CHAPTER 2: INTONATION ANALYSIS

2.1 ACCENT

It is given different names (prominence, stress, accent) but plays a role, one way or another, in all analyses of intonation: some syllables are heard as more prominent than others, and this distinction functions meaningfully and more or less systematically in the organization of speech.

But much further than this agreement does not go in the literature. To what extent the prominences are predictable by linguistic rule, how the speaker creates them, why the listener hears them as prominent, what role they play in speech communication, how the linguist should study them - these are all problems about which there is division of opinion.

It was not my aim to write about the problems surrounding intonation analysis in general, merely to use it as a tool with which to approach a specific problem in the study of media texts, in this case: how do radio announcers accent their speech, which words do they make prominent, how, and why. But without first finding my own position amidst the linguistic controversies surrounding the study of accent, I cannot begin doing so.

In this section of the chapter, then, I will discuss, as briefly as possible, the literature on accent, comparing it to my own observations, and developing, from this give-and-take between the literature and my observations, a method of analyzing accentuation in radio announcing speech.

2.1.1 Accent and stress

Some writers, most notably and eloquently Bolinger, but also others,[1]* have distinguished between stress and accent (or 'word stress' and 'sentence stress', or 'lexical stress' and 'stress accent'). Stress, says Bolinger, is

> ...the term used to indicate that syllable
> that would become prominent.if the word
> were to be accented...
> (Bollinger, 1964, p. 22)

Accent, consequently, is the actual prominence - a definition which, incidentally, excludes the possibility of accenting unstressed syllables, a possibility of which Bolinger himself has, elsewhere, given some convincing examples.[2]

A distinction, then, between, on the one hand, a syllable's potential of acquiring prominence, an abstract feature of syllables (syllables as (part of) lexical entries), and, on the other hand, a syllable's actual prominence, a concrete, observable feature of syllables (syllables as actually uttered). In analyzing samples of actual speech I discovered soon enough that prominent syllables are not always stressed syllables, and that stressed syllables are not always made prominent, as others did before me:

* Notes can be found at the end of the chapter p.100).

...in the vagaries of performance almost
any item (...) can bear phonetic cues to
stress...

(Brown et al., 1980, p. 32)

Bolinger's distinction between stress and accent, therefore,
seems a useful one, and from now on I will speak of accent
to designate the actual prominence of syllables, and reserve
the term stress for the potential of becoming accented which
is accorded to a syllable by a theory of English stress.

For Bolinger, only one rule can be given to relate
the two terms: accents are placed on stressed syllables.
Which of the stressed syllables receive an accent, and why,
is predictable "...only if you are a mind reader...": accent
is "...a performative, realized in the intonation..." (1972,
p. 643).

Other writers, claim that such a thing as 'correct
accent' does exist, that accent can be predicted from stress,
not only when words are uttered in isolation, as specimens
of the correct use of a language, but also in connected,
spontaneous speech. According to them, accent is divorced
from the speaker's intentions, systematic rather than a
'performative'. They give rules based on formclasses -
classes of words which, as it is often put, 'retain their
stress' and classes of words which 'lose their stress' in
the sentence.[3] The exceptions - all too easily observed -
are listed and categorized: rhythm may cause the addition
or deletion of accents;[4] items which are informationally
'given' in the context may not receive the accent predicted
for them;[5] specific syntactic structures may require a
relocation of accents;[6] 'contrastive' intonations (as in

<u>I said IN, not ON</u>) may cause normally un-accentable syllables to receive an accent[7] - to give some of the most frequent examples.

Generative phonologists also predict accent from stress, on the basis of classes of syntactic constituents (which include, of course, word classes) and by means of cyclically operating transformation rules.[8] Exceptions are regarded as phonetic rather than phonological (so that both stress and accent become, in effect, part of the abstract language system),[9] or as "...deviations from the normal pattern...", a satisfactory description of which "...will require the development of methods not currently in use in phonemics..." (Chomsky, Halle and Lukoff, 1956, p. 78).

For those who distinguish, within the boundaries of intonational phrases, 'primary' and 'secondary' accents (or 'nuclear' and 'non-nuclear', or 'tonic' and 'non-tonic' accents), the 'secondary' accents may be predicted by means of formal rules, while semantic, emotive, or 'informational' values determine which of them become 'primary'.[10]

At first sight this problem may not seem immediately relevant to my purpose. Before I can relate stress and accent, before I can reliably establish whether there is a pattern underlying accentuation, (and, if so, which) I must first find a principled method of locating the accents - regardless of whether they occur on stressed syllables or not. A purely phonetic task, it would seem, a question of listening to phonetic cues, or measuring them instrumentally.

Yet the matter is not so simple. There is evidence that the listener's intuitive knowledge of stress interferes

with his perception of actual prominence, that we hear, not only the accents which are in fact realized in the acoustic substance, but also those we know 'ought to be there'.[11] And such non-phonetic factors can interfere even in the systematic and supposedly purely phonetic perception of linguists, as Lieberman (1965) has shown.

This means, I think, in the first place, that an auditory method of locating the accents must be preferred over an instrumental one: machines cannot hear 'what ought to be there'. If, in reality, the perception of phonetic cues is inextricably interwoven with tacit phonological knowledge, it may be unwise and artificial to want to separate what is normally heard as one - a conclusion to which others have also come.[12]

My conviction that this tacit knowledge of stress should be allowed to take its natural course in auditory intonation analysis, however, does not mean that, in locating the accents, I should rely on an explicit theory of the relation between stress and accent. There is a difference between the linguist's knowledge of stress and the normal language user's knowledge of stress, between the explicit, formally taught rules of a linguistic description and the unwritten rules operating in ordinary speech perception, below the level of conscious discrimination.[13] This difference should not be forgotten. To remain true to the natural human perception of speech, auditory analysis should attempt to arrest the normal, fleeting impression of accented-ness in all its ineffability, to bring to the surface what the normal listener hears, without being aware

of hearing it.

2.1.2 The phonetic cues of accent

Experiments have demonstrated that phonetic fore-
grounding of syllables can cause them to be heard as accented.
If one syllable has a higher fundamental frequency (closely
relatable to perceived pitch) or intensity (somewhat less
directly relatable to perceived loudness) or a greater dura-
tion than another, it is likely to be heard as accented.[14]
Vowel quality also plays a role.[15]

Normally all these phonetic cues operate at once,
in various admixtures. Pitch, especially pitch movement or
a step up in pitch, is generally regarded as the strongest
cue, duration as the second most effective.[16]

If one syllable is shorter than another, yet
higher in pitch, which of the two will be heard as accented?
The answer is that the effect of the cues must be set off
against each other, equivalent values of their increments or
decrements computed. That pitch is generally the strongest
cue does not mean that it always is the dominant cue in
connected speech. There exists, between the cues, what
McClean and Tiffany (1973) have called a 'trade off relation-
ship', which may cause, for example, pitch to become less
important as an accentual cue as speech becomes faster, or
softer, and which may cause duration to take over as the
dominant cue towards the end of phrases.[17]

Despite such evidence many forms of intonation
analysis make pitch the dominant or even the only accentual
cue, and reduce the perception of intonation to the perception

of pitch. Pointing out that in whispered speech intonation
patterns are recognizable despite the absence of pitch,
Denes has warned against making this reduction too absolute:

> ...Although intonation patterns are often
> associated with particular changes of
> fundamental frequency, it is still possible
> that variation of other characteristics
> (...) may serve as cues...
>
> (Denes, 1959, p. 107)

Not only do pitch, intensity and duration normally
operate simultaneously, they also appear to the listener
as an indivisible perceptual whole:[18]

> ...the fact that the listener cannot
> isolate single dimensions of the acoustic
> stimulus as they have been encoded into
> continuous speech is no more surprising
> than that he cannot isolate first and
> second formants...
>
> (Martin, 1975, p. 174)

Yet, not a few methods of intonation analysis
appear to assume that these acoustic dimensions operate
separately, and require the analyst to separately assess for
prominence the 'stress' (here meaning loudness) and the
pitch of one and the same syllable - the former as a feature
of the syllable, the latter as a 'point' in the contour.[19]
It is a distinction which is in practice rather difficult
to make, something which Armstrong and Ward noticed more
than fifty years ago:

> ...it is often difficult to decide
> whether stress or intonation or a
> combination of the two is responsible
> for certain effects...
>
> (Armstrong and Ward, 1926, p. 3)

Other methods of auditory intonation analysis
involve an (auditory) distinction between different types

of accent. Crystal, for example, defines the difference
between 'stress' and 'accent' as follows:

> ...in the case of stress the dominant
> perceptual component is loudness, in
> the case of accent (...) pitch...
> (Crystal, 1969, p. 120)

I found it impossible to apply this criterion with any
consistency. Yet, Crystal is not the only one to propose
it.[20]

There is another risk in basing an auditory
analysis of accent on specific accentual cues - the risk of
reducing the number of cues which could in principle cause
a syllable to be perceived as accented to a small, closed
set.

It seems to me that any phonetic cue which can
physically be produced in conjunction with a syllable in
such a way that not too much of the segmental information is
lost, and that that syllable is set off against other,
adjacent syllables, can serve to attract the listener's
attention to it, and hence cause it to be heard as accented.
Some of the more unusual cues reported in the literature
include downwards obtrusions of pitch,[21] laryngealization,[22]
special consonantal allophones,[23] and the pitch substitutes
operating in whispered speech [24]. Even non-phonetic cues
may play a role; kinesic cues, for example: eye-blinks,
head-nods, hand gestures.[25] And, as Martin (1970) has
demonstrated, otherwise identical syllables may be heard as
either accented or unaccented depending on whether they occur
on the ictus of a rhythmic foot or not - an effect wellknown
in the psychology of auditory perception in general, but

less often studied in connection with speech.[26]

Research shows, then, that accent fully shares in the complexities of human perception. It cannot always be related to measurable acoustic dimensions of the stimulus, and it is mixed up with non-phonetic factors, for example the listener's tacit knowledge about stress (cf. 2.1.2 above) and, according to Fry (1958) and others, the listener's kinaesthetic memory of his own production of the syllable he perceives.

There is in all this another argument in favour of the auditory method over the instrumental method: machines perceive separately what normally is perceived as an indivisible whole. We must conclude, with Lehiste, that perhaps

> ...there is no automatic way of recognizing
> stressed syllables...
> (Lehiste, 1970, p. 110)

There is also an argument here against a form of auditory analysis which attempts to turn the analyst into such a machine, and requires him to hear separate phonetic cues when this is not the way accents are normally perceived. In making an auditory analysis, we must listen for global perceptual effects rather than specific phonetic cues. Instead of analyzing before we hear, we must hear (and make ourselves aware of the fact that we hear) before we analyze: once the location of the accents is established in this way, we can begin to investigate the phonetic shape of the accents, we can ask: how did the speaker make them into accents, why did the listener hear them as accents. And for that stage of the analysis instruments can be useful - but as a

means of precisely expressing the results, rather than as a means of identifying the shapes.

2.1.3 The function of accent

If accent cannot be predicted with accuracy, if it can manifest itself in a variety of phonetic shapes, and if even objective prominence need not always be present, we must perhaps define it in terms of the one element which is never absent, the subjective impression of prominence, the global perceptual effect: <u>accent is a way of attracting the listener's attention to certain syllables in a stretch of speech at the expense of other syllables in their immediate vicinity</u>[27] - a definition which can be made into a criterion for the auditory analysis of accent, as we will see in 2.1.5 below.

There is some evidence for the reality of the psychological element in this definition, the term 'attention': from experiments in which the reaction time to accented and unaccented syllables was measured, it has become evident that accented syllables are 'processed' more efficiently than unaccented syllables (Cutler and Foss, 1973); they are probably also retained in our memory for a longer time (Martin, 1972; Nooteboom et al., 1976).

Accent is a way of singling out some syllables as more worthy of (immediate) attention than others, more important to remember. Accent as a facilitator, a structuring which helps the listener to process the incoming speech signal more efficiently, a kind of lubricant.

But the role of accent is not restricted to speech decoding as a _process_, it also pertains to the outcome in the process, the understanding of the message. What is most in need of efficient processing is also what is most in need of getting across to the listener, what carries the greatest semantic weight.

It is not for nothing that the words with which non-linguists speak about accent always denote both its formal and its functional aspect. Words like 'stress'. 'accent', 'emphasis', in common usage designate a process in which the perceptual importance of a word or phrase conveys its importance as information. There is, as Gibbon has said, a 'dominance of the functional mode' in intonation (1976, p. 2) - a fact acknowledged, in different ways, by many linguists: [28]

> ...those words are stressed which are
> felt by the speaker to be important...
>   (Armstrong and Ward, 1926, p. 3)

We have seen that, for some linguists, this applies only in the case of the 'primary accents', while others believe accent to be altogether systematic and formal, rather than governed by the speaker's intention to make this or that word important. Yet, stress itself is part of the language system only to facilitate the efficient recognition of words, and it is placed on those syllables which carry the greatest semantic load. It is only because they do not acknowledge that the efficient recognition of words depends on contextual as well as linguistic redundancy that these writers go too far in conflating system and performance.

Stress signals the _theoretical_ importance of syllables for
the understanding (recognition _and_ interpretation) of words,
accent their _actual_ importance for the understanding of the
message.

The 'importance' signalled by accent is often seen
in terms of information - the 'given'/'new', or 'theme'/
'propos' distinction.[29] What the speaker takes for granted
(and feels his listener should also (be able to) take for
granted) remains unaccented; what he does not take for
granted (and feels his listener should pay some attention to)
receives an accent.[30] 'Emotive' importance, on the other
hand, is usually viewed as resulting from modifications in
the shape of the accent, or from the unexpectedness
('markedness') of an accentual form in the context.[31]

At this point it is perhaps not yet necessary to
take up this question in detail. As long as we can conclude
that the 'importance' of accent is an importance both in
terms of the efficient processing of speech and in terms of
the message and its meaning, we have some kind of guarantee
that accent can give us the means to study what announcers
treat as important and what they take for granted and assume
their audiences to be able to take for granted as well. If
accent can, as Bolinger says, be predicted 'only if you are
a mind reader', then studying what speakers accent can be
a way of 'reading their minds'. Thus if an announcer accents
a personal pronoun:

[if/ _you're_ looking//] [for a _top_ quality/ _used_/ _car_//]
he does not take it for granted that there is a 'you' who
might feel personally addressed by him and tries to lay more

emphasis on the 'personal' nature of what is, in reality, one of the most impersonal and distant forms of communication the world has ever known. And if this turns out to be a habit common to all announcers, then we can, I feel, interpret it as a strategy of the medium to obscure its impersonal character. More generally, if radio announcers turn out to follow a common accentual style, then this style could help us to 'read the mind' of the institution which prescribes it, the mass medium.

2.1.4 The auditory analysis of accent

The definition of accent given above implies that an auditory analysis must involve a comparison - a point also made by Lehiste:

> ...the stressedness of a vowel cannot be
> established without comparing it to
> another segment in the sequence...
> (Lehiste, 1970, p. 3)

Yet, there are instances in which such a comparison is impossible. As Martin (1970) has shown, it is possible for an accent to be recognized only because it occurs on the ictus of a rhythmic foot. This means that it is also possible for a series of accents to occur without intervening unaccented syllables - even though speakers do what they can to avoid this, as Bolinger has so admirably shown (1965).

In such cases - and they occur with some frequency in radio announcing - another perceptual effect, that of isochrony (to be discussed more fully later) must form an additional criterion for the analysis.

Subjective prominence and perceptual isochrony –
they may seem rather vague and imprecise criteria for an
auditory analysis.[32]   Yet, they need not lead to lack of
agreement when applied to samples of actual speech by
different analysts. To lend some credibility to this
assertion, I asked a group of 24 undergraduate students of
linguistics[33] and their phonology lecturer (a highly
qualified phonetician) to mark the accents in the transcrip-
tion of a radio news item and a segment from a more or less
spontaneous interview with the same speaker. I had
previously marked the accents myself.

The group was allowed to listen to the recording
of these items as often as they wished, and given these
instructions:

> Underline all the accented syllables, that
> is, all the syllables which you hear as
> somehow more prominent than other syllables
> in their immediate environment – however
> weak this prominence may be in absolute
> terms, and whichever the phonetic cues that
> cause it to be perceived.
>
> Rhythm may play a role in your judgment if
> difficulties arise. For example, in:
>
> It's ten/ past/ three
>
> ten, past and three may all be of equal
> prominence, or, alternatively, may all
> appear prominent but to a slightly
> different degree. If these syllables are
> also of approximately equal duration,
> constituting 'rhythmic feet', they can all
> be regarded as accented, despite the fact
> that, for example, past may not be more
> prominent than its immediately neighbouring
> syllables, or may even be somewhat less
> prominent.

Agreement between my own judgments and those of
the lecturer was, in 2 subsequent transcriptions, made with

an interval of 4 days, 90.7% and 83.7%. Agreement between
my judgments and those of the students was 82.6%. Statis-
tically, this establishes the reliability of my method of
marking accents at an 0.83 on average.[34]

* * *

2.2 GROUPS

Intonation divides utterances into groups ('tone-
groups', 'tone-units', 'phonological phrases', 'phrase-units')
- groups separated by junctures ('intonation breaks', 'phrase
boundaries', 'terminals'), displaying a pattern of some
kind, usually described in terms of pitch ('tune', 'sentence
melody'), and often divided into components ('pre-tonic' and
'tonic'; 'pre-head', 'head', 'body', 'nucleus' and 'tail').
Groups, finally, of which the intonation is either 'closed'
('declarative intonation', 'tune I') or 'open' ('question
intonation', 'continuative intonation', 'tune II'), depending
on the direction of pitch movement at the end of the group.

On the details of all this there is no agreement.
To what extent the group boundaries can be predicted by
linguistic rule; how many group patterns there are and what
their significance may be; whether they should be divided
into subcomponents and, if so, how; by which means the
speaker signposts their boundaries and in which way the
listener is able to perceive them; how the linguist should
identify them in spontaneous speech - on all these questions
a divergence of opinion exists.

Without first clearing a track through the proliferation of theories and methods of description, I cannot adopt or develop an approach to the analysis of intonational groups. In this section of the chapter I will try to clear that track, dealing, in turn, with group junctures, with the contours that make up the groups, and with the problem of 'open' and 'closed' groups.

2.2.1 Intonational junctures and syntactic junctures

Many writers on intonation consider juncture an intonational device to mark syntactic structure: for them, the boundaries of syntactic groups and the boundaries of intonational groups coincide[35] - although often the qualification is made that not every syntactic boundary needs to be 'marked' by an intonational boundary. In casual styles of speech, for example, less syntactic boundaries are intonationally marked than in more formal styles - yet, in casual speech too, every intonational break coincides with a syntactic boundary.[36]

Others see the location of the intonational junctures as less fully predictable by linguistic rule: there is, at most, a tendency for intonational and syntactic junctures to coincide. The intonation group is, in the first place, an 'information unit' (Halliday, 1967), a 'sense group', in which the words "...are felt as closely associated and are so pronounced..." (Bernard and Delbridge, 1979, p. 102).[37]

It is my own observation that pronounced intonation breaks can and do occur almost anywhere, in spontaneous speech

as well as in the somewhat less than spontaneous speech of
radio announcing: in the middle of noun-phrases, between
prepositions and articles, between adjectives and the nouns
they modify, before and after hesitation noises - the list
is long and will be drawn up more fully in chapter 4. And
I am by no means the only one to have discovered this.
Mitchell and Delbridge (1965) report that in their corpus
of Australian adolescent speech:

> ...long pauses are often placed at points
> not at all related to the syntactic
> structure of the utterance...
> (Mitchell and Delbridge, 1965, p.65)

And Brown et al. (1980), in transcribing the
intonation of their corpus, stumbled on the same problem:

> ...in many cases we found that syntactic
> or semantic criteria would not enable us
> to make a principled decision as to where
> to assign a tone group boundary...
> (Brown et al., 1980, p. 42)

Because of this, Brown et al. decided to retreat to a safe
position, recognizing only 'pause-defined units' as intona-
tional groups - and it is indeed tempting to conclude that,
if syntax cannot provide criteria for intonation analysis,
groups should be delimited on the basis of clearcut phonetic
cues. To relate syntax to the boundaries established in
this way can always be relegated to a later stage of the
analysis.

Yet, it may be that non-phonetic factors, including
syntactic structure, interfere in the normal perception of
intonational breaks. Experimental results point in this
direction. When Fodor and Bever (1965) and Garrett, Fodor
and Bever (1966) superimposed clicks on the boundaries between

syntactic groups, the clicks were accurately perceived.
Superimposed on syllables in the middle of syntactic groups,
however, they were heard as closer to the group boundaries
than they in fact were. According to an experiment by
Abrams and Bever (1969), it takes subjects longer to react
to clicks placed at syntactic boundaries than to clicks
placed in the middle of syntactic groups - the 'processing
load' is greater at these boundaries, so Abrams and Bever
suggest, the sense of the preceding group must 'fall into
place' there. The point is that these effects also occur
in the perception of intonational groups, as Wingfield (1975)
found when he repeated these experiments, using both
boundaries which were intonational as well as syntactic,
and boundaries which were only intonational. Again, while
de Rooy (1975) shows that listeners can perceive 'syntactic
boundaries' from intonational information alone, Lehiste
(1973) shows that syntactic information alone can cause
listeners to hear what is normally regarded as an intonational
juncture cue: her subjects judged the final rhythmic foot
of a syntactic group to be longer than the feet which
preceded it - even though it was, in fact, of equal duration.

Though not predictable by means of explicit
linguistic rules, intonation is, nevertheless, somehow bound
up with syntax. We should be careful not to accept too
uncritically the view that the perception of intonation
boundaries must necessarily result from observable, 'objec-
tively' present juncture cues. We should also abandon the
hope that such boundaries can be located by means of
instruments - or by means of methods of auditory analysis

which fail to take account of the fact that, in normal perception, the listener's tacit, informally acquired knowledge of syntactic structure is inseparably mixed up with the perception of phonetic juncture cues, methods which require the analyst to establish boundaries only on the basis of specific phonetic phenomena. Lieberman (1965) showed that even linguists who believed they were judging intonational breaks purely on phonetic grounds, were, in fact, influenced by syntax and made quite different judgments (more intonational breaks) when the speech was made 'content free':

> ...the linguists' ears were remarkably
> good so long as they did not hear the
> words of the message...
> (Lieberman, 1965, pp. 49-50)

But normally we do hear the words.

Informal, subconscious syntax, then, should perhaps be allowed to play its natural role in the auditory analysis of group juncture, so that the analysis can recover the kind of global perceptual effect which also operates in normal speech perception.

2.2.2 The phonetic cues of juncture

The speech pause,[38] the lengthening of the final syllable or syllables of a group,[39] and the occurrence of a number of specific pitch movements[40] have been shown to function as junctural cues in experiments on the perception of intonation.

On the other hand, it has also been found that none of these cues can be expected to always occur, and the

assessment of their relative importance varies widely. Some linguists state that "...sense groups are always separated by pauses..." (O'Connor and Arnold, 1961, p. 4), others that "...the tone-group boundary does not imply any break in the structure, and there is no pause between successive tone-groups..." (Halliday, 1970, p. 51). Experimental phonetics has supplied evidence both for the superior importance of the pause (or its substitute, the lengthening of the group-final segment) (de Rooy, 1976), and for its significantly lower importance (Isačenko and Schädlich, 1965).

Neither can pitch movement be said to be a criterion on which one can always rely. Crystal (1969), who claims to use a 'wholly phonological' definition of the tone-unit boundary (p. 205), admits that he sometimes had to take recourse to grammatical or semantic criteria to place the boundaries,[41] and although he announces that "...each tone-unit will have one peak of prominence in the form of a nuclear pitch movement..." (p. 205), he later admits that 8% of his tone-units did not have such a pitch movement: in such instances, he says, we should look for a tone which is distributionally similar (p. 215). But this is putting the cart before the horse: one cannot begin to look at the distribution of a phenomenon before that phenomenon has been identified. And the matter becomes even more complicated when one reads that these nuclear pitch movements do not always occur at the end of groups, and are, in fact, not always nuclear: there are also 'non nuclear pitch glides', formally similar to nuclear pitch movements, but distribu-tionally different. Crystal's 'wholly phonological' method

in fact mixes phonetic, syntactic, semantic and distributional (i.e. intonation-theoretical) criteria, and is, in the final resort, not even fully auditory: queries, we are told, must be referred to an acoustic record (p. 14). Above all, it operates at once on the basis of a theory of intonational structure and on the basis of phonetic cues, and thus confuses system and performance, phonological form and phonetic realization. Brown et al. (1980) comment in a similar vein on Halliday's tone-group, showing it to be an abstract theoretical unit, even though Halliday himself considers it related to the utterance the speaker actually produces, and finding it, therefore, of little value for the auditory analysis of spontaneous speech (Brown et al., 1980, p. 46).

Moreover, theories of intonational structure, even when they explicitly distinguish between system and performance,[42] almost invariably make pitch the sole criterion - and we have seen that other factors, phonetic as well as non-phonetic, may also play a role in the perception of intonation.

Anything that can physically be inserted between two groups, any feature that can be superimposed over the whole of one group in such a way that it contrasts with the way that same feature is superimposed over another, adjacent group, can, I believe, serve to segment the utterance into groups. Pike (1946) and Crystal (1969) have listed many such features - as 'modifications of contours', or 'polysyllablic effects' rather than as cues which might set off one group against another for the purpose (among other purposes, perhaps) of segmenting the utterance into groups. They

include pitch key, pitch range, rate of utterance, loudness,
voice quality, manner of articulation (e.g. degrees of
'tension', 'clipped' versus 'drawled' speech), and a number
of 'effects' referred to by means of musical terms: legato
and staccato, crescendo and decrescendo, accelerando and
ritardando. Bolinger (1964) noted that narrowing or widening
of pitch scale can separate otherwise equal ranking elements
(p. 25). Fry (1958) speaks of

> ...a strong tendency for a sense group to
> be spoken in one (pitch) key, and for
> musical modulation to take place between
> groups...
>
> (Fry, 1958, p. 141)

It might also be said that the preoccupation with
pitch has led to a neglect of rhythm. Too often rhythm is
seen as entirely separate from intonation, a feature signal-
ling the (kind of) language being spoken,[43] or enabling
stylistic, aesthetic 'effects',[44] or forming the
'linguistic copy of the heartbeat', as Bolinger tagged a
view with which he does not, in fact, agree (1965, p. 14).
And even when it is noted that the effects of rhythm and
intonation are difficult to tell apart,[45] the logical
conclusion, that it might, therefore, be wiser not to want
to tell them apart, is not often made.

Perhaps the reason for this neglect was the emphasis
which the search for 'objective cues' has always had in
intonation research: measurements of the actual duration
of rhythmic feet, supposedly isochronous, showed that isochrony
did not 'objectively' exist.[46] But subjective facts,
human as opposed to machine- or machine-like perceptions,
have a birthright too, and may often be more relevant to the

understanding of speech communication than objective, measur-
able facts. Where this has been admitted,[47] or experi-
mentally verified,[48] it was possible to acknowledge that:

> ...it is the rhythmic units that the speaker
> manipulates to reflect syntactic structure...
> (Lehiste, 1973, p. 1231)

At the group boundaries there may be 'organ-
points', temporary suspensions of the regular rhythm,
resulting from a pause, or from the lengthening of group-
final elements - it does not matter which.[49] Or the rhythm
may change when the speaker adopts a different footlength
as standard for the timing of the intervals between accents -
a cue described, among others,[50] by Bolinger:

> ...in measuring the length of feet (....)
> one should start over if a pause
> intervenes (...) and not assume that the
> speaker will resume at the same pace (...)
> it is already a large enough assumption
> that he will not change the gait of his
> horse in midstream...
> (Bolinger, 1965, p. 165)

When establishing the location of junctures, I
can, of course, not fail to detect, from time to time,
junctural cues such as described in the literature. But if
I do not detect such a cue, I should place some trust in my
initial perceptions - my perceptions as a language user,
rather than a linguist - and refrain from erasing the
boundary, or from looking in every direction for a reason
why it might be there. Looking for reasons can be delayed
until later. Auditory analysis should concern itself with
hearing.

2.2.3 The function of juncture

If specific phonetic cues cannot be relied on,
and if the relation between syntactic and intonational
boundaries cannot be expressed in terms of formal syntactic
theory, juncture must, perhaps, be defined in terms of the
subjective impression, the perceptual effect without which
it has no existence: <u>juncture is a way of making the listener
perceive that one group of words has come to its conclusion
and another is about to begin</u>.

This definition, again, has the advantage that it
can be made into a criterion for the auditory analysis of
juncture. And there is, as we have seen, some evidence for
its psychological reality.[51] Like accent, juncture serves
to help organize the act of speech decoding. It tells the
listener: now is the time to collate the information
extracted from the preceding group of words, and to store
that part of it that must be retained for further 'process-
ing', at the next higher level of information collation.

But the act of speech decoding has as its ultimate
aim the understanding of the message. The 'processing' I am
speaking about cannot be seen as separate from understanding:
it is the process of understanding itself. The group of
words which juncture instructs the listener to collate and
retain in memory as one unit is also a group of words which
belong together semantically, a group intended by the speaker
to form one unit of information, and received by the
speaker as such. Better still is Halliday's expression 'a
move in the speech act' (1967, p. 30), because, although the
group may be demarcating a unit of information (in which

case the boundaries may enclose a syntactically fully stated
group), it may, for example, also be a'move' of signalling
to the listener that the speaker has not yet finished his
turn (in which case the boundaries may enclose a mere
'filler' noise): a classification of micro and macro 'speech
act moves' might be more capable of accounting for intona-
tional boundaries than conventional syntax, and definitions
of the function of juncture should perhaps be broad enough
to include, not only information, but also other possible
'moves' - O'Connor and Arnold's definition, for example, in
which the 'sense group' is described as:

> ...a closely-knit group of words which
> contributes to the situation in which
> we are placed at a given moment...
> (O'Connor and Arnold, 1961, p.3)

The speaker groups together what he feels needs
to be grouped together for his purpose of the moment. In

[and//] [page/ nine//] [of/ Friday's/ Mirror//]

the conjunction and is made into a separate group, not because
it provides information, but to deliberately withhold
information for a moment, to create anticipation, a kind
of 'suspense'. Page nine and Friday's Mirror are also made
into separate groups, 'information' groups, this time,
separate for the sake of clarity, and, perhaps, because the
speaker assumes that, in speaking to a mass audience, one
can take nothing for granted: the mass audience as an
ignorant mob, with an extremely limited capacity of paying
sustained attention to spoken words.

Another pronunciation would have been equally

acceptable:

[and page/ nine of/ Friday's/ Mirror//]

but would not have created the effects of 'suspense' and 'added clarity', and would, perhaps, have sounded more like news-reading than like a commercial.

Grouping, created by intonational juncture, can tell us something about the speaker's intentions and assumptions - and if these have, to some degree, been institutionalized and professionalized, about the intentions and assumptions of the institution on behalf of which he speaks - the mass medium.

2.2.4 The auditory analysis of juncture

My criteria for marking intonational juncture follow from the definition and general considerations discussed above. Here, too, I have tested their reliability by asking others to use them in transcribing the intonation of two segments of speech - the segments and the 'others' are those described in 1.1.5 above. What I asked the group to do was, in fact, a full intonational transcription of the two segments, following a step by step method - that the results are reported piecemeal in this chapter is merely for the sake of convenient exposition.

The instructions I gave the group expressed, as best as I could, the criteria I have applied myself in transcribing my corpus. It should be noted, however, that I asked them to mark juncture after they had already marked

the contour boundaries - an order which I later decided to reverse, for reasons to be discussed in the next section of this chapter. These, then, were the instructions:

> Each of these contours forms, either by itself, or together with an unspecified number of other contours, a higher order unit which we will call a group. Mark the boundaries between these groups with a double stroke (//). One or more of the following cues will often, though by no means always, be present:
>
> (i)   a rising or falling pitch movement on the final syllable(s) of the group
>
> (ii)   a pause
>
> (iii) a considerable lengthening of the last contour of the group.
>
> The most important criterion, however, is that you feel that the syllables in the group are spoken as an indivisible whole.

The results were encouraging. Agreement between my transcription and those of the phonology lecturer was 96.6% for his first transcription and 93.1% for his second, 4 days later. Agreement with the students' transcriptions was, on average, 89.8%. The reliability of my method of analysing intonational juncture can be established at an 0.90 average. [52]

2.2.5 Rhythmic feet versus contours

What I call 'contour' here, is more often called a 'foot' or a 'rhythmic foot'. [53] I chose the word 'contour' because, the more I listened to intonation, the more I became convinced that there is a much closer connection between rhythm and intonation than has usually been assumed. If the intonational group is, as Hultzen has

said, a 'pattern of accents' (1964, p. 86), and if accents
are always also rhythmic accents, then rhythm, rather than
being a kind of neutral substratum, must be integrated with
intonational structure at all levels, also that of the
building blocks of the intonational group. The rhythmic
foot, therefore, should have more relevant features than its
timing alone.

Pike (1946) appears to have thought along similar
lines. Although his units are defined in terms of their
pitch contour, and although 'upbeat' syllables may form part
of them, rather than that they are always initiated by
accented syllables, they appear to be roughly the size of
rhythmic feet, and, like rhythmic feet, consist of one
accented syllable and a varying number of unaccented syllables
- the five 'primary contours' in Pike's wellknown example
show this:

```
The 'boy in the 'house is 'eating 'peanuts 'rapidly
3-   °2-3 3-      °2-  3 3- °2-   3 °2-    3 °2-    4
```

Halliday's solution (1967; 1970) is almost the opposite:
while Pike describes larger rhythm units as a series of
melodic contours, Halliday divides larger melodic units into
a series of smaller rhythmic units, or rhythmic feet.
Combining the two approaches would perhaps lead to a more
complete, a more integrated picture of the intonational
structure of groups.

The question, then, is: does the contour start
at the low point of its dynamic shape, which may be before
the 'upbeat' syllable, or at the rhythmic beat?

The former possibility might lead one to suppose that junctures (less pronounced, perhaps, than group junctures, but similar in kind) might separate the contours within groups - and the literature on 'internal open juncture'[54] suggests that micro-junctures of this kind do, indeed, exist. On the other hand, there is no agreement about this: according to Klatt (1969) the word boundaries within groups are hidden by coarticulations, and O'Connor and Tooley (1964) found that internal open juncture is not very accurately perceived. It may be that the group is, as Pike (1946) has said, a 'single rush of syllables' (p. 34), within which the individual words are identified only by the accents.

Trim (1964) describes both possibilities. His 'minor tone-groups' are separated by junctures (though, in rapid speech, they may be 'fused'), not isochronous, and have boundaries which always coincide with word- or phrase-boundaries. Rhythmic feet, on the other hand, he describes as isochronous and continuous. They may begin in the middle of a word, and, unlike minor tone-groups, always start on the accented syllable. According to Trim, the former operate in German, the latter in English. Minor tone-groups have also been reported for the intonation of a Philippine language, Chamorro (Garvin and Mathiot 1958), and as a result Chamorro speakers sound 'choppy' when speaking English.

My own approach was, at first, to look for units like Pike's contours, or Trim's minor tone-groups, to try and hear, between each pair of accented syllables, a juncture -

without assuming, however, that these junctures would
always coincide with word boundaries. But this part of my
method did not lead to as much agreement as the marking
of group boundaries.

Instructions were as follows:

> Each of the syllables thus underlined forms,
> either by itself or together with an
> unspecified number of unaccented syllables,
> one and only one contour. Mark the boundaries
> of these contours with a single stroke (/).
> One or more of the following cues will often,
> though by no means always, be present:
>
> (i)   a change in pitch direction
>
> (ii)   lengthening of the final syllable of
>   the contour
>
> (iii)   a short pause, or phonetic juncture cue:
>   lack of assimilation; aspiration;
>   glottal stop; greater duration of
>   occlusions, etc.
>
> But the most important criterion is that you
> feel the syllables in a contour belong
> together and are spoken as an indivisible
> whole.

Although there was, in the lecturer's transcriptions,
100% agreement with my contour judgments, those of the
students agreed with mine for no more than an average of
71.9%, which, statistically, indicates an 0.72 average level
of reliability.[55]

It occurred to me that the lecturer and I had,
perhaps, made a phonetic analysis of the speech segments,
rather than the sort of auditory analysis I have been pro-
posing in the course of this chapter, while the students
might have tried to actually hear ('feel') boundaries, as
I had suggested they should. And hearing subtle junctures
of this kind is not easy, as O'Connor and Tooley (1964), who
obtained 66.8% agreement, also found.

As a result I decided to place less emphasis on cues which require a trained ear to be heard, and more on a global perceptual effect, that of isochrony: contours should be marked by placing the boundary before the accented syllable, as if they were only rhythmic feet. I also decided to reverse the order of the second and third step in my analysis, and mark group boundaries before contour boundaries.

Contours, then, are treated as waves, with peaks and troughs, but not separated by junctures. As seamless repetitions of a pattern which perhaps remains identical for a time, perhaps changes gradually, or weaves variations on a given motif, but of which the beginning- and end-points cannot be determined other than in an ultimately arbitary manner.

The decision involved me in a new problem. From time to time one comes across accents which appear not to participate in the perceptually isochronous rhythm of the group - the accents in foreign words, for example:

[and/ then the/ beautiful/ Song from/ Moulin/ Rouge//]

Such accents will, from now on, be called ancillary accents, and they will be indicated with interrupted lines:

[and/ then the/ beautiful/ Song from/ Moulin/ Rouge//]

2.2.6 The function of contours

We can now define the contour: contours are groups
of syllables within the intonational group, forming
perceptually isochronous cycles lasting from the beginning
of one rhythmically accented syllable to the beginning of
the next.

There is evidence for the perceptual reality and
psychological importance of contours: Köhler, for example,
'rhythmicized' 12 mono-syllables by inserting short pauses,
thus creating rhythmic feet of 2, 3 and 4 syllables. He
reports an increase in memory retention for these syllables -
by 49% for the 2-syllable feet, by 38% for the 3-syllable
feet, and by 36% for the 4-syllable feet. Cutler (1975)
asked subjects to recall certain 'target' phonemes in an
utterance, and found that, if he upset the isochrony by
adding a quarter-second pause between mono-syllable feet,
his subjects' reaction time to that syllable slowed down.
From the estimates which have been given for the maximum
length of contours and groups, it would further seem that,
contrary to what Lehiste has said, contours, rather than
groups, correspond to Kozhevnikov and Chistovich's 'syntagma'
(1965), which, so they report, cannot be longer than the
decay-time of 'pre-categorical storage'.[56] Finally, the
existence of contours and groups as two distinct dimensions
of patterning is also supported by the study of speech errors.
Nooteboom and Cohen (1975) report a maximum 'span' of 9
phonemes for errors involving phonemes ('spoonerisms' in
other words), and a span of 12 syllables, more than the

maximum length of the contour, for errors involving the interchanging of words.

The role of contours in the process of speech decoding, then, emerges as similar to that of the group: at the end of a contour 'cycle', the listener must integrate the information contained in the preceding contour, and store that part of it he needs for further, higher level 'processing', at the boundary of the group - 'that part' being, in fact, the accented syllable, to the understanding of which the unaccented syllables have now made their contribution.

This means that contours, like the other intonational elements we described, are important, not only for speech decoding as a process, but also for the understanding of the message. An importance which, in this case, follows from what we have already said about the function of accent (in 2.1.4) and to which we will return once more, when discussing the concept of ranking, in 2.4 below.

2.2.7 The auditory analysis of contours

The revised method of auditory contour analysis, then, is a matter of listening for rhythm only, of tapping the beat to the accents within and only within each group, of marking those which do not coincide with the beat as ancillary accents, and of placing boundary marks before each rhythmic accent. As such it uses criteria similar to those used in the auditory analysis of accent and group juncture, criteria which are not too alien to the normal perception of speech.

2.2.8 Syntax and group closure

A description of the pitch configurations (levels, movements) at the end of groups is part and parcel of most forms of intonation analysis, linguistic as well as phonetic, and, more often than not, the configurations are connected with specific functions. Apart from serving as juncture cues (which, as we have seen, raises the question of their relation to syntax), they are said, by some, to be determined by mood and modality of the sentence (and to belong, therefore, to the realm of syntax - cf. Halliday, 1970, p. 21), by others to express attitudinal meaning, by again others to be governed by a combination of the two. When an intonation normally associated with a particular type of sentence nevertheless occurs on another kind of sentence (a 'question' intonation on a sentence which, syntactically, is a statement or command, for example) it is often said to be a 'marked' intonation, and this 'marking' adds attitudinal meaning to the sentence. [57]

Such marked intonations must occur fairly often, if we are to judge by studies such as those of Fries (1964), who found that in a corpus of 2561 'yes/no' questions, 38% did not have the predicted pitch rise at the end of the question; or Adams (1969), who, for Australian English, reported that, in a corpus of interview speech, 44% of statements lacked the expected statement intonation; or Burgess (1973), who, in a detailed examination of 12 sentences, found that only half the statements terminated in a falling pitch.

Perhaps the distinction between 'marked' and 'unmarked' rises and falls is only apparent: Cruttenden (1970), for example, believes that falling pitch always signifies 'definiteness', rising pitch always 'incompleteness' - whether or not they coincide with the completion of a sentence or lack thereof, whether or not the grammatical forms of statement and question accompany them. For Bolinger (1964), falling pitch has passed from actual conclusiveness into figurative conclusiveness, and its occurrence can therefore no longer be predicted by linguistic rule. Lieberman (1967), on the other hand, while acknowledging that, for example, actors can sometimes express 'assertiveness' by using a fall where the grammatical structure does not predict it, believes that this can happen only when the constituent structure of the sentence unambiguously expresses the actual lack of finality.

Faced with such a divergence of opinion, and observing, in my recordings, large numbers of rises at the end of statements and countless falls in the middle of sentences, it would, once again, seem prudent not to allow explicit linguistic rules to play too much of a role in the auditory analysis of group closure. Yet, here too, our experience as language users, our normal, tacit knowledge of what should sound continuative or final, is inseparable from our perceptions of the pitch configuration. Hadding-Koch and Studdert-Kennedy (1964) discovered this in an experiment designed to elicit 'question' and 'statement' judgments from listeners. Their subjects heard rises at the end of questions, and falls at the end of statements, even

where these pitch movements did not, in fact, occur, and
they perceived the pitch direction of pure tones much more
accurately than the pitch direction of speech sounds.

To mark what native speakers would ordinarily hear
as a statement (as falling pitch, even) as a rise because
'objectively' there is a rise, or because the specialized
ear can detect one, can only lead to conclusions and inter-
pretations too far removed from the realities of speech
communication. The intonation analyst must somehow try to
recreate normal hearing conditions, capture the normal,
fleeting speech impressions, and suspend his knowledge of
what should, theoretically, occur, or of what does, objectively,
occur.

.

2.2.9 The phonetic cues of group closure

Different linguists and phoneticians have discovered
different numbers of group-final pitch configurations.
For some, only two configurations exist: rise and fall.[58]
Others recognize a third, distinguishing between low
(continuative) rises, and high (interrogative) rises.[59]
Again, others add a fourth by acknowledging 'non final falls'
as well as 'final', low falls.[60] Sometimes the group-
final, 'mood marking' pitch direction is conflated with the
pitch movement of the nuclear accent - in which case there
may be up to 28 group-final tones.[61] At other times
'prominence-lending' and 'non-prominence-lending' rises are
distinguished, the latter responsible for mood and modality,
the former for the nuclear accent.[62] Where such a dis-
tinction is not made, the final accent in the group almost

automatically becomes also the nuclear accent, however small
its importance according to any other criterion - this,
perhaps, accounts for the lack of 'strong stress' (signified
by ") of the 'slightly higher' (↑) and 'narrowly rising' (n)
final nucleus in this example of Crystal's:

"How many "CHILdren have you ↑nGÓT

In the literature on group closure we find that
pitch is once more enthroned as the single determining cue,
whatever the refinements of this basic proposition. [63]
Other cues are rarely mentioned. Lieberman (1967) says that
in the 'marked breathgroup' both pitch and intensity refrain
from drifting downwards, so that these groups end, not only
on a rising pitch, but also with a sharp cut-off in loudness.
Crystal and Davy (1969) report that, in conversational
speech, tone-units may simply 'tail off' in loudness (p. 107),
rather than that they terminate in a specific pitch configura-
tion. Burgess (1973) speaks of the durational structure of
one of his sentences as 'falling'. The role of cues other
than pitch, then, is generally seen as very limited.

Accordingly I felt that there was, in this case,
no alternative but to take pitch movement as a criterion -
and I hoped that this would encompass other, non-phonetic
factors also, as it did for Hadding-Koch and Studdert-
Kennedy's subjects. Testing the method with the group of
linguistics students, I proposed three pitch movements - the
movements which I found most easy to identify in my recor-
dings. I used the following instructions:

> Mark a rising pitch movement on the final
> syllable of a group with an acute accent (´),
> and a falling pitch movement with a grave
> accent (`). Only low falls, 'sounding like
> a full stop', should be taken into account,
> and other types of fall left unmarked.

But the test had rather negative results:
agreement between my judgments and those of the lecturer
was, in his first transcription, only 25%, 4 days later
57.1%. Agreement between my transcription and those of the
students averaged no more than 45.49%, and average reliability
can be calculated as 0.45.[64]

The result persuaded me to reduce the marking of
group-final configurations to a binary choice, and to drop
pitch as a criterion: a group is either open because 'it
sounds as if there is more to come' or closed because it
sounds as if there is not - whatever 'it' is: a topic or
subtopic of discourse, a whole utterance, or a contribution
in an exchange of utterances. And, although I will continue
to use acute and grave accents to indicate openness and
closedness, this should no longer be taken as implying any-
thing with respect to the manifestation of group closure:
whether pitch does or does not play a role (and whether this
role is the 'expected' one or not) is a question we can
consider after the groups have been marked as open or closed.

2.2.10 The function of group closure

These decisions contribute the elements of a
definition: openness is what makes the listener perceive
a group as one which is not final in an utterance or self-

contained portion thereof; closedness is what makes the listener perceive a group as one which is final in an utterance or selfcontained portion thereof.

There is not much direct evidence for the function of group closure in the process of speech decoding, but I would like to believe that openness and closedness involve different stages in the act of speech decoding. When perceiving the boundary of a closed group, the listener collates and stores information from a number of previous groups (the group of groups which has come to its completion) - a process which mirrors, at a higher level, what takes place at every boundary. Instead of selecting, from among the contours of the preceding group, those that have attracted the most attention (the accents) for storage, as happens in the case of the open group, he now selects from the groups in the preceding higher order unit those groups which have most attracted his attention, and stores these in his memory as the final impression of the utterance, or for further processing, at the next higher level, if there is one.

It follows that group closure serves, not only to organize the incoming speech signal for the sake of 'efficient processing', but also to demarcate topics or subtopics in a monologue, and to mark the end of a turn in an exchange of utterances. And, once again: in doing this, the speaker is not bound by grammar - his manipulation of group closure can betray the intentions and assumptions underlying his overt message. When a newsreader repeatedly uses closed groups in the middle of a sentence:

[a/ <u>twen</u>ty one year old/ mán//][is in in/ <u>ten</u>sive/ càre//]
(in/ <u>Syd</u>ney's/ <u>Mo</u>na Vale/ Hospitál//] [after / <u>break</u>ing
his/ bàck//] [in an/ <u>acc</u>ident at a/ <u>beachs</u>ide/ <u>swim</u>ming
pòol//]

he not only takes advantage of the fact that his listeners
cannot interrupt him, but also adds a sense of self-sufficient
definiteness to each separate unit of information, contri-
buting, in this way, to the effect of disconnectedness, of
'surprising and miscellaneous events coming in, tumbling
over each other from all sides' (Williams, 1974, p. 116),
which characterizes news as a form of information.

When a speaker ends his answer to a question with
an open group:

[<u>Wèll</u>//] [you/ <u>got</u> to be a/<u>ware</u> of the éh//] [<u>of</u> the/
<u>time</u> of/ dáy//] [<u>ob</u>viouslý//]

he communicates his feeling that he may not have said the
last wórd on the matter yet, and he makes a covert appeal
to the listener to settle it by agreeing with him. And this
can even happen when no listener is present, the implied
'you know?', 'you remember?' creating a bond which, in fact,
does not exist, and suggesting a conversation where there
is, in reality, only a monologue:

[<u>let's</u> take/ <u>this</u> one/ nòw//] [from the/ <u>group</u> called
the/ <u>Carpen</u>tèrs//] [<u>formerly</u>/ <u>mem</u>bers of the/ <u>Mickey</u>/
<u>Mouse</u>/ clùb//] [on A/ <u>me</u>rican/ <u>te</u>levisíon//] [<u>Ka</u>ren and/
<u>Richard</u>/ Carpentér//] (...)

Or he can use an open group to obscure the transition between two different programme segments, to merge the announcer's role as 'a friendly voice to keep you company' (here, the 'time call') and the announcer's role as a voice for the advertiser:

[Now the/ tíme//] [is/ seven to e/ levén//] [on/ Monday/ mornîng//] [and if/ your special/ taste in/ seafòods//] (...)

If such habits turn out to be general characteristics of radio announcing speech (or of one or more of its varieties), rather than individual peculiarities, then group closure, like accent and juncture, can lead us to, e.g., the attitudes to the news, or the assumptions about the mass audience which have grown up in and around the mass medium of radio.

2.2.11 The auditory analysis of group closure

The auditory analysis of group closure, then, must aim at a subjective impression - aim to hear, at the boundary of every group: does it sound as if there is more to come, or not - whichever the phonetic cues present (if any), and whether or not the topic, or subtopic, or utterance, or exchange of utterances in fact continues.

In my transcriptions the acute and grave accents are arbitrarily placed on the final syllable of the group. However, this should not be taken to mean that phonetic cues, if present, are necessarily executed on this final syllable only. They may stretch over the final contour as

a whole, or even be an indissoluble part of the pattern of the entire group - this problem can be taken up later, once it has been established which groups are open and which are closed.

In this way the auditory analysis of group closure adds another perceptual effect to those already discerned: the contrast between openness and closedness, continuity and finality.

* * *

2.3 HIGHER ORDER UNITS

The discussion of group closure led me to the assumption that there may be higher level 'moves in the speech act', units which combine a number of groups into a whole - a whole which the speaker feels to belong together, and which the listener also interprets as such.

As will become clear in the course of this chapter, it is an assumption in which I am not alone. Daneš, relating it only to pitch, says that intonation structures not only the pitch of 'rhythm-units', but also:

> ...the general level of pitch in a rhythm
> unit in relation to neighbouring units
> (and the) general trend of pitch in whole
> utterances or longish parts of it...
> (Daneš, 1960, p. 45)

Schubiger (1958) notes that long sentences consist of several tone-groups, and that one of these is 'nuclear', so that the whole construction behaves as a higher-level tone-

group. Crystal expects that:

> ...it is likely that further research will
> demonstrate the existence of tone-group
> sequences (tonal paragraphs, let us say)
> which would require at least one intermediate
> stage between utterance and tone-group...
> (Crystal, 1969, p. 202)

Yet, intonational structure above the level of
the group has received comparatively little attention, and
I know of only one experimental study specifically devoted
to it, an investigation of the intonation of paragraphs by
Lehiste (1975).

In this section of the chapter, then, I attempt to
develop, against the background of what literature there is
on the subject, an approach to the auditory analysis of the
junctures which permits the perception of these higher level
units, discussing, in turn, the question of junctural cues,
and the function of the higher order units.

2.3.1 The phonetic cues of higher order juncture

The most frequently mentioned cue for the percep-
tion of higher order juncture is the speech pause. As
Bolinger says:

> ...I am sure that speakers do use pause
> in this way, fitting it to the needs of
> the moment, making a section pause longer
> than a paragraph pause, a paragraph pause
> longer than a sentence pause, and, within
> the sentence, adjusting the pauses in
> similar fashion...
> (Bolinger, 1961b, pp. 24-25)

Other writers also note degrees of pause-length, and
experimental evidence seems to confirm their observations.[65]

The 'closed' group ending forms, as we have seen, another cue for the perception of higher order juncture. Described in terms of falling pitch, rather than as a perceptual effect of 'closedness' it is mentioned as a higher order juncture cue by several writers.[66]

Just as there are, according to Bolinger, degrees of pause, so there may also be degrees of 'closedness', falling pitch configurations of different magnitude, taking the form of a gradual lowering of pitch level (and loudness) towards the end of the unit,[67] or a 'depressing of the pitch span',[68] or the incidence of 'extra low, extra wide-falling contours' at the end of groups.[69]

> ...In a series of sentences, each of which
> ends in a low pitch, one usually detects
> an overall lowering at the end, signifying
> the closing of a particular topic of
> discourse...
>   (Bolinger, 1970, p. 138)

The 'new start' following the higher order juncture forms another cue. According to Brown et al. (1980, p. 24) it is marked by a high pitch and an increased loudness - and Lehiste (1975) has demonstrated this experimentally: the onset of the same sentence is higher in pitch when the sentence forms the beginning of a 4-sentence paragraph than when it is spoken in isolation.

But the list of possible cues should not be limited to the most frequent and obvious ones. 'Polysyllabic effects' such as those we discussed in 2.2.2(p. 54) above may stretch over the whole of a higher order unit, to contrast it to adjacent units. Crystal (1975, p. 120) gives an example in which a polysyllabic effect is superimposed

only on the final unit of the higher order unit, but never-
theless over the whole of it:

/ ships// towers// domes// theatres// "and/ temples lie//"

The section between quotation marks displays the feature
'crescendo', while the double underlined final syllable
is 'drawled'.

Nor should the durational structure of the higher
order units be neglected. If the paragraph in spontaneous
speech is indeed something like a 'topic of discourse', it
is likely to display the characteristic changes in rate of
utterance which Henderson et al. (1967) have described.[70]

Finally, kinesic cues, though of course irrelevant
to the analysis of radio announcing speech, play an
increasingly important role at these higher levels. Scheflen
(1964) describes them as follows:

> ...when an American speaker uses a series
> of syntactic sentences in a conversation,
> he changes the position of his head and
> eyes every few sentences (...) these shifts
> (...) mark the end of a structural unit
> (...) called the point because it corresponds
> crudely to a point in a discussion...
> (Scheflen, 1964, p. 231)

The next higher level he calls a 'position'. It is a
sequence of 'points' which

> ...corresponds roughly to a point of view
> that an interactant may take in an inter-
> action (and is) marked by a gross postural
> shift involving at least half the body...
> (Scheflen, 1964, p. 232)

His 'presentation', finally, is the largest unit, comprising
the

> ...totality of one person's positions
> in an interaction (and terminated by)
> a complete change in location...
> (Scheflen, 1964, p. 234)

There is, again, a proliferation of possible cues.
As Uldall has said:

> ...One of the great difficulties in dealing
> with real intonations set in real contexts
> (...) is that the same kind of information
> is carried by several systems all present
> at all times: pitch, voice quality, tempo,
> gesture, facial expression, any one of
> which may be dominant at a given moment...
> (Uldall, 1962, p. 783)

In the case of higher order junctures, then, as
in the case of contour- and group-junctures, specific cues
should not be used as analytical criteria, not even the
speech pause: radio announcers often move from, e.g., the
back-announcement of a song to a live commercial without
pausing, thus treating two distinct programme segments as if
they were of the same order (in part, at any rate, for in
other respects there may be a noticeable shift in speech
style).

In marking higher order juncture, I cannot, of
course, stop myself from noticing that the boundaries I
establish coincide, at times, with the boundaries of sen-
tences, or subtopics or topics or higher order 'moves in
the speech act' of a different kind - yet, I should not
hesitate to mark intonational boundaries which I cannot
immediately identify as being also the endpoints of syntactic
or semantic units. I will probably also detect, at the
boundaries, quite a few phonetic cues of the kind we have
just discussed - the higher the level, the more easily they

are noticed. Yet, even when I cannot, at first, discover
any, I should confidently mark the boundary - on one condition:
that it strikes me as somehow more definite, more pronounced,
more final than the preceding group-boundaries.

2.3.2   The function of higher order juncture

The function of higher order intonational units is
sometimes seen as stylistic. Sentences can either be com-
posed of several groups, forming a higher order intonational
unit, or of just one group - depending on the formality of
the style.[71] In Crystal and Davy (1969) and Crystal (1969)
this principle is extended to units on the level of the
paragraph: in conversational speech, they say, levels above
the clause are very rare, but in speech styles like radio
commentary and Bible- or poetry reading intonation:

> ...connects separate items of information
> into quite lengthy, coherent sequences,
> terminated by an extra-low or extra wide
> falling contour...
>   (Crystal and Davy, 1969, p. 131)

While some see higher order juncture as determined
by higher order grammar (as marking, for example, 'parenthetic
utterances' and subordinated clauses),[72] others believe
it to be more independent of grammar and serve to mark
'major' and 'minor information points',[73] or to divide
longer speeches into subtopics and topics, and to regulate
turn-taking in an exchange of utterances.[74]

For the purpose of my analysis I would like to
define the higher order units as perceptual effects -
perceptual effects facilitating the act of speech decoding,

but (what can be decoded must also have been encoded) participating also in the construction of the message, not according to rules which relate higher order grammar to higher order intonation, but according to the speaker's intentions (whether or not he has subjugated them to the demands of a social institution), and not necessarily only by enclosing 'topics of information', but possibly also by demarcating other kinds of higher level 'moves in the speech act'.

I will discern three orders: <u>sequence-level juncture is what makes the listener perceive that one sequence of groups has come to its conclusion and another is about to begin. Paragraph-level juncture is what makes the listener perceive that one group of sequences has come to its conclusion and another is about to begin. Section-level juncture is what makes the listener perceive that one group of paragraphs has come to its conclusion and another is about to begin.</u>

The final level, then, is that of the utterance itself - and not every utterance necessarily displays all the levels just defined. Indeed, an utterance may consist of only one syllable - just as groups, sequences, paragraphs and sections - in ascending order of improbability, perhaps - may consist of only one syllable. The boundaries, in such cases, are likely to be 'upgraded', to display the characteristics of the higher order to which they belong, and our method of analyzing juncture step by step will guarantee that what was first marked as a group boundary will later be assigned to the relevant higher order.

Higher order boundaries, as we have argued before, facilitate different stages in the act of speech decoding. At the sequence boundary the listener selects from among the groups in the preceding sequence those that have most attracted his attention, those that are ranked most highly (see the discussion of 'ranking' in 2.4), and he integrates them into a unified whole which can be stored for further processing at the paragraph boundary, if there is one. This process, if still higher levels exist, repeats itself until, at the utterance-final boundary, we are left with a final impression, woolly and indefinable, but nevertheless sub-jectively distinct - and it depends, perhaps, on the amount of attention with which we have listened (as well as on other pragmatic factors) how far we will be able to trace the path back, how much we will be able to recall consciously.

This does not mean, of course, that the listener is at the speaker's mercy, that what he will be able to recall necessarily corresponds with what the speaker has fore-grounded - our attention is attracted, not only by what the speaker has encoded as important, but also by what we already find important. We may remember, next to the general sense of the message and the elements from it which the speaker has intonationally foregrounded, also other, apparently disconnected bits and pieces which are important in another system, a discourse already present in our mind, rather than in the system of the message. But even when the listener disagrees with what the speaker has emphasized, or isolated, or connected, he will still register how the speaker intended him to interpret the message. He will

still understand the speaker, not only on his own terms, but also on the speaker's terms. If communication between speakers of the same language (between a radio announcer and people who have grown up with the medium of radio, for example) is 'ineffective', it is likely to be so, not because the speaker is not understood, but because the listener does not agree with him, and hence fails to fall in line, fails to respond in the way the speaker intended - a distinction which mass communication theory, concerned as it is, at times, with the 'response' to messages, and with improving the 'effectiveness' of messages, too often neglects to make.

2.3.3 The auditory analysis of higher order juncture

In analyzing higher order junctures, then, the analyst should scan, one by one, the already established group junctures, until he finds one which strikes him as somehow more prominent, more pronounced than the preceding boundaries, and this boundary he should mark as a sequence juncture.

Again, in marking paragraph boundaries, the sequence boundaries thus established should be scanned, one by one, until one is observed which obtrudes itself as more definite, more final than the preceding sequence boundaries - whichever the phonetic cues present, and whether or not the boundary concludes, by some other, non-perceptual criterion, a 'topic of discourse' or a 'point of view', or another kind of higher order 'move in the speech act'.

The same procedure is, if necessary, followed once more for the section-boundaries, by scanning, one by one,

the paragraph boundaries.

To mark the boundaries, groups are enclosed in square brackets, sequences in double square brackets, paragraphs in triple square brackets, and sections in quadruple square brackets, while the whole utterance is also enclosed in square brackets - see the example in 2.5.1 below.

I have not assessed the reliability of this part of my method by testing it out on the group of phonology students and their lecturer: the two segments of speech I used in the tests consisted of only one sequence each, and hence contained only group-level and utterance-level boundaries.

2.4 RANKING

Whether expressed in terms of 'levels of stress', or in terms of the distinction between 'nuclear' and 'non-nuclear' ('tonic' and 'non-tonic') accents, 'ranking' plays a role in all forms of intonation analysis: accents do not all attract the listener's attention to the same degree, they differ in amount of subjective prominence.

But when it comes to details, a variety of approaches exists. How many levels should be distinguished, how they are realized phonetically, how the listener perceives them, how levels should be assigned to accents in an auditory analysis - to all these questions more than one answer exists.

In this section of the chapter I will discuss these answers, and compare them to my own observations, in the hope

of developing, in the process, my own answers, and my approach
to the auditory analysis of ranking.

2.4.1 Levels of prominence

Some writers distinguish only two levels: the
normal (rhythmic) 'stress' of the 'secondary' ('non-nuclear')
accents, and the 'stress', augmented by pitch movement, of
the 'primary' or 'nuclear' accent. Ranking, for them, is a
matter of selecting the nucleus and making it the most
prominent syllable of the group, the 'intonation centre'.[75]

Other distinguish more levels. The syllables of
Halliday's (1967; 1970) tone-groups, for example, may be
assigned any one of 4 degrees of prominence: 'salient, tonic';
'salient, non-tonic'; 'weak, non-reduced'; 'weak, reduced'.
Crystal (1969) discerns not only 9 types of primary accent
(which he does not relate to degrees of prominence as such,
though one would expect nuclei with complex pitch movement
to cause greater foregrounding than nuclei with simple
pitch movement), but also 7 different non-nuclear accents
(defined in terms of the listener's expectation of pitch:
'as expected'; 'lower than expected'; 'very much lower
than expected'; 'continuing at the same level'; 'slightly
higher than expected'; 'much higher than expected'; 'very
much higher than expected'), as well as three degrees of
stress ('normal'; 'strong'; 'extra strong').

With so many fine distinctions, so many components
of prominence, so many phonetic cues, it becomes difficult
to assess the overall prominence of a syllable, its global
perceptual effect, and this causes theories like these to be

somewhat removed from the realities of listening to speech. For, in reality:

> ...the listener is never concerned exclusively with a single one of them (i.e. the phonetic cues) (and makes) judgments which are determined by their interactions...
>
> (Fry, 1958, p. 127)

One wonders, for example, whether, in a transcription according to Halliday, a weak, non-reduced syllable at a high pitch point in the contour of the tonic would be more or less prominent, perceptually, than a salient, non-tonic syllable at a low pitch point. Or whether, in a transcription according to Crystal, a narrowly falling, very strongly stressed, very much higher than expected non-nuclear syllable would be more or less subjectively prominent than, say, an extra wide rising, normally stressed, slightly higher than expected non-nuclear syllable. Such questions show that transcriptions of this kind are not only difficult to interpret (and somewhat removed from the normal perceptual effect of intonation), but also difficult to use for the auditory analysis of intonation. In the case of the Trager and Smith method (1951; 1964), a method which distinguishes 4 levels of pitch and 4 degrees of stress, the latter problem has been well documented: Lieberman (1965) found the Trager-Smith transcriptions of two linguists to differ by 60%, while Hadding-Koch (1961) found that only primary stress was perceived with more than random consistency. Bolinger, as always, is persuasive when he says that:

> ...the relative height for relative importance among the accents themselves (is) a characteristic of English that is still to be explored...
>
> (Bolinger, 1964, p. 26)

'Relative height for relative importance': perhaps we should abandon the idea of a fixed number of levels, perhaps we should do no more than comparing the accents within (and only within) each group for their relative subjective prominence, and simply create, in each case, as many levels as there are accents in the group. That such subjective levels of prominence exist is made plausible, for example, in a paper by 't Hart and van Katwijk (1969), who describe an experiment in which listeners were asked to mark stressed syllables without consideration of level. The 'stimuli' had been prepared to represent levels of stress as discriminable steps in all three of the main parameters of accent - pitch, duration, and intensity. As it turned out, these steps corresponded to the degree of agreement in the subjects' marking of the 'stress', from which the authors concluded that 'levels of conspicuity' do indeed exist in the perception of 'stress'.

Can agreement higher than that of Lieberman or Hadding-Koch be reached if the levels are not made absolute, if they are tied, neither to the listener's expectation, nor to specific cues, if the accents within each group are simply compared for relative subjective prominence, and if as many levels are created as there are accents in the group? Testing this with the group of undergraduate linguistics students and their lecturer, I used these instructions:

Determine the relative prominence of the accents in each group separately. Mark the most prominent syllable with the number 1, the next most prominent syllable with the number 2, etc., as follows:
[it is/ ^{2}ten/ ^{3}past/ ^{1}three//]

Agreement between the lecturer and myself was 83.3% for both of the lecturer's transcriptions. Agreement with the students' transcriptions, however, was random - 52.6% on average - reliability can be established at an 0.53 average. It should be noted, however, that agreement is higher if only the primary accents are taken into account: 85.7% for the lecturer, 67.1% for the students.[76]

The conclusion must, I think, be this: though even the marking of primary accents does not lead to particularly convincing agreement, it is, at least, not entirely random. For other, lower levels, however, a global perceptual effect does not appear to exist: only the specialized ear can detect the differences. I have made the system too complex, and:

> ...a complex symbol system, burdensome as
> it is to the user, may reflect either subtle
> perception or confusion about the function
> of prosody...
>   (Harris, et al., 1981, p. 3)

Perhaps the disjunction is incorrect: over-subtle, over-phonetic perception itself reflects confusion about the function of prosody. If listeners who can perceive accents and junctures with reasonable accuracy cannot be made to hear the difference between these accentual levels, then it must be doubtful whether they have, as distinct levels, a very important role to play in speech communication. Going by my results, I should, in contrast to Bolinger's suggestion, select, in each group, just one accent as more prominent than the others, as the intonational core of the group - just as is done in forms of intonation analysis which assign

one and only one nuclear accent to each group.

2.4.2 The function of ranking

Just as accent selects, in each contour, one
syllable as more worthy of attention, more important to
remember, and more vital for the understanding of the message,
so group-level ranking selects, in each group, one accent
(one might also say: one contour) as more important than
the others. That one accent then becomes a core to which
the preliminary assessment of the meaning of the entire
group attaches itself, a token for the group as a whole,
which, by the time it is stored, has already changed from
signifier into signified, become a mental representation of
meaning as well as mental representation of sound.

Ranking, accordingly, can be defined as a way of
attracting the listener's attention to one element in a unit
at the expense of the others. The definition can be used
at different levels, as a re-formulation of the definition
of accent, for example: accent is a way of attracting the
listener's attention to one syllable in a contour at the
expense of the others. Or as a definition of group-level
ranking: group-level ranking is a way of attracting the
listener's attention to one contour from among the contours
in a group at the expense of the others. Or as a definition
of sequence-level ranking: sequence-level ranking is a way
of attracting the listener's attention to one group from
among the groups in a sequence at the expense of the others —
but higher order ranking will be discussed in more detail
below, in 2.4.4.

The speaker does not select this primary accent because it is an obligatory element of the formal structure of intonation groups, and he is not compelled to place it on the final lexically stressed syllable, as some writers say.[77] Rather, the primary accent is also the 'semantic peak', the main 'information point' in the group,[78] not in terms of what is 'given' or 'new' (in the sense of what he or his interlocutor(s) have mentioned earlier), but in terms of what he considers most important for his message -for whichever reason: emotive, informational, or otherwise. He has the freedom to choose. If he allows his intonation to become 'fossilized' (Bolinger, 1958b), or stereotyped, he has only himself to blame, not his native language. In this way, ranking, like accent, can be a way of 'reading the speaker's mind'. The placement of the primary accents in this example, for instance, could be said to foreground what is already evident from the information the announcer chooses to give: that he attaches more importance to the box-office success of a film, or to the sales of a record, than to the film or the music itself:

[it's made a/ ^{1}lot of/ money in A/ ^{2}mericá//] [and it's a/ really/ ^{1}big/ movie in/ ^{2}sydney//]

If this would turn out to be a common habit of, for example, commercial radio announcers, it might be said to betray a priority of commercial radio as an institution: the commitment to an exclusively quantitative and profit-oriented evaluation of the music it plays, of the information it provides, of the ideas it discusses - and, ultimately, of

its own role as a communicator.

2.4.3 The auditory analysis of ranking

The auditory analysis of group-level ranking differs
from the analysis of accent in one respect: in contour-
level ranking, rhythm enables us to anticipate the accents,
so that actual comparison of syllable prominence becomes
almost redundant. In group-level ranking, we must, retro-
spectively, compare up to 7 accents for their relative
subjective prominence, selecting from among them the one
which most strongly obtrudes itself - and this is, as we
have seen, not as reliable a procedure as the marking of
accent. Yet, it is the only method I can find to rank the
accents: the instructions given in 2.4.1 must stand, with
the one proviso added that only the most prominent accent
(the 'primary accent') and the second most prominent accent
(the 'secondary accent') should be marked, and the others
left unspecified. I decided to also mark secondary accents
in order to make it possible for example, to investigate
whether what Crystal calls 'bi-nuclear tone-units' exist,
and whether there might be a pattern in the placement of the
secondary accents, but I will have to keep in mind that my
marking of these secondary accents may not be very accurate
and that they do not necessarily function independently of
the other accents.

2.4.4 Higher order ranking

Just as it is possible to rank the contours of a

group according to their subjective prominence, so it is
also possible to rank the groups of a sequence - and the
literature indicates, in different ways, that I am not the
only one to have come to this conclusion. To draw attention
to the difference in pitch-range between 'subordinate' and
'superordinate' tone-groups, for example, or between
'parenthetic groups' and the groups in which they are embedded,
is to recognize, de facto, if not de jure, a form of
sequence-level ranking.[79]   The same can be said for
Halliday's (1967; 1970) 'tonal concord' (in sequences of two
tone-groups one will contain the 'major information point'
and execute its tone over a wider pitch-range than the other,
which will contain the 'minor information point').[80]   Or
for Crystal's (1969) view that 'subordination' can extend
over more than 2 tone-units - his 'theory of subordination'
extends, in effect, the principle of group-level nuclear
tone placement to the level of the sequence: only one of
the units in a sequence can become 'superordinate', just as
only one of the accents in a tone-unit can become nuclear.
A similar notion can be found in Schubiger (1958).

The prominence of 'superordinate groups' is usually
ascribed to an overall widening of the pitch-span of the
group, although Crystal sees it as a widening only of the
pitch movement of the nucleus of the superordinate group.
I decided to follow him in this, because I found the pro-
minence of primary accents easier to compare than the
prominence of whole groups. The reliability test was based
on the assumption that, here too, it would be possible to
discern as many levels of relative subjective prominence as

there are groups in the sequence. These were the instructions:

> Enclose the groups in square brackets and determine the relative prominence of their most prominent syllables, that is, compare the prominence of all the syllables marked with the number 1. This can, once again, be indicated by means of numbers, 1 being the most prominent syllable, 2 the next most prominent, etc., as follows:

$^1[^1$strange$//]$ $^2[$per$/^2$haps it$/^1$is$//]$

And I proposed that the sequences within paragraphs and the paragraphs within sections could be ranked along the same lines. However, the speech segments used for the test consisted of no more than one sequence each, so that only sequence-level reliability could be tested.

While there was still some regularity in the selection of primary groups (100% agreement with the lecturer, in both his transcriptions; 61.1% agreement with the students), sequence-level ranking did not, in general, lead to more than random agreement. The lecturer marked 6 out of 9 groups in accordance with mine in his first transcription, 4 out of 9 in his second (and these 4 were the ones I had marked as primary and secondary groups). Agreement between my ranking and that of the students averaged 48.1% (average reliability 0.48).[82]

The results of this test resemble those of the group-level ranking test: only primary groups are perceived with some minor degree of regularity. At lower levels a perceptual effect is apparently absent. Accordingly, I will rank higher order units in the same way as I have ranked the

accents within groups: by marking only the primary and
secondary groups with the relative index numbers, and leaving
the other groups unspecified.

The function of higher order ranking resembles
that of group-level ranking: higher order ranking makes it
possible for the essential meaning of a sequence to become,
retrospectively, embodied and concentrated in its primary
group, after which that group can be stored in memory as that
element of the sequence which is most vital for the under-
standing of the message, and for further processing, at the
next higher level, if it exists.

Here too, the speaker has 'freedom of intonation':
in selecting a primary group, he can achieve 'subordinations'
quite different from those recognized in formal grammar. He
is free to foreground that part of the sequence which he
considers most important, however deeply embedded it is in
the grammatical structure, and whichever the nature of this
importance, emotive, informational, or otherwise.

Because of this, higher order ranking, too, can
be a way of 'reading the speaker's mind'. In the example
below a 2CH announcer not only pronounces the adjective
'marvellous' as a separate group, but also gives that group
the highest ranking in the sequence. If this would turn out
to be a habit common to all 2CH announcers rather than an
isolated example, we could, perhaps, interpret it as an
indication that at this station, with its computer programming
of the playlist, and its preference for watered down versions
of the pops as well as the classics, the overall 'sound of
the station' rates as more important than the individuality

of the music and its performers:

^{1}[[That/^{1}time we/^{2}heard fróm//] ^{2}[^{2}Frank/ Pour/

^{1}cél//] [^{1}and éh//] ^{1}[a/^{1}marvellous éh//] [in/

^{2}terpre/^{1}tatión//] [of/ that/^{2}great/^{1}Beatles/

túne//] [^{2}Hey/^{1}Jùde//]]

* * *

2.5 SUMMARY

The method of intonation analysis proposed in this
chapter is auditory rather than instrumental, although the
use of instrumental analysis has not been excluded altogether:
it has its place as an aid to the precise expression and
the detailed study of the phonetic manifestation of the
functional units established auditorily.

It is also proposed that an auditory analysis should
attempt to arrest the fleeting distinctions which normal
users of the language make during speech perception, without
realizing that they are making them, below the surface of
conscious discrimination. Auditory analysis should capture
'global perceptual effects', rather than constitute a detailed
phonetic analysis.

The perceptual effects I propose to use for this
purpose are: isochrony, juncture, closure, and ranking.
They function between and within groups of sounds of differ-
ent magnitude: syllables, contours, groups, sequences,
paragraphs, sections, and, of course, the utterance as a whole.

Juncture divides utterances into groups, and, depending on the complexity of the utterance, into further, higher order units: sequences, paragraphs and sections; perceptual isochrony divides groups into contours; closure determines whether groups are 'open' or 'closed'; ranking selects accents from among the syllables in a contour, primary accents from among the accents in a group, primary groups from among the groups in a sequence, primary sequences from among the sequences in an intonational paragraph, etc., depending on the amount of levels in the utterance.

Auditory analyses of this kind specify how (specific) intonations work, rather than how they sound, distinguish the functional elements and units within them. At the same time, my approach to auditory analysis provides the outline of a theory of intonation which looks at intonational function both from the listener's and from the speaker's point of view.

From the listener's point of view, intonation facilitates the 'efficient processing' of speech: juncture, retrospectively, creates units which enable speech process-ing to take place at different, hierarchically ordered levels, so that, at each level, only the most important element of the preceding unit needs to be retained for further process-ing, at the next higher level - an element to which by then the sense of the whole unit has attached itself. Ranking selects this element from among the elements of the unit. From the speaker's point of view, juncture allows the creation of distinct 'moves in the speech act' of different orders of magnitude (hence also the joining of what is

considered to belong together), while ranking allows the
selection of one element in each 'move' as the most impor-
tant, for whichever reason.

In this way it becomes possible, both for the
listener (although he may not (always) consciously do so)
and for the analyst, to reconstruct the speaker's priorities,
to draw his conclusions from what the speaker considers to
belong together, what to lack connection, from what he treats
as important and what as less important, so that the analysis
can fulfil the criterion formulated in the introduction:
that an analysis of intonational form should lead to an
analysis of intonational function, and that an analysis of
intonational function, in turn, should lead directly (on
the basis of the functional definitions themselves) to
interpretations which can relate the speech to its wider
social and historical context.

2.5.1 · Transcription

The transcriptions, accordingly, do not represent
how intonations sound (whether boundaries are accompanied
by pauses or not, whether 'closed' group endings display a
falling pitch or not, etc.), but how they function. A summary
of the symbols used is given in the table below:

‾‾‾‾‾‾‾	= rhythmic accent
--------	= ancillary accent
^{1}‾‾‾‾‾	= primary accent
^{2}‾‾‾‾‾	= secondary accent
/	= contour boundary
//	= group boundary
´	= open group
`	= closed group
[ ]	= group
^{1}[ ]	= primary group
[[ ]]	= sequence, or utterance consisting of one group only
^{1}[[ ]]	= primary sequence
^{2}[[ ]]	= secondary sequence
[[[ ]]]	= paragraph, or utterance consisting of one sequence only
^{1}[[[ ]]]	= primary paragraph
^{2}[[[ ]]]	= secondary paragraph
[[[[ ]]]]	= section, or utterance consisting of one paragraph only
^{1}[[[[ ]]]]	= primary section
^{2}[[[[ ]]]]	= secondary section
[[[[[ ]]]]]	= utterance consisting of several paragraphs, or of one section only
[[[[[[ ]]]]]]	= utterance consisting of several sections

A full transcription is given below - of a popular music announcement by an ABC announcer. To make recognition of the sequences and paragraphs easier, they are also set apart by the layout: each sequence begins on a new line, and extra spacing separates the two paragraphs.

^{2}[[^{1}[^{1}Fascinating. to/^{2}watch the éh//] ^{2}[ca/ reer of/^{2}Linda/ ^{1}Ronstàdt//] [who/^{2}started/^{1}óff//]   [you/^{2}might re/^{1}càll//] [with. the/^{2}Stoned/^{1}Ponìes//]]

^{2}[a/^{1}bóut//]   [^{1}óh//] ^{1}[^{2}ten/^{1}years a/gò//]]

^{1}[[^{1}then éh//] [on the/^{1}record/^{2}labéls//] [it be/ came the/ Stoned/^{2}Ponies with/ Linda/^{1}Ronstádt//] [^{2}Linda/ Ronstadt with the/^{2}Stoned/^{1}Poniés//] [^{1}ánd//] ^{1}[have you/^{1}heard of the/ Stoned/ Ponies/^{2}recentlý//] [^{2}I/ bet/^{1}nòt//]]

^{2}[[^{1}théy've//] ^{2}[faded/^{1}out of the/^{2}scéne//] ^{1}[^{1}alto/^{2}gethèr//]] [^{1}[but/^{2}Linda/^{1}Ronstádt//] ^{2}[she's/^{2}right on/^{1}tòp//]] [[go into a/^{2}record/ store these/^{2}dáys//] ['n you're con/^{1}fronted bý//] ^{1}[in/^{2}numerable/^{1}albúms//] [with/^{1}her éh//] [^{1}smilíng//] [^{1}count'nánce//] ^{2}[^{1}on the/^{2}covèr//]]

^{1}[^{2}[^{2}[^{1}Linda/ Ronstadt/ off the/^{2}albúm//] ^{1}[^{2}Poor/ Poor/ ^{1}Pitiful/ Mè//]] ^{1}[^{2}[and her/^{1}current/^{2}sóng//] ^{1}[^{1}Blue/^{2}Bayòu//]]]]

* * *

NOTES

(1) Thus Newman (1946) sees stress as "...not a constant, but merely a potential feature of the word..." (p. 175); Lieberman (1967) defines it as an 'abstract entity' and contrasts it with 'prominence'; for Schmerlinger (1976) stress rules apply only in 'unanalysable situations' (p. 55); Jassem and Gibbon (1981) call stress an 'abstract lexical category' and accent an 'observable, textual category'. For Armstrong and Ward (1926), MacCarthy (1944), Trim (1964), Bolinger (1964) the semantic or emotive importance of a word decides whether stress will be realized in the sentence or not.

(2) For example in Bolinger (1965), in connection with the rhythmically motivated 'backshifting' of stress (as in 'the cómplex problem' vs. 'the próblem is compléx'). Brown et al. (1980) also recognize the possibility, ascribing it to 'the vagaries of performance'. In metrical theory the same possibility is more often acknowledged, e.g. in Chatman (1965), who gives lines of Milton and Coleridge as examples ('To bottomléss perdition there to dwell', and 'I see, not feel, how beautifúl they are'). See also Léon (1971).

(3) E.g. Ward (1939), Pike (1946), Trager and Smith (1951), Kingdon (1958), Jones (1962), Schubiger (1958). Halliday (1967; 1970) and Crystal (1969; 1975) only insofar as the 'non-nuclear' accents are concerned.

(4) E.g. Armstrong and Ward (1926), Ward (1939), Schubiger (1958), Jones (1962).

(5) E.g. Jones (1962) - others only where the 'nuclear accent' is concerned.

(6) E.g. Kingdon (1958), Jones (1962).

(7) E.g. Kingdon (1958), Jones (1962) - others only where the 'nuclear accent' is concerned.

(8) Sentence stress sometimes falls on the leftmost, sometimes on the rightmost lexical stress of a constituent (depending on the type of constituent). The rules are applied cyclically, beginning with the smallest constituents, and assigning 1-level stress to them. When the next higher level constituents are subjected to the rule, the leftmost (or rightmost, as the case may be) constituent receives 1-level stress, while the stress assigned by the previous application of the

rule is demoted by one degree, and so on, until the
level of the sentence is reached and the nuclear
accent determined. If follows that there is, in
principle, no limitation to the number of stress-levels
which can be created - although Chomsky, Halle and
Lukoff (1956) and Bierwisch (1968) have argued in
favour of limiting this number to 5 or 6 degrees at
the most.
Among the generative phonologists, Kiparsky (1965)
allows rhythm to be, at times, a determiner of accent,
while Bierwisch (1965) allows for exceptions on the
basis of 'contrastiveness', the 'given/new' distinction,
and 'emphasis'. A discussion around certain specific
exceptions not acknowledged in Chomsky and Hall (1968)
led to further studies by Bresnan (1971), Berman and
Szamosi (1972) and Lakoff (1972), as well as to an
interesting and challenging response to these studies
by Bolinger (1972).

(9) Bierwisch (1965).

(10) Although others claim that the structure of the
     intonational group itself determines which accent
     becomes primary: the last lexically stressed syllable
     in the group. Exceptions acknowledged include expres-
     siveness (Newman, 1946), linguistic inventiveness
     (Bolinger, 1972; Crystal, 1975), contrastive intona-
     tions (Newman, 1946; Trager and Smith, 1951; Schubiger,
     1958; Crystal, 1975), stylistic habits (Schubiger,
     1958; Bernard and Delbridge, 1979), novelty, i.e.
     the 'givenness' or 'newness' of the information contained
     in the word to be accented (Schubiger, 1958; Hultzen,
     1959 and 1964; Halliday, 1967 and 1970), emphasis
     (Ward, 1939; Kingdon, 1958), specific syntactic
     structures (Crystal, 1975). The rule may also be
     called a 'statistical tendency' (Cruttenden, 1970;
     Bolinger, 1972; Crystal, 1975).
     In my view the exceptions form the rule: the primary
     accent or nucleus always expresses, contrasts, informs,
     emphasizes - except when there is nothing to express,
     contrast, inform people about, or emphasize. This
     view can also be found in Daneš (1960), Trim (1964),
     Malmberg (1968), Bolinger (1972), Schmerlinger (1976).

(11) Cf. Studdert-Kennedy and Hadding-Koch (1973). Also
     Léon and Martin (1969): "...the listener mentally
     inserts an accent he expects, even when no acoustic
     prominence asserts itself..." (p. 19, my tr.).

(12) Arguments in favour of auditory analysis over instru-
     mental analysis can be found also in Schubiger (1958),
     Daneš (1960), Fauré (1962), Delattre et al. (1965),
     Crystal (1969), Léon and Martin (1969).

(13) Brown et al. (1980) give a good account of this difference (pp. 48-49).

(14) Cf. Fry (1955; 1958), Lieberman (1960; 1967), Bolinger (1958a), van Katwijk (1964), Isačenko and Schädlich (1965).

(15) Cf. Fry (1958); Lehiste and Peterson (1959).

(16) Cf. references in (14) above.

(17) Cf. Black (1961), McClean and Tiffany (1973), Brown et al. (1980).

(18) Cf. Fry (1958), Denes (1959), Lehiste (1970).

(19) E.g. in Wells (1945), Pike (1946), Trager and Smith (1951).

(20) Schubiger (1958), Kingdon (1958) and various other members of the British school of intonation analysis.

(21) E.g. in Bolinger (1958b; 1970) and Bailey (1971b).

(22) E.g. in Lehiste and Peterson (1959).

(23) E.g. in Armstrong and Ward (1926), Lehiste (1970).

(24) E.g. in Meyer-Eppler (1957), Kloster-Jensen (1958), Denes (1959) and Hadding-Koch (1961).

(25) Armstrong and Ward (1926), Jones (1962) and Crystal (1969) also draw attention to this.

(26) In the perception of in all respects identical and identically spaced clicks, people hear the clicks in groups of 2 to 6, with the first click of each group perceptually stronger than those which follow it (cf. Allen, 1975).

(27) Martinet's definition (1955, p. 12) is close to mine, except in that he specifies the phonetic cues. Bernard and Delbridge's definition (1979, p. 94) is closer still, except in that 'prominence' is defined as the ability of a sound to attract attention, rather than the ability of a speaker to attract the listener's attention to a sound.

(28) MacCarthy (1944), Schubiger (1958), Kingdon (1958), Bolinger (1961a; 1964; 1972), Trim (1964), Cruttenden

(1970), Ladefoged (1972), Crystal (1975).

(29) For example by Bally (1941), Bolinger (1958b and elsewhere), Schubiger (1958; 1964), Hultzen (1959; 1964), Daneš (1960), Halliday (1967; 1970), Léon and Martin (1969).

(30) I agree with Schmerlinger (1976) who underlines that the relevant distinction is a pragmatic one: what matters is not whether the information is given or new in some objective 'intra-textual' sense, but whether or not it is taken for granted by the speaker, for whichever reason. Halliday and Hasan (1976) make the same point: what counts is what the speaker treats as non-recoverable by the hearer.

(31) Cf. e.g. Newman's (1946) 'expressive' and 'rhetorical' accents, or the remark by Schubiger (1958) that increase in pitch range adds the emotive factor. 'Marked' intonation plays a key-role in Halliday (1967; 1970) and in Adams' (1969) analysis of Australian English intonation.

(32) But not vague in practice. Not vaguer, for example, than Jones' (1962) most excellent advice to listen for the intonation 'aimed at by the speaker' - I have always found it necessary to mentally repeat intonations after listening to them in order to be able to determine accents, boundaries, etcetera.

(33) Like Harris et al. (1981) I needed subjects who could accomplish a fairly complex and lengthy task, and who would understand its purpose and treat it seriously. The students had completed a full year course in phonetics and were, at the time of the test, following lectures about intonation, as part of a course in phonology. They were, therefore, not linguistics specialists, but normal listeners, interested in and informed about their language. The test was taken during two 1-hour tutorials, each attended by 12 students, and taking place with an interval of 4 days. For additional interest the lecturer took the test during both these tutorials.

(34) Assuming that every syllable has an equal chance of being selected as either accented or unaccented, and given that there were 43 syllables in the speech sample, the probability of any transcription fully agreeing with mine by chance (P_e) is 0.0000000000006%. Reliability was calculated for each pair (i.e. for each student's agreement with my transcription, expressed as a decimal), according to the formula $\frac{P_o - P_e}{1 - P_e}$, and then averaged.

Reliability figures will be based on the transcriptions of the students as well as the transcriptions of the lecturer (26 transcriptions in all) throughout the chapter, except where otherwise indicated.
It is interesting that on one syllable there was only 54% agreement. This was a hesitation noise, pronounced with a strong articulatory gesture as well as occurring on the ictus of a rhythmic foot, and hence marked as accented in my transcription. Although, to me, such a syllable conveys a message ('Wait, I have not yet finished'), many students apparently considered it a meaningless noise, at least in their conscious appraisal of its importance, which shows that 'semantic importance' influenced their decisions. The lecturer, incidentally, marked the syllable as accented in both his transcriptions.

(35) For most members of the British school, intonation patterns appear to coincide with sentences (Armstrong and Ward, 1926; Ward (1939), Schubiger (1958), Jones (1962). Trager and Smith (1951), Hultzen (1964), Trim (1964) and Crystal (1975) explicitly claim that intonational and syntactic boundaries coincide, a claim which is implicit in much intonation research by phoneticians, e.g. Isačenko and Schädlich, (1965), Lehiste (1973), Klatt and Cooper (1975), Wingfield (1975), de Rooy (1975; 1976).

(36) A claim made by members of different schools: Armstrong and Ward (1926), Ward (1939), Bierwisch (1965), Crystal and Davy (1969), Bailey (1971a).

(37) Thus, Halliday (1970), 't Hart and Cohen (1967), Crystal (1969), Cruttenden (1970). Lieberman (1980) puts it a different way: "...intonation segments the sentence-like sequences of words that form the data base for the sentences of formal grammar..." (p. 188). In my view, grammars would have to change a great deal if they were to really depart from the facts of intonation - and they would, in the process, lose much of their ability to account for what we understand by 'grammaticality'.

(38) Cf. Pike (1946), Trim (1959; 1964), Crystal (1969), Bailey (1971b), Lehiste (1973), Klatt and Cooper (1975), de Rooy (1975; 1976), Nooteboom (1976), Streeter (1980).

(39) Cf. references in (38) above. Also Umeda et al. (1981).

(40) Cf. Trim (1964), Isačenko and Schädlich (1965), Studdert-Kennedy and Hadding (1973), Collier and 't Hart (1975), de Rooy (1975; 1976), Streeter (1980).

(41) Namely when "...a tone-unit with a rising tone and a tail of more than one syllable precedes a tone-unit with a relatively long pre-head..." (Crystal, 1969, p. 207).

(42) O'Connor and Arnold (1961), for example, distinguish between 'tune' and 'tone-group', the former being the physical manifestation of the latter, which, in turn, is the former's abstract, underlying form. For Bailey (1971a) 'patterns' realize 'phonological phrases'. For Cohen and 't Hart (1967) 'intonational blocks' are manifested by specific 'patterns'.

(43) Allen (1975).

(44) E.g. Crystal and Quirk (1964), Crystal (1969), Crystal and Davy (1969), Bernard and Delbridge (1979).

(45) Mitchell and Delbridge (1969, p. 56): "...it is not easy to separate the effects of rhythm and intonation. The two go together. A variation in stress or rhythm involves a variation in intonation and vice-versa..."

(46) Classe (1939, p. 101): "...perfect isochronism can only be realized when very definite conditions are fulfilled. These are (a) similarity of phonetic structure of the group, including number of syllables, (b) similarity of grammatical structure of the group, and similarity of connexion between the groups...". See also Bolinger (1965), O'Connor (1965), Shen and Peterson (1968), Allen (1968a; 1968b).

(47) Laver (1970, p. 69): "...the tone-group has a simple correspondence with units of rhythm..." Pike (1946, p. 34): "... rhythm units are sentences or parts of sentences spoken with a single rush of syllables with long or short pauses in between...". Trim (1959; 1964) speaks of 'tone-cum-rhythm groups' and treats rhythm as a junctural feature: "...it may be objected that rhythm groups are not to be identified with tone-groups (but) it has so far proved possible and useful to represent both closely related systems by a single notation..." (1964, p. 29).

(48) E.g. in Martin (1975), Allen (1975). Lehiste (1973) elicited listener judgments about 'longest' and 'shortest' contours and interpreted the random results as evidence for perceptual isochrony. See also Stone (1981) whose study of jaw movements reaches the conclusion that "...rhythm is the structure upon which the other aspects of the speech signal (...) are superimposed..." (p. 110).

(49) Fauré et al. (1980)   define pause in terms of rhythm,
     as "...any break in the rhythm of the utterance,
     including any type of segementation, with or without
     accompanying silence..." (p. 72).

(50) Lunney (1972) believes that, at the boundaries of
     groups, we "...'switch off' or direct attention away
     from the metrical underlay in our mind and allow it
     to resume when the speech begins again..."

(51) Cf. Fodor and Bever (1965), Garrett, Fodor and Bever
     (1966), Dittman and Llewellyn (1968), Abrams and
     Bever (1969), Huggins (1972), Lehiste (1973), Wingfield
     (1975), de Rooy (1975), and our brief discussion of
     some of these studies, in 2.2.1.

(52) Assuming that every word boundary has an equal chance
     of either being selected as a group boundary or not,
     and given that there were 29 such inter-word spaces
     (not counting the beginning of the first word of the
     utterances and the end of the last), the probability
     of a 100% agreement with my transcriptions occurring
     by chance (P_e) is 0.0000000009%. Reliability was
     again calculated for each pair and then averaged. See
     for further details note 34.
     Lack of agreement usually resulted from the students
     making less, rather than more boundaries than I did.
     Only 20% of the 'deviant' boundaries were 'extra'
     boundaries.
     One boundary was marked in accordance with my tran-
     scription by less than half of the students. The
     speaker said '...a picture of an individual...' with
     what I considered a clear juncture between of and an.
     That this particular boundary was perceived less than
     accurately shows, perhaps, the 'corrective' influence
     of syntax - on the students' conscious discrimination
     of the boundaries, at any rate.
     Three of the student transcriptions agreed 100% with
     my transcription.

(53) E.g. in Abercrombie (1964), Trim (1964), Chatman (1965),
     Bolinger (1965), Halliday (1967; 1970), Lehiste (1973),
     Malmberg (1968) calls it a 'syllabic group'. Leech
     (1969) a 'measure'.

(54) Cf. Lehiste (1960).   O'Connor and Tooley (1964) report
     that perception of these boundaries is not very
     accurate.

(55) Assuming that every inter-syllable space has an equal
     chance of being selected either as a contour-boundary
     or not, and given that there were 34 such spaces,
     the probability of any of the subjects' transcriptions
     agreeing 100% with mine by chance is 0.00000000002%.

Reliability was, again, calculated for each pair and then averaged - see details in note 34.

(56) Ondráčková (1962) reports, for Czech, that, in a corpus of 65000 rhythmic feet, 97% of the feet had 2, 3, or 4 syllables. Only one foot in her entire corpus had 9 syllables. Halliday (1967; 1970) quotes a length of up to 7 syllables for his rhythmic feet. Nooteboom and Cohen (1975) give the duration of a foot as 1-2 seconds, citing support from other studies. Fonagy and Magdics see intonational phrases as between 1 and 10 syllables long. Halliday's tone-group consists of up to 10 rhythmic feet. Crystal's tone-unit ranges from 1 to 7 words, but averages 5 words. Lehiste (1970) discusses Kozhevnikov and Chistovich (1965) equating their 'syntagma' of up to 7 syllables with Trager and Smith's 'phonetic clause' and Halliday's tone-group. The latter equation seems hardly correct, as Halliday's group may maximally consist of 10 feet which, each, may consist of up to 7 syllables. The fact that the contours and phonological phrases of American intonation analysis are generally much shorter than the tone-units, tunes and tone-groups of British intonation analysis is too often ignored.

(57) Most writers see closure as blending, one way or another, the attitudinal and the mood- and modality-marking function. While for Armstrong and Ward (1926) and, more recently, e.g., Daneš (1960), Malmberg (1968), Bernard and Delbridge (1979), the choice of tune is fully grammatical and overall range, or the precise behaviour of the unaccented syllables adds attitudinal meaning, writers like Kingdon (1958) and O'Connor and Arnold (1961) blend the grammatical and the attitudinal by providing each permutation of the basic tones with tages like 'cool, phlegmatic statement', 'reserved, flat question', 'perfunctory statement', etc. Schubiger (1958) acknowledges only the basic opposition assertive/continuative. Hultzen (1964), Halliday (1967; 1970) describe attitudinal meaning as resulting from 'marked' intonations. For Crystal (1975), as for Bolinger (1970) closure is both syntactic and attitudinal, part grammatical, part ungrammatical.

(58) E.g. for Armstrong and Ward (1926), Ward (1939), Jones (1962), Daneš (1960), Hultzen (1959; 1964), Bolinger (1964), Lieberman (1967).

(59) E.g. Trager and Smith (1951), Trim (1964), Delattre et al. (1965), Bernard and Delbridge (1979) - who recognize rise, fall, and level. Hirst (1977) discusses the problem at some length.

(60) Pike (1946), Isačenko and Schädlich (1965), Collier
     and 't Hart (1975).   Isačenko and Schädlich in fact
     distinguish, for German, between rises with post-ictic
     onset (questions) and rises with pre-ictic onset
     (continuations), falls with post-ictic onset (non-
     complete utterances) and falls with pre-ictic onset
     (finality).

(61) O'Connor and Arnold (1961) have 6 tones; Kingdon (1958)
     modifies 5 tones by means of the dimensions 'high' and
     'low', 'emphasis' and 'non-emphasis', and, for 'complex
     tones', 'divided' and 'non-divided', obtaining, in
     this way, 28 tones.   Crystal (1969) has 9 tones,
     Adams (1969) 13 'intonemes', Halliday (1967; 1970)
     uses the dimensions 'high' and 'low' to modify 7 tones,
     thus obtaining 14 tones in all.

(62) E.g. in Collier and 't Hart (1975).

(63) Studdert-Kennedy and Hadding-Koch (1973) found that
     group-final pitch rises, in order to elicit consistent
     (90% or better) 'question' jugdments, needed to rise
     to a pitch level higher than that of any previous point
     in the sentence. Falls, in order to be perceived as
     'statements', should fall lower than any previous
     pitch-level in the sentence. They had earlier emphasized
     the importance of the whole pitch pattern of the group
     for the elicitation of consistent 'question' and
     'statement' judgments (Hadding-Koch and Studdert-
     Kennedy, 1964).

(64) Every group boundary presented three options. But in
     each transcription only those group boundaries could
     be taken into account which agreed with my own boundary
     markings, and, as we have seen, not all students marked
     all boundaries in accordance with mine. The lowest
     number of boundaries any student had in common with mine
     was 5, the highest 9, and the probability of 100%
     agreement occurring by chance varied therefore between
     0.037% for 5 boundaries, and 0.000051% for 9 boundaries,
     and had to be established separately for each pair.
     Reliability was calculated for each pair accordingly,
     and then averaged.
     Just how difficult it is to judge pitch direction, is
     shown also by the fact that the lecturer perceived,
     the first time he transcribed the material, only one
     pitch rise, the second time 7 pitch rises, one of
     which corresponded to a low fall in my transcription.

(65) E.g. Daneš (1960), Bolinger (1961b), Crystal (1975).
     Experimental evidence in Goldman-Eisler (1972) and Lass
     and Deem (1972).

(66) E.g. by Daneš (1960), Garvin and Mathiot (1958),
     Bierwisch (1965), Crystal and Davy (1969).

(67) Cohen and 't Hart (1967), Bolinger (1970), Lehiste (1973), Brown et al. (1980).

(68) E.g. Brown et al. (1980).

(69) Crystal and Davy (1969), Crystal (1975).

(70) An alternation between periods of planning and organization, during which speech is temporarily slow and hesitant, and periods of relative fluency. Henderson et al. represent it by means of a characteristic graph.

(71) E.g. Armstrong and Ward (1926), Ward (1939), Trim (1964), Bailey (1971a). Bierwisch (1965) refines the distinction: it is possible to treat parts of sentences intonationally as sentences ('Rhetorische Auflösung') and to treat sentences as phrase-units ('Rhetorische Bindung').

(72) Palmer (1922), Armstrong and Ward (1926), Schubiger (1958). Particular concatenations of tone groups mark particular grammatical structures, e.g. co-ordination is marked by a series of groups of same type, subordination by series of groups of different types.

(73) Trim (1964), Halliday (1967; 1970), Daneš (1960).

(74) Brown et al. (1980).

(75) E.g. Schubiger (1958), O'Connor and Arnold (1961), Trim (1964), Malmberg (1968), Cruttenden (1970), Bolinger (1972).

(76) In this part of the test only 10 student transcriptions could be used. There were two reasons for this. First of all, the comparison with my transcription could be made only in the case of intonation groups of which not only the boundaries but also the accents were marked in accordance with my transcription. I decided not to use transcriptions in which less than 5 groups agreed with mine in this way, but to end up with a reasonable number of transcriptions I disregarded minor differences in boundary placement - cases in which, for example, an unaccented preposition or hesitation noise was included in the tail of one group rather than in the precontour of the next, or the reverse - I reasoned that such differences would not influence a comparison of the relative prominence of the accents. Secondly, a number of students had not understood the instructions and marked several accents within the same group with the same index-number.

To calculate expected probability, I assumed that
every accent in a group with two accents had an equal
chance of being ranked as either 1 or 2, every accent
in a group with 3 accents an equal chance of being
ranked as 1, 2, or 3, etc.   Expected probabilities
were computed accordingly, for each pair separately.
The highest probability for any student's 'correctly
marked' groups being ranked in accordance with my
transcription by chance was 0.005%. Average reliability
was then established, again for each pair separately,
and averaged.

(77) For some this is a rule with exceptions (see note 10),
     for others one without exceptions (see note 8).
     Cruttenden (1970) and Crystal (1969; 1975) consider it
     a tendency rather than a rule.   In Crystal's corpus
     80% of the nuclear accents were placed on the group-
     final lexically stressed syllable.

(78) Daneš (1960), Schubiger (1964), Halliday (1967; 1970).

(79) See notes 72 and 73.

(80) In Halliday (1970) he adds a refinement: the items of
     information contained in a sequence of two falling
     groups are independent of one another, those in a
     sequence of a low rising group followed by a falling
     group are interdependent, and those in a falling group
     followed by a group with a fall-rise are circumstantial
     to one another.

(81) E.g. Daneš (1960), Bolinger (1964), Lehiste (1975).

(82) For this part of the test 15 student transcriptions
     were usable: those which had, in at least one of
     the two speech segments, the same group boundaries
     (disregarding minor differences of the kind described
     in note (76)), and within the group, the same primary
     accents as my transcription. For students in whose
     transcriptions this condition was fulfillled only for
     the segment of news reading, the probability of 100%
     agreement occurring by chance was 0.0039%, for those
     who had marked only the segment of spontaneous speech
     'correctly' 0.000064%, for those who had marked both
     segments 'correctly' 0.0000002%.   To calculate this,
     I assumed that in a segment of 4 groups every group
     had an equal chance of being assigned any one of 4
     degrees of prominence, in a segment of 5 groups any
     one of 5 degrees of prominence.
     Reliability was calculated accordingly for each pair,
     and then averaged.

* * *