Lingua 19 (1967) 177–192, © *North-Holland Publishing Co., Amsterdam*

ON THE ANATOMY OF INTONATION

A. COHEN AND J. 'T HART

Studies devoted to the description of prosodic features have always shown awareness of the problem of dealing with a highly complex phenomenon, one in which physical features such as frequency, intensity, and time, as well as their psychological counterparts (pitch, loudness and duration) and their interaction all play a part.

In recent years, there has been a tendency to concentrate on studying one of these features in the hope that, by this very reduction, a better understanding of the role of prosody in general could be obtained.

Without going into bibliographical details, we would like to single out a few recent works illustrating ways of tackling the study of pitch comparable to ours. These are, in chronological order: Bolinger, 1958 and 1962, Delattre et al., 1965, Hadding-Koch and Studdert-Kennedy, 1964, Isačenko and Schädlich, 1964 and 1965, Lieberman, 1965 and Mattingly, 1966. In Section 4, Discussion, we will have occasion to point out where we part company with some of these authors.

Without specifically mentioning any of the older workers in the field, we would like to remark that the outcome of our efforts portrays something very similar to Tune I, established by Daniel Jones for English intonation.

With most investigators in the field we share the common belief that reduction of the problem of intonation to a study of pitch phenomena would provide the most fruitful approach. The next step is to select a method that will enable us to establish those pitch movements that are interpreted as relevant by the listener.

In line with what was said at the Fifth International Congress on Acoustics at Liège (Cohen and 't Hart, 1965) we believe that these

relevant pitch movements are related to corresponding activities on the part of the speaker. These are assumed to be characterized by discrete commands to the vocal cords and should be recoverable as so many discrete events in the resulting pitch contours, which may present themselves at first sight as continuous variations in time.

In the following pages we will report on a method capable of providing a means of checking the hypotheses involved in this assumption. Such a method had to be suitable for establishing rules that seem to control the activities of the speaker in generating these contours as well as their interpretation by the listener perceiving them.

In Section 2 a number of techniques, together constituting what we call the method of 'perceptual analysis', will be described.

This method will be applied in a study of Dutch intonation in a way that makes the analysis objectively manageable and the validity of the rules to be derived therefrom experimentally verifiable.

The results to be given in Section 3 have been selected to illustrate the outcome of this approach and should be interpreted as a reflection of work in progress.

In order to provide an outline against which the actual investigations should be seen, a number of considerations should be mentioned. These are constituted by the inclusion or exclusion of notions, derived from others as well as developed by ourselves.

1. Basic considerations

A major consideration is that it is possible to make use of the internal representation of language patterns in a native speaker. This faculty, which will be employed throughout our investigations of pitch phenomena, can be used at the outset to provide a criterion for the selection of speech material to be studied.

This consideration implies that linguistic structuring as such is not at issue, which means in practice, that the actual material chosen need not in all respects correspond to a particular linguistic unit, such as a sentence.

Another important consideration is that exclusive regard for the distinctive function cannot possibly be expected to result in the establishing of features that are characteristic of a particular

language. In fact, we believe that study of the non-distinctive features is more promising for an understanding of what really goes on in intonation. Thus we are in agreement with the intuitive notions about the important contribution of intonational features towards characterizing a particular language, a fact amply confirmed by Delattre et al. (1965).

Some notions that we originally held were rejected in favour of ideas suggested by others: this applies to the consideration put forward by Lieberman (private communication) that a speaker has an idea of what he is about to say and, in modulating his voice, will therefore carry out pitch movements at the beginning, anticipatory to those that are still to follow. This notion is contrary to the one we entertained at one time, which was to the effect that from a study of the intonation of words in isolation, characterized by different accent patterns, one could hope to set up rules from which could be derived corresponding rules accounting for the intonation of larger stretches.

Another notion we cherished, namely, that in spite of the continuity with which pitch contours present themselves to the investigator, they may yet be ultimately derived from a number of discrete events in time, found experimental support in the work of Isačenko and Schädlich (1964). Working with vocoder processed speech characterized by steplike changes in F_0 in otherwise monotonous stretches, they were able to have their subjects respond in a way they could have done to corresponding natural pitch patterns.

Finally, to obviate the seemingly hopeless task of inductively finding regularities in pitch measurements on a large amount of speech material, we consciously looked for a method of an 'active' analysis on a perceptual basis. This principle was partly inspired by seeing Mattingly's work in progress at the J.S.R.U.

2. Method

2.1. *Analytic listening*

The method of perceptual analysis has three aspects, the first of which, analytic listening, was treated more fully at the 5th ICA at Liège. In this technique, successive 30 ms gated voiced parts of utterances were matched for pitch by ear with similar synthetic-

ally generated acoustic signals of adjustable and measurable frequency.

2.2. *Active analysis*

The second aspect is termed active analysis, thus indicating a procedure in which ideas about the origin of the phenomena to be measured are introduced in order to play an active part in the analysis itself. This procedure can take different forms: in our case, a preliminary set of notions that were supposed to account for the measurements were formulated explicitly, so as to give rise to predictions about the pitch patterns of utterances for visual comparison with actual measurements. (About this technique cf. Cohen and 't Hart, 1965). A second form of active analysis, subsequently employed by us as a logical consequence of the first, consists in using a tool, to be called an Intonator (cf. Cooper, 1961), constructed for the purpose of generating the predicted pitch patterns. This procedure is largely similar to the principle of analysis by synthesis as used to check the acoustic theory of speech production to account for the spectral composition of vowel-like sounds.

Since the predicted pitch patterns could be formulated basically in terms of a continuous and gradual fall-off during the utterance, interrupted by rises and falls at discrete moments, the tool to be constructed should contain a function generator meeting these demands: unless directed otherwise it should generate a gradual downward slope, and moreover be capable of producing steep rises or falls of prescribed slopes when triggered to these ends at specific moments in time. For the latter purpose an active counter is used.

This generator controls the repetition rate F_0 of an artificial voice source forming part of the synthesis end of a convential analysis-synthesis circuit as known from telephony. The circuit is fed by natural speech, recorded on a loop of tape.

The result is that the original intonation of the recorded utterance is replaced by an artificial intonation, while the remaining overall acoustic characteristics of speech, spectral, temporal, and dynamic, are preserved well enough for giving the output speech a natural character.

This set-up enabled us to judge the thus processed speech utterances for acceptability with respect to pitch contours. In this way, adhering to the practice of analysis by synthesis, we substitute for

comparison by eye of measured and predicted curves, as plotted on paper, an auditory comparison of internally represented and synthetically generated patterns.

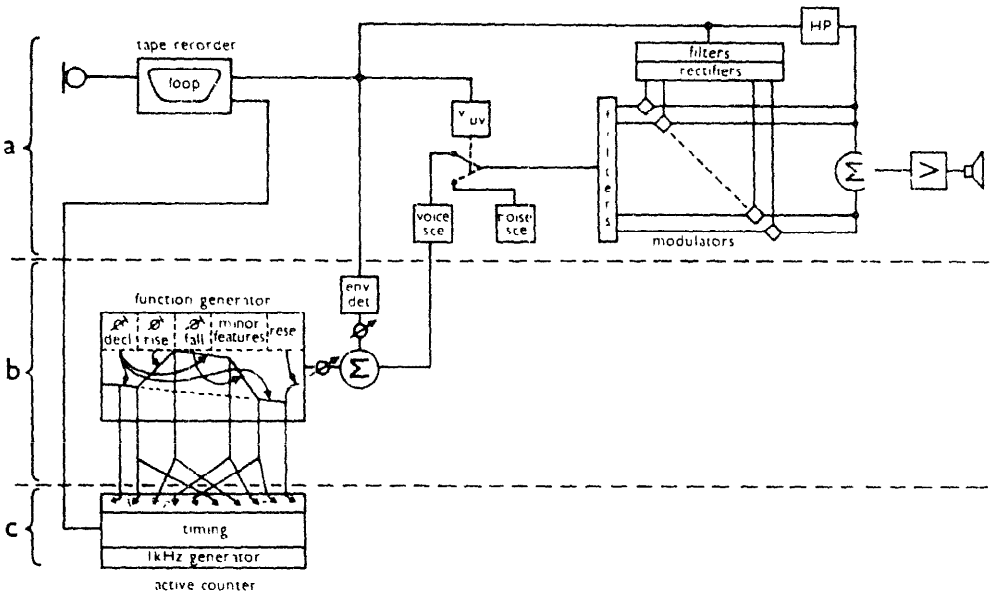

Fig. 1. Block diagram of the Intonator. In (a) the word content of the original speech message is preserved while its pitch information is lost; in (b) a new pitch contour can be generated by a function generator controlling an artificial voice source; in (c) the commands in time to the function generator are selected with an active counter.

Fig. 1 gives an outline of the apparatus needed for the purpose mentioned. A set of filters, each followed by a rectifier, measures the spectral composition of the input speech in terms of so many DC-voltages as functions of time. Meanwhile, a voiced-unvoiced detector serves to determine the choice of either a voice or a noise source. The signals from these sources are led to a second bank of, in principle, identical filters, whose outputs are regulated in amplitude by modulators, controlled by the DC-voltages obtained in analysis. The voice source is controlled by a function generator, by which the repetition rate is varied as logarithmic functions of time. To these functions can be added, to a variable degree, the output of an envelope detector, causing the frequency to go up and down with the amplitude of the speech input. A 1600 Hz high-pass filtered shunt is included in the system to improve intelligibility, mainly with respect to stops and fricatives. Experimental evidence shows that pitch information present in this higher frequency region leaves the pitch impression derived from the artificial voice source completely unaffected. It can be easily demonstrated e.g., that the system, set for constant F_0, gives rise to a monotone impression in spite of the natural inflections contained in the input speech.

The device termed active counter above provides, within a maximum cycle duration of 10 seconds, a number of pulses at moments in time accurate to the nearest millisecond, related to a zero moment at reset given by a pulse on the second track of the tape loop. These moments, at which the various ramps of the function generator are triggered, can be varied at will and on the spot.

On the basis of data obtained from the prediction method, it was thought feasible to impose a relationship between the originally independent ramps. Within this relationship, the rate of change can be controlled as a whole.

The essential task of the Intonator consists in providing satisfactory contours for perceptual comparison with internally represented pitch patterns, rather than in copying in detail the pitch of the input speech. Hence it becomes possible to look consciously for a manageable reduction of the number of degrees of freedom of the system. The restriction we adopted is calculated to preserve the freedom to vary the time pattern of the pitch configuration, which is deemed essential in our general approach.

2.3. *Synthesis by rule*

The third aspect of perceptual analysis can be designated synthesis by rule (cf. Liberman et al., 1959). This term is used to distinguish a procedure of speech generation based on an inventory of prescriptions from one in which the parameters are derived from acoustic representations of actually spoken utterances. Normally, synthesis by rule constitutes an ideal which, if fulfilled, would enable the investigator to show, by using the explicit knowledge in quantitative terms, that he has succeeded in giving an adequate analysis of the phenomenon he has been studying. In our case this would mean application of a readily available set of rules which would result in making audible satisfactory intonation patterns for any Dutch utterance.

In practice, even though we are still far removed from this ideal, we have nevertheless been able to apply this technique. Indeed, in the process of active analysis, in order to study the effect of changing one parameter, it is desirable to lay down, for the time being, specific values for the remaining parameters. Thus, e.g., in the stage where we made predictions, the moments where rises and falls should occur were taken for granted in an effort to determine their respective slopes, whereas, in a later stage, we included the knowledge obtained about these slopes in the Intonator, in order to examine their locations in time.

3. Results

The presentation of the results will reflect the operations we carried out on the pitch material through a number of sections, each laying bare characteristics which can be distinguished as being superimposed one upon the other.

Three classes of phenomena can now be described. First, a basic class, to be called 'major', characterized by a pitch pattern recoverable from any Dutch speech input material; by the same token it should provide the basis for the synthetic pattern to be generated. Second, a class containing features depending in the majority of cases on the incidental word usage, to be called 'minor'. Both major and minor pitch movements can be taken to result from the speaker's commands, contrary to the features belonging to the third class, to be called 'micro' which are solely a consequence of physiologically determined factors. The latter statement does not exclude the presence of involuntary phenomena within the other two classes.

3.1. *Major class phenomena*

The canonical pattern of the major class was found to resemble a hat (see Fig. 2). It is composed of, from left to right, an initial

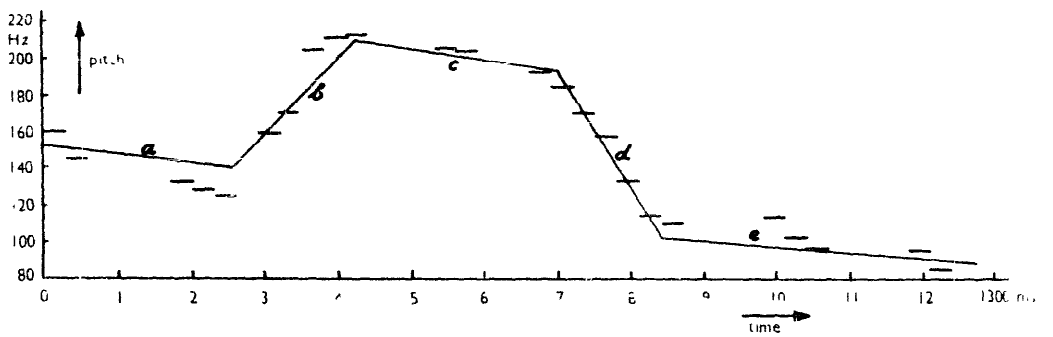

Fig. 2. The 'hat pattern' constituting the canonical form of the major class of Dutch intonation as seen against actual measurements of the utterance *ik moet eerst m'n kleren pakken* (I've got to fetch my clothes first). This pattern consists of three segments of declination (a), (c), (e), rise (b), and fall (d).

For the convenience of readers unacquainted with Dutch a comparable example for English is added underneath the phonetic transcription of the Dutch sentence. (The hat pattern illustrated on Dutch material is *not* supposed to represent the proper intonation for the English example).

gradual fall-off, to be called declination; a steep rise, an upward shifted segment of declination line, a steep fall, and a final declination which is the extension of the initial declination line.

From this pattern, depending on the presence or absence of the three subsections of the declination, seven other forms can be derived (see Fig. 3).

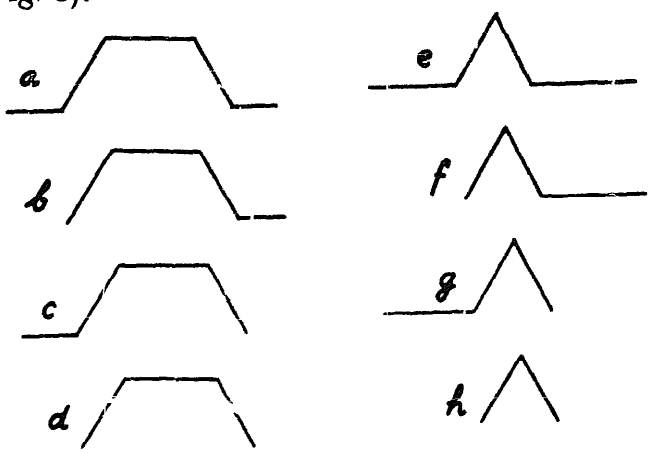

Fig. 3. Various forms of the hat pattern. Left column: rise and fall on separate syllables with various types of occurrence of declination. Right column: the same, with rise and fall on one syllable.

The slope of the declination line turned out to be 3%, that of the rise 25%, and that of the fall 50% every 100 ms in a normal situation, whereas in emphatic conditions these slope values are increased proportionally.

The duration of a rise is normally about 100 ms, and for a fall we find about 75 ms. Again, in emphatic conditions, these values may be increased.

The position of rise and fall coincides with prominent syllables of those words that play a dominant part in the utterance. In most cases, the vowel part of the prominent syllable contains at least a part of these movements.

The perceptual tolerance of the precise location of rises and falls has been determined experimentally by means of the Intonator. The most important conditioning factor in this respect was found to be the nature of the surrounding consonants, a rise, for instance, tending to take shape earlier with unvoiced prevocalic consonants than with voiced ones.

The forms, denoted as e–h in Fig. 3, imply coincidence of both rise and fall on a single syllable.

Although the major pattern, the hat, is presented here as a continuum, it is, of course, interrupted in reality by unvoiced parts of the utterance.

3.2. *Minor class phenomena*

Having found a single pattern characterizing the major class, we have thus far found four unrelated features belonging to the minor class.

As illustrated in Fig. 4, the hat pattern can be preceded by an

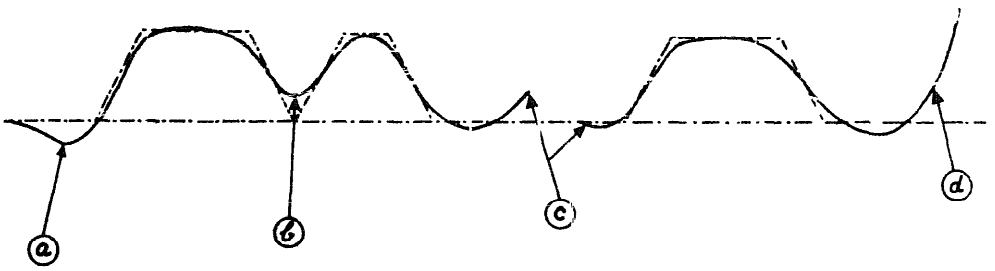

Fig. 4. Four features of the minor class (shown by the continuous line): anticipatory dip (a); dent in the hat pattern (b); end wriggle, non-final (c); end wriggle, final (d). As a reference, the dotted line represents the construction according to the rules.

anticipatory dip, (a). A second feature is to be found between the rise and the fall, causing a dent in the hat pattern, (b). A third feature, in complexes of hat patterns, appears as an upward wriggle followed by resumption of the pitch as an extension of the declination line, (c). The fourth feature consists of a terminal wriggle at the end of a hat pattern, after which the pitch is resumed independently of the previous contour, (d).

As for a quantitative specification of these features, this can as yet be given only for the dent, the shape of which can be described in terms of a fall and a rise, which may be separated by a segment of declination line. The rise coincides with a dominant word and has, just as the fall, the same slope as those in the hat pattern. In spite of the outward resemblance of Fig. 4b to a complex of two hat patterns, a number of characteristics of the first fall plead for a description as a dent in a single hat pattern. These characteristics, as opposed to those of the second fall, are listed below.

First fall	*Second fall*
Non-dominant word	Dominant word
The immediately preceding word is dominant, and bears a rise on its prominent syllable	The immediately preceding word need not be dominant
Start of fall preferably in unvoiced interval	Start of fall preferably in vowel
Fall tends to undershoot the extended declination line.	Fall tends to overshoot the extended declination line.

3.3. *Micro class phenomena*

In registrations of pitch measurements we found slight elevations coinciding with all full vowels. Moreover, the form of these elevations was found to depend on the nature of the surrounding consonants. Thus, e.g., the pitch contour of the vowel in this layer was found to be different after an unvoiced preceding consonant from that after a voiced one (see Fig. 5).

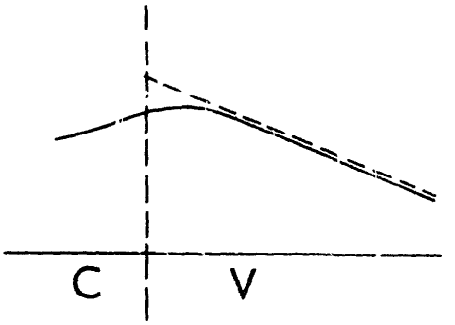

Fig 5. Illustration of a typical micro feature: the effect on vowel pitch under the influence of preceding consonant, ——— when voiced; ------- when voiceless.

4. Discussion of results

In the following discussion, certain points will be raised which are directly related to the results mentioned above.

In all three aspects of perceptual analysis described under Section 2 'Method', analytic listening, active analysis and synthesis by rule, the speech material used was preferably derived from utterances observed in natural speech situations (conversation, lectures, etc.). In the majority of cases we ourselves were the only subjects, both in measuring and judging natural and synthetic intonation contours. Whenever we enlisted other subjects, who

likewise fulfilled the condition of being native speakers of Dutch, no significant differences were found.

The idea of trying to describe intonation patterns in terms of a number of classes in a certain hierarchical order, as also advocated by Fónagy (1965), originates from the consideration that, in producing voice inflections, several controls function simultaneously, reflecting the intricacies of linguistic structure as well as the physiologic mechanisms. However reliable acoustic methods of measuring pitch contours may be, one cannot avoid being confronted with a superposition of features in these various postulated classes.

The stylized hat pattern, presumably representative of the intonation contour of a great number of utterances, fulfils a function in the description of Dutch pitch contours similar to Delattre's 'bird' for German (1965).

The concept of declination line has come about as a result of a conscious search for a way of relating up or down excursions to an average reference voice level. Such an average could not be established as long as we were thinking in terms of a horizontal line, whose existence one tends impressionistically to assume. The solution of this problem, viz., opting for a slightly tilted line, turned out to correspond very well with the running down pattern that Bolinger suggests as a language universal in pitch contours.

Introduction of this declination made it possible to accommodate rises and falls as discrete events in time, resulting in an overall contour which, when synthesized, sounds quite natural. Although one is usually and understandably inclined to associate pitch with voiced parts of the utterance, commands controlling rises or falls to take place in unvoiced parts of the utterance are not to be excluded. As a consequence, rises or falls may not be found in the pitch contour as such.

Although the notion of dominance still requires a proper definition, we have been able to handle it on an impressionistic basis, which by no means excludes a subsequent linguistic interpretation.

While the hat pattern has already proved its use as a descriptive device, it constitutes an interesting challenge with respect to its foundation in a linguistic or psychological reality.

So far only qualitative results have received detailed mention; as for the quantitative results the following should be pointed out: the values given for the slopes of rise and fall must not be taken as

playing a decisive part in the pattern. It is rather the location of these movements that is highly critical, as could be established from tests for perceptual tolerance by means of the Intonator.

Thus, for instance, when confronted with a fall, one has to determine its character as either the end of a hat pattern or the beginning of a dent within. It turned out that under otherwise similar conditions a shift of the inception of the fall over several tens of ms could determine the character of the fall either way, as illustrated in Fig. 6.

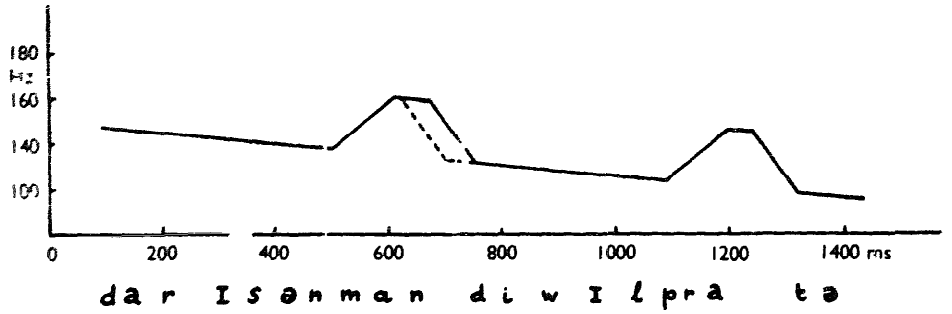

Fig. 6. Illustration of the way in which a shift over several tens of ms of a fall may result in shaping either a final or a non-final fall, thus giving rise to interpretation of the utterance in terms of two sentences (There is a man. He wants to talk.) or one respectively (There is a man wishing to talk).

When the fall at issue constitutes the end of a hat pattern, it is called 'final', in the other case, when part of a dent, 'non-final'. We have every reason to believe that the final fall fulfils the function of announcing the end of a unit of speech.

We are well aware of the severe limitations the introduction of the hat pattern presents with respect to actual intonation. For that reason we have striven to accommodate as many minor features as had to be accounted for from a perceptual point of view. This approach implies that the feature inventory obtained so far is by no means exhaustive, as becomes apparent from considering speech utterances showing all the diversifications to be observed in spontaneous speech.

Two important minor features require further comment. What appears as the rise in Fig. 4c can be considered an 'inter-hat' feature, since it anticipates a continuation of the original line of declination. As for the rise illustrated in Fig. 4d, this can be interpreted as a 'post-hat' rise. The difference between the two is

inherent not so much in the curve shapes but rather in the contours that follow: therefore, 4c can be designated as a feature of continuation, and 4d of non-continuation.

Instances of the latter rise can be found at the end of utterances known as a certain type of interrogative sentence. This so-called question rise need not occur in dominant words or even on prominent syllables, as opposed to final falls. In other words, this rise should not be taken to replace a final fall, but must be seen as an added feature.

As for the micro class, past experience with the generation of synthetic speech had already taught us the importance of such slight F_0 movements towards obtaining naturalness of inflection such as could be derived from a coupling of the F_0 control with the intensity. In actual speech, similar movements observed in pitch contours may well be explained as caused by the influence of the subglottal pressure. The effect illustrated in Fig. 5 can be adduced in support of this hypothesis.

5. Critical evaluation

In the next part of our discussion we would like to compare a number of salient points in our approach with those of others. We have no reason to quarrel with Delattre's statement: 'The value of an investigation depends on its research technique'.

We share his insistence on the importance of the non-distinctive characteristics of intonation. We therefore object to exclusive attention being paid to the distinctive function of intonation, as is done in the work of Isačenko and Schädlich (1964, 1965).

Our experience in measuring pitch curves (matching technique) has shown three discrepancies between physical reality and subjective interpretation. First, the presumed pitch levels extending over entire syllables did not materialize. Second, in quite a few cases parts of contours that seemed to obtrude themselves as unmistakably high, proved to consist merely of rapid falls. This kind of outcome is in line with results Lieberman (1965) obtained in asking trained linguists to judge pitch contours. Third, the experimentally established phenomenon of declination seems to elude normal observation.

As major rises and falls in our treatment have been correlated with prominent syllables of dominant words, one may well ask how

prominence comes about in non-dominant words. We would like to hypothesize that in these cases the otherwise concomitant features, such as duration, loudness and timbre, carry the brunt of conveying prominence in the absence of a dominating pitch accent. This would mean that the quest for hierarchical order among prominence features at the level of longer utterances has thus become largely irrelevant.

Within the confines of the major class, two ways seem to be open for emphatic purposes: increase of the slope values as well as prolonged duration. In our opinion no purpose will be served in trying to opt for one of the two, since perceptually both pitch movements may have the same effect.

At this juncture it seems opportune to state explicitly that the way followed by us is opposite to that followed in the majority of linguistic studies of pitch. The hope of correlating semantic categories and intonation contours is fed by the criterion of distinctiveness, and leads to the search for appropriate 'tunes' that in themselves seem to have the required semantic function. This approach, in our opinion, has too often led to disappointing results, as can be seen from the following instances:

Investigators have always been at pains to show that the end of a sentence coincides with a final pitch fall. However, we have found, in line with Isačenko, that such a fall need not occur at the end of an utterance, but may be followed by one or more syllables on the line of declination.

Another instance is provided by the vexed problem of a special interrogative intonation. For one thing, the semantic concept of a question is hard to delineate; for another, the so-called typical question as manifested by a rising pitch at the end, need not be representative of questions in general (in the presence of other means, such as word order, question particles, etc.), whereas such a rise may be found in cases where no question is signalled.

Some of these objections may be raised against the way Delattre prejudges the analysis of intonation contours by imposing concepts like 'finality, major continuation and minor continuation' on his material.

Having discovered patterns of rises and falls in analytic listening, we subsequently were able to predict such rises and falls based on the notions of dominance of words in the utterance. This particular

notion, which owes its origin to an impressionistic analysis of which any language user is capable, is now being investigated to give it firmer linguistic support. Tests carried out so far give us every reason to believe that some such relation between dominance and syntactic structuring will indeed be found. For instance, by shifting, with the help of the Intonator, the location of rise or fall, it is possible to bring about a corresponding shift in dominance, which may result in a change of interpretation on the part of the subject.

These tests dealing with phenomena of the major class can be extended to include the contribution of the features present in the minor class. Thus a dent in the hat pattern can be straightened out so as to lead to a construction of a simple hat pattern, and both patterns can be presented for perceptual evaluation.

6. Conclusion

Our aim has been to study intonation features in a way by which we could dissociate ourselves from various more or less traditional approaches, which either show results that are largely uncorrelated, or run the risk of remaining within the limits of an impressionistic method. The method of perceptual analysis advocated in this paper seems to be promising in so far as it enables us to systematize a limited set of preferential patterns, the underlying rules of which can be verified experimentally.

In our opinion, reduction of the seemingly capricious pitch movements in speech to such rather simple patterns has been made possible largely through our assuming the operation of discrete commands on the speaker's part in generating intonation.

Institute for Perception Research,
Eindhoven, The Netherlands

Bibliography

Bolinger, D. L., 1958. 'A theory of pitch accent in English', *Word* 14, 109–49.

Bolinger, D. L., 1964. *Intonation as a universal*, Proc. 9th Int. Congr. of Linguists, Cambridge 1962. The Hague, pp. 833–44.

Cohen, A. and 'T Hart, J., 1965. *Perceptual analysis of intonation patterns*, Proc. 5th Int. Congr. of Acoustics, Liège, A 16.

COOPER, F. S., 1962. *Speech synthesizers*, Proc. 4th Int. Congr. of Phon. Sc., Helsinki 1961. The Hague, pp. 3–13.

DELATTRE, P., POENACK, E. and OLSON, C., 1965. 'Some characteristics of German intonation for the expression of continuity and finality, *Phonetica* 13, 134–61.

FÓNAGY, I., 1965. *Zur Gliederung der Satzmelodie*, Proc. 5th Int. Congr. of Phon. Sc., Munster 1964. Basel, pp. 281–6.

HADDING-KOCH, K. and STUDDERT-KENNEDY, M., 1964. 'An experimental study of some intonation contours', *Phonetica* 11, pp. 175–85.

ISACENKO, A. V. and SCHÄDLICH, H., 1964. *Untersuchungen über die Deutsche Satzintonation*. Akademie-Verlag Berlin, 66 pp.

ISACENKO, A. V., 1965a. *Deutsche Satzintonationen als Konfigurationen zweier diskreter Tonebenen*, Proc. 5th Int. Congr. of Phon. Sc., Munster 1964. Basel pp. 371–2.

LIEBERMAN, P., 1965. 'On the acoustic basis of the perception of intonation by linguists', *Word* 21, 40–54.

LIBERMAN, A. M., INGEMANN, F., LISKER, L., DELATTRE, P. and COOPER, F. S., 1959. 'Minimal rules for synthesizing speech', *J. Ac. Soc. America* 31, 1490–9.

MATTINGLY, I. G., 1966. 'Synthesis by rule of prosodic features', *Language and Speech* 9, 1–13.