SUPERPOSITION AND SUBORDINATION IN INTONATION A NON-LINEAR APPROACH

Nina Grønnum
Institute of General and Applied Linguistics
University of Copenhagen

ABSTRACT

After a brief presentation of the preliminaries and my presuppositions, a model of speakers' production of standard Danish intonation is presented. Its basic property is the layered, superpositional organization of its components - where the manifestation of components at lower levels of the linguistic hierarchy is subordinate to components at higher levels.

INTRODUCTION

A symposium at the quadrennial international congress of phonetic sciences is an appropriate occasion to remind oneself of and clearly state the underlying assumptions, the tacit goals, the ultimate ambition, and the implicit restrictions of our descriptions of intonational structure. They almost certainly are not identical across authors, and so mine are presented below. They are to be understood as programmatic in nature, rather than axiomatic, and - for reasons of space - they are also rather telegrammatic.

Since other views of intonational structure are presented and summarized by the other participants at this symposium I shall concentrate on my own, and only note in passing that similar views about intonation as a layered structure are to be found in [3, 5, 6, 8, 11, 12, 15]. Furthermore, the documentation for the descriptive adequacy of the model does not find room here and will have to be sought in previous publications, all of which are referenced and most of which are summarized in [7].

PRELIMINARIES

Intonation

'Intonation' encompasses all the linguistically relevant, suprasegmental, non-lexical aspects of the fundamental frequency (F_0) variation - or its perceptual correlate, the pitch variation - through the course of spoken utterances. 'Suprasegmental' removes the intrinsic F_0 characteristics of segments and their coarticulation. The segmentally induced variation is generally supposed to be beyond the speaker's conscious control, although it is equally generally assumed to be perceptually relevant for the identification of segments, and probably should be included in the low level rules for synthetic speech, if it is to sound natural. 'Non-lexical' excludes syllable tones and word accents.

The crucial term here is 'linguistically relevant'. But the line cannot always be unambiguously drawn between linguistic and para-linguistic in intonation. Prototypical functions in either domain are easily established: linguistically relevant is the cuing of (1) various types of prominence (reduced stress, normal stress, default sentence accent, focal accent, emphasis for contrast), (2) prosodic boundaries at various levels and (3) speech act function (imperative, declarative, interrogative). Typical para-linguistic meaning conveyed by certain aspects of F_0/pitch variation (inter alia) will be everything that characterizes an individual speaker, like sex, age, and present state of health and mind. But - to mention just one example - how to classify speaking style? Is that a linguistic or a para-linguistic parameter? Or is that not a decision which should be made universally? I have excluded speaking style from amongst the intonational parameters of my model, but more out of necessity - for want of relevant data, than for any more principled reason. -- Beckman [this symposium] takes a different attitude to para-linguistic aspects.

Models

Intonation can be modelled from various perspectives, for various purposes: we can aim at speakers' production or their perception; or models may be adapted to the demands of synthetic speech or automatic speech recognition,

respectively, procedures which do not necessarily parallel human processes. Furthermore, speaker and listener behaviour may be modelled at various levels of abstraction. All of us, in this symposium, have been mainly concerned with the modelling of production data, and mainly acoustic data at that, and the following is limited to models of human production of intonation.

To be a model, a description of intonation must do more than merely depict or replicate singular events, individual items and individual speakers - it must at least be a generalized account, generalized over typical speakers of a given speech community speaking in a given speech style and also over different exemplars of a given utterance type. I think it uncontroversial to also demand that it account not only for available data, but correctly predict new utterances as well. - Over and above that, the level of ambition may be highly variable, in terms of multiplicity of input to the model and also in claims about its universality. - Models, accordingly, are representations or formulae which mediate, back and forth, between a stage upstream in the human production of linguistic utterances where all the morpho-syntactic and lexical information has been supplied, and the next one where significant and distinctive intonational information is inserted, and from which the phonetic F_0 implementation can be derived by rule at a lower level and presumably be translated into neural commands to activate the physiological production system, alternatively into commands to drive a speech synthesizer. This level is often referred to as the 'phonological level'. It is an unfortunate term, however, with its connotations of 'minimal units to bring about a difference in (lexical) meaning' and - particularly - 'the double articulation' of language. Firstly, I do not think it feasible or expedient to phonologize differences in F_0 or pitch contours which are merely the acoustic or perceptual correlates of a contrast in another linguistic dimension, namely stress. Secondly, outside the realm of stress, intonation does not differentiate but carries meaning, though not of a lexical sort, cf. above. But it does not do so autonomously, only in an intricate interplay with syntax, semantics and pragmatics. Finally, I think it highly probable that intonation, except explicit local boundary phenomena, is perceived in relational terms as holistic gestalts. All this makes an analogy to paradigmatic segmental contrast rather forced. If 'intonational phonology' is merely synonymous with 'abstract representation of intonation', then the latter is perhaps a more fortunate term.

Universals

The general search for universals in linguistics has had an impact in intonation research as well. I am perhaps more sceptical than most about statements to the effect that, e.g., "all languages have boundary tones; all languages have sentence accents; questions end in final rises", etc. Languages sound so vastly different, intonation-wise, and I do not see why these huge impressionistic differences should not be reflections of fundamental principled differences in the abstract representation of their intonation, in the underlying components which speakers manipulate and the way they are organized. But to the extent that a model can be adapted without undue complication to typologically different intonation systems, it is of course the more powerful one.

Adequacy

There are several routes to models of intonation, from divine inspiration, through qualified introspection if the language is a familiar one, to laborious and time-consuming analyses of acoustic data, or a mixture of these. Likewise, there are several measures of a production model's adequacy. First of all, does it produce an acceptable output? (Note that it may do so without laying any claims to psychological reality.) This can be tested in speech synthesis. Then, how likely is it as a representation of speakers' behaviour, granted what we know about the physiology of F_0 production and other aspects of speech? How cognitively real is it? How well does it

lend itself to typological research, i.e. how easily are parameters added or deleted or modified without debilitating the model's functioning?

A MODEL OF STANDARD DANISH INTONATION

The model presented here is derived from acoustic analyses of a considerable amount of speech material from a fair number of speakers, and some perceptual experiments. It is a generalized description of the central linguistically relevant aspects of F_0 contours in short texts in informal but distinct monologue. It is also a hypothesis about speakers' internal representation, about the nature of the components and how they interact to turn a string of semantically coherent sentences into a series of prosodically coherent utterances. It lays no claim to universality, though by uncomplicated incorporation of extra parameters and by proper quantification and adjustment, it will account for a rather rich variety of Danish regional languages. It produces acceptable sounding synthetic speech.

The ultimate ambition, of course, is to be able to account for any style of speech and any syntactic structure in Danish. -- The model's strongest present limitation is its restriction to informal but distinct monologue, i.e. one-way communication, and read speech at that, based as it is on analyses of speech produced under laboratory conditions. However, two fundamental features, the recurrence of the F_0 pattern associated with stress and the quasi-rectilinear slope of utterance intonation contours have been demonstrated also in informal spontaneous speech in interviews ([4]), and I think it justified to assume that the model can serve at least as point of departure for the description of Danish speakers in any speech style. -- Syntactic boundaries below the sentence level are not incorporated in the model either. Again, this is not from any a priori theoretical preclusion of their relevance, but for lack of relevant data. From an early study it appeared that syntactic boundaries within simple, though long, sentences have no direct reflection in the intonation, but obviously there are prosodic boundaries affiliated with a number of syntactic ones in complex sentence structures, and the model should be expanded accordingly when the necessary experiments have been performed.

The textual contour

Short texts are characterized by a gradual global descent, the textual contour, *T* in the figure. Its onset and offset are defined by the first and last stressed syllable in the first and last utterance, respectively. It typically spans half an octave. Superposed on *T* are the individually sloping utterance contours, *U1-3*. Utterance onsets (defined by the first stressed syllable) lower gradually through the text and typically span 3 semitones. Utterance offsets (defined by the last stressed syllable) likewise lower gradually but typically span only 2 semitones, to the effect that utterance slopes are slightly steeper initially than finally in the text - in utterances of equal length and function. In texts of more than four utterances the medial part of the textual contour levels out. It is unreasonable to expect that speakers be able to manipulate lowering of successive utterance onset to a degree finer than 1 semitone, particularly in view of their relatively large mutual temporal distance.

Coordinate main clauses are less slanted relative to the global textual contour than are a sequence of terminal declaratives (not depicted in the figure).

The utterance contour

Utterance contours are defined solely by the string of stressed syllables, because: (1) Local rise-falls depend for their existence upon the presence of unstressed syllables in the stress group. In a succession of stressed syllables there is no upwards deflection of the F_0 course between them. (2) The stressed syllables are frequency scaled in relation to each other without regard to the presence or not of any rise-falls in the surroundings. For a given intonation type (cf. below), the range spanned by the contour is constant. In utterances of up to five stress groups they are equidistantly spaced in frequency, in intervals which are inversely proportional to their number.

But their timing will depend on stress group length (syllable structure and number of unstressed syllables), cp. *a1* and *a2* in *U3*. The local deflections (the 'highs') in the F_0 course have no independent role in the shaping of grosser trends in F_0 contours.

Further characteristics of utterance slopes are: If the utterance is long, its contour will not lower continuously but be broken into prosodic phrase contours, with a slight resetting between them, cf. *U2*, granted that the break does not cut up the utterance in an unacceptable manner, cf. below. The organization of the phrase contours relative to the superordinate utterance contour is analogous to the organization of utterances in the text: phrases descend along the utterance contour while each phrase is associated with its own slope, so the first phrase onset (defined by the first stressed syllable) is higher than the last phrase onset, and the first phrase offset is higher than the last phrase offset, cf. *P1* and *P2* in *U2*. Intermediate phrases have intermediate onsets and offsets, but above four prosodic phrases the utterance contour must level out medially, so as not to frustrate speakers' control over step down magnitude between phrase onsets.

As mentioned above, the syntax-prosody interface in complex sentence structures is largely unexplored in Danish, though see [14]. But the following is valid for simple sentences in isolation or in combination. (1) The syntactic structure of short sentences is not reflected in their intonation contour. The order of constituents does not matter, nor their internal structure. (2) Longer utterances are produced as a descending sequence of sloping prosodic phrases, but the conditions governing the location of the breaks are complex. (a) A prosodic phrase must contain at least two stress groups. (b) Prosodic phrases tend to be of equal length. (c) But this tendency is easily overruled: the prosodic boundary cannot occur within a syntactic constituent, nor between syntactic constituents which are semantically coherent. Thus, "Der går mange store Røde Kors busser til Grosny i Tjetjenien i aften." (Many big Red Cross buses depart for Grozny in Chechenya this evening.) will most likely be produced in one sweep, in spite of its length, and in spite of the formal boundary before the place complement, because buses are intimately associated with their destination (or their point of departure). (3) Coordinate main clauses are more likely to be separated by a resetting of the intonation contour than subordinate and main clause constructions (irrespective of their ordering). (4) Individual utterance contours are steeper, also the text final one, and demonstrate a greater amount of resetting between them, in a succession of terminal declaratives than in a corresponding string of coordinate main clauses. This difference, induced by different lexico-syntactic boundary conditions, at least hints at a solution to Ladd's *ABC*-problem [this symposium].

Utterance contours vary between most steeply sloping (in declarative sentences used conventionally) and horizontal (in sentences which are not marked lexically or syntactically for their interrogative function). Other questions and non-final clauses fill in the intermediate space, with a clear tendency for a trade-off between lexical/syntactic markers of their function and the slope of the intonation contour, cf. *a*, *b* and *c* in *U1*. The steepest contours typically span 4 semitones initially in the text and 3 semitones finally. The choice of contour slope for a given utterance is determined by syntactic and - not least - pragmatic factors, in accordance with, I propose, principles of markedness and typicality as follows: By definition, unmarked intonation is associated with syntactically unmarked sentences used conventionally. That makes the steepest contour - which accompanies conventional declaratives - unmarked and any less falling contour marked. Typical intonation is the contour which accompanies any given sentence type when it is used conventionally. Thus, a conventional Yes/No question will have a slope somewhere between unmarked and the horizontal, i.e. it is marked but typical. Any deviation from the typical intonation will have im-

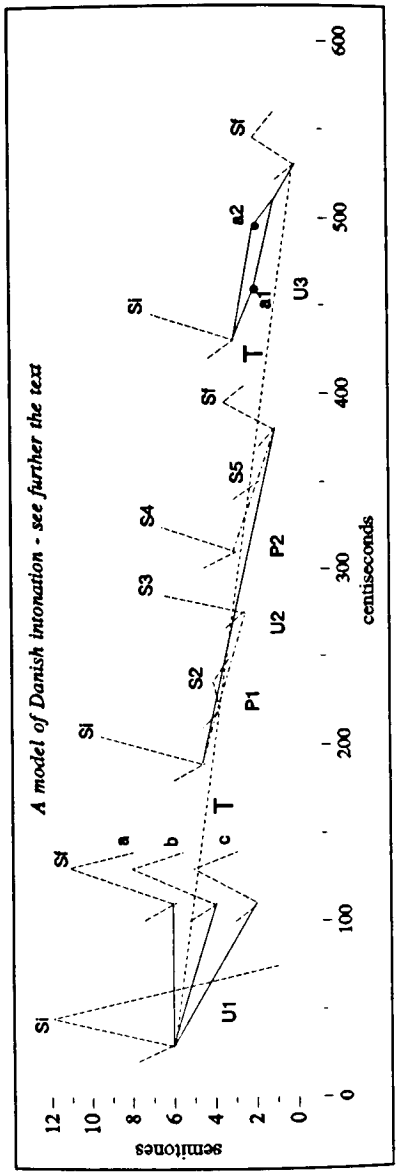
plications for illocutionary force and go counter to conventional usage. Thus, the most strongly marked intonation contour, the horizontal, will turn any sentence into an interrogative speech act.

The stress group pattern

F_0 is the most explicit among the acoustic cues to stress and prominence. The onset of any stressed vowel coincides with the onset of a recurrent melodic pattern which extends over all succeeding unstressed syllables within the same sentence, irrespective of their morphological or syntactic affiliation, until the onset of the next stressed vowel. This stress group pattern is subject to truncation or extension: A maximally developed pattern describes a brief initial fall succeeded by a steep rise to the first post-tonic and a steep fall through succeeding post-tonic syllables, cf. the initial stress group in *U1*. The shorter the stress group, the less extensive the F_0 pattern, so with a short stressed vowel and absence of post-tonic syllables, all that remains is the slight and brief initial fall, cf. *S5* in *U2*.

The implementation of F_0 patterns is extensively sensitive to their prosodic environment. (a) Rises are higher on marked contours, ceteris paribus, cp. *a*, *b* and *c* in *U1*. (b) Rises are successively lower from initial to final utterances, ceteris paribus, cf. the initial and final stress groups, respectively, in *U1-3*. (c) Rises lower progressively through an utterance, but any further differentiation according to prosodic phrasing - in the shape of higher post- than pre-boundary rises - is absent, cf. *S3* and *S4* in *U2*. (Either because it is too taxing for the production system or because a prosodic boundary per se is not intended.)

It is hard to say how much of this variation is directly speaker controlled, introduced by phonetic rule, and how much can be ascribed to general speech production principles which reduce articulatory explicitness through time or in unmarked vs. marked contexts. But whichever the output control mechanism, the variation is predictable, and though stress group patterns are an integral part of a model of speaker performance, they



are not part of the abstract underlying representation properly speaking. It is equally hard to determine the degree to which stress group pattern variation is perceptually relevant or redundant. In principle it can be entirely automatic and

yet perceptually relevant, whereas the reverse - rule governed but perceptually redundant - is perhaps less likely. I opt for perceptual relevance, because: The patterns described above pertain to utterances where the stressed syllables are equally prominent perceptually. When prominence varies, within the realm of normal stress, so does the magnitude of the rise: the higher the rise, ceteris paribus, the more salient the stressed syllable. Of course, this can only work on a background of *expected* neutral rise magnitudes in the various contexts.

Focus, on the other hand, is cued by suppression of the succeeding rise-fall, cf. *S2* in *U2*. Suppression of both preceding and succeeding rise-falls will create an emphasis for contrast (not displayed). -- In spontaneous speech focusing can also take another shape: the whole stress group is lifted out of (above or below) the contour, but the stress group patterns are not modified ([4]).

Note that syntactically or prosodically determined sentence accents, in the shape of a particularly prominent F_0 excursion finally (whatever the constituent), are absent in Standard Danish.

Intonation cues are global, not local

Standard Danish does not exhibit specific local tonal cues to either speech act function or boundaries (whatever the unit), in terms of final highs or lows. Intonational markedness and completion are inherent in the global course of utterance contours, possibly in conjunction with the derived variation in the magnitude of stress group pattern excursions. This property is shared by the majority of regional Danish variants.

Subordination and superposition; look-ahead and non-locality

Why this layered system of simultaneous, interacting, non-categorial intonational components of varying structural and temporal scope, where larger scope components carry and set the scale for smaller scope ones? And where the implementation of F_0 events is performed on the basis of upcoming as well as preceding events and is sensitive to syntactic and semantic structuring? Why not an abstract representation in terms of a linear sequence of categorially different, non-interacting pitch accents whose manifestation is exclusively locally determined, implemented on a left-to-right basis? And where there is no intonation component separate from the accents, and where global trends thus are the result of iterative application of local downstep rules ([1, 10, 13]) or a downstep morpheme ([2]) or locally determined range reduction ([9])? It would definitely be computationally simpler.

First of all 'pitch accent', with its connotation of phonologically distinct differences, is not an appropriate term for stress group patterns in Danish, because (a) there would be only one category and it would always align in the same fashion with the segmental material and (b) its phonetic manifestation is entirely predictable. In other words, a speaker cannot - at least not in the speech style analyzed so far - make a choice between various types of pitch accent. And when the magnitude of stress group pattern rises is manipulated to cue varying degrees of prominence, we are dealing with a scalar, not a binary, phenomenon. Speakers are not making a choice in those circumstances either of a particular pitch accent from a set of phonologically distinct ones, but are simply subordinating the manifestation of stress to the demands for signalling more or less prominence. -- I have seen Bornholm speakers invert their slowly falling-rising F_0 patterns across a whole utterance with a resulting change in perceived speech style or register, and I am certain a similar mechanism, a 'long component' or 'setting', is operative in Standard Danish. But that does not prove the existence of another pitch accent. It only shows that the manifestation of stress interacts with parameters at other (para-linguistic?) levels of description. Finally, if pitch accent were phonological stress would not be, a proposal with no serious merit. So, once more: stress group patterns are not part of the abstract representation of intonation. Therefore, with proper scaling the model will cover the majority of regional vari-

ants of Standard Danish as well, namely all the 'global' types, cf. above. The principal difference among them is in the shape and the magnitude of the stress group pattern and its alignment with segments. Whatever the patterns, they are superposed on a contour defined by the stressed syllables, without interfering with its course, but the intonation contour (its location in the text, its slope, its length), on the contrary, is decisive for the manifestation of stress group patterns. Furthermore, the manifestation of prominence degrees above normal stress demonstrates yet more syntagmatic interaction: stress group patterns are shrunk before emphasis for contrast, but not before a mere focus, i.e. the realization depends on the nature of the succeeding stress group. If all of this is not a paragon of superposition and subordination, I do not know what is.

Secondly, utterance and text intonation contours cannot be computed on a purely left-to-right basis either. As mentioned above, slope varies in inverse proportion to length, or - in other words - stressed syllables in an utterance are more closely spaced in frequency in longer utterances, or - in yet other words - the frequency location of a stressed syllable is sensitive to the number of succeeding ones. So, e.g., the second stressed syllable is higher in an utterance of four than one of three stressed syllables, ceteris paribus. A similar look-ahead and pre-planning is operative in the combination of utterances into prosodically coherent texts, cf. above. In this manner, utterance and text intonation are characterized by compression and expansion in time (contrary to stress group patterns which are truncated or extended), a process which is inconceivable without pre-planning. It is also attested in the difference between co-ordinate main clauses vs. a sequence of terminal declaratives, the former being less slanted relative to the global textual contour. -- Look-ahead and interaction turn up in the temporal structure as well: Speaking rate is somewhat accelerated before a (non-initial) focused item.

Thirdly, intonation contours, except the maximally marked one, are *always* associated with a more or less negative slope. Thus, e.g., there are no declaratives without a global downtrend.

These facts should make an account of sentence intonation in Danish in terms of (varying degrees of) local downstep or range reduction, triggered by certain pitch accent configurations, a formal possibility but empty of the significance it carries in, e.g., tone languages. -- Note that the superposition principle does not preclude features of a local kind, like Ladd's 'edge tones' [this symposium]. Thus, there are Danish regional varieties that definitely have a very local final high or low (as the case may be) which does not interact with what precedes it. The point is that edge tones are not universal, and when their corresponding function is carried by global trends, as in a number of Danish varieties, the linear representation is at a loss. Whereas their existence and incorporation is rather straightforward in a superposition model, precisely because they occur at domain boundaries. Likewise, a hierarchical organization does not, of course, preclude lexical tonal differences, cf. [6].

To sum up: the superpositional and sequential models do not differ in the acknowledgement of look-ahead, but they differ in its representation. For example, in the linear sequence model longer utterance onsets are higher than shorter ones but pitch accent relations are unaffected. In the superpositional approach utterance onsets need not vary, but the stressed syllables are less descending, the slope is less steep in longer than in shorter utterances. In [10 (p. 231)] the authors stated that a pitch accent can only look back to a previous pitch accent, a phrase to a previous phrase, etc., and apparent instances of anticipation should be explained by, e.g., feature spreading or temporal overlap in the realization of the segments in question [which really only amounts to passing the buck, NG]. The hierarchical concept entails a more direct interaction between subconstituents and superordinate structures.

Psychological reality?

A linear representation is computationally somewhat simpler and thus easier to implement in rule synthesis, than layered structures involving look-ahead, but that is not necessarily indicative of how speakers and listeners process intonation. And look-ahead and pre-planning are definitely operative and evidenced in other aspects of speech (syntax, slips of the tongue, 'phrasal stress reduction'), so why not in intonation?

Secondly, we are both speakers and hearers, and although we do not a fortiori produce and perceive in the same terms, it is at least not unlikely. I hypothesize that utterance intonation in Danish is holistically conceived, and that the concept is cognitively simpler, to the speaker and listener, than the summation of its atomic elements (the local ups and downs). Anecdotal evidence hints that this hypothesis would be worth testing: When linguistically naive Danes are asked to characterize the melody of various Danish utterances, they typically provide overall shapes. They have to be pushed hard - with exaggerated patterns - to hear that there are local pitch movements associated with stressed syllables. The local humps are not conceived as part of the melody and listeners seem to disregard the contribution to pitch that comes from stress.

REFERENCES

[1] Beckman, M. and J. Pierrehumbert (1986), "Intonational structure in English and Japanese," *Phonology Yearbook*, vol. 3, pp. 255-309.

[2] van den Berg, R., C. Gussenhoven and T. Rietveld (1991), "Downstep in Dutch: Implications for a model," in *Gesture, Segment, Prosody* (eds. G. Docherty and D.R. Ladd), Cambridge University Press, pp. 335-359.

[3] Bruce, G. (1977), *Swedish Word Accents in Sentence Perspective*, Lund: Gleerup.

[4] Dyhr, N.J. (1992), "An acoustical investigation of the fundamental frequency in Danish spontaneous speech," in *Nordic Prosody VI* (eds. B. Granström and L. Nord), Stockholm: Almqvist & Wiksell, pp. 23-32.

[5] Fujisaki, H., K. Hirose and K. Ohta (1979), "Acoustic features of the fundamental frequency contours of declarative sentences in Japanese," *Ann. Bull. Res. Inst. Logopedics and Phoniatrics*, vol. 13, pp. 163-173.

[6] Gårding, E. (1994), "On parameters and principles in intonation analysis," *Working Papers, Dept. of Linguistics, Lund University*, vol. 40, pp. 25-47.

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

[8] 't Hart, J. and R. Collier (1979), "On the interaction of accentuation and intonation in Dutch," *Proc. Ninth Int. Cong. Phonetic Sciences*, vol. II, pp. 395-402.

[9] Ladd, D.R. (1993), "In defense of a metrical theory of intonational downstep," in *The Phonology of Tone. The Representation of Tonal Register* (eds. H. van der Hulst and K. Snider), Berlin: Mouton de Gruyter, pp. 109-132.

[10] Liberman, M.Y. and J. Pierrehumbert (1984), "Intonational invariance under changes in pitch range and length," in *Language Sound Structure* (eds. M. Aronoff and R.T. Oehrle), Cambridge, Mass.: M.I.T. Press, pp. 157-233.

[11] Möbius, B. (1993), *Ein quantitatives Modell der deutschen Intonation*, Tübingen: Max Niemeyer Verlag.

[12] Öhman, S.E.G. (1968), "A model of word and sentence intonation," *STL-QPSR, Royal Institute of Technology, Stockholm*, pp. 6-11.

[13] Pierrehumbert, J.B. (1980), *The Phonology of English Intonation*, M.I.T. Doctoral Dissertation.

[14] Reinholt Petersen, N. and P. Molbæk Hansen (1994), "Fundamental frequency resettings, pauses, and syntactic boundaries in read-aloud Danish prose," *Acta Linguistica Hafniensia*, vol. 27, 2, pp. 383-401.

[15] Vaissière, J. (1983), "Language-independent prosodic features," in *Prosody: Models and Measurements* (eds.: A. Cutler and D.R. Ladd), Berlin: Springer-Verlag, pp. 53-66.