SUPRASEGMENTAL TRANSCRIPTION*

NINA GRØNNUM THORSEN

This chapter deals with problems in the transcription of duration/length, stress, and intonation whereas tones in tone languages are left out of consideration. The emphasis will be on theoretical issues and rather less space will be devoted to purely technical/notational aspects. I shall also sidestep the more philosophical issue: what it is to transcribe at all. In other words, I am assuming a consensus about both the necessity and the feasibility of transcribing suprasegmental phenomena, but I will discuss such problems as degree of abstraction, descriptive or prescriptive transcription, validity and reliability, and reader target groups.

You will note that I have chosen to understand by "suprasegmental" the rather strictly linguistic phenomena. Thus, voice quality (which may, in fact, serve distinctive, linguistic purposes, cf. Ladefoged, 1980) and the various other aspects of speech production (like variation in loudness and tempo, pauses, etc.) which are a necessary part of discourse analysis, are left out of consideration here. They are treated in separate chapters in this volume.

Even though prosodic analysis has received considerable attention in the literature over the past decade or two, relatively little has been published about suprasegmental transcription, whereas - in later years - a number of publications have dealt with the transcription of segments. To a large extent, of course, segmental and suprasegmental notation pose the same type of general problems (validity, reliability, target readers, etc.) which have accordingly been treated in the literature, and which are also the subject of Vieregge's and Tillmann's contributions to this volume. Left to me are those considerations which are specific to transcribing length, stress, and intonation.

If I have anything to say at all that will not be trivial to most readers, it is that I do not think it meaningful to transcribe a language or a dialect without knowing (speaking) it,

at least rudimentarily, and that a useful transcription must be
based on some (hypothetical) model for prosody in the language
in question. (In fact, some notational systems - particularly
the 'digitalized' ones - have implicit in them such a model,
but the model is often not defended on independent grounds, so
that the notation becomes the model becomes the notation....)
Transcribing without knowing the speaker's intention may be a
good exercise in ear-training but the output is rarely of any
use for linguistic purposes. Thus, it is entirely possible to
transcribe in such a way that the notation may be said to re-
represent acceptably the "noise" on one's tape, but a speaker of
the language, whether phonetically trained or not, will not be
able to recognize the message behind the transcription. A few
examples will illustrate my point. Speakers of a language
that has high rounded front vowels will be tempted to tran-
scribe some of the very fronted varieties of (British and
American) English /u:/ as [yː] or [yːː]. This is quite adequate
phonetically, but it makes no sense to the native speaker, who
would probably choose to modify the [uː]>[uːː]. Most foreigners
transcribing Danish will render Danish /ø/ as [ʌ]. Again, this
can be defended phonetically, since Danish /ø/ lacks friction,
i.e. it is phonetically a sonorant, and it also has a rather
considerable velarization. True, there is no contact between
apex and/or tongue blade and the alveolar ridge in Danish /ø/,
but this is also true of many varieties of velarized [ʌ].

If the listener is not familiar with the phonetic cues to
stress in the language he is transcribing, he may place his
stress marks on the wrong syllables. Thus, I have seen tran-
scriptions of Russian which consistently had stress marks on
the syllable immediately succeeding the one which the speaker
intended to be stressed. The post-tonic syllables carried the
relatively highest pitch, and it is not unreasonable to associ-
ate this tonal prominence with stress, unless you are aware
that there are other cues to stress in Russian, namely vowel
duration and vowel quality. A further difficulty in a com-
pletely "naïve" transcription is the lack, or the arbitrar-
iness, of word boundary assignment. A model for prosody ensures
that the transcriber has at least been obliged to consider the
number and type of elements which make up the prosodic system
of the language, and he will also have had the opportunity to
consider how these elements interact in their phonetic mani-
estation. (Needless to say, some or all of these preconceived
ideas may have to undergo revision, as the transcription and
analysis progress.) For one thing, this will prevent an
otherwise rather common confounding of stress and intonation,
in languages where one of the cues to stress is pitch vari-
ation. It is also evident that transcriptions of suprasegmental
phenomena should rest upon a theoretical basis when they serve
comparative and/or didactic purposes, i.e. when they are in-
tended for other linguists or for language learners. - The
importance of knowing why and for whom you transcribe, where
segmental notation is concerned, is treated by Martinet (1946).
There are two more provisos in this chapter. I am assuming that the transcriber is an experienced one and that she is transcribing tape-recorded speech. Problems specific to field work with informants, without recourse to taped material, as well as problems that arise due to lack of sophistication in the transcription, are dealt with in Rischel's contribution.

I also wish to state something which is repeated in the various sections, but which is important enough for the exposition as a whole to deserve to be underlined here: I believe prosodic transcription to be a vastly more complicated task where spontaneous speech is concerned as compared with read, or otherwise monitored, speech material. This is primarily because duration and stress variation will be shaded to more and finer degrees, to suit the pragmatic purposes of enunciation. We may also expect that the pitch patterns characteristic of monitored speech will be less explicit in a situational context, where they are not alone in carrying the communicative burden.

Finally, the reader should note that I do not - in this chapter - restrain myself to the conventions of the International Phonetic Association, mainly because the guidelines laid down in the IPA-conventions are not sufficient to meet the demands of all the rather different prosodic systems one may encounter.

LENGTH

Length or duration of sounds is a relative measure. This is a trivial observation and we never attempt to mark absolute durations. But for a given rate of delivery, a given speech tempo, some sounds are longer than others. Several factors contribute towards this variation.

PHONOLOGICAL DISTINCTIONS

The language may have a phonological distinction between short and long sounds. Some languages know only of short and long vowels. In Danish, the majority of the short/long vowel pairs have identical quality. Thus, e.g. [viːls]/[vɪls] and [ɛluːls]/[ɛluːls] ('to rest/wild' and 'ball/cold'). In English and German the short vowels have a more centralized quality. Thus, e.g. [biːd/biːd] and [wʊd/wʊd] ('bead/bid' and 'wood/wood'); [biːtən/biːtən] and [fpuːkən/fpuːkən] ('to bid/to ask' and 'to haunt (the place)/to spit'). In the latter case, you may choose to indicate both the durational and the qualitative differences, as I have done here. But one may also wish to introduce a certain degree of abstraction, according to the phonological analysis of the vocalic systems of, in casu, English and German. If the length difference is considered phonologically primary, the qualitative differences may be omitted in the transcription (to render [biːd/biːd - wʊd/wʊd - biːtən/biːtən - fpuːkən/ fpuːkən]); and vice versa, if the qualitative difference is considered primary, the length mark may be omitted from the long vowels (to render [biːd/biːd - wʊd/wʊd - biːtən/biːtən - fpuːkən/
(Note that I am not here discussing the criteria by which you arrive at one or the other solution, whether they be purely phonological/historical or based on auditory cues to the identification of vowels in the language.) The choice of transcription will also depend on the target readers. Germans learning English may not need to be reminded of the difference in quality accompanying the difference in length (or vice versa), whereas it is essential to Danish learners of both English and German to be made expressly aware of the centralization of the short vowels. Mixed into, or cutting across, these considerations of the degree of abstraction away from the physical reality and of the choice of feature to be considered redundant is the question, whether the transcription is to be purely descriptive or prescriptive/normative. Into these considerations is also mingled a decision about how broad the transcription is to be, in terms of geographical area, i.e. whether more than one variety of the language is to be encompassed. For example, a transcription which is to represent both Standard North German and South German will disregard the qualitative difference in the long and short vowels (South German does not have the centralization of the short vowels which is characteristic of North German). On the other hand, a prescriptive/normative transcription, such as one would write in a pronouncing dictionary which indicates the pronunciation to be recommended for foreigners, will want to point to precisely this difference between the Standard German and other norms.

A number of languages also have long and short consonants. In Swedish and Italian, vowel and consonant length are in complementary distribution in stressed syllables, i.e. a long vowel is succeeded by a short consonant, and vice versa. Thus, Swedish [vi:la/vi:la] and [fe:t/fe:t:] ('to rest/villa' and 'fat (adj.)/fat (sb.)') and Italian [ɛ:ko/eks:o] and [fa:to/ fat:o] ('echo/her'e and 'fate/fact'). In Swedish the short and long vowels may differ only slightly in quality (this is true of short and long /i/) or considerably (this is true of e.g. short and long /u/). Here again, one may choose to note the duration of both vowels and consonants (as well as the qualitative differences in the vowels), or one may consider the variation in consonant duration to be a concomitant feature of vowel length and thus leave consonant duration unspecified, or vice versa, depending on the analysis. (The facts of Swedish and Italian phonology are more complicated than one would be led to think from these examples, but that is of no consequence for the present exposition.) In a language like Finnish any combination of short and long vowels with short and long consonant may occur. Thus [mu:ta/mu:ta/mu:ta/mu:t:a] ('mud/(something) else/but/without (something) else').

Needless to say, phonological distinctions in length must be captured by the transcription, irrespective of its purpose or target group. In the examples above I have employed the IPA convention for marking length, a colon after the sound in question. This gives you the opportunity to modify the notation of "halflong" sounds, which are marked with only one dot.
Likewise, "overlong" sounds may be noted with double colons (::). The distinction between short, long, and overlong may be phonological. Thus, in Estonian [sada· - sa:da· - sa::da] and [lina· - li:n:a· - li:n::a] ("one hundred/send" (2. ps.sg. imp.) (to) become" and "linen"/genitive of 'town' /illative of 'town'). The half-length on the word final [a·] in the first two words in each series is determined by the structure of the preceding syllable - it is a bound variation (Diana Krull, personal communication).

There are other ways to mark length, of course. You may double the symbol of the long sound (e.g. Danish [viilø]). However, insofar as each vowel symbol traditionally constitutes a syllable, this is not the most fortunate of conventions, unless one or the other of the two be marked for non-syllabic (e.g. [viiø] or [viilø]). I do not know of any non-arbitrary way to decide whether the first or the second part of the long sound is the best candidate for this semi-vowel status, since it does not make any sense phonetically unless the vowel is actually diphthongized. The gemination of symbols may, however, be adequate in a phonological transcription where, likewise, one may find long vowels denoted as a sequence of vowel plus consonant (/w/, /j/, /h/- see, e.g., Trager and Smith 1957).

STRESS

Another factor which influences the relative duration of sounds is stress. Sounds - particularly vowels - are longer in stressed syllables than in unstressed ones, ceteris paribus. The lengthening of vowels in stressed syllables may be particularly pronounced in languages that do not have a phonological distinction between short and long vowels, like Spanish and Portuguese. Thus, Portuguese ['fa:bi:kə/fe'bri:kə] ("factory/ (he) manufactures"). Insofar as these durational variations are one of the auditory cues to the identification of stress, we may wish to indicate them in the transcription. On the other hand, if the transcriber and the reader, both, are familiar with this effect of varying the duration of vowels according to the degree of stress, such a notation may be considered redundant, and the extra length be contained within the stress mark. See further below.

SENTENCE ACCENT

In languages where sentence accent is an obligatory phenomenon, sounds may be extra lengthened in the syllable which receives this special prominence. That is, a stressed syllable will have even longer sounds, ceteris paribus, when it occurs under sentence accent. Thus, English [ə 'blæ:k *baːd / ə 'baːd iz *blæ:k] ("the black bird / the bird is black" - the star denotes the sentence accent, see further below). However, the extra length may also be omitted and considered part of the realization of [*].
POSITION

Position in the utterance is another variable. A number of languages have final lengthening. The sounds in the last syllable(s) before a phrasal (or stronger) boundary are longer than in other positions, ceteris paribus. This will be most apparent in syllables which are not already lengthened for other reasons, e.g., in unstressed syllables. Thus, English [ðæ 'jeləʊ *bæːd / ðæ *bæːd iz 'jeləʊ-ðæ] (‘the yellow bird / the bird is yellow’). Again, one may wish to note this lengthening or not, according to the purpose of the transcription and the degree of sophistication of one's readers (that is to say: their degree of familiarity with the language). Thus, it is useful to mark final lengthening in the teaching of English to Danes who lack this phenomenon in their mother tongue.

PHONOLOGICAL SURROUNDINGS

Phonological surroundings also interfere. There are languages where vowels are perceptibly longer before voiced consonants than before unvoiced ones. This is true of, e.g., English and French. Thus [bɪd/bɪt] and [ɛrˈɪz/ɛrɪz] (‘bead/beat’ and ‘raise/race’); [ɡeɪr/ɡeɪʃ] and [boʊɡ/boʊk] (‘strike (sb.)/ scion’ and ‘roe/rock’). However, the degree to which the voicing of homosyllabic succeeding consonants lengthens the preceding vowel (and possible intervening sonorant consonants) is language specific. Thus, the lengthening is much more pronounced in English than in French. It is, once more, a matter of choice and evaluation whether these durational variations are to be captured in the transcription. If one is comparing English and French, e.g. for didactic purposes, it may be of the utmost importance for both sets of learners that the transcription of English clearly indicates the more considerable lengthening before voiced consonants, compared with French.

INTRINSIC DURATION

Intrinsic duration is the term given to the phenomenon that certain sounds are inherently longer or shorter than other sounds, ceteris paribus. For instance vowels with high tongue/jaw position are shorter than vowels with low tongue/jaw position. Fricatives are longer than unaspirated or weakly aspirated stops, ceteris paribus; apical consonants are shorter than at other places of articulation, etc. These durational variations go completely unnoticed by the listener. Either because they are below the 'just noticeable difference' for duration of speech sounds, or because of compensatory perceptual mechanisms. Thus, every listener is also a speaker, and as such she "knows" that in the production of a certain (sequence of) sound(s) an intended sameness in duration is blurred by constraints in the peripheral speech production mechanism, and she therefore overhears, ignores, the (not inconsiderable) difference in duration between, e.g., high
and low vowels. (This difference is of the order of magnitude of 50 msec, and thus well above the difference limen for duration.)

It should be clear from the above that the choice of framework for notation of duration is heavily influenced by one's analysis and the reader target group. Thus, a lot of the variation in duration may be rule governed and can be taken care of in the introductory notes to one's transcription, provided the rules have been discovered and can be formulated clearly (which of course presupposes an earlier stage in the analysis with a narrow transcription which includes all the perceptible durational differentiation in the material). The extreme case of abstraction or simplification is the purely phonological transcription.

PERCEPTUAL ILLUSIONS

When we transcribe segments we are prone to perceptual illusions as well as phonetic/phonological expectancy, cf. Oller and Eilers (1975). It is perhaps less evident that "we hear what we expect to hear" where duration is concerned. This would mean, for example, that we perceived a long vowel when a short vowel was actually pronounced, or vice versa. I know of no empirical data to support my contention, but I think it much more likely, given my experience with listening to e.g. Danish spoken by foreigners or by hearing-impaired speakers, that deviations from the expected durational norm are noticed as precisely such by native speakers of a language. This is undoubtedly true when phonological distinctions are thereby lost, but also such phenomena as the wrong amount (whether too much or too little) of lengthening of sounds in stressed syllables, in sentence accented syllables, or in pre-phrasal boundary syllables are noticed by native speakers of a language. Thus, Englishmen and Swedes react to Danes and think their English or Swedish speech abrupt because, inter alia, Danish does not have final lengthening as part of the prosodic system. I wish to claim that native speakers of a language have a high sensitivity to those durational phenomena which are part of their prosodic system. But it is equally true that listeners in general are deaf to durational phenomena in foreign languages which do not occur in their mother tongue. Thus, Danes have a hard time perceiving long consonants, which do not occur in Danish, except across word boundaries. We are likewise insensitive to final lengthening, and durational variation due to voicing in succeeding consonants may also be troublesome for a Dane. This only goes, of course, for naïve listeners. One of the distinctions of a phonetically trained, experienced transcriber is that she is much less restrained by the phonological system of her mother tongue and less prone to suffer from the phonetic/phonological expectancy syndrome.

Reliability of length notation can best be checked by the degree of consensus reached by a group of trained transcribers. Acoustic analysis may be of some aid, but does not solve any and every
dilemma, because it is true of duration - as of other phenomena - that the relation between acoustics and perception is not bi-unique. Listeners are bound by perceptual limitations and by the indoctrination which being simultaneously speaker and listener subjects us to. Another check on the reliability and validity of one's transcription would be to try and synthesize the transcribed text in a speech synthesis system where intrinsic durational variation is supplied "internally", i.e., by the system itself. Gross discrepancies between the original speech and the re-synthesized version are likely to be detected that way. Such a check will always, however, be reserved for a few test cases. First of all, you need access to synthesis facilities and, secondly, it is a time consuming procedure.

STRESS

Stress is a far more controversial phenomenon than length/duration. Among other things because the articulatory, acoustic, and perceptual cueing of stress is not one-dimensional. Length has one correlate, namely time, but stress has several. Relative prominence of one syllable over others may be achieved by variation in fundamental frequency, in duration, in sound quality, and maybe in intensity/loudness, or by any combination of these parameters, cf. Berinstein (1979). Not only is the cueing complex but our perception of prominence is heavily dependent on expectancy and on a semantic/syntactic analysis of what is being said, i.e. on the understanding of the utterance. It is also my experience that stress is one of the hardest concepts to teach people. Thus, it is difficult - in e.g. a class of first year language students - to reach a consensus among native speakers of Danish about the number, let alone the location, of the stressed syllables in short Danish utterances, and the bewilderment increases when they are dealing with a foreign language. I am speaking here only of recognition of stressed versus unstressed syllables. The task becomes nearly hopeless when such practically linguistically naïve listeners are requested to distinguish more than two degrees of stress.

STRESS VARIATIONS

These observations will collide with the apparent confidence with which many writers about (American) English phonology and prosody operate with stress variations involving four (or more) degrees of relational stress (see e.g. Bloch and Trager 1942, Trager and Smith 1957, and Liberman and Prince 1977). The phonetic reality of these analyses is seldom, if ever, put to any empirical test, and the one experiment (Lieberman, 1965) which tests the ability of a trained linguist to hear stress gradations when he could not resort to a linguistic analysis, goes largely unnoticed. The outcome of Lieberman's experiment was that when utterances are stripped of their content (through a synthesis procedure which replaces the original signal with a series of [a]'s, retaining the amplitude and F0 contours of the original) a trained linguist could only reliably recognize
the stressed and unstressed syllables. Reduced main stress and secondary stresses went largely untranscribed. This is in contradistinction to the transcription of the complete utterances, where four stress degrees were recognized according to the Trager and Smith (1957) tradition (main stress, reduced main stress, secondary stress, and no stress). Lieberman suggests that only two degrees of stress may have acoustic correlates independent of vowel quality. This is in accord with my own qualms about perceiving stress gradation beyond two degrees, in non-emotional, pragmatically-neutral speech. (I disregard the special prominence attached to the 'sentence accent' in languages where this is a relevant parameter, see further below.) Danish phoneticians are perfectly able to assign four degrees of syllable weight in Danish compounds, for instance. Thus 'landmand' ['lanman'] ('agricultur'-literally: 'landman') has the word stress on its first syllable, whereas the second member of the compound is subjected to a stress reduction, which is mainly signalled through the F0 contour. It behaves tonally like an unstressed syllable, like in 'landlig' or 'landet' ('rural, the country'); that is to say: F0 rises immediately after the stressed syllable, and the second syllable lies at the peak of a rising movement, which then falls again if more unstressed syllables succeed. You will note that the stød2 of 'mand' in 'landmand' is retained. In 'landmandsliv' ['lanmansliu'] ('rural life'), the second syllable is further reduced through its loss of stød, and if we add the definite particle: 'landmandslivet' ['lanmansliuet'], it is not without meaning to speak of four degrees of syllable weight, as follows: 1 3 2 4, where the distinction between the lowest degree (4) and the higher ones is carried by the distinction non-full vowel quality versus full vowel quality, and where the distinction between the highest degree (1) and the lower ones is carried by the course of F0. The distinction between the two intermediary degrees may have an acoustic correlate (presence versus absence of length and/or stød), but not invariably, i.e. not if the syllable in question has a short vowel and never takes stød. Thus, there is probably no acoustic distinction between the second syllable of 'landluft' and 'trykluftbor' ['lanluft' and 'treyklufbor'] ('country air' and 'pneumatic drill' - literally: 'pressure-air drill') which can be directly referred to a difference in stress (2nd versus 3rd degree). However, it is possible that differences may occur which have to do with the different number of unstressed syllables (one versus two) - i.e. differences which are not specific to specific types of composita. Insofar as Danish listeners would still assign different syllable weights to the two 'luft's, my claim is that such an analysis derives from a knowledge of the syntactic composition of the word. - I have carefully avoided the term 'stress' here, because I expressly do not think that we are dealing with four degrees of stress in words like 'landmandslivet' or in 'ěvělátor-őpérátor' (which would digitalize into 1 4 3 4 2 4 3 4), to quote a famous example from Bloch and Trager (1951, p. 48). The listener's experience - in neutral speech - of some syllables as heavier than unstressed but lighter than main stressed may be a real one, but it derives from properties in
the speech signal which are not independently controlled by
the speaker, in the same way the difference stressed/unstressed
is. The sensation of a further gradation is a by-product of
the segmental and syntactic composition of the word/phrase/
utterance. In other words, the purportedly suprasegmental
hierarchies of generative phonology (e.g. Chomsky and Halle,
1968) and metrical phonology (Liberman and Prince, 1977) are
syntactic, not prosodic, ones. Furthermore, the lack of a bi-
unique syntactic/prosodic relation makes it impossible to read
off the syntactic structure directly from the prosodic surface
manifestation (the distribution of stresses of varying degrees),
as also pointed out by Rischel (1972). Simultaneously I wish
to say, though, that I think this is something which deserves
empirical verification in a somewhat larger-scale experiment
than Lieberman's (1965) one.

You will note that I have carefully avoided to note anything
but the main stresses in the examples given above. This is in
line with the claim that lower degrees of stress are deductible
from the syntactic and segmental make-up of the word or phrase.
However, it is of course perfectly possible to mark a perceived
syllable weight between main stress and no stress, and even to
make a gradation like in [ˈlanˌmænsˌliˈæʊs]. There are also
other means than vertical strokes to note stress. Accent marks
above the vowels are quite common, as in Bloch and Trager (1951)
'élève̩tor-ôperator', and may be a good solution for its typo-
graphical simplicity and for the fact that one does not have to
make any decisions about syllable boundaries. On the other
hand, the same accents are used in the notation of syllable
tones or word tones, and insofar as tones and stress may exist
simultaneously in a language, as is certainly the case in
Norwegian and Swedish, a confusion between tones and stress
may arise. If the transcription indicates more than two de-
grees of stress, digitalization is possible, although I think
that it much too easily effects an impression in the reader
of absolutes and objectivity, notions which have no foundation
in physical facts.

When speech is transcribed from spontaneously spoken texts -
where the semantic and pragmatic context determines a variation
in prominence among the stressed words, i.e. among syllables
with main stress - the question of how many degrees to take
down in the notation becomes more preponderant, but I do not
know how to answer it non-arbitrarily.

Before I proceed, I want to point out that stress is apparently
not a relevant parameter - neither phonologically nor phonetic-
ally - in all the languages of the world. It is hard to assign
any independent meaning to stress in Greenlandic, for instance.
In other words, there is no systematic variation in segment
and syllable duration beyond that conditioned by syllable struc-
ture and syllable sequence. Likewise, there is no systematic
variation in Fo, except that the three final morae before a
phrase boundary are characterized by a (high-)low-high Fo pat-
tern, which lends a certain prominence to the antepenultimate
(Mase 1973, Mase and Rischel 1971, Rischel 1974, p. 91ff,
Jacobsen 1986).
SENTENCE ACCENT

We will return now to languages that have sentence accent as an obligatory feature of their prosodic systems. The phenomenon has many names: sentence accent, sentence stress, primary accent, primary stress, nucleus, nuclear accent, focal accent, Satzakzent. There is not complete agreement across authors and languages about neither the semantic/pragmatic function, nor the phonetic manifestation of sentence accent (which is the cover term I employ here), but - roughly speaking - it is the label given to the somewhat greater prominence attached to one stressed syllable over other stressed syllables in a phrase or an utterance. The greater prominence is generally achieved by a more elaborate F₀ movement and greater duration than in the surrounding stressed syllables. If nothing else is specified by the context, this extra prominence will be located on the last lexical item in the phrase or utterance. Languages which lack sentence accent as an obligatory prosodic element, such as Standard Danish, do not have any special phonetic prominence attached to the last or any other lexical element in the phrase or utterance in non-emotional, pragmatically neutral speech.

I will not be concerned, in the following, with the function of sentence accent, whether or not it focalizes semantic material, to what extent it reflects a theme/rheme structure, or what (pragmatic and other) rules govern its occurrence for the speaker. I will limit myself to a discussion of the perception and transcription of sentence accents when and where they occur, regardless of the why's and wherefore's. - The perception of sentence accent is not as straightforward as one may be led to believe from the description of the phenomenon given in most textbooks and introductions to the phonetics of specific languages. Thus, Brown et al. (1980) describe two experiments, where trained phoneticians, linguists and professional language teachers - who were all familiar with the concept of tonic - marked tonic placement in a set of Scottish English sentences, read aloud and in spontaneous speech, respectively. The disagreement across transcribers about both number of tonics and their placement was rather remarkable, even in short utterances. Thus, in the read sentence 'There is my house' (where acoustic analysis revealed 'there' as having the highest F₀, the largest F₀ movement and the highest intensity) 12 judges marked 'there' as the only tonic, 14 marked 'house' as the only tonic, and 3 judges marked both 'there' and 'house' as tonics. The authors conclude about the read sentences, inter alia, that "In no case, even in two-word sentences, was only one tonic identified. In all cases the judges, between them, identified at least two tonics and usually three, four or five. Any item perceived as stressed seems to be at risk. Judges reported that they found the task a very difficult one - this was true even of seasoned phoneticians who have been teaching intonation for years." (p. 146). In the spontaneous speech material, the phonetic cues to tonicity (as defined by the authors: F₀ peak and range, and intensity) tend to cumulate on one item, and the judges, con-
currently, identify fewer tonics and also disagree less. In
the read material it appeared that the last lexical item con-
stitutes the default case since "it is regarded as being the
tonic by right of being the last lexical item if some other
item is not heavily marked phonetically as being in competi-
tion." (p. 146). In contradistinction, in the spontaneous
speech material, the authors conclude that where contrast
(determined by the semantic and pragmatic context) is involved,
"only phonetic cues are available to mark it. The phonetic
cues therefore cumulate on this item and the judges regularly
recognise only one tonic, the contrasted element, even in a
long structure." The first observation lends justification
to the prescription of sentence accent finally in context-
free, neutral sentences, for learners of the language. The
second observation indicates that contrast may justifiably be
considered a phenomenon different from focalization and rHEME-
signalling, see further below.

In the examples given in the section on length above, I marked
the sentence accent with a star. This is to indicate that it
may be advantageous to consider sentence accents as being dif-
f erent in kind, rather than just in degree, from stressed syl-
lables in general. I will try and justify this position in
the following, but wish to point out right away that this is
an area where a good deal of empirical research is called for,
and it may turn out that my claim is not tenable. First of
all, the perception of sentence accent need not be triggered
by any phonetic cues at all, cf. Brown et al.'s results above
and their conclusion about last lexical items. Secondly, it
is said about British English in numerous publications, e.g.
O'Connor and Arnold (1961), that no accented syllables can
follow the nucleus. Thus, in their description, a distinction
is being made between accented syllables, which are stressed
syllables that are also made prominent by tonal means (a skip
up or down or an extensive movement), and stressed syllables
which have no such tonal prominence. And their claim is that
the nucleus can be preceded by accented (tonally prominent)
syllables but accented syllables cannot succeed it. This makes
the nuclear accent distinct from other accents and stresses
which do not constrain their surroundings in this manner.
Unfortunately, I know of no empirical acoustic evidence for
these statements, and they are also somewhat at odds with the
Brown et al. (1980) conclusions. If nuclear accents were so
unambiguously signalled they ought not to be so hard to reach
a perceptual consensus about. We may be dealing here with a
rather characteristic difference between prescriptive and de-
scriptive analyses: the prescriptive analysis which serves
didactic purposes represents an idealized picture of reality,
which is often more complex than the prescription would have
you believe. - On the other hand, there are data from German,
an acoustic analysis of a material read aloud, that do bear
out my initial statement about sentence accents as a special
phenomenon. Bannert (1985) publishes F0 tracings of German
sentences which clearly demonstrate that stressed syllables
before the Satzakzent are associated with F0 movements, they
are accented, but stressed syllables after it run smoothly to
the end of the utterance without any F0 deflections. The
SUPRASEGMENTAL TRANSCRIPTION

Satzakzent itself need not have any particularly elaborate \( F_0 \) excursion. The German material was read sentences, which furthermore were presented in a context that left no choice on the part of the readers as to Satzakzent assignment, as opposed to the Brown et al. sentences which were presented to the readers without any context. This may account for the ambiguous acoustic and perceptual results from Scottish English.

To conclude about sentence accent: I believe that in non-emotional, neutral speech, controlled for semantic and pragmatic effects, sentence accents can probably be recognized by the phonetic cues supplied by the speaker, i.e. increased duration and increased \( F_0 \) excursion, and by the lack of any \( F_0 \) perturbations after it. I also believe that it is a phenomenon apart from ordinary stress, and should be noted as such, by a special symbol in the transcription. The German sentence 'Der französiche König war ein launischer Geselle' (in answer to the question 'Was für ein Geselle war der französische König?') would transcribe as follows: [dɛ̃ ʃwanˈtseːzifə 'kʰɪŋːɪç va ˈain *laʊnɪç ɡeˌzɛlə] (the example is from Bannert 1985 and translates as follows 'The French king was a moody fellow'). Note that such a transcription involves a good deal of abstraction. It requires an understanding and interpretation of what is being said, and relies on the presence of a context. It does not indicate the difference between stresses before and after the [\*], but see further below. - I am less sure that sentence accents can be easily defined and perceived in spontaneous speech, as an 'otherness', i.e. as a different category, in relation to stressed syllables in general (cf. the Brown et al. 1980 results).

EMPHASIS FOR CONTRAST

Emphasis for contrast, or simply contrast, is another phenomenon within the realm of neutral speech, i.e. speech devoid of any particular speaker attitude or emotion, though it does demand a context, as the term implies: one of the words in an utterance is contrasted with an element which is explicitly mentioned or is implicit in the larger textual context.

From Jones (1960) we learn that "Contrast emphasis is expressed mainly by intonation. The special intonation may be accompanied by extra stress or length, but these are secondary." (§ 1047). It further appears from his text and examples that the \( F_0 \) movement associated with contrast is more extensive than that associated with sentence stress and that "the only syllable with a really strong stress is the stressed syllable of the emphatic word. Other syllables may have a medium or fairly strong stress, but they have the intonation of unstressed syllables." (§ 1050). Bannert (1985) also notes that contrast is associated with larger \( F_0 \) movements than Satzakzent. He does not say, but it appears from his figures, that there is a further difference between Satzakzent and contrast: the \( F_0 \)
movements preceding the contrasted syllable are partially suppressed or completely deleted, so the only clear $F_0$ excursion is the one associated with the contrast. That ties up with Jones’ observation quoted above, and is similar to results I have obtained in an analysis of emphasis for contrast in short utterances in Danish (Thorsen, 1980): Emphasis for contrast will make the stressed syllable of the emphasized word stand out clearly from the surroundings, which is brought about by an $F_0$ raising of and/or an elaborate $F_0$ rise within that syllable and by a deletion of the $F_0$ deflections in neighbouring stress groups, to the effect that the immediate surroundings, except the first post-tonic syllable, fall away sharply from the emphasized syllable. In an informal experiment with LPC-synthesis of the sentence 'Det er sidste bus til Tiflis' [deː eː sɪːsdə bʊs tʰeː tʰɪfliːs] ('It is the last bus for Tiflis.') I found that it is the shrinking of the $F_0$ movements associated with 'sidste' and 'Tiflis', rather than the higher location in the frequency range associated with 'bus', that will make 'bus' appear as the contrasted element.

The semantic distinction between sentence accent and contrast emphasis is not always clearcut, there will be many instances in spontaneous speech where one cannot decide whether a particular prominence is to be classified as one or the other, because phonetic and contextual cues are open to both interpretations. But when the distinction can be made, emphasis for contrast and sentence accent should be transcribed differently. Again, I think it is appealing to treat emphasis for contrast as being different in kind from stress in general, mainly due to its effect on surrounding stressed elements. In English and German, for instance, where stressed and accented syllables are distinguished, contrast emphasis de-accent both preceding and succeeding stressed syllables. In Danish, it would be possible to operate with a similar distinction, namely between stressed syllables which are associated with $F_0$ changes and such that are not. This would be an innovation in the terminology in Danish phonetics and phonology. It is customary to speak of main stress and secondary stress, in which case one would say that emphasis for contrast in Danish leads to a reduction to secondary stress of the stressed syllables surrounding the emphasized element. In either case, the de-accentuation or de-stressing of the surroundings can be deduced from the presence of the mark for contrast, so it may still be appropriate to transcribe the Danish example given above as follows: [deː eː sɪːsdə bʊs tʰeː tʰɪfliːs]. It is equally possible to note the change in the surroundings, and for this purpose the lowered stress stroke may be used: [deː eː sɪːsdə bʊs tʰeː tʰɪfliːs]. That would make the German mood word ‘with sentence accent and contrast emphasis, respectively’ appear as follows: [deː fɔʊnːtʰoːziːʃa kʰɔːnɪːɡ vo ɐɪn *lɔːmɪːɡe ɡɔ,ɛːzɪə] and [deː fɔʊnːtʰoːziːʃa ,kʰɔːnɪːɡ vo ɐɪn *lɔːmɪːɡe ɡɔ,ɛːzɪə] which makes explicit the difference in the effect that contrast and sentence accent have on preceding stressed syllables.
PHRASAL UNIT ACCENTUATION

There is yet another type of stress which deserves mention here, the so-called phrasal unit accentuation, applied under certain syntactic conditions to a group of words which, in a vague formulation, can be said to constitute a single semantic concept, see further Rischel (1983). (The phenomenon has received modest attention in the literature of other languages, and the following examples are all from Danish. However, I suspect that phrasal unit accentuation is not an exclusively Danish phenomenon, though the syntactic conditions which trigger it may be different and less extensive in, say, Swedish, German, and English.) This type of stress does not involve any extra phonetic prominence on the stress bearing element (in contradistinction to sentence accent and emphasis for contrast) but is characterized solely by a downgrading of another element. Thus, e.g. 'læse romaner', 'købe hus' (to read novels, to buy a house) have main stress on 'romaner' and 'hus', respectively, and reduced stress on the verbs. The reduction is always and at least signalled with tonal means in Danish, i.e. the reduced stresses are not associated with any $F_0$ excursions (in German and English terminology, they are de-accentuated). Stød is normally lost, too, but the length of long vowels may be retained, according to criteria that I do not think have been established yet. In this case, I do not find it justified to mark the stressed member of these units in any special way, but rather to note the reduction. Whether to employ a lowered stress stroke or to leave the reduced items without any stress mark at all, [ˌlɛsə ʁoːmɑ̃ːˀnə] or [ˌlɛsə ʁoəmɑ̃ːˀnə], depends on the purpose and the target group of readers. Both are acceptable phonetically, except maybe that no independent meaning can be ascribed to the lowered stress stroke, it is redundant. It can be defended in a transcription which involves a certain degree of abstraction - on the grounds that it reflects the process involved in phrasal unit accentuation.

It is apparent that transcription of stress, even in non-emotional, neutral speech, almost invariably involves some degree of abstraction, and relies rather heavily on the listener's comprehension of what is being said and on the choice made in the analysis of such phenomena as stress reduction in compounds, sentence accent, emphasis for contrast, and phrasal unit accentuation. I have suggested the minimal solution in the transcription of non-emotional, neutral speech, i.e. a restriction to a distinction between stressed and unstressed syllables, with a further separate notation of sentence accent, when and where that is relevant, and of emphasis for contrast. This is due primarily to a reservation with regard to listeners' ability to reliably identify more stress degrees. Two degrees of stress will probably also satisfy most descriptive and prescriptive purposes. - I wish to repeat, though, that the transcription of spontaneous, pragmatically unrestrained speech most likely will presuppose a renewed consideration of the possibility of a further gradation among the prominent syllables.
Objective checks of the validity of one's stress notation are not available. Partly because several parameters are involved ($F_0$, duration, sound quality, and - possibly - intensity), but primarily because the relation between any single one of these acoustic parameters and perceived prominence is not bi-unique. Thus, it is a fairly common mistake that intensity directly reflects stress. It cannot, because intrinsic properties of sounds intervene and other parameters than physiological effort influence physical intensity. High vowels have lower intensity than low vowels, ceteris paribus. Sounds pronounced on a high fundamental frequency have higher intensity than on low fundamental frequency, ceteris paribus. Long sounds and vowels with non-reduced vowel quality may well exist in unstressed syllables. - In languages where stressed syllables are associated with a higher $F_0$ than their unstressed surroundings, stress and intensity may correlate fairly well (which does not make intensity a primary stress cue, however), but when stressed syllables are associated with lower $F_0$, unstressed syllables may well appear in intensity registrations with higher intensity than stressed syllables.

INTONATION

Intonation, understood narrowly here as "speech melody", is in one sense as straightforward as length/duration: its physical correlate is unambiguous, namely rate of vocal fold vibration, i.e. fundamental frequency, $F_0$. In this section I will maintain a distinction between the physical parameter, $F_0$, and its perceptual correlate, pitch. I will employ "tonal" indiscriminately about both, when no such distinction is relevant. The complexity in the transcription and analysis of intonation arises from the fact that several linguistic (as well as extra-linguistic) phenomena are signalled in this medium. In languages where pitch change is one of the cues to the perception of stress, part of the variation in the $F_0$/pitch course of an utterance is stress-determined. When sentence accent and contrast are signalled by tonal movements, this is equally true. Phrase and sentence junctures may also be tonally cued. Besides, there is the more strictly intonational phenomenon which has the whole phrase or utterance as its domain (such as the distinction between terminal and non-terminal, declarative and interrogative). In brevity, the complex course of $F_0$/pitch through an utterance may contain elements of sentence function, sentence accent, contrast, stress and juncture (and word or syllable tone). Furthermore, acoustic $F_0$ registrations contain "noise" which is due to the intrinsic properties of sounds, and to variations arising at the boundary between sounds, see further Thorsen (1979). These latter disturbances are not heard and identified by the listener as prosodic variations, because - as with duration - the listener uses her knowledge as a speaker in her interpretation and compensates perceptually for such $F_0$ differences and variations which are due to constraints in the speech production mechanism. (However, $F_0$ variations caused by intrinsic properties and by coarticulation may play a role in the identification
of the segments of speech, see further Reinholt Petersen, forthcoming."

NOTATIONAL CONVENTIONS

Methods to convey graphically the ups and downs of pitch abound in the literature. A few examples will have to suffice: Trager and Smith (1957) employ a digitalized system, where [4] indicates the highest and [1] the lowest pitch level. Within each level, four varieties are distinguished: [.] for the lowest, [.] for the next higher variety, [^] for still higher, and [-] for the highest:

/[^]jëw [^]ây [^]ây [^]ây[^]/ ('how do they study?', p. 42)

where the final 'minus' indicates a terminal fall in pitch. Pike (1945) also transcribes in terms of four pitch levels, but here 1 is the highest and 4 the lowest level:

"Two times three plus two is ten.
"2-4-3/ 4- 2 1- -4-3/ 3- 2-4/ (p. 33)

where " " denotes the beginning of a primary contour, "/" indicates a tentative, and "/" a final pause.

Fries (1964) employs horizontal lines cutting through the printed text:

I'll go if you want me to go (p. 246)

Lee (1960, quoted from Hartvigson and Jürgensen 1971, p. 19) indicates rises, falls and high and low levels as follows:

\Now Frank, are you \coming on the /train or \cycling back?

Schubiger (1964) uses accent marks, placed above or below the line:

Unfortunately he \died without having \made a \will. (p. 257)

Bolinger (1970) lets the print rise and fall on the page:

Hand

me that little knife of yours. (p. 141)

Armstrong and Ward (1926) use an interlinear transcription which has been widely employed since

\'wot a ju 'going to 'du: about it? (p. 6)

Musical notation can be seen in, e.g., Fönagy and Magdics (1963).
The Trager and Smith (1957) and Pike (1945) transcriptions reflect a phonemic analysis of intonation in terms of four distinctive pitch levels. Whether or not you agree with this analysis, you may still consider their digitalization as a notational system, but as such I think it suffers from a pretended objectivity which is not justified by listeners' perception, though it may serve its purpose in a descriptive account of intonation. However, once the notation is no longer intended to depict the result of a particular phonological analysis, one may question the number four. Why not six, or eight? Whatever we arrive at, it is an arbitrary decision.

Lee's (1960) and Schubiger's (1964) systems may be criticized on the grounds that they are really not very explicit, and thus demand a good deal of sophistication on the part of their readers, but they are appealing in their simplicity, and easy to handle typographically.

Bolinger's (1970) transcription is remarkably transparent, but difficult to handle practically, and I much prefer the inter-linear transcription of Armstrong and Ward (1926) which contains just as explicit intonational information, with the added advantage of distinguishing stressed and unstressed syllables.

Musical notation is difficult to employ for anyone not musically trained. Furthermore, speech is not necessarily produced in the semitone intervals of the twelve tone octave.

Whichever notational system we employ, there are a number of factors that are left out of consideration. Just as duration is a relative measure, so is pitch. We do not ordinarily concern ourselves with the absolute frequency but with the relations within a given speaker's range. And we know that different speakers speak within different ranges, i.e. higher or lower in the frequency scale. (Men generally have lower pitched voices than women, and children have still higher voices.) Different speakers may also cover differently sized intervals. One speaker may habitually only employ a range of, say, 10 semitones, whereas another may cover 1½ octaves. These differences likewise do not concern us, except as a cue to speaker identity, sex, and age.

I have voiced a preference for the interlinear notation, where the pitch course in the stressed syllables is depicted with straight or curving lines, as the case may be, and unstressed syllables are rendered as points. This warrants two comments. First, the reduction of the course of pitch in unstressed syllables to points in the frequency range is justified by the fact that unstressed syllables will often be too short to allow the listener to detect any pitch movement, even though Fo may perform steep rises and falls (detected in instrumental acoustic analysis). There is a limit to our perception of pitch movements, a limit that is set partly by the duration of the movement and partly by the frequency range it spans, cf. Rossi (1971, 1978). When movements are not perceived as
such, a level pitch is heard which corresponds to the frequency value of the glide at a point in time 2/3 of the distance from the onset of the glide. Second, the notation is discontinuous, although it may be argued that our perception of pitch is continuous, i.e. we fill in - by interpolation - the empty spaces left by unvoiced sounds in the speech chain. On the other hand, it may also be argued that we anchor our perception of intonational phenomena on certain points in the time varying course of pitch and disregard what lies between such fix points. I do not know that anyone has yet settled this argument, or suggested an experiment which can resolve the issue. In the end, we may choose a continuous or discontinuous notation, regardless whether our perception of pitch is discontinuous or not.

At the lowest level of abstraction the transcription renders, as accurately as possible, the perceptual equivalent of \( F_0 \) curves as they are produced by acoustic analyses. This transcription may be checked against a truly objective, instrumentally obtained one, if one remembers, firstly, that perceptual limitations constrain the amount of detail we can perceive and, secondly, that the rather considerable variation introduced by intrinsic \( F_0 \) differences between sounds and by coarticulation goes unnoticed by the listener. Thus, \( F_0 \) may easily be 25-30 Hz higher, ceteris paribus, in [i] than in [a] with female speakers, 10-15 Hz with male speakers, corresponding to roughly 2 semitones (cf. Reinholt Petersen 1978, and Hombert 1978). In other words, [i] and [a] may appear on \( F_0 \) tracings considerably spaced in the vertical dimension and still justifiably be perceived as having the same pitch.

NOTATIONAL SIMPLIFICATIONS

\( F_0 \) tracings, and perceptual equivalents thereof, are generally much too rich in information and detail to be of any use for descriptive or comparative or didactic purposes. The information must be simplified, an abstraction made, irrelevant information or information which can be deduced from the remainder be filtered out. In languages where stress is a relevant parameter and is cued by pitch, the notation may be anchored on the stressed syllables. Their pitch levels and movements should be specified. It is generally also necessary to keep track of the relation between stressed syllables through the utterance, because the overall drift created by these pitch relations may be (part of) the signal for sentence intonation function. The shape and slope of this overall trend is the only prosodic cue to sentence function in some languages (e.g. Danish, cf. Thorsen 1978); in other languages overall trends coexist with special final pitch movements (e.g. Swedish, cf. Gårding 1979, and German, cf. Bannert 1985). These final movements are the "terminal contours" in the analyses of Pike (1945) and Trager and Smith (1957). How much information about the pitch course of the unstressed syllables to include or leave out will depend on the purpose of the transcription, the language transcribed, and on the readers. There are languages
where the unstressed syllables, at least in non-final position, have no independently controlled signalling function, at least not in non-emotional, pragmatically neutral speech. This is the case in Standard Danish, for example, where unstressed syllables describe a high-falling course, relative to the preceding stressed syllable. Readers familiar with Danish need not - for many purposes - be given this information, whereas it may be crucial to supply it for foreign learners, especially since this does not seem to be the most wide-spread pattern in the languages of the world. Stressed syllables commonly are the higher pitched ones.

In this connection the Dutch approach to intonation analysis should be mentioned. 't Hart and Cohen (1973) and 't Hart and Collier (1975) describe a procedure where Dutch utterances are synthesized on the basis of an original, spoken version. The synthetic version retains the spectral (segmental) and durational aspects of the original, but the complex F0 variations can be replaced with stylized and simplified, straight-line approximations to the original. The stylization is performed while retaining a perceptual similarity between the original spoken version and the synthesis, i.e. unto the point where this similarity is lost. The authors - in this manner - arrive at a description of what they consider to be the perceptually relevant pitch movements in Dutch. The procedure has been applied to English by Willems (1982).

With this - and similar synthesis facilities available - you can perform an abstraction of the original to various degrees: you can stylize F0 so as to retain a perceptual identity between the original and the copy. The outcome would correspond to an interlinear transcription performed by a truly expert, acute listener. That is, it will resemble the original, except that those variations which no human listener can detect have been "filtered" out. This "close-copy" can be useful as a spot check on the transcriber's accuracy and reliability/validity. You can stylize further, without losing the perceptual similarity, as 't Hart and co-workers have set out to do, and thus establish classes of contours that sound similar phonetically - without regard to their linguistic function. You can carry the simplification one more step, presumably, while letting go of perceptual similarity but maintaining identical linguistic function. That is, although the two utterances, the original and the synthetic one, no longer sound similar, they still function adequately as, say, terminal declarative utterances with X number of accented syllables.

I do not think that synthetic speech, of this or any other kind, can ever become a primary tool in transcription. For one thing, it cannot be made available to all and everyone who needs to transcribe intonation. Secondly, it is time consuming and cannot be applied to every utterance of large corpora. But it is a sophisticated and excellent tool for the further, phonetic and linguistic, analysis of intonational data, and possibly in the education of transcribers.
ALIGNING PITCH AND SEGMENTS

In some cases the indication of the alignment of segments with pitch movement is decisive. Firstly, pitch movement alone may carry the primary cue to syllabicity. Thus, Danish 'kom' and 'komme' (imperative and infinitive, respectively, of 'come') consist of the same sequence of segments [kʰʌm] when the final schwa is dropped from the infinitive, which is normal in fluent speech. However, the infinitive has a somewhat longer nasal, and a rising pitch movement in the nasal, which distinguishes it from the imperative, approximately thus:

\[
\text{[kʰʌm] [kʰʌm]}\]

In words like 'bonde, bundne, bundene' (peasant, tied (adj., pl.), the bottoms) duration and pitch and timing of pitch movements signal the difference, which phonologically is one of number of segments and syllables: /bɔnə bɔnə bɔnənə/. In the first word the pitch begins to rise at the transition between vowel and nasal. In the second word, the pitch turning point lies in the initial part of the long nasal. In the third word, likewise, the pitch does not begin to rise till after the onset of the nasal, and here the rise is completed and the succeeding fall initiated, before the onset of the schwa, approximately thus (note that these are impressionistic transcriptions, not tracings from \(F_0\) meters):

\[
\text{[ˈbɔnə] [ˈbɔnə] [ˈbɔnə:]}\]

CO-OCCURRENCE OF PROSODIC SIGNALS

In numerous languages the last part of a phrase or an utterance carries a heavy burden of prosodic information: sentence accent may be located here and be signalled with \(F_0\), sentence function likewise, and junctures, as the following example, adapted from Bruce (1977, p. 93 and 94) shows (Swedish is a word accent language, and I have included an example of both accents):

```
\(\text{accent I} \quad \text{sentence accent} \quad \text{terminal juncture} \quad \text{accent II} \quad \text{sentence accent} \quad \text{terminal juncture}\)
```

<table>
<thead>
<tr>
<th>C</th>
<th>V</th>
<th>C</th>
<th>V</th>
<th>C</th>
<th>'V</th>
<th>C:</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>V</td>
<td>C</td>
<td>'V</td>
<td>C:</td>
<td>V</td>
<td>C</td>
<td></td>
</tr>
</tbody>
</table>
For the mere taking down of the pitch course, the separation of these various contributions is not essential, but for the ensuing analysis it is crucial to segregate each element and know that there are three components here, which overlap partially in time and modify each other.

A typographical notation system which satisfies this demand has been developed by Pierrrehumbert (1980) and further modified and expanded by Liberman and Pierrrehumbert (1984). You do not have to agree with their phonological analysis (as I happen not to do, cf. Thorsen 1983 and 1985) to appreciate the elegance and flexibility of the transcription from F₀ traces to a transcription in terms of sequences of high and low tones. On the background of a slightly declining, physiologically determined, baseline (the broken line in the illustration), parameters such as prominence relations between stressed syllables, downstep of high tones in certain tonal contexts, lowering of tones in final position together account for the time varying course of F₀ through phrases and utterances. In the examples below (from Pierrrehumbert, personal communication) you see F₀ traces of two phrases: 'Anna came with Manny' which differ in the degree of emphasis on the two proper names: 'Manny' is focalized in the upper, 'Anna' is focalized in the lower tracings. A star indicates that the tone is associated with a stressed syllable, and "%" denotes a phrase boundary tone. The weighting of prominence was not indicated in the example.
Below are three examples (also from Pierrehumbert, personal communication) of F0 traces and their tonal transcription which show the effect of the baseline declination upon a succession of three high tones (above). The relatively lower medial H* is presumably due to a lesser prominence, which is not specified here. The downstepping of H tones is illustrated in the two lower graphs, where the H's form part of bi-tonal pitch accents, indicated with a linking "+". The rule lowers a high tone by a certain fraction of the preceding high tone in the context H+L* and H L+, see further Pierrehumbert, 1980, and it applies iteratively, that is, every time the condition for its application is met, also if the conditioning tone has itself been lowered.
The downstepping is not explicitated in the notation, but can of course be supplied, if the formulation of the rule is not sufficient for one's purpose. The particular transcription of the examples above is directly related to the phonological analysis. For instance, the shallow local low points in the upper tracing is not given any tonal representation, whereas the mere levelling out of the $F_0$ course in the middle tracing has warranted "L*" marks. That need not concern us here. I do believe that the system may be used as a notational tool, adapted - as far as complexity or simplicity is concerned - to one's needs, with more or less information specified in the notation or supplied by rules, respectively. The problem with such a notation is not inherently larger or smaller than with any other intonation transcription, where reliability is concerned. Intonation is notoriously "difficult" to transcribe, i.e. it is hard to reach agreement even among experienced transcribers, if not about the gross trends, then at least about finer details in the pitch course. And it is particularly important, therefore, that - where instrumental checks are not possible or practicable - the transcription be compared across two or more transcribers. (Note, however, the reservations voiced above about the lack of a linear correspondence between fundamental frequency and perceived pitch.)

NOTES

1. 'fundamental frequency' refers to the rate of vibration of the vocal chords. I will abbreviate it "$F_0$" in the following.

2. 'stød' is a kind of creaky voice which is assigned to certain syllable types, under certain conditions, see further Basbøll (1972). It attacks the final part of a long vowel or - if the vowel is short - the succeeding voiced consonant.

ACKNOWLEDGEMENTS

Thanks are due to Catherine Høgsbro Holm, Birgit Hutters, Niels Reinhold Petersen, and Wilhelm Vieregge for criticism and comments to the manuscript.

REFERENCES


Chomsky, N. and Halle, M. 1968: Sound Patterns of English (New York)


't Hart, J. and Collier, R. 1975: "Integrating different levels of intonation analysis", J. Phonetics 3, p. 235-255

Hartvigson, H. and Jürgensen, B. 1971: Engelsk Intonation, (Copenhagen)


Liberman, P. 1965: "On the acoustic basis of the perception of intonation by linguists", Word 21, p. 40-54


Reinholt Petersen, N. 1978: "Intrinsic fundamental frequency of Danish vowels", J. Phonetics 6, p. 177-189


Rischel, J. 1974: Topics in West Greenlandic Phonology, (Copenhagen)

Rischel, J. 1983: "On unit accentuation in Danish - and the distinction between deep and surface phonology", Polia Linguistica XVII, p. 51-97

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
Rossi, M. 1978: "La perception des glissandos descendants dans les contours prosodiques", *Phonetica* 35, p. 11-40


Thorsen, N. 1978: "An acoustical investigation of Danish intonation", *J. Phonetics* 6, p. 151-175

Thorsen, N. 1979: "Interpreting raw fundamental frequency tracings of Danish", *Phonetica* 36, p. 57-78


Willems, N. 1982: *English Intonation from a Dutch point of view*, (Dordrecht).