 4-5, Fig. 5 may be summarized thus:

UNEVEN			BC	EVEN
DISCONNECTED			SNC	CONTINUOUS
CHOPPING	ST		BC	FLUENT
LIFELESS	S	N	T	VIVACIOUS
MONOTONOUS	A		T	VARIED
DRAWLING	A	T	C	ENERGETIC
HEAVY	AB		CN	LIGHT
SLOW	AB	T		QUICK
DULL	ABNSC			INCITING
DREARY	AC			GAY

I had some expectations from the outset:

(1) Scores would be similar within each of three groups of adjectives (which I conceive of as being near synonymous in this context): (I) EVEN, CONTINUOUS, FLUENT; (II) VIVACIOUS, VARIED, ENERGETIC; (III) LIGHT, QUICK, INCITING, GAY. That is flatly denied.

(2) Since ˈV and C are generally highly correlated in production, the ˈV/C ratio would be the decisive factor for the perception of foot-internal rhythm, and thus Næstved (high ˈV/C) and Sønderborg (low ˈV/C) should, as mentioned above, be distinguished among themselves and from the rest, in either of two ways

(2a) The higher the ˈV/C ratio (Næstved), i.e. the longer the 'beat' (the stressed vowel), the higher the score on EVEN, CONTINUOUS, FLUENT (group I) and the lower the score on the remaining adjectives, and vice versa (Sønderborg). *Or*

(2b) ˈV/C ratios around the average (Bornholm, Copenhagen, Aalborg) would score high on group I, and low on the rest, whereas extreme ratios would score low on group I; high (Næstved) and low (Sønderborg) ratios would score low and high, respectively, on group II and III.

(2b) seems to be the better prediction for average ˈV/C ratios: Bornholm and Copenhagen (but not Aalborg) score high on group I, and Aalborg (and Bornholm) score(s) low on group II and III. But it fails for the extreme ˈV/C ratios. Thus, none of my predictions are consistently borne out, nor do I see any other clear correlations between the durational data and the rhythmical evaluations.

Listeners are apparently not reacting to durational relations per se to any great or consistent extent. (However, they were *not* asked to judge *durations*, and I cannot really know how they understood 'foot internal rhythm'; nor how, exactly, they interpreted the adjectives.) The other obvious candidate then is Fo/pitch which can in fact – perhaps together with duration – explain some of the auditory evaluations.

Stress group patterning in the six varieties

Four of the six samples share most intonational properties (Copenhagen, Næstved, Aalborg, Tønder), whereas Sønderborg and Bornholm stand out with respect to sentence intonation signalling, final lowering, completion cues and (where Bornholm is concerned) also sentence accentuation (Thorsen 1988a, 1988b, Grønnum (Thorsen) 1989, Grønnum 1990, 1991, 1992). However, one feature ensures a differentiation across all six regions and may well be the one primarily responsible for listeners' ready identification of speakers from these regions: the Fo pattern associated with the prosodic stress group. Fig. 6 depicts stylized model Fo patterns of the six variants. There are differences in shape, in location in the overall range, in range spanned by the pattern, in rise and fall slopes, and Bornholm expands/compresses Fo patterns

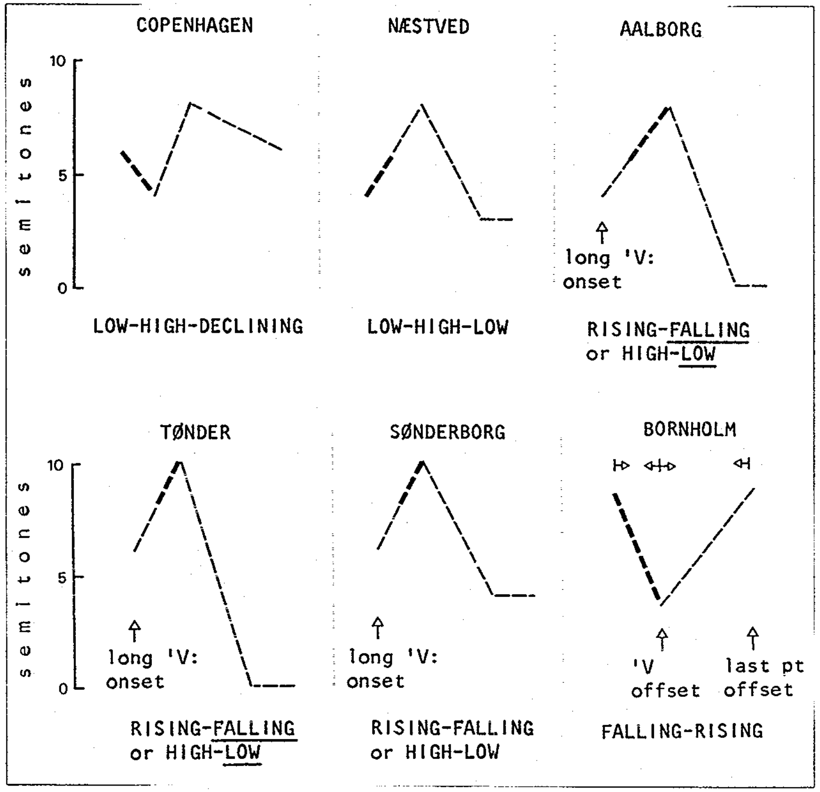

Figure 6. *Model stress group patterns from the six varieties investigated. Heavy lines denote stressed vowels. Horizontal arrows delimit movements which may be expanded or compressed in time. Vertical arrows depict turning points which anchor segmental events. The patterns are characterized auditorily beneath each frame.*

in unison with foot duration, rather than simply truncating or completing, respectively, an otherwise invariant pattern shape.

Since the impression of foot-internal rhythmical differences cannot be due to objective and measurable differences in 'V~C~$_0$V durations in the stress group, the explanation must be sought elsewhere. Perceived rhythm may well be the result of an *interplay* between segment duration and its associated Fo movement, granted that Fo movement (level versus dynamic, rising versus falling, extensive versus less extensive Fo range) may make segments appear different in duration, ceteris paribus, cf. Lehiste (1976) who found that listeners

judged segments with dynamic Fo to be longer than static Fo segments. It has also been shown that rising tones take more time to produce than falling ones, and for rising tones – but not for falling ones – their duration is longer the larger the range covered, cf. Hombert (1977), Sundberg (1973). If this intrinsic difference in duration is compensated for in perception, then – for tones of equal physical duration – rising Fo movements will appear to be shorter than falling ones, and more extensive rises will appear to be shorter than less extensive rises. Cf. also that Carlson et al. (1979) hint that a mismatch in synthetic speech between the durational and Fo relations characterizing normal speech may give rise to unexpected perceptual preferences for segment durations.

However, perceived rhythm may not even be any direct correlate of 'pitch corrected' durations. I do not know how exactly to quantify the influence from Fo on perceived duration in this material. But if pitch corrected durations were a major factor, I would at least expect that Bornholm and Aalborg, which cluster around average *acoustic* 'V~C~$_0$V relations, would be distinguished *auditorily*, granted their directly opposite Fo patterns, cf. Fig. 6: Bornholm should lean more towards a long~short~short, and Aalborg more towards a short~long~long type. But they are not really distinguished rhythmically, in fact they are both judged equally HEAVY, SLOW, DULL. Correcting for pitch would not greatly alter the distinct acoustic differences in relative segment durations between Næstved, Tønder and Sønderborg, since they share a rising-falling Fo pattern, and thus would not bring us any nearer a reconciliation between (differences in) durational data and (similarities in) rhythmical evaluation.

If duration – whether acoustic or perceived – taken separately fails, there really is (among the easily quantifiable candidates) only pitch patterning left to appeal to for an explanation of listeners' evaluation of foot-internal rhythm. (But I do also suspect that much less easily quantifiable, speaker-idiosyncratic factors come into play, like voice quality and tempo.) So, on the above assumptions, let me attempt an outline of an explanation of the more pertinent facts. But note that no single aspect can unambiguously account for the scores in Fig. 5. And note also that Fo patterns per se cannot account for rhythmical evaluations. That should make Tønder (with its quick rise-fall) and Bornholm (with its slow fall-rise) clearly distinguished in Fig. 5, which they are not: Perceived 'V~C~$_0$V durations *and* the *total* Fo pattern are probably evaluated holistically, and thus there will of necessity be apparent contradictions in the statements to follow:

Tønder is CHOPPING, VIVACIOUS and VARIED due to its very lively and wide-ranged Fo pattern. Thus, the rise is steep and the 'low' is generally reached within the first post-tonic syllable, which makes the Fo patterns sound like repeated sequences of 'high-low- ... -low'. The short $_{\circ}$V may also be responsible. But Tønder is neither ENERGETIC nor QUICK perhaps for its intermediate or long V.

Sønderborg is CONTINUOUS, LIFELESS and DULL perhaps for its long $_{\circ}$V, but CHOPPING for its steep rise in the stressed vowel, which makes it appear auditorily as 'high' rather than 'rising'.

Næstved is CONTINUOUS and DULL both for its long V, the limited Fo range and the modest slopes, and perhaps it is LIGHT because – like Copenhagen – the first post-tonic vowel is near the peak of the pattern.

Copenhagen is EVEN, CONTINUOUS, FLUENT both for its intermediate segment durations, its limited Fo range and the slow fall through the post-tonics; ENERGETIC due to the fact that the two significant events, the Fo turning points, are both located comparatively early in the sequence; LIGHT because the first post-tonic vowel is at the peak of the pattern; but DULL and DREARY for its initial fall, the limited range and the modest slope of the fall.

Bornholm is EVEN and FLUENT due both to the intermediate segment durations and to the frequency constant 'high-low-high', but HEAVY, SLOW and DULL by virtue of its not very large range, thus rather slow movements, and particularly its pronounced initial fall.

Aalborg may be MONOTONOUS, DRAWLING, HEAVY, SLOW, DULL and DREARY for both its intermediate segment durations and its slow Fo movements, particularly the rise, and perhaps also because the *slow* fall lands at the bottom of the range (in contradistinction to the quick fall of Tønder).

Final remarks

In and of itself this little experiment naturally does not tell us very much besides the fact that perceived rhythm is not directly correlated to durational relations in the foot. Nor does it pretend, of course, to exhaust questions of timing in the six varieties. Obviously, other phonological structures would have to be included. But the results raise a host of questions – with obvious implications for speech synthesis – which can only be answered through a multitude of experiments with synthetic manipulation of duration and Fo. They would demonstrate the requirements for appropriate realizations of stress group patterns (in terms of magnitude, slopes and timing of Fo events relative to the segmental structure) in conjunction with the proper durational relations, for any given variant.

Unfortunately, Win van Dommenlen's paper "Does dynamic Fo increase perceived duration? New light on an old issue" in *Journal of Phonetics (1993) 21*, 367-386, only arrived after this manuscript went to the printer's, and I have not been able to make use of his very interesting results.

References

Allen, G.D. (1975) "Speech rhythm: its relation to performance universals and articulatory timing", *Journal of Phonetics 3*, pp. 75-86.

Bruce, G. (1984) "Rhythmic alternation in Swedish", in: *Nordic Prosody III* (Umeå Studies in the Humanities 59, eds.: C.-C. Elert, I. Johansson, E. Strangert), Stockholm, Almqvist & Wiksell, pp. 31-41.

Carlson, R., Granström, B. and Klatt, D.H. (1979) "Some notes on the perception of temporal patterns in speech" *Proceedings of the Ninth International Congress of Phonetic Sciences, Copenhagen 1979*, vol. II, pp. 260-267.

Dauer, R.M. (1983) "Stress-timing and syllable-timing reanalyzed", *Journal of Phonetics 11*, pp. 51-62.

Dauer, R.M. (1987) "Phonetic and phonologic components of language rhythm", *Proceedings of the Eleventh International Congress of Phonetic Sciences, Tallinn 1987*, vol. 5, pp. 447-450.

Eriksson, A. (1991) "Aspects of Swedish speech rhythm", *Gothenburg Monographs in Linguistics 9*, Department of Linguistics, University of Göteborg.

Fant, G. and Kruckenberg, A. (1989) "Preliminaries to the study of Swedish prose reading and reading style", *Speech Transmission Laboratory, Quarterly Progress and Status Report* 2/1989, Royal Institute of Technology, Stockholm, pp. 1-83.

Fischer-Jørgensen, E. (1955) "Om vokallængde i dansk rigsmål", *Nordisk Tidsskrift for Tale og Stemme* 15, pp. 28-48.

Fischer-Jørgensen, E. (1982) "Segment duration in Danish words in dependency on higher level phonological units", *Annual Report of the Institute of Phonetics, University of Copenhagen* 16, pp. 137-189.

Fowler, C.A. (1979) "'Perceptual Centers' in speech production and perception", *Perception and Psychophysics* 25, pp. 374-388.

Grønnum (Thorsen), N. (1989) "Stress group patterns, sentence accents and sentence intonation in southern Jutland (Sønderborg and Tønder) – with a view to German", *Annual Report of the Institute of Phonetics, University of Copenhagen* 23, pp. 1-85.

Grønnum, N. (1990) "Prosodic features in regional Danish – with a view to Swedish and German, in: *Nordic Prosody V* (eds.: Wiik, K. and Raimo, I.). University of Turku, pp. 131-144.

Grønnum, N. (1991) "Prosodic parameters in a variety of regional Danish standard languages, with a view towards Swedish and German", *Phonetica 47*, pp. 182-214.

Grønnum, N. (1992) *The Groundworks of Danish Intonation*, Museum Tusculanum Press, Copenhagen.

Grønnum, N. (in print) "Standard (Copenhagen) Danish intonation", in: *Intonation Systems: A Survey of Twenty Languages*, (eds.: Hirst, D. and di Cristo, A.). Cambridge: Cambridge University Press.

Hoequist, C. Jr. (1983) "Durational correlates of linguistic rhythm categories", *Phonetica 40*, pp. 19-31.

Hombert, J.-M. (1977) "Difficulty of producing different Fo in speech", *Working Papers, University of California at Los Angeles 36*, pp. 12-19.

Huggins, A.W. (1972) "On the perception of temporal phenomena in speech", *Journal of the Acoustical Society of America 51*, pp. 1279-1290.

Klatt, D.H. (1976) "Linguistic uses of segmental duration in English: acoustic and perceptual evidence", *Journal of the Acoustical Society of America 59*, pp. 1208-1222.

Klatt, D.H. and Cooper, W.A. (1975) "Perception of segment duration in sentence context", in: *Structure and Process in Speech Perception* (eds.: A. Cohen and S. Nooteboom), Berlin, Springer-Verlag, pp. 68-89.

Kohler, K.J. (1981) "Perceptual centres?", *Arbeitsberichte, Institut für Phonetik der Universität Kiel 20*, pp. 207-228.

Kohler, K.J. (1983a) "Stress-timing and speech rate in German. A production model", *Arbeitsberichte, Institut für Phonetik der Universität Kiel, 20*, pp. 5-54.

Kohler, K.J. (1983b) "Fo in speech timing", *Arbeitsberichte, Institut für Phonetik der Universität Kiel 20*, pp. 55-97.

Lehiste, I. (1970) *Suprasegmentals*, Cambridge, Mass., MIT Press.

Lehiste, I. (1971) "Temporal organization of spoken language", in: *Form and Substance* (eds.: L.L. Hammerich, R. Jakobson and E. Zwirner), Copenhagen, Akademisk Forlag, pp. 159-169.

Lehiste, I. (1976) "Influence of fundamental frequency pattern on the perception of duration", *Journal of Phonetics 4*, pp. 113-117.

Lehiste, I. (1977) "Isochrony reconsidered", *Journal of Phonetics5*, 253-263.

Lehiste, I., Olive, J.P. and Streeter, L.A. (1976) "Role of duration in disambiguating syntactically ambiguous sentences", *Journal of the Acoustical Society of America 60*, pp. 1199-1202.

Marcus, S.M. (1979) "Perceptual centers (P-centers)", in: *Proceedings of the Ninth International Congress of Phonetic Sciences, Copenhagen 1979*, vol. I, p. 238.

Miller, M. (1984) "On the perception of rhythm", *Journal of Phonetics 12*, pp. 75-83.

Morton, J., Marcus, S.M. and Frankish, C.R. (1976) "Perceptual centers (P-centers)", *Psychological Review 83*, pp. 405-408.

Ohala, J.J. (1975) "The temporal regulation of speech", in: *Auditory Analysis and Perception of Speech* (eds.: G. Fant, and M.A.H. Tatham), London, Academic Press, p. 431-453.

Strangert, E. (1985) *Swedish Speech Rhythm in a Cross-Language Perspective* (Umeå Studies in the Humanities 69), Stockholm, Almqvist & Wiksell.

Sundberg, J. (1973) "Data on maximum speed of pitch changes", *Speech Transmission Laboratory, Quarterly Progress and Status Report 4*, Royal Institute of Techonology, Stockholm, pp. 39-47.

Thorsen, N. (1984a) "Variability and invariance in Danish stress group patterns", *Phonetica 41*, pp. 88-102.

Thorsen, N. (1984b) "Fo timing in Danish word perception", *Phonetica 41*, pp. 17-30.

Thorsen, N.G. (1988a) "Intonation on Bornholm – between Danish and Swedish", *Annual report of the Institute of Phonetics, University of Copenhagen 22*, pp. 25-138.

Thorsen, N.G. (1988b) "Stress group patterns, focus signalling and sentence intonation in two regional Danish standard languages: Aalborg and Næstved" *Annual Report of the Institute of Phonetics, University of Copenhagen 22*, pp. 145-195.

Uldall, E.T. (1971) "Isochronous stresses in RP", in: *Form and Substance* (eds.: L.L. Hammerich, R. Jakobson and E. Zwirner), Copenhagen, Akademisk Forlag, pp. 205-210.

Woodrow, H. (1911) "The role of pitch in rhythm", *Psychological Review 18*, pp. 54-71.

Woodrow, H. (1951) "Time perception", in: *Handbook of Experimental Psychology* (ed.: S.S. Stevens), New York, Wiley, pp. 1224-1236.

AN INTRODUCTORY STUDY OF TONE AND INTONATION IN A LAO DIALECT

by

EVA GÅRDING AND JAN-OLOF SVANTESSON

Jørgen Rischel is admired not only for his work in Scandinavian phonology and phonetics but also for his analyses of non-Indoeuropean languages, in recent years those of South East Asia. It is therefore perhaps appropriate to dedicate to him an acoustic study of tone and intonation in a Northern Lao dialect, which to our knowledge has not been analysed before.

The Lao language belongs to the southwestern group of the Tai languages, which apart from Thai, the main language of Thailand, also includes several minority languages of Burma, Thailand, Laos, Vietnam and China (e.g. Shan and Lü). Thai and Lao form a dialect continuum and the language of northeastern Thailand, although officially regarded as a Thai dialect, is very close to Lao (for a survey see Li 1977 or Svantesson 1991).

The syllable type plays an important role in the description of tones in this area. As in Thai, there are two types of syllables in Lao, traditionally called 'live' and 'dead'. Live syllables are open or end in a sonorant (nasal or glide: *m, n, ŋ, j, w*) and dead syllables end in a voiceless stop (*p, t, k, ʔ*). A long/short vowel contrast occurs in both types, except that open syllables always have a long vowel and syllables ending in /ʔ/ have a short vowel.

Most often the tones are described separately for live syllables, short dead syllables and long dead syllables (for different analyses see e.g. Gedney 1972, Brown 1976, Strecker 1979). In the tone-count, the tones of the live syllables are listed first (e.g. 1-6), and the dead-syllable tones are simply added (e.g. 7-10). This is the convention followed in the phonological analysis which our investigation is based on. Another convention, used in the traditional analysis of Standard Thai, is to recognize the complementary distribution of tones in the two sets of syllables and regard the dead-syllable tones as contextual variants of the live-syllable ones (see Gedney 1972 p. 424 for a discussion).