last page

Papers from a symposium, Stockholm, August 12 - 14, 1992

something which has to be adhered to, even when the speaker speaks distinctly enough to overcome the 'sluggishness' of his speech production mechanism. In the same vein I interpreted the less drastic reduction of rise intervals found on less falling contours (cf. Fig. 3), as an expression the markedness of less sloping contours vis-à-vis more sloping contours. The more marked contours are supposed to be subject to a more careful control by the speaker, and the reduction in gestural distinctness with time is accordingly less pronounced, which is reflected in steeper - and thus also higher - rises, ceteris paribus.

CONCLUSION

All these arguments become circular if we do not determine what was first: the hen or the egg. And it is my contention that in the beginning was production, with its constraints and natural tendencies. Perception adjusted to the inherent properties of speech signals, which have therefore become part of the linguistic code, and which accordingly have to be produced, even when articulatory constraints have been partly or completely neutralized or suspended. If "we speak in order to be heard in order to be understood", then the speaker has to adjust to the listener's expectations, which holds true not just at the level of the phrase or sentence/utterance, but also in the relatively small details of temporally restricted structures. The alleged invariance in prosodic stress group patterns in Danish is thus a perceptual, rather than an articulatory invariance.

REFERENCES

Dyhr, N. (1990) "The activity of the cricothyroid muscle and the intrinsic fundamental frequency in Danish vowels", *Phonetica 47*, pp. 141-154.

Fischer-Jørgensen, E. (1990) "Intrinsic Fo in tense and lax vowels with special reference to German", *Phonetica 47*, pp. 99-140.

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

Petersen, N.R. (1978) "Intrinsic fundamental frequency of Danish vowels", *Journal of Phonetics 6*, pp. 177-189.

Petersen, N.R. (1979) "Variation in inherent Fo level differences between vowels as a function of position in the utterance and in the stress group", *Annual Report of the Institute of Phonetics, University of Copenhagen 13*, pp. 27-57.

Petersen, N.R. (1980) "Coarticulation of inherent fundamental frequency levels between syllables", *Annual Report of the Institute of Phonetics, University of Copenhagen 14*, pp. 317-354.

Petersen, N.R. (1986) "Perceptual compensation for segmentally conditioned fundamental frequency perturbation", *Phonetica 43*, pp. 31-42.

Thorsen, N. (Grønnum) (1980) "Neutral stress, emphatic stress, and sentence intonation in Advanced Standard Copenhagen Danish" *Annual Report of the Institute of Phonetics, University of Copenhagen 14*, pp. 121-205.

Thorsen, N. (Grønnum) (1982) "On the variability in Fo patterning and the function of Fo timing in languages where pitch cues stress", *Phonetica 39*, pp. 302-316.

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

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

B. Granström & L. Nord, eds.
Almqvist & Wiksell, Stockholm 1993

PERCEPTUAL INVARIANCE IN DANISH STRESS GROUP PATTERNS

Nina Grønnum

INTRODUCTION

This paper summarizes observations made in previous investigations and tries to see what impact they may have on the account of the production and perception of the tonal patterns associated with the prosodic stress group in Standard Danish. I should point out from the outset that I have *not* carried out the perceptual experiments that would have been natural follow-ups to the acoustic observations. - For a general account of Danish intonation, see Grønnum (1992).

THE PROSODIC STRESS GROUP

The stress group pattern in Standard Danish

A prosodic stress group in Standard Danish consists of a stressed syllable and all succeeding unstressed ones, independently of intervening syntactic boundaries (within the same intonation contour). This unit is the carrier of a recurrent tonal pattern, which in Standard Danish has a low falling start, a quick rise, which generally reaches its peak during the first post-tonic syllable, and a fall during succeeding post-tonics. It is important to note that syntactic boundaries cannot be traced in this pattern. It reaches across any boundary within the simple sentence. Thus the pattern in Fig. 1 would describe any one of the three underlined sequences in HÅNDBOLDSPIL ER MEGET ANSTRENGENDE (handball-playing is very strenuous); BANANERNE I KASSEN ER RÅDNE (the bananas in the box are rotten); ALLERGI ER EN INFAM SYGDOM (allergy is an infamous desease) (stressed vowels are in fat script. Note that the prosodic stress group onsets with the stressed vowel, and any syllable initial consonants attach to the preceding stress group, cf. Thorsen 1984b). The course of this pattern is independent of the segmental composition of the stress group, i.e. it is independent of the duration of the stressed vowel and independent of the voicing of consonants. The segments simply line up on the pattern (or leave holes in it if unvoiced) according to their duration, as exemplified in Fig. 2 which depicts (stylized) Fo patterns of DAMEFRAKKE (ladies' overcoat) and ANSTRENGENDE (strenuous). This is important for the argumentation to follow.

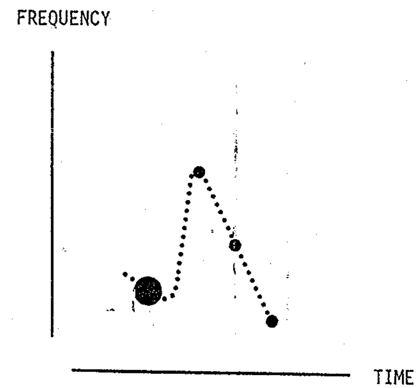

Figure 1. *Prosodic stress group pattern in Standard Danish*

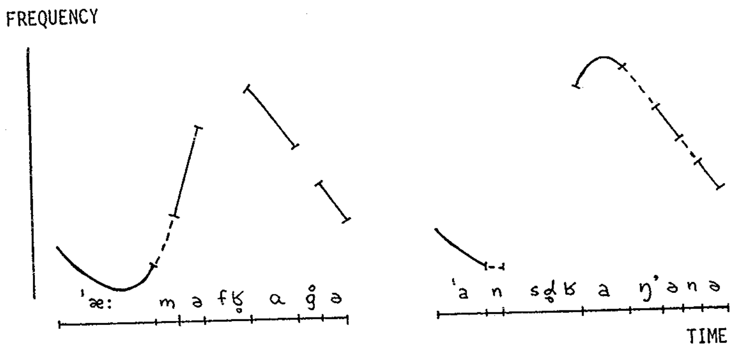

Figure 2. *Two prosodic stress groups with different segmental structure - see further the text.*

When the stress group is too short to complete af full rise and fall, the Fo pattern is truncated, and in the extreme case - a monosyllable with a short vowel surrounded by unvoiced consonants - all that is left of the pattern is the brief initial fall, cf. Thorsen (1982, 1984a). In other varieties of Danish the stress group pattern is different, and there are also different strategies for handling variation in the total duration of the stress group, cf. Grønnum (1992, pp. 63-65).

The rise from low to high

The stress group pattern is in most variants of Danish qualitatively constant, i.e. you find the same low-plus-high-falling pattern (or whatever the pattern may be) in all positions, disregarding cases where extra prominence of some kind occurs, cf. Grønnum (1992). But the pattern varies quantitatively. This is particularly so for the magnitude of the rise from low to high, cf. Fig. 3. (The following statements should all be read under a 'ceteris paribus' condition:) (1) The rise varies with the quality of the stressed vowel, being smaller after high than after low vowels. (2) The magnitude of the rise varies with position in time: the rise diminishes progressively from start to end of the utterance or phrase. (3) The magnitude of the rise varies with the intonation contour it rides upon, being smaller on more declining contours.

Intrinsic Fo differences between vowels of different jaw/tongue height

It is a very well established fact that vowels with high versus low jaw/tongue positions have intrinsically different Fo levels. There is a vast literature on the subject which space does not allow me to review. Readers are referred to Fischer-Jørgensen (1990) and the references there. Intrinsic Fo differences between high and low vowels have been attested also in Danish, cf. Petersen (1978, 1979, 1980, 1986) and Dyhr (1990). Petersen (1979) found that the difference is larger in stressed than in unstressed vowels, and it is larger early in declarative utterances than late, ceteris paribus. These differences are not negligible. They amount in stressed vowels to 10-12 Hz in male and 20-25 Hz in female voices, respectively (Petersen 1978), corresponding to about 2

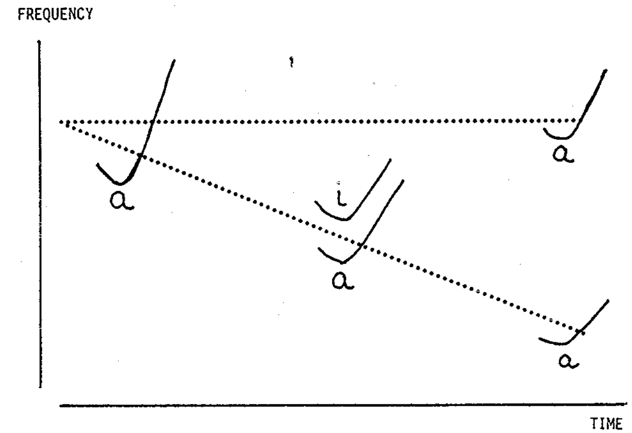

Figure 3. *Examples of stress group pattern rises as a function of the quality of the stressed vowel, position in time and position on more or less sloping contours.*

semitones, which is well beyond the 'just noticeable difference' for perception of pitch differences in the frequency range concerned. Petersen (1986) established that these Fo differences - though perceptible - are compensated for, parsed out, in the decoding of the signal by the listener, so that an [i] with an objectively higher Fo than an [a] is heard as having the same pitch as [a] (granted, of course, that the difference corresponds to what we measure in actual speech), or - inversely - that an [i] and an [a] with the same Fo are heard as having different pitches, the [a] appearing higher than the [i]. He found, furthermore, that when spectral cues to vowel quality are ambiguous, higher Fo lends preference to identification with higher vowel quality, and inversely. Thus, if the spectrum is ambiguous between [u] and [o], higher Fo will yield more [u]-identifications, and vice versa. There is thus no doubt that intrinsic Fo differences are an acoustic as well as perceptual reality. They are something speakers produce and listeners interpret.

Several explanations have been offered in the literature over the years, some based on articulatorily constraints, some on acoustic facts, and some based on perceptual considerations, explanations which are summarized in Fischer-Jørgensen (1990). If the phenomenon derives from articulatory or acoustic constraints, the intrinsic Fo differences between different vowel qualities must be regarded as automatically conditioned by the peripheral speech production mechanism or by impedance relations between the supra- and subglottal systems. They are thus beyond the speaker's immediate control. If the phenomenon has a perceptual basis, it must be considered speaker controlled: it is something the speaker produces to achieve a certain perceptual goal. An 'automatic' explanation - on the one hand - is not tantamount to saying that the differences must therefore be imperceptible and that the listener accordingly cannot compensate for them in his decoding of the signal: the intrinsic Fo differences are produced involuntarily by the speaker but they are perceived by the listener, who proceeds, however, to parse them out, to ignore them. *Because* the listener is also a speaker and thus "knows" that, when he himself intends to produce an [i] and an [a] on the same pitch they come out with different Fo's, then - in analogy - he will also "know" that an [i] and and [a] with (appropriately) different Fo levels were intended to have the same pitch by the speaker, and it is thus he will obligingly interpret them. On the other hand, a perceptually based explanation does not contradict the fact that the phenomenon may derive originally from articulatory constraints. Somewhat simplified: increased jaw and/or tongue height yields higher Fo, ceteris paribus, which has a favourable influence on the perception of vowel quality, the higher Fo enhancing the low F1 of high vowels. Thus, the physiologically conditioned Fo difference is perceptually useful. When these facts become embodied in speakers' and listeners' (unconscious) linguistic knowledge, speakers may actively utilize the Fo differences which articulation (or impedance relations) creates, and perhaps they may even exaggerate them, i.e. carry the magnitude beyond the values which physiology or acoustics dictates, in the same way that in English increased vowel duration before voiced sounds has been enhanced beyond what mere articulatory constraints prompt. In brief: whatever the reason, or reasons, since several factors may collaborate, we can state that an [i] becomes more [i]-like if it is produced on a higher Fo than an [a].

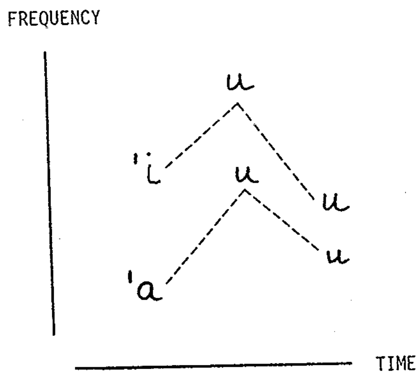

Figure 4. *Stylized stress group patterns with different stressed vowels - see further the text*

Assimilation of intrinsic Fo differences

Petersen (1980) found that intrinsic Fo differences in stressed vowels spill over to succeeding post-tonics. I.e. the same post-tonic vowel, say an [u], has a higher Fo after an [i] than after an [a], cf. Fig. 4. The difference is smaller than the difference between stressed [i] and [a], but it is statistically significant. In the second post-tonic vowel there are reminiscences of the difference between the preceding stressed vowels, but here the difference is no longer statisticallly significant. These facts have consequences for the magnitude of the rise after the stressed vowel. The rise from low to high has several functions. It signals stress, i.e. it singles out the low syllable as the stressed one. And everything else being equal, the *magnitude* of the rise will signal the *degree* of prominence on the stressed syllable: higher rises lend more prominence. Thus, for a given degree of prominence, say normal neutral main stress, the rise must - for a given position in time and for a given intonation contour slope - have a certain perceived magnitude (leaving room for some variability, of course).

I will assume that if the listener compensates for, parses out, intrinsic differences in the stressed vowel, he will also parse out the effect these differences have on succeeding vowels. Petersen (1980) offered a tentative explanation for this spill-over

to succeeding post-tonics in terms of an attempt by the speaker to maintain a rise from low to high of a magnitude which ensures that the stressed vowel still be perceived as stressed. I.e. he saw it as governed by perceptual demands, namely the need for appropriately signalling stress location and degree of prominence. I do not think this is very likely. It would make good sense, *if* the intrinsic difference between the two stressed vowels was transferred 100% to the post-tonic, i.e. if the rise after [i] was of the same magnitude as after [a], but this is not the case. I think it more likely that we are having to do with true coarticulation, or assimilation. Petersen's explanation is also contradicted by the fact that we *do* compensate perceptually for intrinsic differences in stressed vowels. Granted that we do that, i.e. if we hear the pitch of [i] as being lower than its Fo dictates, then we should also, as a matter of consequence, interpret the physically smaller rise from stressed [i] to a post-tonic [u], the smaller interval, as being larger than it is physically. Accordingly, there would be no need for a controlled spill-over. If - on the contrary - we interpret the effect on the post-tonic vowel as a true coarticulation, it becomes comprehensible (1) that a post-tonic vowel should exhibit differences in Fo depending on the height of the preceding stressed vowel and also (2) that the intrinsic difference is not transferred 100% to the post-tonic, it grows progressively smaller until the function dies out. Thus, whether the difference between stressed [i] and [a] be controlled or automatic, it is - like so many articulatory phenomena - not a temporally very sharply defined on/off function. Once the process has been initiated it exhibits a certain inertia, as in all other progressive assimilation.

Recapitulate: *if* the intrinsic Fo difference between stressed vowels were assimilated 100% to the post-tonic, we would expect the rise from stressed [i] to post-tonic [u] to be of the same physical magnitude as that from stressed [a] to post-tonic [u], and we would expect these rises to be perceived as having the same pitch interval, but the whole Fo course would be positioned higher up in the speaker's range. If *no* assimilation to the post-tonic occurred, we would expect the rise from stressed [i] to post-tonic [u] to be smaller than the rise from stressed [a] to post-tonic [u], and the difference would equal that between stressed [i] and [a]. If assimilation is only *partial*, the rise from stressed [i] must be smaller than the rise from stressed [a], but the difference in magnitude must be less than the difference between stressed [i] and [a]. The difference in the magnitude of the rise could be brought about either by adjusting the slope of the rise (cf. Fig. 5a) or by manipulating its duration (cf. Fig. 5b).

My own data (from 1984a) seem to indicate that it is the situation sketched in Fig. 5b we are dealing with, cf. Fig. 6. I.e. the smaller rise after high vowels is accomplished by arresting the rise earlier in time, rather than by adjustment of its slope. Note that the differences observed in Fig. 6 are *not* due to differences in duration between (shorter) high and (longer) low vowels. Firstly, we are well into the post-tonic syllable when the high turning-point is reached. Secondly, there is nothing to indicate that turning points in the Fo pattern are lined up with certain segment boundaries, cf. Fig. 2. It is easier to account for the shorter and thus smaller rises after high vowels as an attempt on the part of the speaker to keep the *perceived* rise of constant magnitude, in face of the fact that speakers and listeners compensate for intrinsically determined Fo differences in the stressed vowels, and granted also that these differences are partially assimilated to the succeeding post-tonic vowel. This supports an assumption that we *are* in fact dealing with a speaker-controlled phenomenon, an adjustment to the demands that the listeners' interpretation of the signal make.

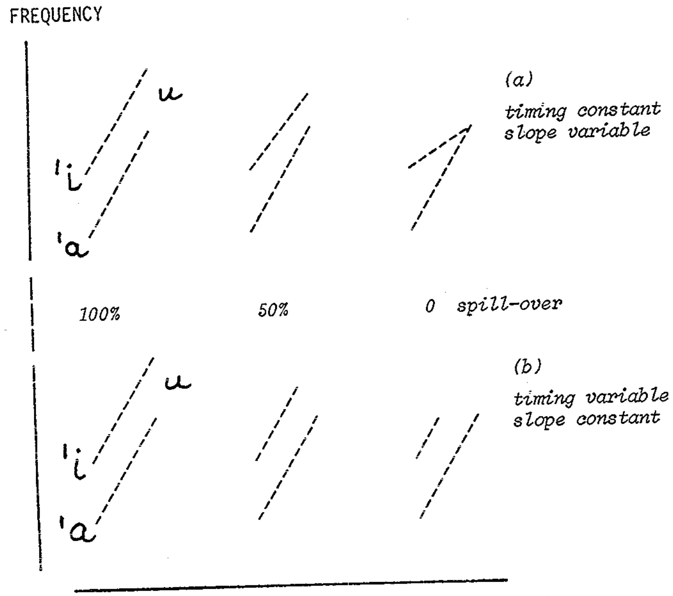

Figure 5. *Stress group pattern rises under different assumptions - see further the text.*

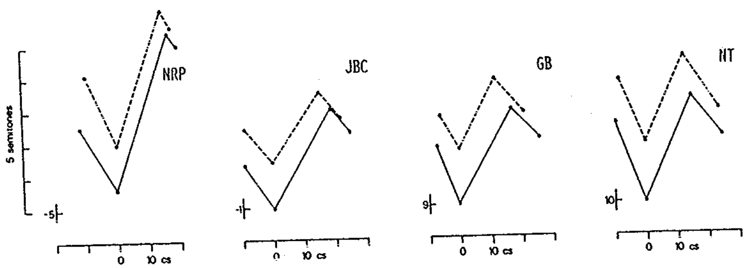

Figure 6. *Average Fo patterns of six words with low stressed vowel (solid lines) and six words with high stressed vowel (dashed lines), by four speakers.*

Intrinsic Fo and position in the utterance

Petersen (1979) observed that intrinsic vowel differences are smaller later in the utterance than earlier. Therefore the difference in rise magnitude will also be smaller, cf. Fig. 7. He saw this as an expression of an underlying attempt to produce rises of

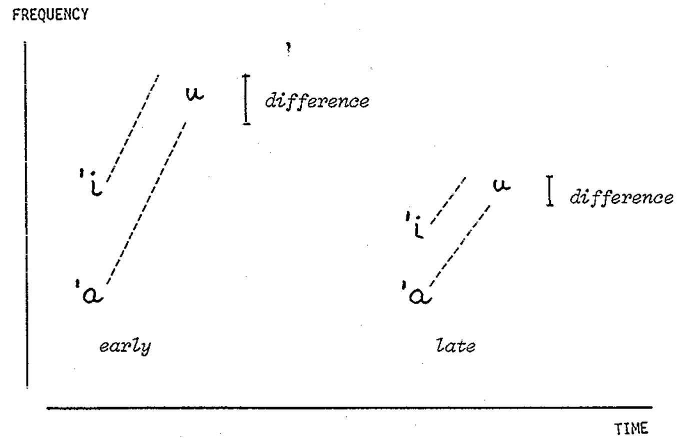

Figure 7. *Stylized stress group pattern rises as a function of stressed vowel height and position on the intonation contour.*

constant magnitude, but suggested that in initial position we are dealing with a target undershoot in the rise from stressed [i]. I cannot wholly ascribe to that explanation either. Firstly, in normal declarative utterances the first high peak is well below the upper limit of the speaker's range, and speakers easily accomplish high rises initially in the utterance. Secondly, there are several indications that articulatory gestures are more distinct initially in the utterance than finally, a fact which should favour maximal rises initially. However, it is of no consequence for the matter at hand - the variation in the magnitude of the rise from stressed to post-tonic vowel - whether its variation be speaker controlled or due to articulatory constraints. Given (1) that intrinsic Fo differences between vowels of different tongue/jaw height exist, (2) that they prove to be compensated for, parsed out, perceptually, and (3) when these differences furthermore assimilate partially to the succeeding post-tonic, then these facts demand of the speaker that he produce rises which - in accordance with the embodied linguistic knowledge of speakers and hearers - are perceived with the adequate pitch intervals. It seems then as if - for a given position in time - the speaker is less likely to change the rise of the slope than to change the timing of the high turning point. This last observation - that speakers apparently adjust the duration of the rise rather than its slope - may find its explanation in the fact that, ceteris paribus, rises grow progressively less steep through the utterance, cf. Fig. 3. I.e. for a given position in time, on a given intonation contour slope, the slope of the rise from stressed to post-tonic is to be held constant, and that leaves only rise duration to be adjusted if its magnitude is to be varied.

I have suggested previously (Thorsen 1980, Grønnum 1992) that progressively less steep and smaller rises through the utterance are to be seen as an expression of the fact that all articulatory gestures - including prosodic ones - are less distinct at the offset than at the onset of the utterance. Like other "natural" articulatory phenomena this too becomes part of the linguistic code, part of the signal for utterance onset and offset, and

[last page at the front]