(ARIPUC 11, 1977)

ON THE INTERPRETATION OF RAW FUNDAMENTAL FREQUENCY TRACINGS[1]

Nina Thorsen

Abstract: This paper describes a procedure with which complex raw fundamental frequency (Fo) tracings can be dissolved into their constituent components, namely 1) the intonation contour of the sentence, 2) the Fo patterns associated with stress groups, 3) Fo movements pertaining to individual segments (vowels), 4) intrinsic Fo differences between segments (vowels), and 5) coarticulatory Fo variations. The problems inherent in such a procedure are discussed as are assumptions of underlying physiological production models of and perceptual correlates to the physical signal.

A phonetic investigation of fundamental frequency (F_o) may concern phenomena at various levels, such as the over-all intonation contour(s) of sentences, the F_o patterns associated with stressed and unstressed syllables (in languages where stress and F_o are interrelated), intrinsic F_o differences between segments, F_o movements associated with individual segments, and F_o variations arising from the interaction between neighbouring segments (coarticulatory variations). The contributions from each one of these phenomena may vary between languages, between dialects, and between speakers but, to some extent, they all interfere with each other to make for the rather complex F_o courses we observe in the outputs of our analyzers, and thus, no matter what the focal point of interest may be, the effect of the others upon it should be accounted for, in order to arrive at a stringent and relevant description and also to create a certain simplicity in the frequency dimension.

---

1) Paper read at VIII^{èmes} journées d'étude sur la parole, Aix-en-Provence, 25-27 Mai 1977.

We may attain simplicity in the time dimension by disregarding parts of the F_o course, which presupposes a reliable delimitation. Segmentation is also necessary to create line-up points for the averaging of several repetitions of a given utterance and is in itself a problem which, however, is not considered any further here.

The outcome of the processing of F_o tracings is often (tacitly) assumed to reflect either an underlying physiological production model or a perceptual correlate to the physical signal - but, clearly, either of these goals can only be approached in a roundabout fashion which involves guess-work and inferences made from observations of similarities and dissimilarities among and between sounds in various environments.

I shall try to exemplify some of the problems inherent in separating, in raw F_o tracings, the contributions from each of the factors mentioned in the first paragraph, stressing the fact that they are relevant to "manual" as well as to computerized treatment of data. The problems are the experimenter's and the answers to them can only be supplied by him.

In a recent investigation (Thorsen 1976) of Danish intonation (in non-emphatic, non-emotional speech) - which was also to illuminate the relationship between linguistic stress and F_o - it was found necessary, first of all, for the averaging of repetitions of the same utterance, to line up the traces to each new beginning of a stretch of voicing in the sentence. The time-axis is thus, of course, distorted.

Only the F_o course of vowels and (voiced) sonorants were included in the subsequent treatment, assuming that the F_o course in voiced obstruents is irrelevant for the perception of pitch patterns and -contours (associated with stress groups and sentences, respectively), the ultimate goal of the investigation being a description as close to the perceived pitch course as possible.

The assumption of the irrelevance of the course of F_o in <u>voiced obstruents</u> will have to be tested in perceptual experiments,

but it seems justified in the light of a case like the one below (which is but one out of many): The word stavelserne ('the syllables' (the stress is on the first syllable)) may appear in any of the three shapes depicted in Fig. 1. (Note that Danish does not have an opposition between unvoiced and voiced alveolar fricative, but /s/ may be voiced between voiced sounds, especially in rapid speech.)

(a) a rise in the first two syllables, continuing through the [s̬], followed by a fall in the last two syllables,

(b) a rise in the first two syllables, a fall and a rise in the [s̬], and a fall in the last two syllables, which corresponds to the fall in (a),

(c) a rise in the first two syllables, 'silence' during the [s], a jump upwards to the fall in the last two syllables, which corresponds to (a) and (b).

The basic production model might look somewhat like (a) and is realised as such (in rapid speech) when the glottis remains "closed" and the vocal cords vibrate during the [s] and, probably, when at the same time the constriction at the alveolar ridge is fairly loose. It is modified as in (b) when the vocal cords vibrate around a somewhat more open position during the [s], a modification which is not voluntary but due to mechanico-acoustico-aerodynamical factors. Finally, the model may be realised as in (c) if the glottis is wide open during the [s], impeding vocal cord vibrations. Unless one's attention is specifically drawn towards the detection of voicing in an /s/, the three editions are likely to be perceived in the same fashion, i.e. there is as much information for the listener in (c) as in (a), and the F_o course in the (voiced) obstruent is thus irrelevant in the sense that it passes unnoticed - and (c) may thus be an approximation to the perceptual model.

In vowels one often finds, after an aspirated stop or an [s] or [f], i.e. unvoiced consonants produced with an open glottis and with forceful airflow, one or two vibrations which are

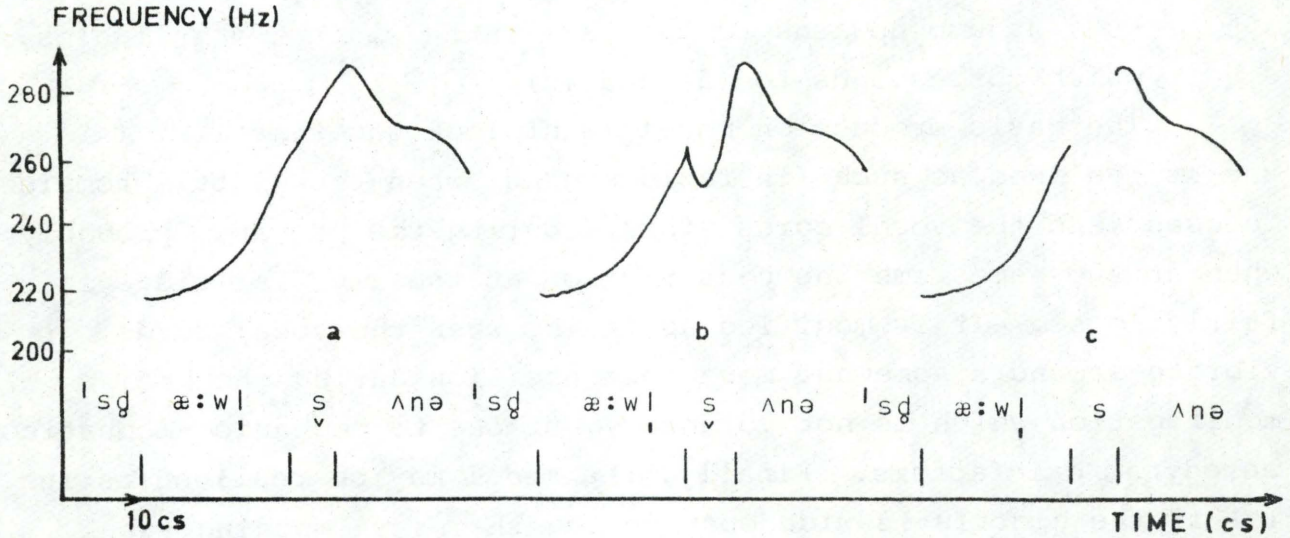

Figure 1

F_o tracings of three recordings of the word stavelserne (female subject)

between 10 and 40 Hz above the succeeding ones. Such vibrations were left out, cf. the line of argument above for /s/.

The pruning of the F_o course is, regrettably, not uncontroversial. How much of a movement in a vowel surrounded by (voiced and unvoiced) consonants can be ascribed to influence from these consonants? In other words, what does the intended F_o course - the production model - look like before it is superimposed by involuntary modifications? And, once again, what are the perceptual correlates (which are not necessarily identical to the production model)? To my knowledge, there is no sure way to compensate accurately for such coarticulatory F_o variations since, as yet, too little is known about the physiological and aerodynamic mechanisms involved to allow us to quantify these variations. - But there are ways around this problem. One (which I chose to follow) is to disregard those portions of the F_o course where coarticulatory variations occur, i.e. mainly at the boundaries between sounds of the obstruent and sonorant classes, and prune away the obstruents plus bits of the adjoining sonorants (but how large bits is difficult to decide). This is probably permissible if the outcome of the processing is to reflect a perceptual correlate to the signal. Another possibility is to interpolate an F_o course through all voiced obstruents so as to smoothly connect the courses in the adjoining sonorants and thus approach what may be the underlying production model.

The F_o patterns associated with stressed and unstressed syllables were attacked next - on the basis of a material of reference words of two and three syllables ([bʰibʰi(bʰi)]) where only the position of the main stress varied (they are nonsense words but they are all possible words in Danish). These words were embedded, in various positions, in declarative sentences. By employing the same consonants and vowels in the reference words, observed differences in F_o courses must be due to different stress distributions. (It turned out that a stressed syllable and all succeeding unstressed syllables in the same non-compound

sentence, irrespective of intervening morphological and syntactical boundaries, constitute the unit for an F_o pattern, which can then be described as a low stressed syllable followed by a tail of higher and falling unstressed syllables, cf. the dotted line in Fig. 2,d.) But the problem remains to demonstrate that words of different and varying segmental structures exhibit the same F_o patterns as do the reference words, since intrinsic and coarticulatory F_o variations interfere.

A tabulation of intrinsic F_o differences between segments (vowels) must be performed for each individual in an experiment since the magnitude of these differences varies. It may be as large as 35 Hz for females, and as small as 10 Hz for males, between [i:] and [ɑ:], cf. Reinholt Petersen (1976). But the magnitude of these differences seemingly also varies with degree of stress (being much larger in stressed syllables), with the short/long distinction (being larger in long vowels), and with position on the intonation contour (being larger at the top than at the bottom of the contour).

Once these intrinsic differences in various conditions have been established for each subject, it becomes possible, in sentences consisting solely of short vowels and unvoiced consonants, to simply move the F_o course in a given vowel up or down the frequency axis, as the case might be, - and only then does the similarity to the F_o patterns of the reference words become apparent. - But sentences consisting also of words of long vowels and of vowels surrounded by sonorants, which exhibit longer stretches of voicing (abbreviated "voiced words" in the following) have to be dealt with as well. It is possible to reduce the long continuous stretches of voicing in "voiced words" to a succession of shorter ones if we assume that identical tonal patterns shall recur in "voiced words" and reference words. (This assumption must of course be verified or falsified in perception experiments.) One can compare (under identical conditions) the F_o courses in "voiced words" and reference words with identical

stress patterns after corrections for intrinsic F_o differences have been performed. Parts of the F_o courses in the "voiced words" are then concurrent with the F_o courses in the reference words, and if the remainder of the F_o courses in the "voiced words" is disregarded, "voiced words" are seen to follow the same F_o patterns as exhibited by the reference words (and other words with short vowels and unvoiced consonants). (This is in itself an indirect support of the assumption of the recurrence of identical tonal patterns in reference words and "voiced words".)

F_o movement associated with individual segments (vowels) must be accounted for, but in this case results will probably be valid for the whole ensemble of speakers of the same dialect. (In Advanced Standard Copenhagen Danish all short vowels exhibit falling movements and all long vowels exhibit falling-rising movements, but the fall is of greater extent than the rise. The only exception seems to be the first stressed syllable in a sentence, which may be rising. The direction of the movement is unaffected by surrounding consonants, but the extent of the fall is slightly greater after aspirated stops and unvoiced fricatives than after unaspirated stops and sonorants, cf. Reinholt Petersen (1976).) If a movement is not heard as such, but rather as a level pitch, it can be represented as such, but should the level then correspond to the beginning, the middle, or the end of the movement? This is, of course, a problem which has engaged phoneticians for years, cf. e.g. Rossi (1971). - If a movement is perceived as a movement, it may be preferable to preserve it in the description, especially if it is to serve comparative purposes. Different vowel F_o movements may well be one of the important prosodic distinctions between different Danish dialects.

When F_o patterns associated with stress groups are recurrent entities (allowing, however, for context dependent modifications, as long as they are predictable), the intonation contour can be defined narrowly as the course described by the stressed syllables alone, cf. the "crossed" line in Fig. 2,d.

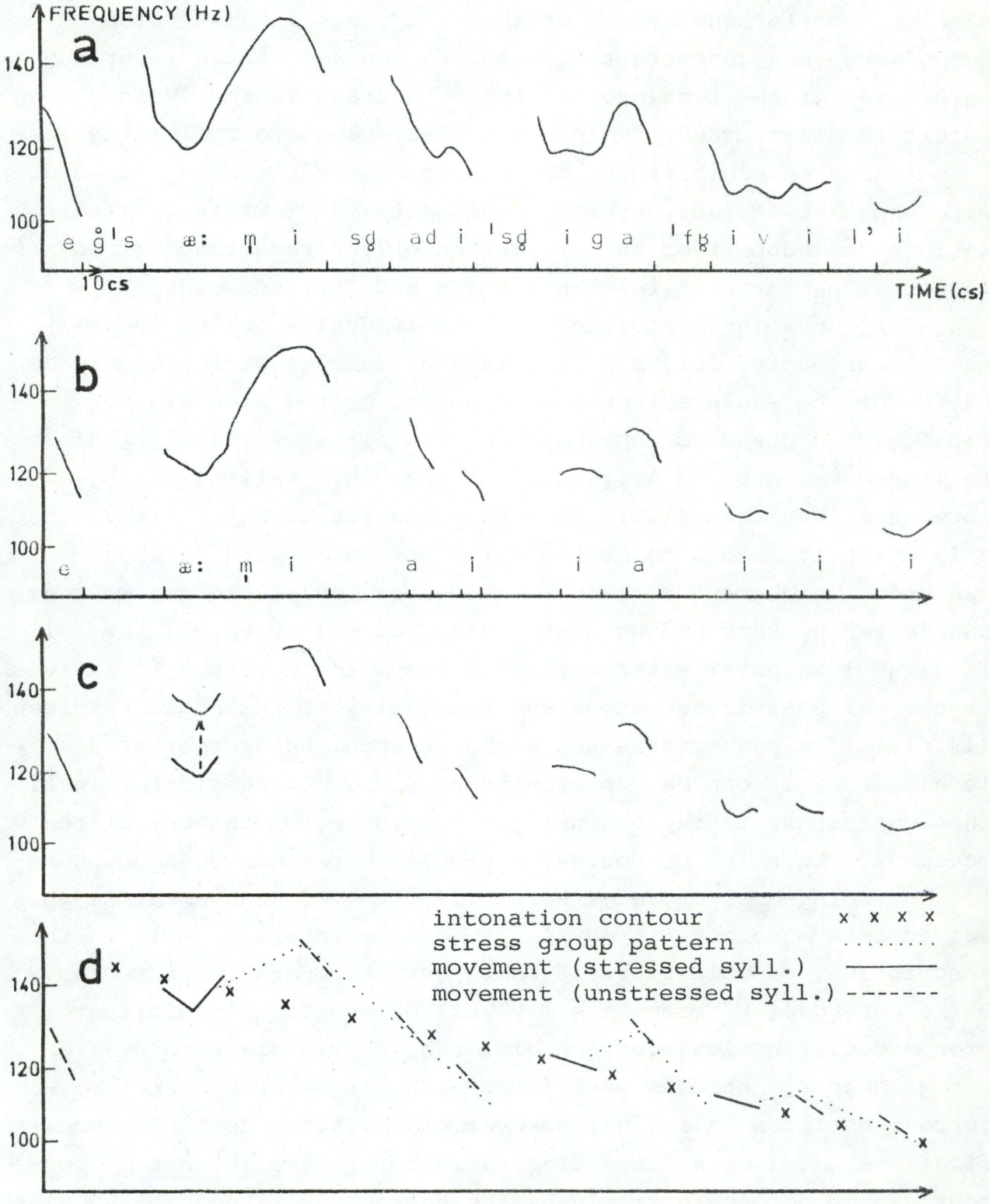

Figure 2

F_o tracing of a sentence Eksamen i statistik er frivillig ('The examination in statistics is non-compulsory') (Male subject.) a: raw signal, b: pruned signal, c: reduced and corrected signal, d: stylized signal.

Given the results of experiments and considerations as outlined above, a raw F_o tracing can be broken down into constituent components, as illustrated in Fig. 2. First "irrelevant" passages (cf. above) are pruned away (b). Corrections for intrinsic F_o differences between (stressed) vowels are performed and, simultaneously, longer voiced stretches are reduced (via comparisons with reference words in identical context) to a succession of shorter ones (c). In this case movements associated with individual vowels are preserved but it may be justified to reduce them to level courses. Finally, the tracings may be stylized (d).

On the basis of a relatively comprehensive material it has been possible, following the procedure just described, to postulate a preliminary model (Fig. 3) for F_o in various types of short sentences in Advanced Standard Copenhagen Danish.

A scale has been tentatively indicated. For X equal to one it would fit a rather low male voice. For other speakers, X will have to be determined, which is probably not a simple procedure. A very coarse approximation is $X = \frac{a}{100}$ where 'a' is the frequency in Hz of the second (of three) stressed syllable(s) in a declarative sentence. This value of X, however, predicts too high values for the highest unstressed syllables for female voices.

The model is supposedly predictive as well as descriptive. Given a sentence with a certain intonation contour (determined by the syntactical and semantical nature of the sentence) the stress group patterns are superimposed on the contour and the appropriate vowel movements are added (if they are not prescribed by the model). The result is modified by intrinsic F_o differences pertaining to individual segments (vowels) and, finally, is turned into a continuous course by smoothly connecting the vowels, with interruptions, of course, for unvoiced consonants (and with dips for voiced obstruents, but this last step needs further clarification). For a more detailed account, see Thorsen (1976).

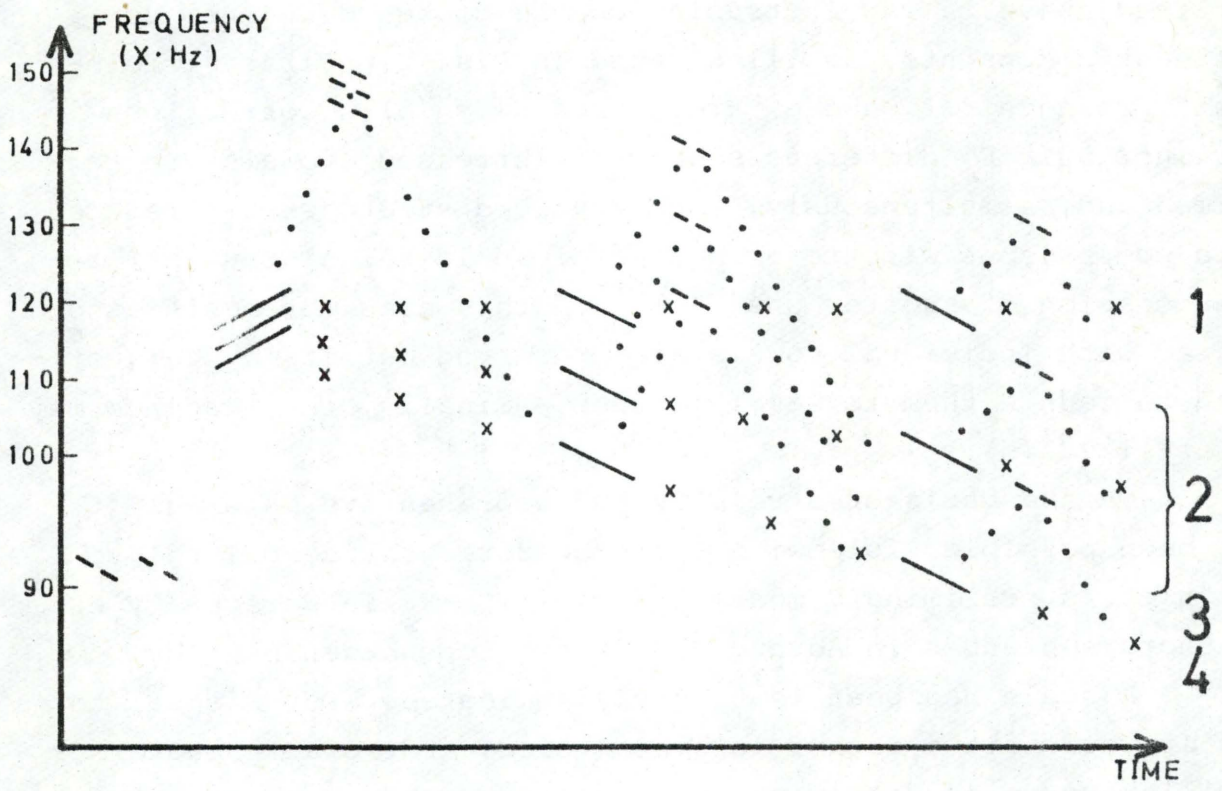

Figure 3

Model for F_0 in short sentences in Advanced Standard Copenhagen Danish. 1: Intonation questions, 2: Questions with word order inversion and Non-terminal periods (variable), 3: Questions with interrogative particle, 4: Declarative sentences. See further the legend to fig. 2,d.

References

Reinholt Petersen, N. 1976: "Intrinsic fundamental frequency of Danish vowels", ARIPUC 10, p. 1-28 (also to be published in JPh.6)

Rossi, M. 1971: "Le seuil de glissando ou seuil de perception des variations tonales pour les sons de la parole", Phonetica 23, p. 1-33

Thorsen, N. 1976: "An acoustical investigation of Danish intonation: preliminary results", ARIPUC 10, p. 85-148 (also to be published in JPh.6, in a slightly revised form)