ARIPUC 17, 1983 p. 1-20

TWO EXPERIMENTS IN THE PERCEPTION OF FØ TIMING IN DANISH

NINA THORSEN

The perceptual relevance of rather finely timed differences in the occurrence of fundamental frequency (Fø) rises as observed in natural speech is established in listening experiments with synthetic speech stimulus material. (1) When vowel and consonant duration cues are ambiguous between /ku·lə/ and /kulə/ an early Fø rise, relative to the vowel-consonant boundary, will tend to favour identification of stimuli as /ku·lə/, while a later Fø rise increases the number of /kulə/ judgments somewhat. (2) When vowel duration cues are ambiguous between /ˈbilisd/ and /biˈlisd/ an Fø rise before the intervocalic consonant will significantly increase the number of /ˈbilisd/ identifications, whereas an Fø rise after the consonant yields more judgments of /biˈlisd/. The latter result corroborates the acoustic observation that initial voiced consonants in stressed syllables dissociate tonally from the vowel and join up with the preceding material, if any, to the effect that tonally the stress seems to begin with the vowel.

I. INTRODUCTION

A. VOWEL LENGTH AND FØ TIMING

On previous occasions (e.g. Thorsen 1982a and 1982b) I have suggested a description of the Fø pattern characterizing prosodic stress groups in Danish in terms of an essentially invariant Fø wave upon which the segments and syllables are superposed. This will make intrasyllabic Fø movements predictable from the shape of the wave where the syllables hit it (falling, rising, etc.). Any Fø movement may then of course be modified

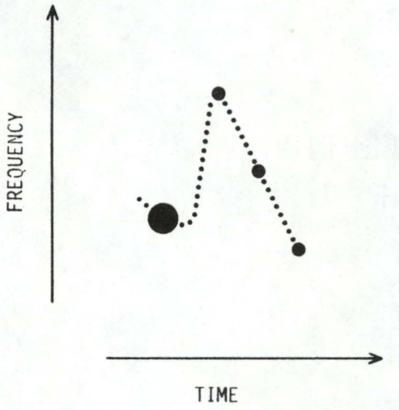


Figure 1

Stylized fundamental frequency pattern associated with the prosodic stress group in Standard Danish. The heavy dot denotes the stressed vowel, the light dots the unstressed syllables. The slope of the fall through the posttonics is subject to individual variation.

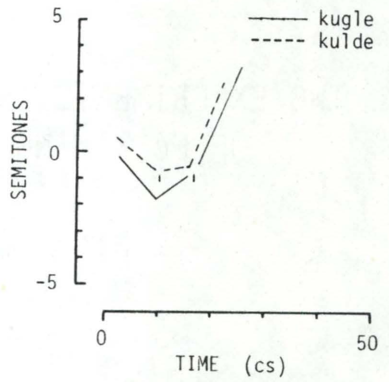


Figure 2

Fundamental frequency tracings (averages over 5 recordings) of two utterance medial words kugle and kulde [g̊ʰu·l̇ g̊ʰulḷ]. *The vowel-consonant boundary is indicated with a vertical stroke. Zero on the logarithmic frequency scale corresponds to 100 Hz. Male speaker.*

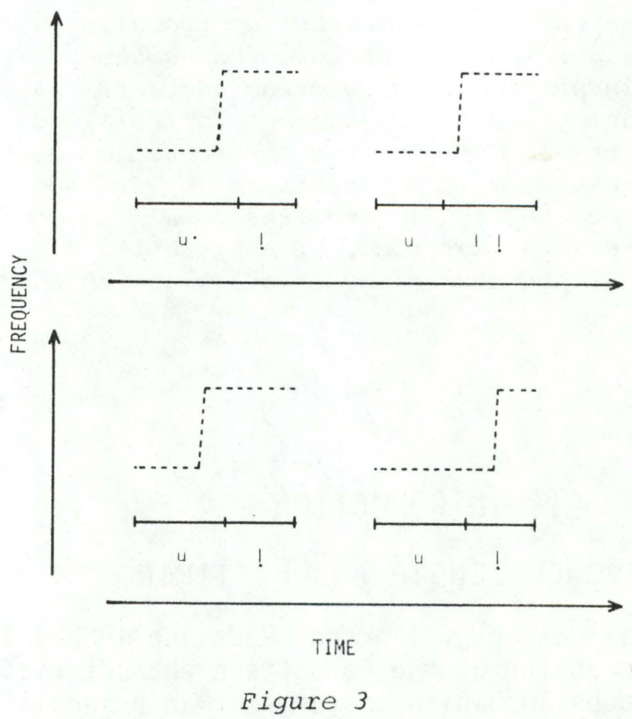


Figure 3

Stylized tracings of the normal alignment of segments with fundamental frequency in /-'V·C-/ and /-'VC-/ words (top) and of two different alignments of fundamental frequency with identical sequences of segments (bottom).

by microprosodic (segmental) phenomena. The shape of the Fø wave and the timing of stressed and unstressed syllables with respect to its trough and peak vary across dialects, cf. Thorsen and Nielsen 1981. In Standard Danish the stressed vowel hits the Fø wave in the trough before the fairly steep rise to the peak, so that generally short vowels have falling movements, whereas long stressed vowels and diphthongs will be falling-rising, cf. figure 1. With this account of Fø patterns in Standard Danish, a word pair like *kugle-kulde* (/ku·lə - kulə/ 'ball - cold') with assimilated schwa ([g̊ʰu·l̩ - g̊ʰull̩][1]) should be distinguished acoustically, inter alia, by an earlier Fø rise relative to the vowel/consonant boundary in the word with long vowel (where Fø rises before the boundary) than in the word with short vowel (where Fø rises at or after the boundary), and so is generally also the case, cf. figure 2. This account tallies well with Brink and Lund's (1975, p. 197) description of the consequences of schwa assimilation in these word types. Thus, if in figure 3 the two upper tracings are stylized illustrations of the normal alignment of segments with the Fø pattern, you would expect the two words in the lower tracings (where vowel and consonant durations are identical and presumably ambiguous) to be perceived as having a long vowel (left) and short vowel (right), respectively, if Fø rise timing has any perceptual cue value.

B. STRESS LOCATION AND Fø TIMING

Under certain circumstances in Standard Danish, a difference in stress location in otherwise comparable words can be seen, inter alia, as a rather finely timed difference in Fø movement. This is not only a question of timing the turning point in bidirectional Fø patterns (as in /ku·lə - kulə/ above) - quite the contrary. Take minimal stress pairs like *bíligst*[2] - *bilíst; Pállas - paláds; Nánna - Naná* (/'bilisd - bi'lisd -- 'palas - pa'las -- 'nana - na'na/ ['b̥ilisd̥ - b̥i'lisd̥ -- 'b̥ʰalas -'b̥ʰa'las -- 'nana - na'na] 'cheapest - motorist'; the Greek goddess - 'palace'; Danish and French girl's name). When they are preceded exclusively by unstressed material they will all have an Fø rise from the first to the second syllable. *bílligst*, *Pállas* and *Nánna* have an Fø rise because this is the normal Fø pattern on a succession of a stressed plus unstressed syllable; *bilíst*, *paláds*, *Naná* have an Fø rise because utterance initial unstressed syllables are often lower than the first stressed syllable. The three leftmost columns in figure 4 depict Fø tracings of these words from utterance initial position by four speakers, averaged over 6 recordings. The difference in placement in a subject's Fø range (higher or lower) between the members of a pair is to be expected if the stressed syllables are to have approximately the same pitch. The difference in intrinsic Fø level between the high vowels and the [l] in *bílligst* - *bilíst* is barely noticeable with the males (NRP and JBC) but very apparent with the females (BH and NT), even though the scale is logarithmic. (I do not know why this should be so - such a difference between males and females is not to be expected from Reinholt Petersen's (1976) investi-

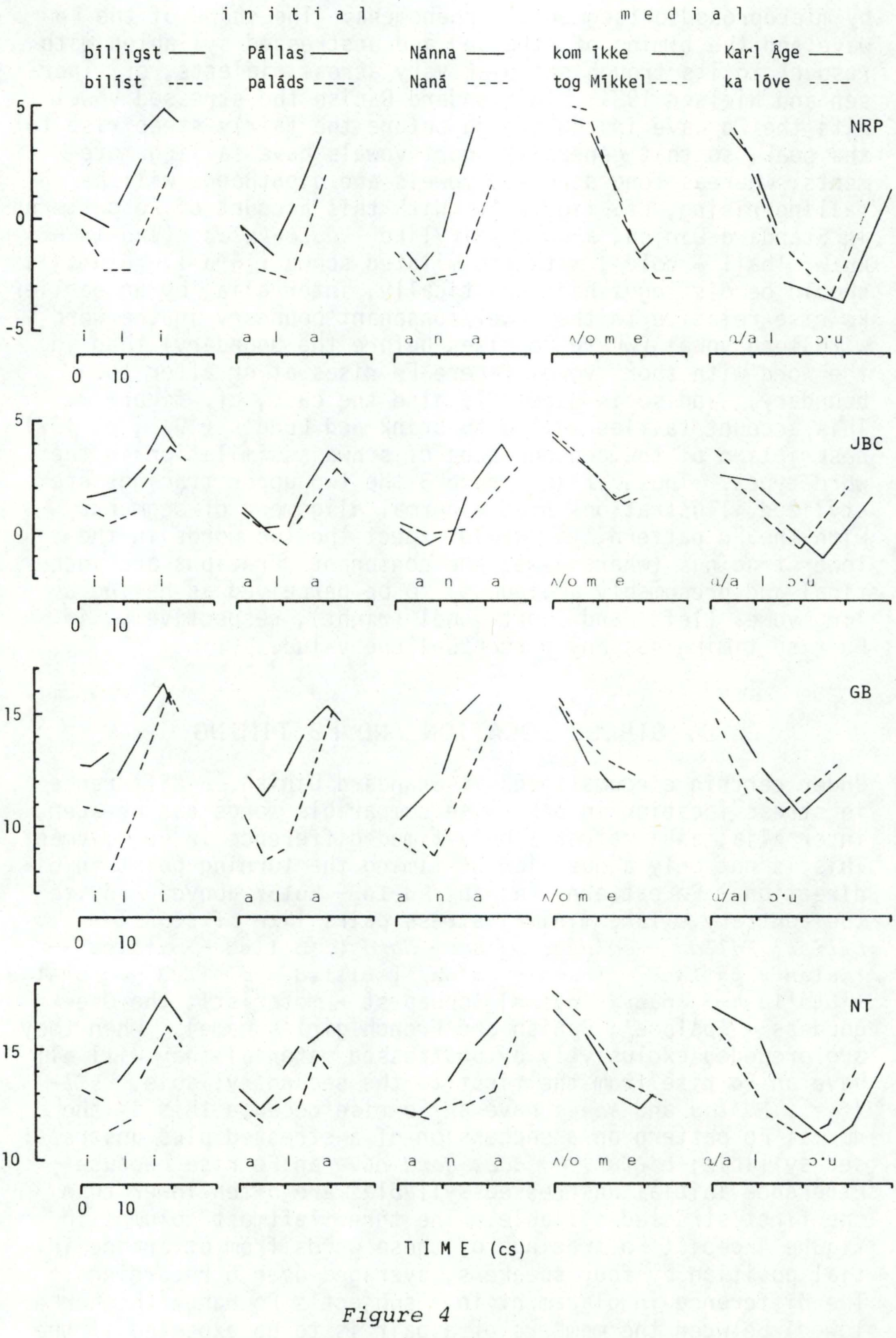


Figure 4

Fundamental frequency tracings (averages over 6 recordings) of three utterance initial word pairs (left) and two utterance medial sequences (right). The three initial word pairs, identified at the top of the figure, differ only by their stress placement. The two medial sequences differ by the location of the word boundary in a /-VCV-/ structure. Two male speakers (top) and two female speakers (bottom). Zero on the logarithmic frequency scale corresponds to 100 Hz.

gation.) Apart from differences in range and intrinsic Fø, there is a systematic difference in the relation between the vowels and the intervocalic consonant in words with first syllable stress (full line) versus words with second syllable stress (dashed line). The major Fø rise or step up begins after the consonant in the words with second syllable stress (dashed lines) but before the consonant in words with stress on the first syllable (full lines); or, in other words, the intervocalic consonant remains low and more or less at a level with the first, pretonic vowel in words with second syllable stress, whereas it rises away from the first, stressed vowel in words with initial stress. - In a pilot experiment (Thorsen 1982a) on the perception of Fø rise time differences in otherwise identical [bilisd̥]-stimuli, the location of a 2 semitone rise proved sufficient to shift listeners' location of the stress in the absence of other (e.g. durational) cues.

I take the difference in consonantal Fø movement to be an expression of a tonal dissociation between the syllable initial consonant and the stressed vowel in *bilı̈st, paláds, Naná*. (This interpretation will be supported by further examples below.) The rise during the consonant away from the stressed vowel in *bı̈lligst, Pállas, Nánna* is probably not to be seen in a similar manner as a tonal dissociation from the stressed vowel but rather as a consequence of the way Fø is patterned in the prosodic stress group in Standard Danish, being low on the stressed syllable and high-falling on the post-tonics, cf. figure 1. Furthermore, the rise is not always coincident with the vowel-consonant boundary, cf. section A and figure 2 above.

Tonal dissociation between an initial consonant and stressed vowel is not confined to utterance initial position. The two rightmost columns in figure 4 show Fø tracings (averages) of the underlined medial sequences *Fár kom ı̈kke méd til fésten. Mór tog Mı̈kkel méd til fésten.* ('Father did not make it to the party. Mother brought Mikkel along to the party.', i.e. [ʌm 'e]/[o 'me]) and *Jeg trór Karl Åge får góde karaktérer. Jeg trór jeg kan lóve du får góde karaktérer.* ('I think Karl Åge will get good grades. I think I can promise you will get good grades.', i.e. [ɑ·l 'ɔ·u]/[a 'lɔ·u]). The word boundary lies after and before the [m] and [l], respectively, but there are no apparent systematic differences in the course of Fø within each pair of tracings. (The sharp discontinuity in NRP's --*Karl Åge*-- can be ascribed to a brief glottal stop before the stressed vowel. Such boundary signalling is optional, but not very frequent in pragmatically neutral speech styles.) The tracings confirm a previous observation (e.g. Thorsen 1980) that word boundaries as a rule leave no trace in the course of Fø, and in this particular case the sonorant consonants smoothly connect the unstressed with the succeeding stressed vowel in /-- VC+V́--/ and /-- V+CV́--/ alike. However, these examples are not conclusive as to a tonal dissociation between initial consonant and stressed vowel (except maybe negatively, because there is no dissociation between the previous unstressed vowel and the initial consonant either). Such an interpretation only lends itself clearly from consideration of sequences of stressed vowels with less intervening segmental material. Figure 5 depicts the course of Fø in the underlined sequence of *Dén øl ér lúnken.* ('That beer is tepid.', i.e. ['øl ?'æʌ 'lɔ].) The initial [l] in *lúnken* behaves very much as if it were part of the preceding prosodic stress group; together with the preceding stressed vowel it performs the characteristic low + high-falling pattern. (The high consonant is not due to intrinsic Fø level differences between the [l] and the vowels because, if anything, the consonant has an intrinsically lower Fø.) In fact, the initial [l] in *lúnken* behaves qualitatively exactly as the final [l] in *øl*. Its more ample rising movement

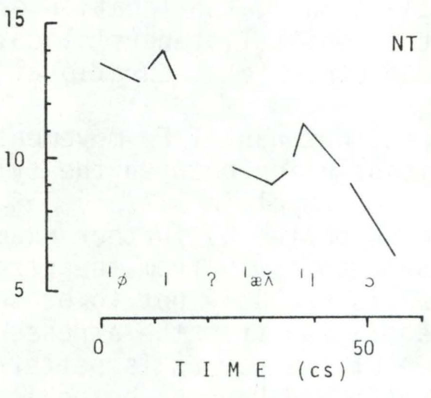


Figure 5

Fundamental frequency tracing of the underlined sequence in the utterance Dén øl ér lúnken. Zero on the logarithmic frequency scale corresponds to 100 Hz. Female speaker.

is due to the longer total duration of the sequence ['æʌl] versus ['øl], which allows for a more complete gesture before the downward course towards the next stressed vowel. In other words, the "tonal syllabification" seems to be V̂C-V́ rather than V̂-CV́, irrespective of phonological and morphological/syntactic boundaries. Note, however, that tonal dissociation between initial consonant and stressed vowel does not always make itself noticed. A consonant will not go out of its way to dissociate from the succeeding homosyllabic stressed vowel, cf. the rightmost examples in figure 4, but when two stressed vowels are sufficiently spaced in frequency (as they will be in short terminal declarative utterances), and when there is not sufficient segmental material to connect them smoothly, an initial consonant can be seen to team up with the preceding prosodic stress group.

Whether a stressed vowel can repel more than one initial consonant is hard to say, due to the phonetics and phonotactics of Danish. The sonorant consonants do not combine initially. The obstruents are all unvoiced, except [v]. [v] combines initially only with [ʁ] as in *vride* ('to wring') and with [l] in foreign names like *Vladimir*, so the material for investigation is limited, and it may be difficult to interpret due to intrinsic and coarticulatory effects of the [vʁ] and [vl] combinations.

To summarize the acoustic observations: An initial voiced consonant in a stressed syllable will be dissociated tonally from the stressed vowel and join up with the preceding material, if any. Postvocalic consonants after a stressed vowel behave tonally in an unmarked fashion, and their relation to the preceding vowel is determined by the Fø pattern which characterizes a prosodic stress group in Standard Danish. - Under certain circumstances, the relation between an intervocalic voiced consonant and the surrounding vowels may be the only tonal difference between words with different stress locations.

To what extent tonal dissociation between initial consonants and stressed vowels is manifest in other varieties of Danish, I cannot say, but I doubt that the phenomenon should be an exclusively Standard Danish one, or even a Danish phenomenon, for that matter. It matches the observation from some tone languages that postvocalic consonants may carry the final part of the distinctive tonal course but prevocalic consonants do not seem to be included (Pike, 1948, p. 10, 30; Selmer, 1928).

II. PROCEDURES

On the hardware parallel synthesizer (Rischel and Lystlund 1977) at the Institute of Phonetics, I synthesized two sets of "words", [g̊ʰul] and [b̥ilisd̥]. In figure 6, top left, the four variables in the [g̊ʰul̥] stimuli are shown schematically. (1) Vowel duration varies in three steps: 100, 140, and 180 ms;[3] (2) consonant duration varies in three steps: 100, 120, and 140 ms; (3) the timing of the Fø rise occurs at the vowel-consonant boundary or 60 ms before the end of the consonant, i.e. it is constant with respect to the end of the consonant; (4) the rise is 3 or 4 semitones high, and is accomplished in 5 ms. Only 26 (of a possible 36 total) stimuli were included in the test.

Five parameters varied in [b̥ilisd̥], cf. figure 6, top right. (1) the timing of the Fø rise, which occurred immediately after the first vowel or immediately after the [l]. That is to say, either the [l] was low - at the level of the first vowel - or it was high - at the level of the second vowel. Both vowels were level pitched and had identical intensities; (2) the magnitude of the Fø rise, which was 0, 2, 3 or 4 semitones high; (3) placement in the Fø range of the whole word, that is to say, the first vowel was at 89 Hz or at 100 Hz; (4) duration of the first vowel varied in three steps: 60, 80 and 100 ms; (5) duration of the second vowel varied in three steps: 80, 100, and 120 ms. These were values which I myself found would create definite /ˈbilisd/ and /biˈlisd/ stimuli, as well

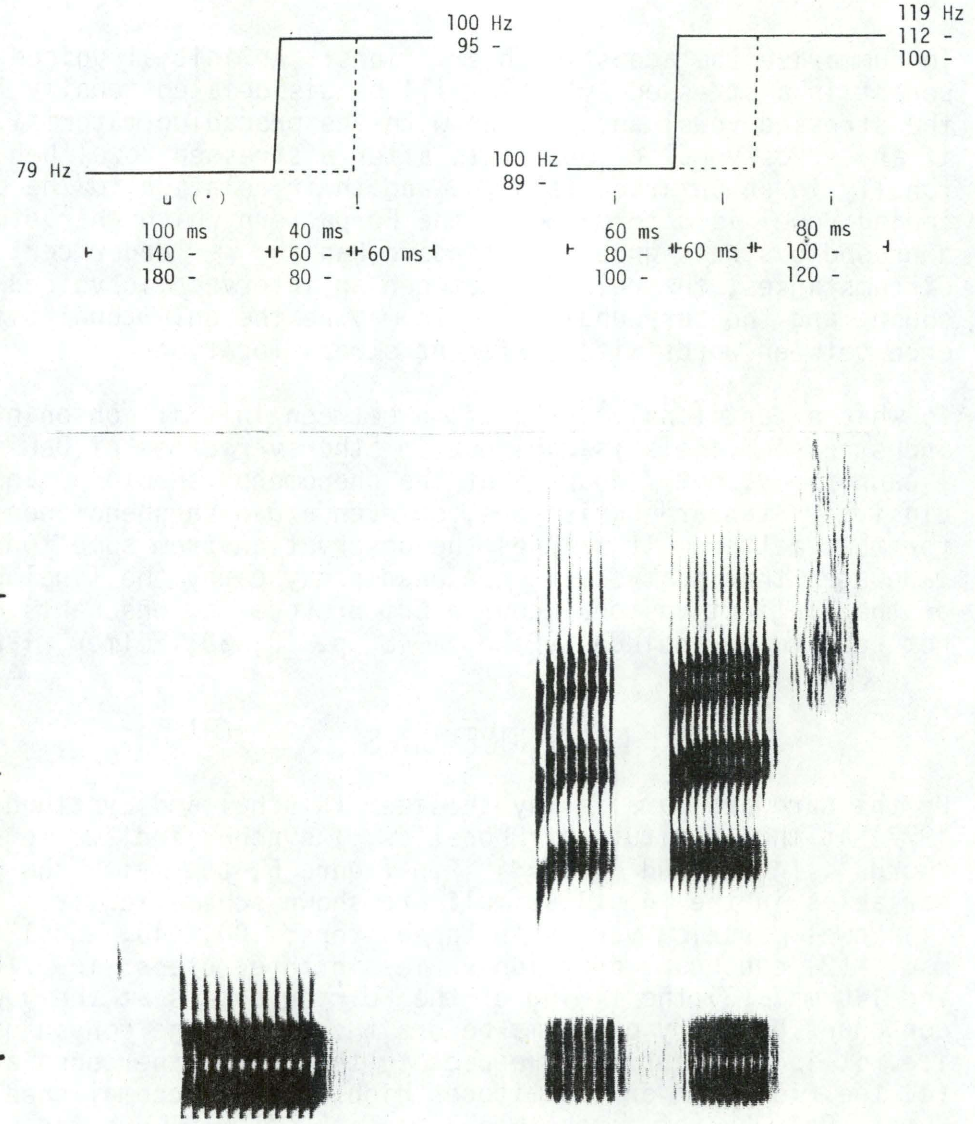


Figure 6

Scheme of the variables in two synthesized stimuli [g̊ʰuḷ] *and* [b̥ilisd̥] *(top), and spectrograms of two sample stimuli (bottom). See further the text.*

as ambiguous ones. From a possible 126 different stimuli, 63 were selected for the test. (Actually, 12 more words were included, where [l]'s Fø was lower than the first vowel. Since these stimuli neither add to nor subtract anything from the conclusions to be reached from stimuli where [l]'s Fø is identical with the first vowel, they are disregarded in the following.)

The total of 101 (26 + 63 + 12) different words were randomized five times, each randomization on a separate tape. Every tape was introduced by the same two repetitions of three [b̥ilisd̥] and three [g̊ʰul] which swept the whole range of variation and were intended for practice and not to be responded to. The 101 test items were led by five and succeeded by four dummies, giving a total of 110 items to be identified. Each stimulus occurred once only, with 4 sec intervals between stimuli. Leader tape after every ten words automatically stopped the tape recorder, to help subjects keep track of their progression. One run of a test tape lasted 11 minutes.

Subjects were instructed (in writing) about the composition of the tapes. They were asked to identify each word as either *kugle* or *kulde* or as *billigst* or *bilist* (forced choice) by ticking off the appropriate box on prepared answer sheets. The tapes were presented to subjects individually, over headphones. Some listened to one tape a day, five days running, others distributed their trials over a couple of weeks. One subject heroically worked his way through all five tapes in one session. Nine colleagues and students at the Institute of Phonetics, representing different age groups and dialectal background, took this test.

III. RESULTS

A. STATISTICAL TREATMENT

The only suitable test of independence between response distributions such as depicted in figure 7 and 9 is the rather conservative, non-parametrical χ^2. χ^2 values are given in the lower right of each rubric and the confidence level, p, is indicated if it is 0.05 or better. (In a one-tailed test it takes a χ^2 above 13.36 to be significant at the 0.05 level at 8 degrees of freedom and above 10.64 at df = 6.) N is the total number of responses behind each data point.

The significance of the difference between scores on any pair of stimuli (or collapsed stimuli) may be tested in a student's t-test, from the score variances, according to

$$\sigma^2 = n \cdot p \cdot q$$

where σ^2 is the score variance, n is the total number of presentations of the stimulus (or collapsed stimuli), p is the proportion of responses in one of the two categories, i.e. p = score/n, and q = (1-p). When a score difference is statistically significant at the 0.05 level, or better, this is indicated with a star just below the zero line in figure 7 and 9.

For example: the two response distributions in figure 7 are not significantly different (χ^2 = 5.73, df = 8), but two stimulus pairs did receive significantly different scores.

Thus, the stimuli in the fifth pair received 13 and 19 *kugle* identifications, respectively. The score variances are 6.24 and 4.56, and t = 1.83. At df = (2·25-2) = 48 this value is significant beyond the 0.05 level (one-tailed test).

B. *kugle-kulde*

To four of the nine subjects vowel duration seems to have been the only perceptual cue to the identity of the test words. At the shortest vowel duration (100 ms) only *kulde* is heard, and at the two longer vowel durations (140 and 180 ms) only *kugle* is identified, irrespective of the duration of the consonant and of the timing of the Fø rise with respect to vowel-consonant boundary. In other words, the range of vowel durations is ill suited to these subjects, and they are therefore disregarded here.

Responses from the remaining five subjects to 9 stimulus pairs (those with a 3 semitone Fø rise) are shown in figure 7 where the percentage *kugle* responses (of a possible 25 total) is depicted. There is an effect from all three parameters, vowel duration, consonant duration, and Fø rise time, but the latter two are of little consequence only when the vowel is short (100 ms - no. 1-3) and long (180 ms - no. 7-9). At the intermediate vowel duration (140 ms - no. 4-6) there are more *kugle* identifications when the consonant is short and with an early Fø rise relative to the vowel-consonant boundary. The overall effect of Fø rise time is weak and not statistically significant, and in natural speech there is every reason to believe that vowel (and consonant) duration alone decide(s) the issue.

C. *billigst-bilist*

1. RESPONSES

Two subjects consistently identified very nearly 100% of the [b̥ilisd̥] stimuli as *billigst*, and their data are disregarded in the following. However, their behaviour illustrates how different subjects may choose different (internal) judgment criteria, and how the choice of response category label therefore determines identification. Both subjects had participated in the pilot test six months previously (Thorsen, 1982a). One had then given balanced and consistent responses in both categories, and the other had surprisingly given almost exclusively *bilist* responses. The latter subject had explained her reaction to the pilot test in the following manner: she had heard some adequate *bilist*s and some which perhaps were not good but which she would not identify as *billigst* at any rate. In other words, *billigst* seems to have been the marked member of the pair to this subject in the pilot test, since her requirements for *billigst* identification did not allow her to put doubtful stimuli in that slot. In the present experiment she turned the tables (but not consciously, because she

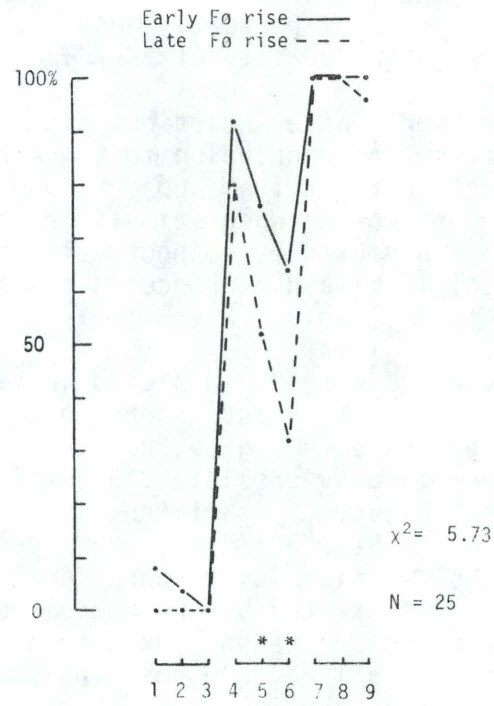


Figure 7

Identification functions (proportion of long vowel judgments, i.e. kugle) of two sets of synthetic [g̊ʰu(·)l̩] *stimuli which differ with respect to the timing of the fundamental frequency rise from the stressed vowel to the post-tonic syllabic consonant. Vowel duration increases from left to right thus: 100 ms (no. 1-3), 140 ms (no. 4-6), 180 ms (no. 7-9). Within each triad consonant duration increases from left to right thus: 100 ms (no. 1,4,7), 120 ms (no. 2,5,8), 140 ms (no. 3,6,9). Fundamental frequency rises 3 semitones in 5 ms, either at the vowel-consonant boundary ("early") or 60 ms from the end of the consonant ("late").* χ^2 *on the two response distributions is given in the lower right. The total number of responses behind each data point is 25 (5 subjects times 5 responses). A star beneath a pair of data points indicates that those two stimuli received significantly different response scores.*

did not remember what her choice had been on the first occasion), and she now confessed to have heard some adequate *billigst*s and some which were really not very good, but none that could have passed as good *bilist*s. The other "uncooperative" subject described her reaction to this test in a similar manner. Now, clearly, if subjects had been asked to sort out the acceptable *billigst*s, i.e. if response categories had been labelled *billigst/not billigst*, these two subjects' response distributions might have been radically altered. But then, so would probably some of the other subjects' responses: at least two of them (ND and OT) declared to have been quite certain to have heard a number of perfectly good *bilist*s but they were less satisfied with the quality of *billigst*.

The fact that some subjects never heard but one member of the pair does not invalidate the conclusion to be reached on the basis of response distributions from subjects who had identified occurrences of both words: When stimuli are heard predominantly as *billigst* and *bilist*, respectively, this difference may be caused solely by a difference in the timing of the Fø rise between the two syllables.

The results from seven subjects are depicted in figure 8. Subjects are ordered from left to right according to the total number of *billigst* responses they gave, ranging from 154 (ND) to 252 (PM) of a theoretically possible 315 total (63 stimuli times 5 responses per subject). Apart from an overall smaller or greater inclination to hear *billigst*, there seems to be a difference in subjects' sensitivity to durational variation, which is most clearly illustrated by the bottom row, where Fø is constant and only first and second vowel duration vary: AM and PM have identified all seven words consistently as *billigst*, JB has identified the last five stimuli in the septet, i.e. those with medium and long first vowel as *billigst*, irrespective of the duration of the second vowel. With JR and OT the increase in the duration of the second vowel can be seen as a decrease in *billigst* responses at medium first vowel durations (no. 3, 4, and 5 in the septet).

Despite quantitative differences between subjects, and different sensitivity to some of the parameters under investigation, subjects do not exhibit systematic qualitative differences that prohibit consideration of their pooled data. Accordingly, figure 9A-D and F present the responses to individual stimuli as the percentage *billigst* identifications (of a possible 7x5 = 35 total). Figure 9E,G, and H present the data collapsed in various ways, see further below.

2. INTERPRETATION

Figure 9A-E shows clearly that, everything else being equal, an [l] at the same Fø level as the preceding vowel (broken lines) yields a smaller proportion of *billigst* answers than an [l] at the level of the succeeding vowel (full lines). In other words, when the consonant is more closely associated tonally with the first vowel, the second vowel tends to be perceived as stressed and vice versa, when the consonant is more closely associated tonally with the second vowel, the first vowel tends to be perceived as stressed. The effect is dependent on duration, being generally nearly nil at the longest first vowel duration (no. 6 and 7 in each septet). The cue inherent in duration alone is apparent from figure 9F which depicts responses to stimuli with a monotone Fø at 100 Hz. That figure resembles very closely the distribution of collapsed responses

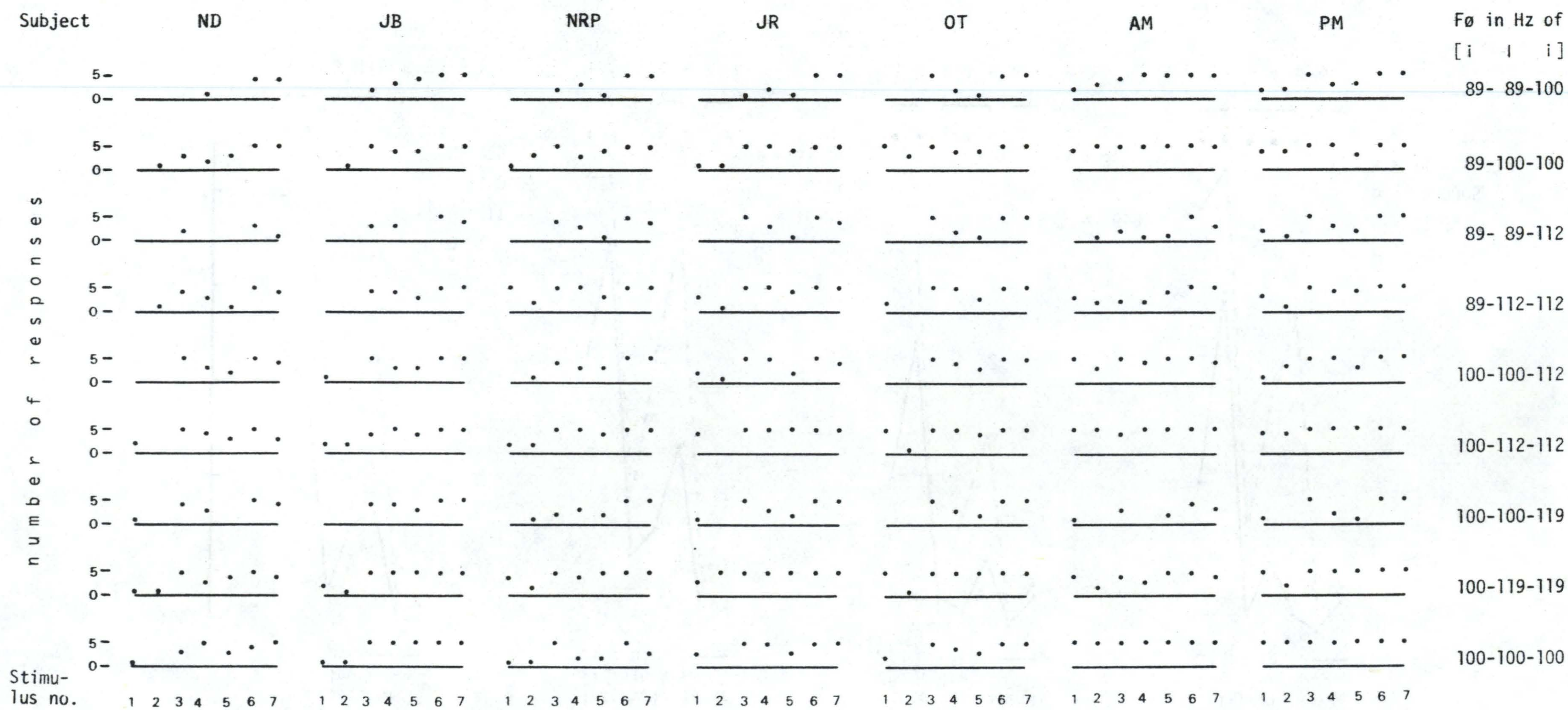
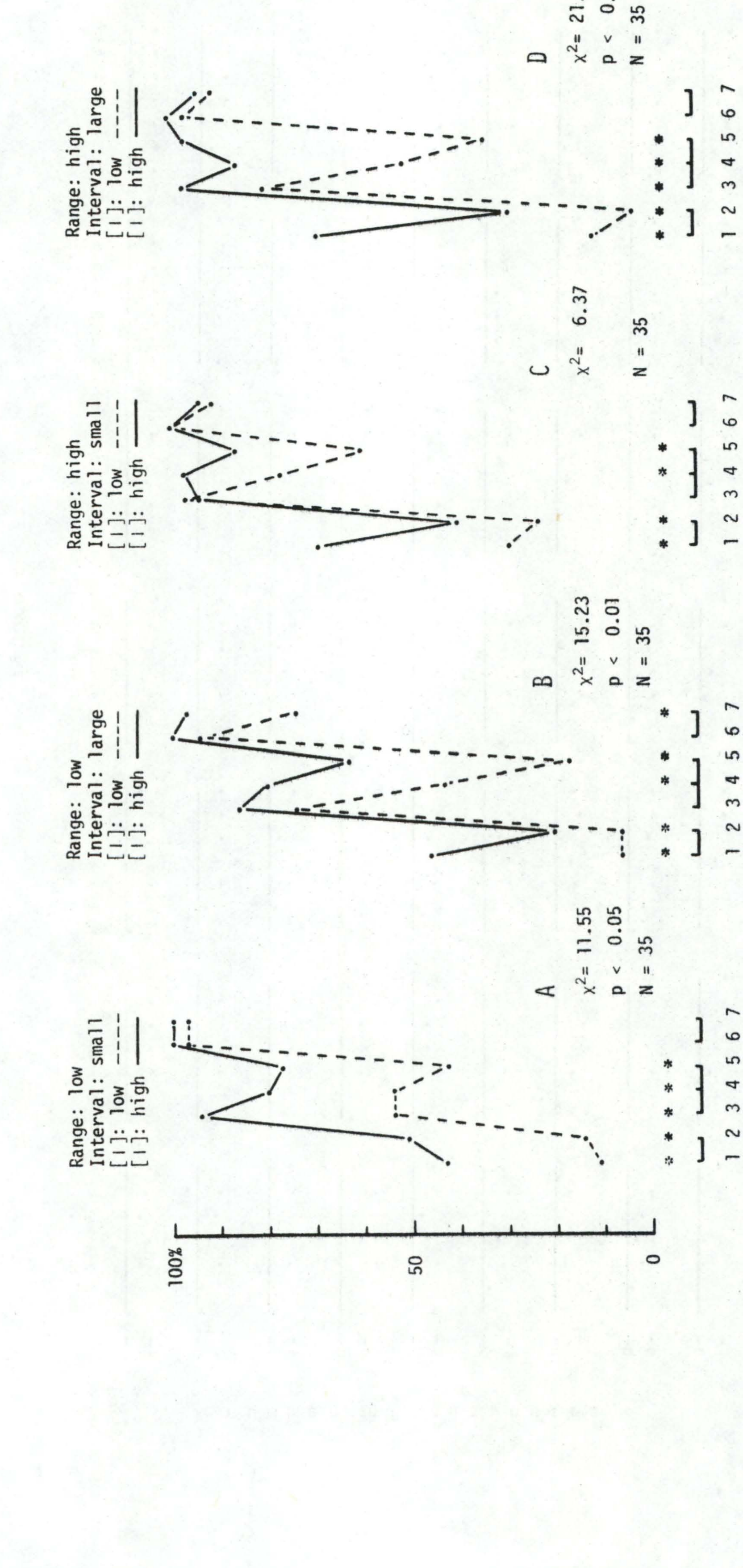


Figure 8

Seven subjects' judgments of initial stress (i.e. billigst) to 63 synthesized [b̥ilisd̥]-stimuli. There are 9 different fundamental frequency conditions, as indicated to the right in the figure. Within each septet of stimuli the duration of the first and second vowel varies thus: no. 1 - 60/100 ms; no. 2 - 60/120 ms; no. 3 - 80/80 ms; no. 4 - 80/100 ms; no. 5 - 80/120 ms; no. 6 - 100/80 ms; no. 7 - 100/100 ms. Subjects are ordered from left to right according to the smaller or larger total number of initial stress judgments they gave.


100%
50
0
Range: low
Interval: small
[I]: low ----
[I]: high ——
A
χ²= 11.55
p < 0.05
N = 35
1 2 3 4 5 6 7
Range: low
Interval: large
[I]: low ----
[I]: high ——
B
χ²= 15.23
p < 0.01
N = 35
1 2 3 4 5 6 7
Range: high
Interval: small
[I]: low ----
[I]: high ——
C
χ²= 6.37
N = 35
1 2 3 4 5 6 7
Range: high
Interval: large
[I]: low ----
[I]: high ——
D
χ²= 21.34
p < 0.0005
N = 35
1 2 3 4 5 6 7

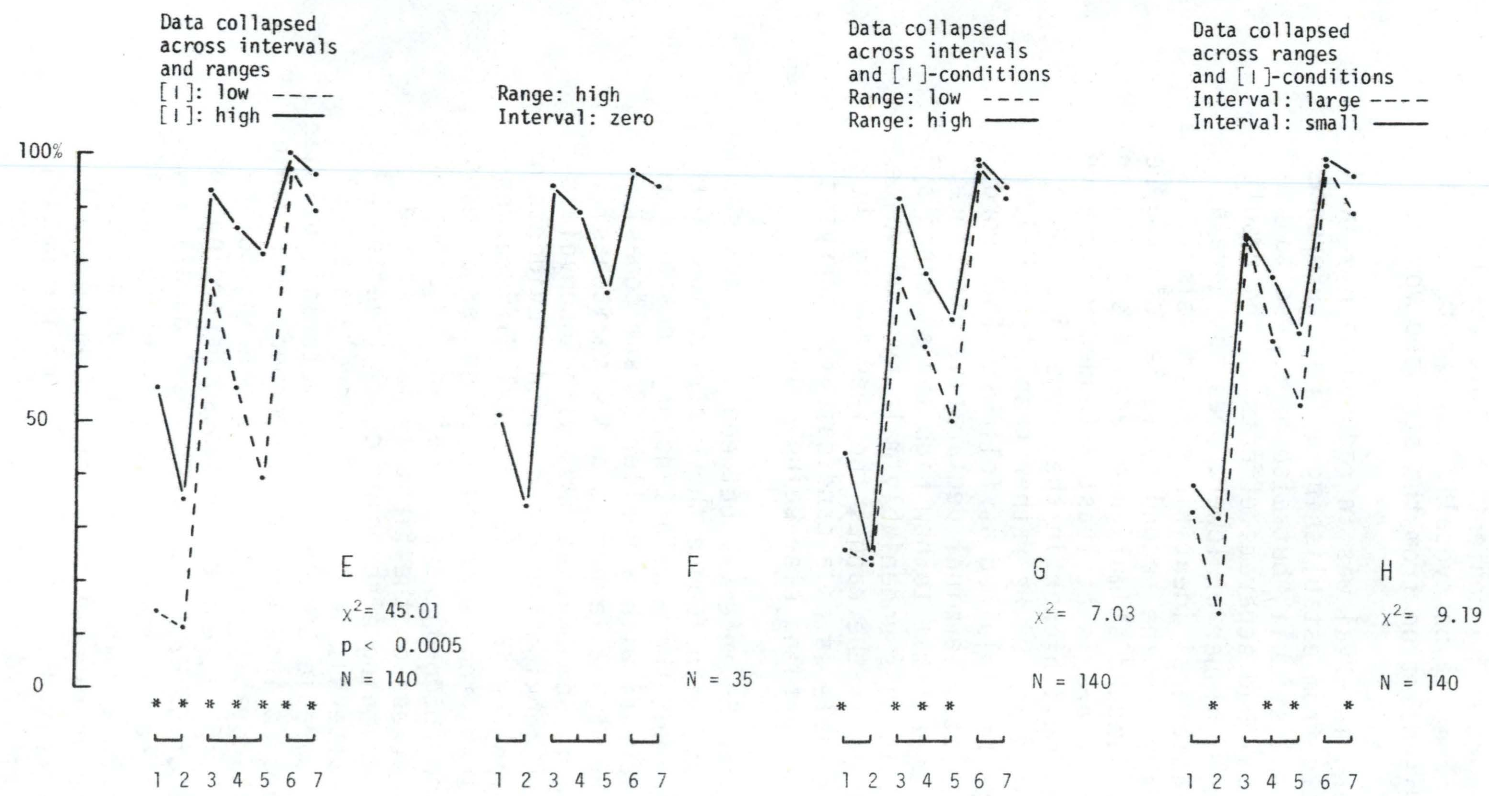


Figure 9

Identification functions (proportion of initial stress judgments, i.e. billigst) of 9 sets of synthesized [bilisd]-stimuli, collapsed across seven subjects. See further the legend to figure 8. In rubrics E, G, and H the data are collapsed across intervals, ranges and [l]-conditions as indicated at the top. χ^2 on each pair of response distributions is given in the lower right, and p is indicated if 0.05 or better. N is the total number of responses behind each data point. A star beneath a pair of data points indicates that those two stimuli received significantly different response scores. "Range" refers to the frequency of the first vowel (89 Hz: low; 100 Hz: high). "Interval" refers to the frequency interval between the first and second vowel (2 semitones: small; 3 or 4 semitones: large). "low" and "high" [l] refers to the location of the consonant in the frequency range (on the same level as the first vowel: low; on the same level as the second vowel: high).

obtained from stimuli with an [l]-condition which favours *billigst* responses (9E, full line). This can be taken to mean that an Fø rise timed immediately after the vowel is the "unmarked" Fø condition, since it does little to change listeners' identification of stimuli where Fø cues are totally absent, and it is a further support of my contention above (section I.B) that the rise performed by postvocalic consonants is not to be seen as a tonal dissociation from the stressed vowel.

Variation in Fø range and interval was introduced mainly in order to prevent subjects from establishing a fixed reference against which to judge all stimuli, but also of course to see whether range and interval had separate effects on the results. Placement in the Fø range, higher or lower, does not have a significant effect on listeners' identification of stimuli as bearing stress on the first or the second syllable, cf. figure 9G, but four pairs of collapsed stimuli did receive significantly different scores, and there is at least a tendency for more *billigst* responses to stimuli higher in the range (i.e. when the first vowel is located at 100 Hz rather than at 89 Hz). This trend might perhaps be explained as follows. Each word is simultaneously a complete, terminal declarative utterance. There is therefore an upper limit to how high the stressed syllable can plausibly be, in this undeniably male type of voice, and the higher [b̥ilisd̥] is in the range, the less likely the second, and highest, syllable is as a candidate for carrying (non-emphatic) stress, everything else being equal.

Likewise, magnitude of the Fø interval between first and second vowel does not significantly influence the distribution of responses, cf. figure 9H. But there is a tendency towards more *billigst* responses to stimuli with a smaller (2 semitones) interval than to stimuli with a larger (3 and 4 semitones) interval. This finding is somewhat at odds with Rosenvold's (1981) results. She was looking for - and found - evidence of a perceptual compensation for intrinsic Fø level differences between vowels of different tongue height. In her experiment synthesized [b̥id̥b̥iɡ̊] and [b̥id̥b̥ɑɡ̊] were perceived, by 9 out of 10 subjects, as having stress on the first syllable at the higher Fø intervals, everything else being equal (the tenth subject reversed this pattern). However, Rosenvold's stimuli are not immediately comparable to the present material, neither where duration nor where Fø interval are concerned. Her range of Fø intervals is considerably larger, corresponding to approximately 0, 2, 3½, 5, 6, 7, and 8 semitones, departing from a constant first vowel of 85 Hz. The first vowel duration was constant at 100 ms, the second vowel varied between 80 and 150 ms in 10 ms steps. If Rosenvold's data are collapsed across different durations and vowel qualities (figure 4a, p. 157), the second vowel is perceived as stressed (85% of the presentations) at intervals of 0 and 2 semitones, the cross-over point is at 3½ semitones, and intervals of 5 and more semitones are perceived as having stress on the first syllable (90% of the presentations). The fact that at the larger intervals, the first vowel is perceived as stressed may be explained along the same tentative lines as the effect of placement higher in the range in this experiment, cf. above: the second syllable may be too high to be a likely stressed syllable in a one word non-emphatic terminal declarative utterance. The fact that the second syllable is perceived as bearing stress when the interval is 0 or 2 semitones may be due to duration, rather than to Fø interval as such: a first vowel at 100 ms and a second vowel varying between 80 and 150 ms could be expected to provoke a predominance of second vowel stress judgements in the absence of other, overriding, cues (such as Fø).

No doubt Fø rise timing is the single most effective stress location cue in this experiment, disregarding vowel duration. The only instance where the difference between response distributions from the two different Fø rise time conditions is not statistically significant is when the stimuli were high in the range, and the Fø interval small, cf. figure 9C, a fact that may be explained as the combined effect of the trend towards a general preference for initial stress high in the range and with small intervals, cf. above. In other words, this is a condition where the tendency is towards perceived stress on the first syllable anyway, and differences in Fø rise timing will be less evident.

3. CONCLUSION

When everything else is equal, differences in the timing of an Fø rise between utterance initial syllables may influence the identification of the location of stress, but vowel duration - particularly the first vowel - may easily override the cue inherent in Fø rise timing. There is hardly any effect from Fø when the first vowel is sufficiently long. In view of the identification of stimuli with a monotone Fø, the proper way to interpret the results is probably to say that when the intervocalic consonant is dissociated tonally from the second vowel, then that vowel is heard as stressed. If not, then the first vowel is stressed.

Naturally, the timing of Fø events at this, segmental, level will hardly ever be the only acoustic cue to stress location. Mainly because in utterances with more than one prosodic stress group in them, the tonal relations to the surrounding syllables will decide the issue. If the stressed syllable in *bi̇̆lligst* and *bili̇̆st* is, e.g., the second one in the utterance, then the stress cuing Fø rise will occur after *bi̇̆l-* in *bi̇̆lligst*, but after *-li̇̆st* in *bili̇̆st*, cf. figure 10. Furthermore, segment duration, vowel quality, degree of aspiration, etc. will help single out the stressed syllables of an utterance.

Tonal segregation of initial consonants may never serve any direct communicative function and may at best be said to be a redundant cue to stress location. But since it occurs and <u>can</u> be perceived - at least in laboratory conditions - there <u>must</u> be an explanation for it. It seems that as far as its tonal manifestation goes, <u>stress begins with the vowel</u>, because - contrary to expectations, perhaps - a stressed vowel

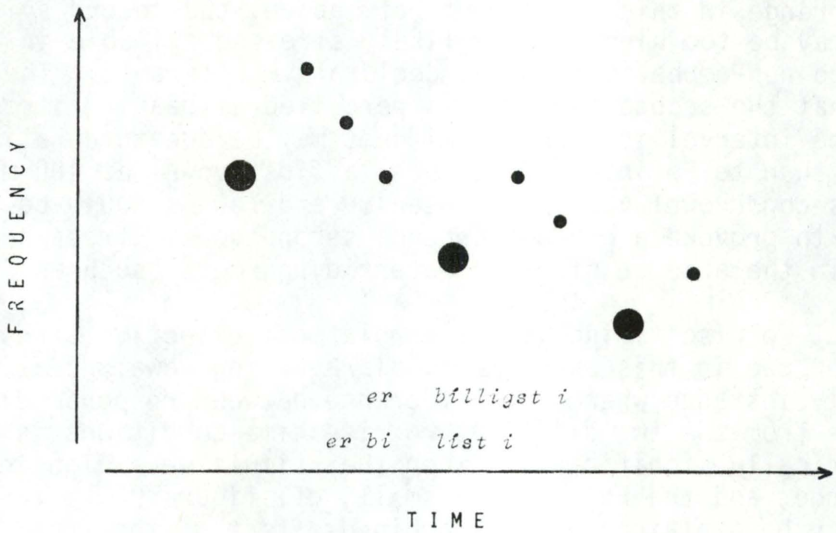


Figure 10

Stylized course of fundamental frequency in a terminal declarative utterance where the second stressed word is bĭlligst and bilĭst, respectively. The heavy dots denote the stressed syllables, the light dots the unstressed ones. The utterances could be, e.g. Várerne er bĭlligst i Ísland. and Ánni er bilĭst i Galĭcien. ('The goods are cheapest in Iceland. and 'Anni is the motorist in Galicia.)

repels rather than attracts preceding homosyllabic consonants. One side-effect of this is that word and syllable boundaries are deleted in the course of Fø.

Prosodic segregation of initial consonants may not be a purely tonal phenomenon. Fischer-Jørgensen (1982, p. 159) presents evidence that under certain circumstances the pattern of segmental duration indicates a boundary between a pre-vocalic consonant and a stressed vowel. Recent experiments on Swedish indicate that the most reasonable account of rhythmic phenomena is achieved if the onset of the rhythmical unit is taken to be the onset of the stressed vowel rather than, say, the onset of the first pre-vocalic consonant. This was one of the conclusions drawn by Lubker et al. (1983) from data on articulatory compensation in bite-block experiments, and Strangert (1983) likewise found that the most appropriate segmentation for an account of inter-stress intervals is the onset of the stressed vowel.

IV. NOTES

1. A dot under a consonant symbol signifies syllabicity here.

2. The stressed vowels are indicated orthographically with acute accents in the following.

3. As on most speech synthesizers, actual segment duration may vary by as much as a whole period both at the onset and offset of voicing, because the settings pertain to the voice source amplitude gate control, and this control is independent of the repetition rate of the voice source pulses. Several repetitions of a given stimulus were recorded on master tape, and from oscillogram and intensity curves, those items were selected which best corresponded to the intended durations. However, measuring accuracy is hardly better than ± 5 msec, and minor deviations from intended vowel duration values may accordingly still occur.

ACKNOWLEDGEMENT

Thanks to Jørgen Rischel for valuable suggestions for improvements in the manuscript.

REFERENCES

Brink, L. and Lund, J. 1975: *Dansk Rigsmål 1-2* (Gyldendal, Copenhagen)

Fischer-Jørgensen, E. 1982: "Segment duration in Danish words and its dependency on higher level phonological units", *Ann. Rep. Inst. Phon., Univ. Copenhagen 16*, p. 137-189

Lubker, J., Lindblom, B., Gay, T., Lyberg, B., Branderud, P. and Holmgren, K. 1983: "Prosody and compensatory motor behavior", paper read at a symposium on *Prosody - Normal and Abnormal*, Zürich, April 6-8

Pike, K. L. 1948: *Tone Languages*, (University of Michigan Press, Ann Arbor)

Reinholt Petersen, N. 1976: "Intrinsic fundamental frequency of Danish vowels", *Ann. Rep. Inst. Phon., Univ. Copenhagen 10*, p. 1-29

Rischel, J. and Lystlund, S.-E. 1977: "The IPUC speech synthesizer", *Hamb. phon. Beitr., Misz IV* (= Band 22), p. 143-160

Rosenvold, E. 1981: "The role of intrinsic Fo and duration in the perception of stress", *Ann. Rep. Inst. Phon., Univ. Copenhagen 15*, p. 147-166

Selmer, E. W. 1928: "Noen bemerkninger om den musikalske aksent i dens forhold til den sterkt og svakt skårne aksent", in *Festskrift til Rektor J. Qvigstad. Tromsø Museums Skrifter 2*, p. 250-262

Strangert, E. 1983: "Temporal characteristics and rhythmic units in Swedish", paper read at the *Third Symposium on the Prosody of Nordic Languages*, Umeå, June 3-4

Thorsen, N. 1980: "Word boundaries and Fo patterns in Advanced Standard Copenhagen Danish", *Phonetica 37*, p. 121-133

Thorsen, N. 1982a: "On the variability in Fo patterning and the function of Fo timing in languages where pitch cues stress", *Phonetica 39*, p. 302-316

Thorsen, N. 1982b: "Selected problems in the tonal manifestation of words containing assimilated or elided schwa", *Ann. Rep. Inst. Phon., Univ. Copenhagen 16*, p. 37-100

Thorsen, N. and Nielsen, B. Jul 1981: "Fundamental frequency patterning and sentence intonation in two Jutlandic dialects - A first approach", *Ann. Rep. Inst. Phon., Univ. Copenhagen 15*, p. 1-12.