nottingham linguistic circular

Volume 7, Number 1
August 1978

EDITORIAL NOTE

Beginning with the present volume (Vol.7), NLC will appear in two issues (Numbers 1 and 2) each calendar year. (Previous volumes have been spread over the period October-September in successive years). Year-affiliations are therefore as follows:

Vol.1	1971-2
2	1972-3
3	1973-4
4	1974-5
5	1975-6
6	1976-7
7	1978
8	1979

References to individual issues of NLC will be made by Volume and Number, eg. the 8th issue of NLC is Vol.4 Number 2. Pagination will in future be based on volumes rather than individual issues.

These changes are designed to bring the format of NLC into line with that adopted by most other academic journals.

PHONOLOGICAL MODELS FOR INTONATION *

A.S. Crompton
Language Centre
University of Nottingham

There exist a number of proposals as to how intonation contours should be represented phonologically. I shall discuss two in particular, with a view to assessing their overall adequacy according to a set of criteria to be given below. The models in question are (1) the system of tonetic stress marks widely used in British studies; (2) the theory of pitch-levels developed mainly in the United States. We shall see that both these approaches display serious shortcomings, and an alternative model will be proposed that will hopefully prove to be more adequate.

1. We begin by proposing a set of criteria according to which the adequacy of a model of intonation may be assessed. Two types of criteria present themselves: one phonetic and one phonological.

First, on the phonetic side we should require of any phonological model of intonation that it be relatable in some fairly straightforward manner to relevant aspects of the phonetic representation of utterances, since it is only in this way that the model can attain a high degree of testability. This immediately raises the question: what are the relevant aspects of phonetic representation? Now although the present state of our knowledge probably does not permit the question to be answered in any fully conclusive and satisfactory way, nevertheless, enough is known about the phonetics of intonation for us to lay down the broad outlines of such an answer. To begin with, it would seem that phonetic representation must contain precise specifications of the physical and perceptual parameters involved in intonation, in particular, fundamental frequency (F_O) and pitch. (Other parameters are doubtless also relevant, such as intensity/loudness, and duration/length, as argued, for instance by Crystal (1969, Ch.4), but we shall not take these into consideration here.) Now it is well known that fundamental frequency is only one of the acoustic determinants of perceived pitch, so that the pitch contour of an utterance need not always correspond in an entirely straightforward fashion to its F_O contour. Nevertheless, it is probably safe, for the purposes of the present discussion, to ignore any possible discrepancies between these two parameters. Assuming, then, that the phonetic representation of an utterance must contain details of its pitch/fundamental frequency contour, let us now ask what the principal characteristics of this contour are, and how they relate to intonation. It is immediately apparent that the pitch/F_O contour of an utterance and its intonation contour cannot simply be equated (even allowing for the idealisations concerning intensity and duration, etc. mentioned above). For there are many aspects of pitch/F_O contours that seem quite irrelevant as far as intonation is concerned, namely those that have to do with the perturbations due to the segmental structure of the utterance (the disappearance of the pitch/F_O contour during voiceless segments, the disruptions associated with consonantal segments in general and due to the effects of coupling between the

* Paper given at the spring meeting of the Linguistics Association of Great Britain, University of Lancaster, 10th to 12th April 1978.

larynx and the supraglottal tract on the rate of vibration of the vocal folds), and also those that have to do with the fine-structure of the contours, as discussed, for instance, by Lieberman and Michaels (1962). It seems necessary to abstract away from such phenomena in order to arrive at a true picture of intonation contours themselves. Having done this, we are left with smooth, continuous contours (presumably something like those represented by Jones in his early work on "intonation curves" (1909)) and it is these that must be output by any adequate grammar of intonation.

Performing this task in an accurate and explicit manner is the first criterion of adequacy that we may impose on a model of intonation.

Turning now to phonological criteria, it is reasonable to impose on grammars of intonation the same requirements as are generally applied to segmental phonology, of which three are of particular importance. First, we may require that the grammar should express all and only the distinctive oppositions in the language. So, for example, just as a theory of distinctive features in segmental phonology must be able to express all the phonemic oppositions in a given languaage, and should not permit distinctions to be expressed that are not found in any natural language, so also a theory of intonation must, say, be able to express all the tonal oppositions found in a given language, and should not be able to express tonal distinctions not found in any language.

Second, we may require of a theory of intonation that it should embody a measure of complexity both for individual intonational units (such as tones) and for classes of such units. Thus, for example, just as we may expect a theory of segmental phonology to say something about the relative complexity of segments such as [t] and [t'] or about the relative complexity or naturalness of classes of segments such as [p,t,k] and [p,n,œ], we may also expect a theory of intonational phonology to have something to say about the relative complexity of tones such as [ˋ] and [ˇ] and about the relative naturalness of sets of tones such as [ˊ,↑‒] and [ˋ,↓ˇ].

Third, an adequate theory of intonational phonology should specify what constitutes a natural or possible rule of intonational grammar. If the analogy with segmental phonology holds true here too, attainment of this goal may prove somewhat elusive.

These, then, are the criteria I shall adopt for assessing the adequacy of the models of intonation to be discussed in what follows.

2. The first of the models I shall consider is the tonetic stress marks system widely used in British studies of intonation, and developed by Kingdon (1958), O'Connor and Arnold (1973), Crystal (1969), and others.

The basis of this approach is, as its name implies, that the intonation pattern of an utterance is as it were build on a framework made up of stressed syllables (perhaps more precisely: accented syllables). Each accented syllable bears a particular tone (rising, falling, static, etc), and once the tonality of the accented syllables has been specified, the intonation patterns over the unstressed syllables is to a very large extent predictable. The tones themselves are represented by means of a set of symbols, such as [ˊ] for a rising tone, [ˋ] for a falling tone, [-] for a static tone on a "nuclear" syllable, ['] for a static tone on a non-nuclear accented syllable, and so on. We thus have transcriptions of the form (1a), corresponding to an intonation pattern of the form (1b).

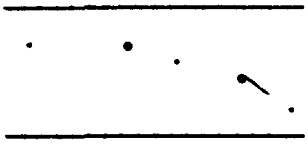
(1) a. The 'prop 'forward 'bit my ˋear

b.

The convention for the interpretation of the symbol ['] states that the immediately following syllable is stressed, that it carries a static tone, that it lies on a slightly lower pitch level than the preceding stressed syllable, and that the unstressed syllables following it are pronounced on approximately the same level. Examples of this type of pattern can be found in Kingdon 1958, §20, O'Connor and Arnold 1973: 37, Crystal 1969, §5.8.

An important feature of the approach is that the representations in terms of tonetic stress marks are systematisations of intonation patterns transcribed by means of interlinear graphs of the sort illustrated in (1b). In these graphs, horizontal lines mark the upper and lower limits of the graphs, and purport to represent to upper and lower limits of the speaker's pitch range. Dots or dashes are placed in between the two horizontal lines in such a way as to represent the pitch level of particular syllables within the total pitch range available to the speaker. This system clearly involves a number of idealisations: the precise interpretation of the notion pitch range is not altogether clear; the degree of precision with which the pitch level of particular syllables is represented is not always made explicit, and the time span over which comparisons between the relative pitch levels of syllables are to be made is rarely stated. More important, perhaps, is the fact that intonation patterns are treated essentially as sequences of syllabic pitch levels. This is not always apparent from the transcriptions with tonetic stress marks themselves, which exploit redundancies in intonational patterning to a very high degree, and thereby avoid having to place a symbol before every syllable (cf. (1a) above, for example). It is, however, quite clear from the interlinear graphs (cf. (1b) above), and also from cases where the pitch levels of individual unstressed syllables are not totally predictable from the tonality of neighbouring stressed syllables and must therefore be marked by means of extra symbols. This is illustrated by the following example, which uses the transcription system proposed by Crystal (1969, §4.8.1).

(2) -He | was-nt RÈAdy



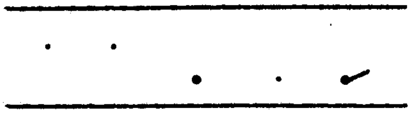
The fact that intonation patterns are treated in this syllable-by-syllable manner is a serious defect in the tonetic stress-marks approach, as will be argued below.

Let us now assess the adequacy of the tonetic stress marks approach in the light of the criteria put forward in §1. Consider first the requirement that a model of intonation should make precise and explicit statements concerning the phonetic properties of utterances, in particular concerning their pitch/fundamental frequency contours. It is obvious that representations of intonation patterns in terms of symbols such as ',ˊ,ˋ, etc. do not in themselves meet this requirement. This means that at the very least, the work done so far within the tonetic stress marks framework must be supplemented by indications of how pitch/fundamental frequency contours may be derived from representations in terms of ',ˊ,ˋ, etc. Such indications have not so far been provided, to my knowledge (but cf. Ashby 1975). This does not in itself invalidate the tonetic stress marks model, of course, but it does at least cast doubt upon its adequacy, since even if the relevant phonetic detail can be specified, it will still for that purpose be necessary to devise some mechanism whereby the representations in terms of ',ˊ,ˋ, etc. can be transformed into representations of pitch/F_o contours proper. Consequently, the tonetic stress marks model would be likely to prove inferior on grounds of simplicity to some other theory which did not require such a mechanism, say, one in which the phonology of intonation was already dealt with in terms of a direct representation of pitch/F_o contours.

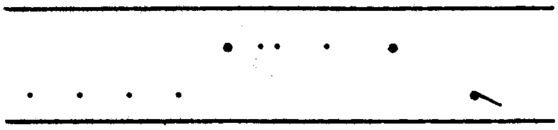
Next, we turn to the phonological criteria, and first, the requirement that a model of intonation should be constructed in such a way that it can express all and only those distinctive oppositions that are found in human languages.

Taken in its strongest form the theory claims that the entire pitch pattern of an utterance is determined by the tonality of the stressed or accented syllables, so that there are no distinctive oppositions affecting only unstressed syllables. It is not clear whether anyone has ever seriously tried to maintain this position, since all the well-known versions of the theory provide symbols that relate specifically to patterns over unstressed syllables. Probably the most familiar of these devices are those that have to do with "pre-heads" in English, ie. syllables occurring before the first accented syllable of the tone group or phonological phrase. Most analysts recognise at least a two-way distinction at this point, namely between a "high pre-head" and a "low pre-head". Cf. the following examples from O'Connor and Arnold 1973: 22-27:

(3) a. ¯But you ˌcan't do ˏthat



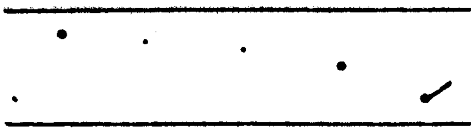
b. It was an un'usually °dark ˎnight



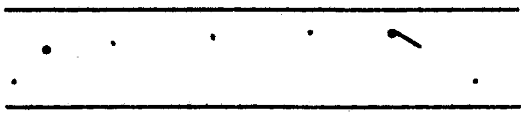
As a result of this situation, the theory is perhaps not very aptly termed the 'tonetic stress marks theory', since tonetic marks other than tonetic stress marks must be permitted.

Once the theory has been weakened in this way, it is not at all clear that there are any conceivable intonation patterns that it could not deal with by some means or other, if necessary by the addition of further symbols to the already existing inventory. In the version elaborated by Crystal, for instance, the symbols [¯] and [_] are available to indicate unstressed syllables that are either higher or lower than the preceding syllable: cf. Crystal 1969: 147. Thus transcriptions of the following sort may be found:

(4) a. I |would_n't _have 'said THÁT



b. I |did¯n't ¯think ¯it ↑ MÀttered

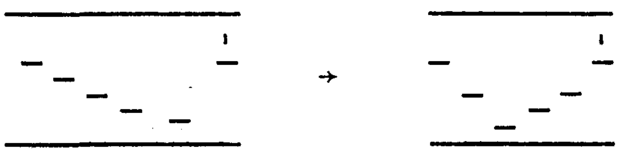


Given such devices, there would seem to be no series of ascending, descending, or level sequences of syllables that could not be accommodated by the system, and this, plus the fact that there seems to be no intrinsic restriction on the inventory of individual tones, forces one to the conclusion that the theory embodies almost no constraints on the notion 'possible intonation pattern'. Although in itself undesirable, this state of affairs does at least have the advantage that the model would presumably have no difficulty in fulfilling the requirement that it should be able to express all the distinctive intonational oppositions found in human languages.

If, on the one hand, the tonetic stress marks model is probably equal to the task of expressing all the distinctive intonational oppositions found in human languages, on the other hand one may ask whether it will express only those oppositions. Consider the following facts of French intonation (cf. Crompton 1978a, Ch.2).

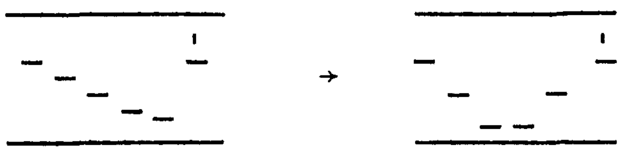
First, there is a process by which a descending intonation pattern over a sequence of unstressed syllables is transformed into a descending-ascending pattern before a "high" accented syllable, as diagrammed in (5a), where ' marks an accented and _ an unaccented syllable.

(5a)



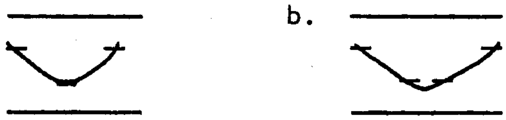
A variant of this process takes the form illustrated in (5b).

(5b)



Now according to the interlinear graph representation used here, the two patterns of (5) must be considered quite distinct: (5a) has a descent followed by an ascent, whereas (5b) has a descent followed by a level sequence followed by an ascent. Consequently, the tonetic stress marks model, being based on interlinear graphs, would also have to treat these two patterns as being different in kind. Such a treatment ignores completely the fact that as far as their pitch/fundamental frequency contours are concerned, the patterns of (5a) and (5b) are identical, since both involve a descent followed by an ascent; the only difference between them is that in (5a) the changeover from descent to ascent takes place within a particular syllable, whereas in (5b) it takes place between two syllables, as diagrammed in (6).

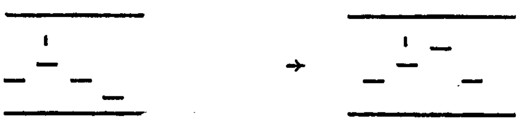
(6) a. b.



Consequently, the tonetic stress marks model is compelled to treat as distinct, patterns that are in fact intonationally identical, and thereby fails to express only the distinctive oppositions found in the language.

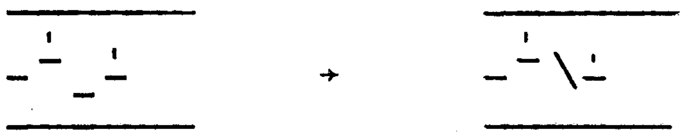
A second case in which the tonetic stress marks model provides an inadequate treatment of intonational oppositions concerns a process known as spreading (cf. again Cromton 1978a, Ch.2). This has the effect diagrammed in (7).

(7)



In most cases, the accented syllable is followed by at least two unaccented ones before the next accent. Sometimes, however, only one unaccented syllable intervenes, and in this case, the effect of spreading is as shown in (8): the unaccented syllable acquires a falling tone.

(8)



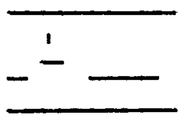
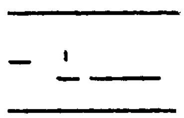
According to the interlinear graphs, the patterns of (7) and (8) are distinct, and they must therefore be transcribed differently in terms of tonetic stress marks. Seen as sequences of ascents and descents, however, the two patterns are identical: there is an ascent from the accented syllable onto the beginning of the following unaccented syllable, and this ascent is followed by a descent; in (7) the descent is spread over two (or more) unaccented syllables, whereas in (8), it is concentrated on a single unaccented syllable. Again the tonetic stress marks model is forced to recognise an opposition where in fact there is none.

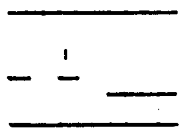
The two arguments just stated point strongly to the inadequacy of the tonetic stress marks model as a system for the expression of distinctive intonational oppositions. In fact, they apply not only to the tonetic stress marks theory, but also to any model which represents intonation patterns in a syllable-by-syllable manner. Intonation is often said to be a suprasegmental phenomenon; an even better term would be 'suprasyllabic'.

A second set of criticisms that may be levelled at the tonetic stress marks model has to do with the specification of natural classes of intonational units. We have, in fact, already seen one case where the model fails to capture a natural intonational class, namely in our discussion of spreading on p. . Here as seen from (7) and (8), patterns that involve (1) a falling tone on a single unaccented syllable, and (2) a descent over a sequence of unaccented syllables are generated as alternative outputs from a single intonational process. The tonetic stress marks model is, however, unable to express this generalisation.

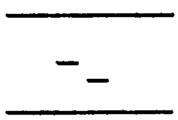
A second case involves various types of accented syllable in French (cf. Crompton 1978a, Ch.2). Accents are classified as 'high', 'low', or 'level' according to whether they begin higher, lower or on the same level relative to pitch level reached at the end of the preceding syllable. High accents and level accents are typically followed by a descending sequence of unaccented syllables, whereas low accents are typically followed by an ascent. It follows that high and level accents must be treated as a class by any descriptively adequate grammar of French. In terms of the tonetic stress marks theory, however, high accents would presumably be represented by means of the symbol [↑'], level accents by means of [→'], and low accents by [↓'] (cf. Crystal 1969, §4.8.1 for a system of transcription that implies this notation). Given representations of this sort, it is not at all easy to see how the relevant classes could be referred to except by means of a list.

Third, we consider the co-occurrence restrictions relating to level stretches in French. Apart from their occurrence in phrase-initial position, these may occur only in the environments shown in (9). (I shall continue to use interlinear graphs for representing intonation patterns simply because they are familiar and convenient for purposes of exposition, even though, as we have seen above, they are theoretically inadequate).

(9) a.  b. 

c. 

(The environment shown in (9c) is not quite correct as it stands, since it appears that the level accent must be preceded by a level stretch; this need not concern us here, however). The common factor in this set of environments is the step down onto the beginning of the level stretch; the differentiating factors are (1) the nature of the preceding accent, and (2) the location of the accent relative to the beginning of the level stretch: in (9a) and (9c), the level stretch follows the accent, in (9b) the beginning of the level stretch coincides with the accent. The least we need to be able to do, therefore, in order to determine the occurrence of level stretches, is refer to a sequence of the form (10), where either the first or the second syllable is accented.

(10) 

(In fact, it is obvious that much more than this is needed, but the further details need not concern us here). However, even this apparently simple task can be dealt with by the tonetic stress marks theory only by means of lists. If the first syllable is accented, then the relevant configuration is presumably something like (11a) (the symbol s stands for any syllable).

(11a) 's _s

If the second syllable is accented, then the relevant configuration will be something like (11b).

(11b) ↓'s

Again, the theory fails completely to capture the relevant generalisation.

Finally, we consider natural classes of nuclear tones in French. These are classified first of all as being high, level, or low (according to the level of their onset relative to the level of the end of the preceding syllable), and as rising, falling or static. The obvious notation within the tonetic stress marks model is that shown in (12):

(12)

	high	level	low
rising	↑ˊ	→ˊ	↓ˊ
static	↑-	→-	↓-
falling	↑ˋ	→ˋ	↓ˋ

According to the analysis of Crompton 1978a, Ch.3, the natural classes among these tones are (1) rising and high static tones, (2) falling and low static tones. (Level static tones are extremely rare, and it is not clear how they pattern). Other than by listing, there seems to be no way of referring to these classes within the tonetic stress marks model.

The results of our assessment of the tonetic stress marks model are as follows: (1) the model is insufficiently explicit phonetically; at the very least a considerable amount of additional machinery must be developed in order to derive satisfactory specification of the pitch/fundamental frequency parameter. (2) The model fails to provide an adequate account of certain distinctive oppositions in the phonology of French intonation. (3) Intonational units whose behaviour with respect to certain processes in French intonation indicates that they belong to a single class can only be referred to by means of lists within the tonetic stress marks model. There are thus strong reasons for seeking a better alternative.

3. We now turn to the "pitch levels" model of intonation developed principally by Pike (1943), Wells (1945), Trager and Smith (1951), and Harris (1944). According to this theory, intonation patterns are represented in terms of a number, usually four, of "pitch phonemes", which are concatenated to form morphemes. The pitch phonemes are claimed to be realised phonetically by varying pitch/ fundamental frequency levels, and are represented by numbers, with, usually, the highest number corresponding to the highest pitch level. For example, a falling intonation pattern might be represented as /21/, /31/, /41/, /32/, etc, a rising pattern as /12/, /13/, etc, a level pattern as /11/, /22/, etc, and so on. In addition to the pitch phonemes themselves, most analysts also posit a set of juncture phonemes, which mark the boundaries between intonation units, and are realised phonetically in a number of ways, including pause, increased duration of preceding syllables, and various more centrally intonational phenomena. The latter include, in particular, effects such as terminal sustention, rise, or fall, so that in the Trager-Smith system, for instance /1#/ represents a terminal fall from level /1/, /2#/ a terminal fall from level /2/, /1||/ represents a terminal rise from level /1/, /3|/ a terminal sustention at level three, and so on (Trager & Smith 1951: 41-52).

An important feature of the pitch levels theory is that intonation and stress are held to be phonemically independent in that the intonation pattern of an utterance in terms of its pitch phonemes may vary quite independently of the occurrence of stressed or unstressed syllables. This is, on the other hand, not to say that intonation and stress are entirely unrelated in the model. According to Pike's analysis, for instance, "A stressed syllable constitutes the beginning point for every primary contour; there is no primary contour without a stressed syllable, and every heavily stressed syllable begins a new contour" (1945: 27). For Trager and Smith, the interactions between stress and intonation are to be found in allophonic variants of the pitch phonemes; they state, for instance, that, "The several [^{2}]s are seen to vary directly with the stress; lower absolute pitch ([$^2_\vee$]) on a weak syllable than on a tertiary ([$^2_\circ$]), higher absolute pitch ([$^2_\wedge$]) on a secondary than on a tertiary. (...) we see that the [$^2_-$], an absolutely highest variety of [^{2}], occurs with the primary stress." (1951: 42).

The pitch/levels model has almost from the beginning been subjected to criticisms denying its validity as a theory of intonation. One of the most persistent critics has been Bolinger, who has taken issue both with the notion of pitch-levels itself (1949, 1951), and with the alleged independence of intonation and stress (1958). A well-known study by Lieberman (1965) showed striking discrepancies between the way single utterances were transcribed by two linguists in terms of pitch-levels and also demonstrated the existence of not easily explainable discrepancies between the intonation patterns in the linguists' transcriptions and the corresponding F_0 contours in the test utterances themselves. Crystal (1969: 197) considers that these counter-arguments "have not so far been answered", and on this basis rejects the pitch-levels model as a viable theory of intonation. Avoiding going into the question of the validity of these criticisms (they are in my view almost all invalid, as I hope to show in a later paper), we may ask how the pitch levels theory rates according to our own criteria, proposed in §I.

We begin with the requirement that an adequate model of intonation should provide an accurate specification of the relevant aspects of pitch/fundamental frequency contours. Now a representation in terms of pitch levels is equivalent to the specification of a step-function, of the sort illustrated in (13) - this example from Pike (1945: 135).

(13) Every man on the force is working on it sir

°3--2 °3--2- °3- -2- °4- -1-2/

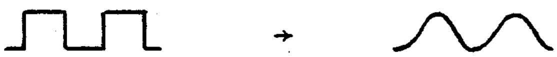
Are step-functions a reasonable representation of pitch/F_0 contours? Even allowing for the idealisations noted in §I, it is clear that they are not. But does this constitute a refutation of the pitch-levels theory? Obviously not: no advocate of the model has ever suggested that it should be interpreted in this way.

The closest thing to the above interpretation that has actually been suggested is found in Trager and Smith:

"Also to be noted is the prelinguistic finding that pitch as used in language is heard around a limited number of points rather than as a continuum."

(1951: 41). It is clear from the following discussion that they consider the number of points in question to be sixteen. This is an extraordinary claim. As far as I am aware it is unique, and is totally un-supported by experimental evidence; it is simply false. The theory of pitch-levels advocated by Trager and Smith is, as it stands, wrong.

However, other interpretations of the model are possible. One would be to assume that the step-functions corresponding to the pitch-levels are related to pitch/F_0 contours by a process like that illustrated in (14).

(14) 

The contour obtained as output from this process bears at least some resemblance to observed pitch/fundamental frequency contours, especially when the necessary abstractions and idealisations mentioned in §1 are taken into account. Although the 4-valued step-functions corresponding to the 4-level theory would almost certainly not be directly transformable into adequate representations of pitch/F_0 contours, it is possible that better results could be obtained by means of a set of allophonic rules that supplied finer variations among pitch levels in much the same way that segmental allophonic rules supply finer variations for parameters relating to vowel quality, nasalisation, and so on. There are therefore reasonable grounds for optimism with regard to this interpretation.

Looking at the version of the model proposed by Pike, it seems that this corresponds quite well to the interpretation just outlined. To begin with, Pike is quite explicit that intonation contours may include many more than four levels; and it is not the level of all syllables that is important, but rather the levels attained at what Pike calls "contour-points" (cf. Pike 1945: 26-27). These are defined in the following way:

"In determining the pertinent level or levels of contour, one does not classify the pitch cr every syllable or part of a syllable, but only those points in the contour crucial to the establishment of its characteristic rises and falls; these may be called CONTOUR POINTS. In any rising or falling contour, two contour points are present: the pitch level at its beginning and the pitch level at its end. For a contour which first falls and then rises (or, very rarely, one which first rises and then falls) a third contour point is always present at the place where the direction of pitch movement changes."

(p.26) Given the contour-points, which define the value of the intonation contour at certain specific points, the values of the contour at the remaining points are obtained by, presumably, some sort of interpolation procedure - Pike is not too explicit here - as the following quotation suggests:

"In the following utterances notice the contour points (numbers will be placed approximately under them).

Tom!	Tom!?	Tommy!	Tommy!?	Margȧret!?	telephone number!	telephone number!?
2-4	2-4-3	2- -4	2- -4-3	2- -4-3	2- -4	2- -4 -3

In the first five samples each syllable had at least one contour point, so that the relative pitch of each syllable was important to the establishing of contours and their meanings. In the last two samples, there were more syllables than contour points. These extra syllables can be pronounced with intermediate pitches in a general descending scale, or with considerable variation in the amount of drop from syllable to syllable."

(p.26) One may therefore conclude that the version of the pitch-levels theory proposed by Pike (i) does not make obviously false claims concerning the phonetics of intonation, in the way the Trager-Smith version does, and (ii) allows in principle for further refinement which might result in an adequate phonetic representation of intonation contours. On the other hand, the further refinement that is required has not so far been provided.

We may note at this point that the foregoing discussion goes at least some way to answering one of Lieberman's (1965) implied criticisms of the pitch levels theory. When he examined the ranges of fundamental frequency values associated with the four pitch levels in the transcriptions of a set of test utterances by two linguists, Lieberman found a considerable amount of overlap between ranges (cf. Lieberman 1965: 47 and Figure 4); he notes that:

"The Trager-Smith pitch levels do not correspond to discrete non-overlapping ranges of fundamental frequency nor do they correspond to discrete relative ranges of fundamental frequency. These comments apply even when we consider the transcriptions made by a single linguist who carefully transcribed the tape recorded sentences of a single talker."

and also that

"The pitch levels of the Trager-Smith system do not even reflect the relative pitch levels of a single utterance of a single talker when it is transcribed by a single linguist. The fundamental frequency that corresponds to pitch level one, for example, may be identical to or greater than the fundamental frequency that corresponds to pitch level two. The pitch levels reflect the relative fundamental frequency only during segments of speech in which there is continuous voicing."

(p.52) Is this state of affairs incompatible with the claims made by the pitch levels theory? I think not. Suppose, for example, we have a /3121/ pattern, whose step-function representation is as shown in (15).

(15) 

The actual realisation of this pattern would on Pike's formulation be something like (16).

(16) 

It is obvious from (16) that the material associated with level /1/ in (15) must overlap with that associated with level /2/. As the situation illustrated in (15) and (16) may also be found, mutatis mutandis, with many other patterns, it is only to be expected that a considerable amount of overlapping will take place. Lieberman's findings are for this reason not prejudicial to the theory of pitch-levels, at least in the sense in which it is interpreted by Pike.

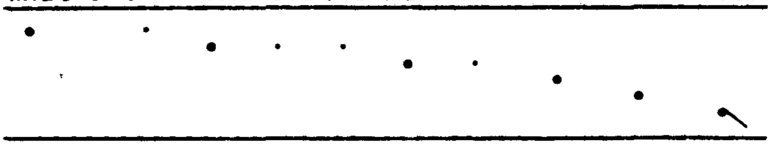
In summary, we may say that no direct refutation of the pitch-levels theory is forthcoming from an investigation of its relation to the phonetics of intonation. On the other hand, it is clear that none of the versions of the theory provides sufficiently accurate and explicit specification of pitch/fundamental frequency so that the model as it stands cannot be considered adequate according to the first of the criteria of §1.

We now move on to consider the adequacy of the model according to the other, phonological criteria of §1, in particular those relating to the expression of distinctive intonational oppositions.

First, we consider the restriction of the number of distinctive pitch levels to four. In principle, this is desirable since it imposes tight constraints on the range of oppositions the language may employ. In practice, however, it is simply incorrect, since there seems to be nothing wrong with the examples of (17a) and (17b) which involve six and seven distinctive pitch levels respectively.

(17) a. <u>But did the machine actually work?</u>

b. <u>What's the name of the man who lives next door?</u>



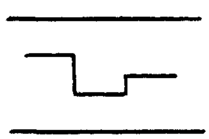
(cf. Kingdon 1958: 25-29, Crystal 1969: 225-233 for further examples of a similar sort).

Does this mean that English has 7 pitch phonemes rather than 4? Presumably not, since it would also be possible to construct examples requiring 8 distinct levels. So does English have 8 rather than 7 pitch phonemes? Presumably not that either. In fact, the whole attempt to fix the inventory of pitch phonemes at one particular number rather than another seems to be misconceived. What is important in the above examples (and others) is the levels of certain syllables relative to one another, and the use of numbers to represent these levels should be seen simply as a notational device for coding these relativities. (Observations such as these have indeed led other people working within the pitch-levels framework to abandon the notion of a fixed number of pitch phonemes as such, and to use numbers simply to indicate the height of one syllable relative to another - cf., for instance, Bierwisch 1966).

A second criticism involves the interaction of the pitch phonemes and the juncture phonemes. In the Trager-Smith system there are three junctures: ||, |, and #. All of these have different effects on the pitch contour immediately preceding them: || is associated with a terminal rise, | with a terminal sustention of pitch, and # with a terminal fall. In Pike's system, there are two junctures, /, associated with a rise or sustention, and // associated with a fall. The difficulty, which is common to both systems, is the following. Presumably, the final pitch movement of each phrase is taken to be a feature of the following juncture. But what happens when the pitch movement associated with the preceding two pitch phonemes is the same as that associated with the juncture, namely in cases like 31# or 23|| in the Trager-Smith system (where higher numbers represent higher pitch) or 2-4// or 4-3/ in the Pike system (where higher numbers represent lower pitch)? In the Trager-Smith system, is 31# distinct from 32# and 3#, and is 23|| distinct from 24|| and 2||? In the Pike system, is 2-4// distinct from 2-3// and 2||, and is 4-3/ distinct from 4-2/ and 4/? Neither Trager and Smith nor Pike make clear statements on this point. I myself find it by no means easy to imagine just what some of these theoretically possible distinctions would sound like, and it seems very likely that both the Trager-Smith and the Pike versions of the theory allow for a greater number of distinctive oppositions than are in fact operative in the language.

The criticisms just stated are in my view sufficient in themselves to raise serious doubts concerning the adequacy of the pitch-levels theory as a means of expressing distinctive phonological oppositions. Nor are they the only criticisms that can be made. One might, for instance, question the claim that three different types of ascent are possible from the lowest pitch level, /1/ in the Trager-Smith system, namely /12/, /13/, /14/, whereas only two are possible from level /2/ (/23/, /24/), one from level /3/, (/34/), and none from level /4/. Similar remarks hold with relation to descents. Or one might question the claim that in a falling-rising-falling pattern of the form (18)

(18)



where the initial level is /3/ and the final level /2/, it is necessarily the case that the middle level is /1/, so that there is no possible opposition between types of descent from the initial level /3/; normally, of course, /31/ and /32/ contrast. Further criticisms could doubtless also be devised.

Taken together, the objections we have raised suggest that the pitch-levels theory, if it is not going to be rejected in its entirety, must at least be seriously overhauled. A last-ditch attempt to salvage the theory might be made in the following way, involving two modifications.

The first, and perhaps less important modification, is designed to counter the second of the criticisms made; instead of representing terminal pitch movements by means of juncture phonemes, we make use of pitch phonemes. Thus in the Trager-Smith system, for instance, 2# and 3# are replaced by 21 and 31 or 32; 2|| and 3|| are replaced by 23 or 24 and 34; 2| and 3| are replaced by 22 and 33; and so on. This means, of course, that some special means must be found for representing 1# and 4||; this problem will be dealt with directly. It should be noted that representations something like those just mentioned would have to be generated at some point in the original system as well. For the purposes of phonetic representation, rising intonation contours are not adequately represented by means of a symbol such as ||, which would therefore have to be mapped onto a terminal rise in the pitch/F_0 contour in any case. Similarly for terminal falls and sustentions. Consequently, the modification we are proposing increases the adequacy of the theory without adding to its complexity.

The remaining criticisms voiced previously all bear on the use of particular numbers in the representation of intonation patterns. The first objection related to the inadequacy of any theory involving a fixed number of pitch phonemes. The last two objections had to do with unjustified restrictions on the range of oppositions available within certain patterns, restrictions arising in arbitrary fashion out of the assignment of particular integer values to pitch levels. In order to overcome these criticisms, one might propose replacing the numbers 1, 2, 3, 4 by variables, α, β, γ, ..., whose relative values were defined by a set of equations and inequalities. For example, in place of /243/ we might have $/\alpha\beta\gamma/$ together with the statement: $\alpha<\beta>\gamma$. This formula would in fact replace not only /243/, but also /132/, and maybe /343/, /143/, /342/, etc., as well. In fact

$/\alpha\beta\gamma/ : \alpha<\beta>\gamma$

would correspond to any rising-falling pattern. Similarly, the formula

$/\alpha\beta\gamma/ : \alpha>\beta<\gamma$

would correspond to any falling-rising pattern,

$/\alpha\beta\gamma/ : \alpha=\beta>\gamma$

would correspond to any level-falling pattern and so on. If it proved to be necessary to express different degrees of ascent or descent (steep vs shallow rises and falls), the system could easily be made to accommodate this, by means of formulas such as

$/\alpha\beta/ : \alpha>\beta$

$/\alpha\beta/ : \alpha>>\beta$

and so on.

The system just described offers at least some hope of salvaging the pitch levels theory. One might, however, ask precisely what is retained of the original approach, and whether there is in fact anything in it that is worth retaining.

The fundamental claims of the pitch-levels model are (i) that intonation is basically a matter of attaining particular levels at particular points in time; pitch movements - ascents, descents, levels - are simply means to this end; (ii) that languages operate in terms of a small fixed number of distinctive (phonemic) levels. The second of these claims has already been shown to be false (at least in the case of English). As for the first, the revised pitch-levels framework just proposed, which seems to me to be the only workable version of the model, raises serious doubts as to its correctness. For when we describe an intonation pattern by means of a formula such as

$/\alpha\beta\gamma/ : \alpha < \beta > \gamma$

we are not specifying particular levels which must be attained, but rather we are saying that between the two points corresponding to α and β the intonation pattern rises, and that between the two points corresponding to β and γ the intonation pattern falls. In other words, the formulae are devices for defining pitch movements rather than pitch levels. This means we have moved right away from the static view of intonation basic to the pitch levels model and towards a radically different view in which the dynamic properties of intonation contours, the ascents, descents and level stretches, are what count. If this alternative view is correct, then the various versions of the pitch-levels model proposed by Pike, Trager and Smith, and others are not only inadequate technically, but they are also based on a fundamental misunderstanding of the nature of intonation. In fact, as I understand it, this is precisely Bolinger's main criticism of the model: intonation is not a matter of levels, but of "configurations", sequences of ascents, descents and level stretches (cf. Bolinger 1951).

This distinction between a level-based and a configuration-based view of pitch patterns is also relevant to the difference between speech and singing. Singing *is* a matter of attaining particular levels at particular points in time; the voice generally jumps relatively cleanly from one level to another (except in glissando) and failure to hit precisely the right level is considered incorrect. In addition, the number of levels in the system is strictly limited to 12 per octave (at least in classical tonality). Singing is therefore just the sort of activity that is suitably described in terms of levels, and it is precisely this that is provided by the musical stave, of course. Speech, on the other hand, is according to Bolinger (and also to me), a very different activity; it is not the levels but rather the directions of movement that are important and this makes the intonation patterns of speech extremely ill-suited to representation by means of a musical stave (although several investigators have tried to make use of this form of notation - with few interesting results, in my view). In fact, at least one study of intonation, M.Liberman 1975, has explicitly taken singing, or chanting, as a point of departure; not surprisingly, the model that emerges is a variant of the pitch levels theory, and contains all the defects mentioned earlier.

If, therefore, we abandon the static view of intonation as a series of levels and replace it by the dynamic view in which it is the configurations that count, it follows that we must also abandon the pitch-levels model, and replace it by one in which the basic building blocks of intonation are ascents, descents, and levels. The presentation of such a model is our task in the following section.

4. In the preceding sections, we have presented a set of criteria for assessing the adequacy of models of intonation, and have considered two such models, the tonetic stress marks theory, and the pitch-levels theory. Both were found seriously wanting according to the criteria proposed. In this section, we shall develop an alternative, hopefully more adequate model, and illustrate its operation with examples from French intonation in particular. The new model is based on two fundamental characteristics of intonation mentioned earlier; namely (1) that intonation is a suprasegmental, or, more accurately, a suprasyllabic phenomenon; (2) that intonation is basically a matter of configurations in Bolinger's sense, ie. of ascents, descents, and level stretches. We shall refer to the new model as the "configurational model".

To begin with the second of the two points just mentioned, the first thing we require is a means of representing the direction of pitch movements - up, down or on a level. We posit two features [Rise] and [Fall] (abbreviated to [R] and [F], and adopt the following specifications:

ascents/rises : |+R, -F|
descents/falls : |-R, +F|
level stretches/statics : |-R, -F|

(The fourth combination, |+R, +F|, is excluded by convention). The features [R] and [F] are suprasyllabic features, in the sense that their domain will in general extend over stretches of utterance containing several syllables, although as we shall see, their domain may also cover stretches the size of a single syllable or less. Sequences of [R]-[F]- specifications can be formed in order to represent complex intonation patterns; for instance, a rising-falling pattern is represented as

$$\left|\begin{array}{c|c} +R & -R \\ -F & +F \end{array}\right|$$

or simply as

|+R | +F|,

a falling-level pattern is represented as

$$\left|\begin{array}{c|c} +F & -R \\ & -F \end{array}\right|,$$

and so on. If it is required to represent varying degrees of steepness in ascents or descents, this is easily accomplished by the use of scalar coefficients of the two features in place of the symbol "+".

We may thus have representations of the form

$\begin{vmatrix} 1R \\ -F \end{vmatrix}$, $\begin{vmatrix} 2R \\ -F \end{vmatrix}$, $\begin{vmatrix} -R \\ 2F \end{vmatrix}$,

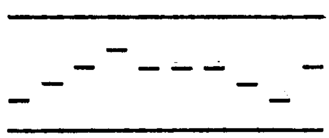
and so on.

The fact that [Rise] and [Fall] are suprasyllabic features means that they have in general very different domains from the familiar segmental features [syllabic], [consonantal], [nasal], and so on. This brings one immediate advantage, namely that it is possible to treat the segmental and prosodic aspects of phonological representations completely independently of one-another. Thus the segmental content of an utterance may vary without its prosodic content being affected, and the prosodic content may vary without its segmental content being affected. This is an intuitively satisfactory and theoretically most desirable situation. Clearly, at the phonetic level, the two forms of representation will have to be combined, for instance by specifying the interruptions in the pitch/F_0 contour caused by voiceless segments, but that is something that need not concern us here.

What we do need to deal with, however, is the way the segmental and prosodic aspects of a phonological representation are alligned temporally; in other words, we need to ensure that the ascents, descents, and level stretches of the intonational representation come in the right places relative to the segmental or syllabic structure of the utterance. There are in fact two separate problems here, one phonological and one phonetic. It is well-known that the allignment of prosodic to segmental structure is rather precisely controlled at the phonetic level (cf. for instance, Collier 1970). On the other hand, what is important at the phonological level is the range of distinctive contrasts available within the system. As far as the present inquiry is concerned, only the second of these problems is of interest.

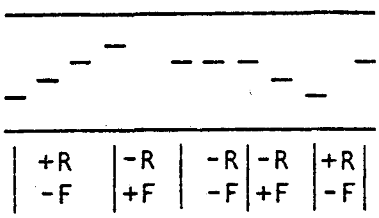
We saw earlier, in our discussion of the tonetic stress marks approach, that a widely used method of transcribing intonation contours makes use of interlinear graphs, in which the pitch level of each syllable is represented by a dot or dash placed between lines delimiting the speaker's pitch range. An illustration is given in (19).

(19)



Now although we have rejected the syllable-by-syllable method of transcribing intonation contours that is embodied in this particular type of interlinear graph, there is nevertheless one feature of this method that is probably correct, and which we can make use of. Namely, that the turning-points in the intonation contour - the points at which the contour changes from ascending to descending, or descending to level, or whatever - are most frequently associated with particular syllables rather than falling between syllables. That being the case, we can represent a contour of the form (19) in the way shown in (20).

(20)



Let us call the vertical bars used in the [R]-[F]-specifications discontinuities. Using the square brackets to represent syllables, we may indicate the allignment of the syllabic structure of an utterance to its suprasyllabic structure in the manner shown in (21), which corresponds to (20).

(21)

```
[ ][ ][ ][ ][ ][ ][ ][ ][ ][ ]
|    +R   |-R|  -R  |  -R |+R|
|    -F   |+F|  -F  |  +F |-F|
```

It is representations of the form just illustrated which form the basis of the configurational theory of intonation. The form of representation just described is adequate to deal with a great many different types of intonation pattern, but not with all. Patterns which cannot be dealt with in this way are those which involve kinetic (moving) tones on individual syllables. Since each segment of the intonational representation is bounded before and after by a discontinuity, in order to deal with [R]-[F]-utterances like those of (22), which involve [R]-[F]-specifications whose domain is a single syllable,

(22)

we must allow representations in which two discontinuities may fall within a single syllable:

(23)

```
[ Who ]      [Hell]      [ Erm ]
 |+R|         |+F|        |-R|
                          |-F|
```

Even this does not go far enough, however, since in some languages, English is one, individual syllables may carry not only rising, falling, and static tones, but rising-falling, falling-rising, and even rising-falling-rising tones as well. This requires representations of the form (24).

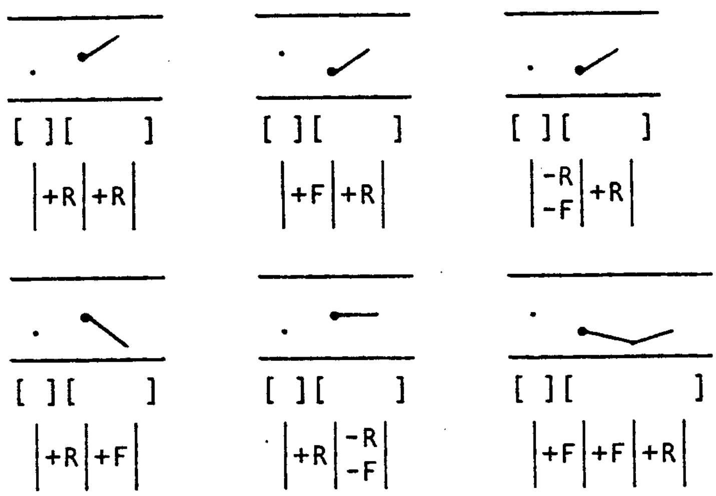
(24)

```
[        ]      [        ]      [           ]

 |+R|+F|         |+F|+R|         |+R|+F|+R|
```

In cases where syllables with their own rising, falling, or static tones are preceded by other syllables, which may be higher, lower, or on the same level relative to the onset of the tone-bearing syllable, we have representations like those of (25).

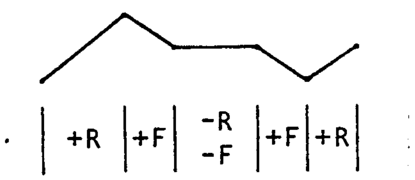
(25)



In case it might be thought that the theory we have proposed rests crucially on the validity of the interlinear graph method of transcription, it should be noted that our use of interlinear graphs and syllable-by-syllable transcriptions has been for ease of exposition only, and that our theory is logically quite independent of them. The essence of the configurational theory lies in the features [Rise] and [Fall], which are suprasyllabic features, and in the way the prosodic structure of an utterance is alligned with its syllabic structure. We have claimed that this is best achieved by locating the discontinuities in the [R]-[F]-patterns within particular syllables. This claim stands or falls not according to the validity or invalidity of the expository devices used to present it, but rather according to its ability to fulfil the criteria of adequacy imposed on theories of intonation in general, namely those outlined in §1. And it is to this that we address ourselves next.

The first question concerns the ability of the model to provide accurate specification of the pitch/fundamental frequency contours of utterances, or more precisely, of the smoothed continuous contours that are obtained when the effects of the segmental structure of the utterance and random variations in the fine-structure of the contours have been removed. Since the features [Rise] and [Fall] are related to the direction of pitch/F_0 movements, the representations they generate are of the form illustrated in (26).

(26)

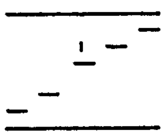


These are almost certainly not adequate in themselves, even allowing for the idealisations just mentioned, since the range of different rates of ascent and descent found at the phonetic level is likely to be much wider than that found at the phonological level. On the other hand, if the theory is able to express all the distinctive phonological oppositions in rates of ascent and descent, there is every reason to suppose that the additional information required for an adequate phonetic representation can be supplied by means of a suitable set of allophonic (detail) rules. Second, the temporal allignment of the prosodic with the segmental structure of utterances has been effected so far by means of the location of the discontinuities in the [R]-[F]- specifications within particular syllables.

We have already mentioned that more precise control is needed for an adequate representation of the phonetic facts, but as before, as long as the phonologically distinctive oppositions are expressible within the theory, any additional information can be provided by low-level rules. It seems reasonable to conclude, therefore, that the model we have proposed does provide the basis for an adequate phonetic representation of intonation contours.

Next, we consider the adequacy of the model according to the specifically phonological criteria set out in §1. The first of these concerns the expression of distinctive oppositions. The theory we have developed so far is capable of expressing any opposition based on rate or direction of pitch movement - ascent vs descents vs level stretches, greater vs lesser rates of ascent or descent. An illustration of the distinctive role of different rates of ascent is provided by patterns of the form (27), found in French

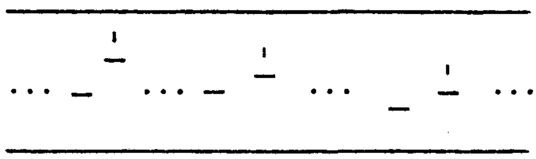
(27)



Here we have a shallow ascent followed by a steep ascent followed by a shallow ascent, and the syllable marked ' is rendered intonationally prominent (hence, if stressed, accented) by virtue of being jumped on to by an extra-steep rise.

One thing our theory cannot do, as it stands, is express distinctive oppositions that relate to the pitch level of individual syllables. We saw in §3 that intonation is not, as the pitch-levels theory claims, basically a matter of attaining particular levels at particular points in time, nevertheless there is considerable evidence that the level of individual syllables is important in certain cases. As far as French is concerned, the cases in question exclusively involve accented syllables, never, to my knowledge, unaccented ones. For instance, it is generally the case in French that each successive high accented syllable is on a lower level than the preceding high accented syllable, so that patterns of the form (28) arise (where _' marks an accented syllable).

(28)

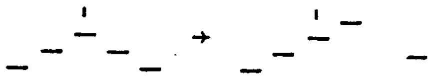


It would not be easy to generate such patterns if the grammar were not allowed to make reference to the pitch/F_0 level of the individual syllables concerned. Moreover, there is a rule in the grammar of French intonation which states that if a word contains two accented syllables, then the second must be higher than the first, unless the first accent is either (1) the first accent in the phrase, or (2) itself higher than the preceding high accent. (cf. Crompton 1978a: 116-118). Again, it would be difficult to deal with such phenomena solely in terms of the features [Rise] and [Fall].

These observations point to the need for an extension of our theory which will make it possible to refer to the level of individual syllables. We therefore posit a syllabic feature [f], with scalar coefficients that correspond more or less directly to pitch/F_0 level. There are interesting interactions between the feature [f] and the features [Rise] and [Fall], as explained in Crompton 1978a, Ch.2, §5.

It might seem that our decision to introduce a feature that refers to the level of individual syllables is almost tantamount to a return to the pitch levels theory, and that if the grammar has to contain the feature [f] in any case, then why not abandon the features [Rise] and [Fall] and work only in terms of [f]? There are two answers to this. First, the arguments produced against the pitch levels theory at the end of the last section remain valid: intonation is still fundamentally a matter of directions of pitch/F_0 movement. Second, the use of the feature [f] within the configurational theory is highly restricted. We have already seen the need for referring to the pitch/F_0 level of individual accented syllables; apart from this, the only other cases in which the level of individual syllables are referred to involve the end-point of nuclear glides (Crompton 1978a, Ch.3, §6) and, maybe, the onset (first syllable) of phonological phrases, as, for instance, in the specification of high vs low preheads in English. The first case (that of the accented syllables) involves processes that apply within phonological phrases (tone groups); the second two involve processes that define relations between phonological phrases. There is therefore the possibility of imposing fairly tight restrictions on the form intonation rules may take; if they are intra-phonological-phrase rules, then reference to the level of individual syllables is confined to accented syllables; if they are inter-phonological-phrase rules, then reference to the level of individual syllables is confined to the first and last syllables (which may be accented or unaccented) and to certain specific accented syllables (eg. the first or occasionally the second accent in the phrase, also the accent which constitutes the highest point of the phrase - cf. Crompton 1978a, Ch.3 for fuller discussion of these cases). Outside this narrowly defined set of cases, rules may not make reference to the level of individual syllables. That is the situation within the configurational theory. According to the pitch levels theory, on the other hand, it should be possible to refer to the level of almost any syllable, irrespective of whether it is accented or not, and irrespective of its position within the phonological phrase; any restrictions in this area would be ad hoc ones. In consequence, the two models embody quite different predictions about the class of cases in which intonation rules may refer to the level of individual syllables: a highly restricted class in the case of the configurational theory, an almost unrestricted class in the case of the pitch levels theory. The question as to which approach is correct is therefore an empirical one, which can be answered by analysis of the intonation systems of languages. The evidence so far (admittedly based on scant data - the analysis of French presented in Crompton 1978a is the only full one that exists to date for any language) points to the correctness of the configurational theory. Other ways of deciding between the two approaches can also be devised. For instance, we saw earlier (p. 6) that the grammar of French intonation contains a rule that has the effect illustrated in (29) (cf. also Crompton 1978a, Ch.2, §4.4).

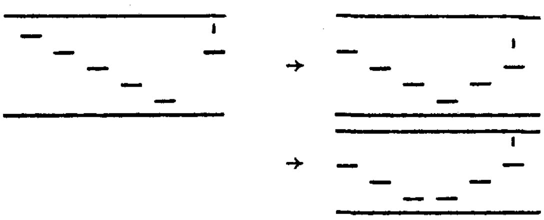
(29)



If the configurational theory is correct, then the only syllable in this pattern whose level may be referred to is the accent (marked ╧). This means that since the level of the accent is not affected by the rule in question, the two patterns of (29) should be treated as equivalent with regard to any processes referring to the level of individual syllables (processes such as those mentioned on p21, for instance). According to the pitch levels theory, on the other hand, the two patterns of (29) must be considered quite different, since the pitch level attained at the peak of one is higher than that attained at the peak of the other. Again, the issue is an empirical one, therefore.

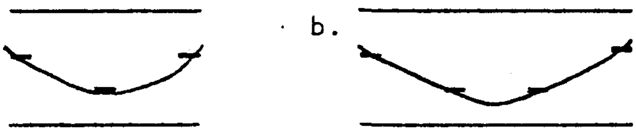
Returning to our discussion of the adequacy of the configurational model with respect to the criteria of §1, consider next the process illustrated in (30), and discussed earlier in connection with the tonetic stress marks theory (cf. also (5) above).

(30)



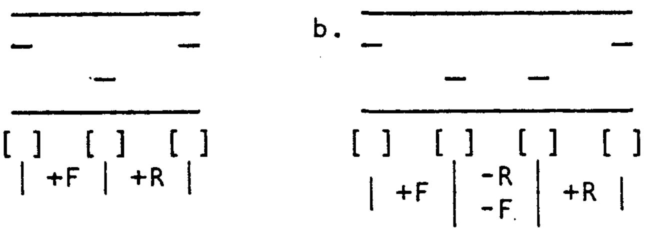
We saw that a defect in the tonetic stress marks approach was that its syllable-by-syllable treatment of intonation necessitated an analysis in which the two patterns on the right of (30) were intonationally distinct, despite the fact that they differ only in the allignment of the pitch/F_0 pattern to the syllabic structure of the utterance, as shown in (31) (cf. (6) above).

(31) a. b.

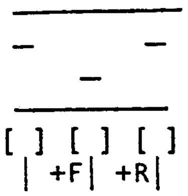


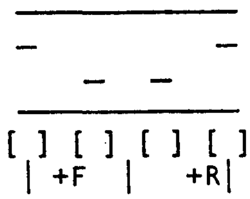
The question now is whether the configurational theory can prove more successful than the tonetic stress marks theory in dealing with such patterns. As our theory stands at present, the only way we can represent (31a) and (32b) is in the manner shown in (32).

(32) a. b.



which is obviously no better than the tonetic stress marks analysis. On the other hand, since the distinguishing factor between (31a) and (31b) is the location of the turning point at which the direction of pitch movement changes from descending to ascending, and since we have represented such turning points by means of discontinuities in [R]-[F]- specifications, the obvious solution is to allow discontinuities to fall not only within, but also between syllables, thereby representing (31a) and (31b) in the manner shown in (33).

(33) a. 

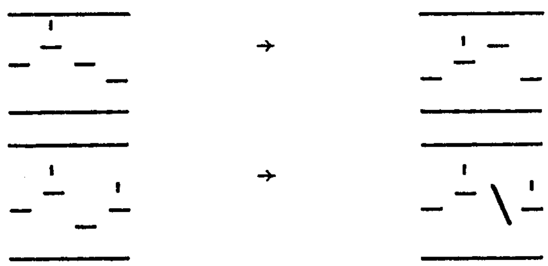
b. 

This extension of the theory results in a significant improvement over the tonetic stress marks approach.

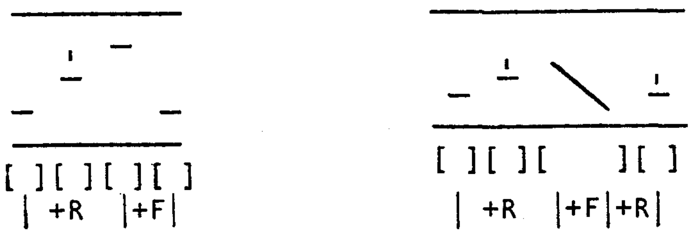
Another case in which the tonetic stress marks theory proved inadequate was in its treatment of spreading, a process illustrated in (34) (cf. (7), (8) above).

(34) a.

b.



As was pointed out before, seen as sequences of ascents and descents, the two patterns on the right of the arrows in (34) are identical. They differ, however, in that whereas the descent in (34a) is spread over two (or more) unaccented syllables, in (34b) it is compressed onto a single unaccented syllable. The representations we require are therefore of the form (35a,b).

(35) 

Again, the alternative proposed here proves superior to the tonetic stress marks theory.

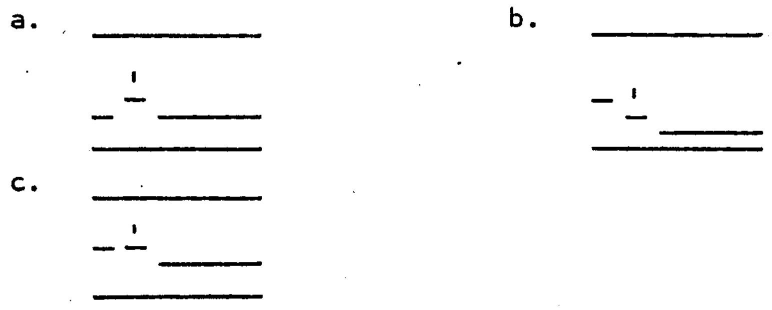
The second of the phonological criteria proposed in §1 concerned the specification of natural classes of intonational units. In §2, we found a number of instances in which the tonetic stress marks theory proved inadequate according to this criterion; in all cases, configurational theory is easily shown to be superior.

The first case involved various types of accented syllable in French. These are classified as 'high', 'level', or 'low', according to their relation to the preceding syllable. High and level accents are typically followed by a descent; low accents are typically followed by an ascent. High accents and level accents must therefore be treated as a class. Under the tonetic stress marks approach, this class could be referred to only by means of a list: {↑',→'}. According to our theory, we refer simply to configurations of the form (36), where "[+A]" stands for an accented syllable.

(36)
[+A]
|-F|

Second, we saw that the tonetic stress marks theory was incapable of producing an adequate account of the distribution of level stretches in French. These occur only in the environments illustrated in (37) (cf. (9) above).

(37)



The tonetic stress marks theory can deal with this phenomenon only by listing the environments concerned, thereby missing the obvious generalisation that in each case, the level stretch is preceded by a descent. For the configurational theory, there is no problem: we simply refer to a configuration of the form (38), where α is a variable over + and -.

(38)

$$\begin{matrix} [\alpha A] & [-\alpha A] \\ \multicolumn{2}{c}{| + F |} \end{matrix}$$

(As noted earlier, this is not the only thing that is required for a proper treatment of the distribution of level stretches in French, but the further details have no bearing on the present discussion).

Our third example concerns the specification of classes of nuclear tones in French. We saw on p 9 that the classes to which reference must be made are (1) rising and high static tones, (2) falling and low static tones. As before, the tonetic stress marks theory can refer to these only by means of lists. According to our theory, the [R] - [F] - specifications for the tones in question are as shown in (39).

(39)

rising tones: $[\ \ ]$ $\begin{vmatrix} +R \\ -F \end{vmatrix}$ falling tones: $[\ \ ]$ $\begin{vmatrix} -R \\ +F \end{vmatrix}$

high static tones: $[\ ][\ \ ]$ $\begin{vmatrix} +R \\ -F \end{vmatrix}\begin{vmatrix} -R \\ -F \end{vmatrix}$ low static tones: $[\ ][\ \ ]$ $\begin{vmatrix} -R \\ +F \end{vmatrix}\begin{vmatrix} -R \\ +F \end{vmatrix}$

To define the set of rising and high static tones, we may draw on the fact that the [R] - [F] - specifications in question terminate in a |-F| segment; we then require that this |-F| specification must be accompanied by a |+R| specification that either precedes or coincides with it. Similarly, to define the set of falling and low static tones, we require a terminal |-R| segment that either follows or coincides with a |+F| specification. Again something that was not possible within the tonetic stress marks theory proves relatively unproblematic for the configurational model.

The third of the phonological criteria set out in §1 was that an adequate model of intonation should specify what constitutes a natural or possible rule of intonational phonology. At present, our knowledge of the sorts of rules that are found in the intonational components of grammars is very limited; in fact, as far as I am aware, the only full-scale treatment that exists for any language is my own analysis of French (1978a). Speculation regarding universal properties of intonation rules is therefore rather premature at present. Nevertheless, it is perhaps worth drawing attention to one rather striking aspect of the intonational phonology of French. The rules fall into two categories: those that assign intonation patterns to strings of syllables previously unspecified intonationally, and those that effect certain transformations of intonational specifications. Illustrations of the sorts of processes involved in the second category have already been given in this paper. The striking thing about these intonational transformations is that nearly all of them express processes of assimilation. In particular, we find that the intonation patterns over strings of unaccented syllables are assimilated to those associated with neighbouring accented syllables. A typical example of this is the process referred to twice already, and illustrated in (38), which can be represented roughly in the manner shown in (40). (For a more precise statement, cf. Crompton 1978a, Ch.2, §4.3).

```
(40)        [+A] .... [-A] .... [-A] [+A]
             |         +F        |  +R |
                            ↓
                       |   +R    |
```

In this case, the |+F| over the final part of the string of unaccented syllables has been assimilated to the |+R| associated with the following accent. Although, as we have said, it is probably still too early to try to construct a theory of language universals with respect to intonation, the fact that rules that conform to a rather restrictive pattern constitute a large part of the stock of transformations of a language such as French allows one to hope that given sufficiently detailed studies of other languages, the model we have proposed will ultimately lead to the development of significant restrictions regarding the formal properties of intonation rules, and thus prove adequate according to the fourth of the criteria we have proposed as well as according to the first three.

References

Ashby, MG, 1975: "Analysis and synthesis of English intonation", paper presented at the 8th International Congress of Phonetic Sciences

Bierwisch, M, 1966: "Regeln fur die Intonation deutscher Satze", *Studia Grammatica* 7: 99-201

Bolinger, DL, 1949: "Intonation and analysis", *Word* 5: 248-254

Bolinger, DL, 1951: "Intonation: Levels versus configurations", *Word* 7: 199-210

Bolinger, DL, 1958: "A theory of pitch accent in English", *Word* 14: 109-149

Collier, R, 1970: "The optimum position of prominence-lending pitch rises", *IPO Annual Progress Report* 5: 82-85

Crompton, AS, 1978a: *A Generative Phonology of French Intonation*; thesis submitted for degree of PhD, University of Cambridge

Crompton, AS, 1978b: "Intonational subordination in English", *Pragmatics Microfiche* 3: A2

Crystal, D, 1969: *Prosodic Systems and Intonation in English*, Cambridge University Press, Cambridge

Harris, ZS, 1944: "Simultaneous components in phonology", *Language* 20: 181-205

Jones, D, 1909: *Intonation curves*, Teubner, Leipzig & Berlin

Kingdon, R, 1958: *Groundwork of English Intonation*, Longman, London

Liberman, MY, 1975: *The Intonational System of English*, PhD dissertation, MIT, Cambridge, Mass.

Lieberman, P, 1965: "On the acoustic basis of the perception of intonation by linguists", *Word* 21: 40-54

Lieberman, P & Michaels, SB, 1962: "Some aspects of fundamental frequency and envelope amplitude as related to the emotional content of speech", *Journal of the Acoustical Society of America* 34: 922-927

O'Connor, JD & Arnold, GF, 1973: *Intonation of Colloquial English*, Longman, London

Pike, KL, 1945: *The Intonation of American English*, University of Michigan Press, Ann Arbor

Trager, GL & Smith, HL, 1951: *An Outline of English Structure*, American Council of Learned Societies, Washington

Wells, RS, 1945: "The pitch phonemes of English", *Language* 21: 27-39

THE 'COARTICULATION RESISTANCE' MODEL OF ARTICULATORY CONTROL: SOLID EVIDENCE FROM ENGLISH LIQUIDS?

F.J. Nolan
Department of Linguistics
University of Cambridge

1. Overview

Coarticulation is often seen as a key to the way the brain organises speech production. Recently it has been proposed by Bladon and Al-Bamerni (1976) that coarticulatory behaviour is best accounted for by an articulatory control principle known as "Coarticulation Resistance"; in section 5 this model is outlined and then further developed. Before that, data from an investigation of syllable-initial English liquids are discussed, and (section 4) a new method of quantifying coarticulation (as inferred from its acoustic reflex) is introduced. The data are compared throughout with those presented in Lehiste (1964) for General American English. Finally an attempt is made to relate synchronic coarticulatory behaviour, in the case of /r/, to historical phonetic change.

2. Experiment

The experiment was organised as follows. A word list was devised which included examples of words beginning with /l/ and /r/. The words chosen provided three examples of both /l/ and /r/ occurring word-initially before each of ten vowel phonemes - /i· ɩ e æ ʌ ɑ· ɒ ɔ· ɷ u·/. Monosyllabic words were chosen as far as possible, but in any case the syllable released by /l/ or /r/ always bore stress. In order to control phonetic environment as closely as possible, the following principles were observed: the syllable released by /l/ or /r/ should be closed by a plosive, or as second choice a fricative; the three examples of each syllable nucleus should be followed by consonants exemplifying the three broad place of articulation categories - labial, dental-alveolar, and velar; and those three consonants should exemplify both fortis and lenis types. Final nasals were rejected, since co-articulated nasality might produce marked changes in the vowel spectrum. Whenever a number of words might have fulfilled these conditions, nouns were chosen in preference to words of other grammatical categories. Phonologically permissible but non-occurring "nonsense" syllables were avoided because of the (orthographic) difficulties of eliciting their "correct" pronunciation from phonetically untrained subjects.

The subjects were 17-year-old males from a residential school; they were chosen because of the likelihood of their having a fairly homogeneous, "educated Southern British English" if not strictly RP, pronunciation. Fifteen subjects recorded the word list, reading each word twice - once preceded by the indefinite article. The recordings were made in a quiet room using a Tandberg Series 15 tape recorder and a Tandberg TM6 microphone.

The recordings were analysed using Cambridge University Linguistics Department's model 7029A Sonagraph together with plug-in unit 6075A Sona Counter. From the wide variety of analyses made available by the possible combinations of filter bandwidth, scale size, and sectioner, a standard format had to be decided on for the analysis. The format finally adopted, which seemed to offer the most useful combination of the merits of wide and narrow filters, involved a three-stage process. Firstly a wideband, scale magnified spectrogram was made of the item - a 5000 Hz range was chosen, as it would include at least the first four formants of /l/ and /r/ and as clearly defined formants would be unlikely above 5000 Hz; secondly a narrowband section on the same scale was taken at a point defined during the consonant of interest, and this was produced on a blank portion of the same piece of paper by sliding it round the drum; and thirdly a section of the same kind was taken at a point after the first one such as to fall centrally within the following vowel. The result was a single sheet for each item which contained a wideband spectrogram and two narrowband sections all to the same scale. Tests performed with pure tones showed that wideband and narrowband scales coincided exactly, so it was possible to measure formants using a combination of information from the time-frequency-intensity display and the two amplitude-frequency sections; the former gives the clearer indication of the general location of the formants, in part by showing their (usually high degree of) continuity through time, while the latter help in assessing the centre frequency of the formant by displaying the relative amplitude of the harmonics.

The Sonagraph's 500 Hz calibration tone was tested by comparing it with tones of known frequency, and found to be accurate to within approximately 2% - that is, a 20 Hz discrepancy at 1000 Hz. This finding was reconfirmed a number of times during the making of spectrograms. Consequently a 500 Hz calibration tone was included on all spectrograms made, and this enabled accurate alignment of the frequency scale of the spectrogram with the calibrated template which was constructed for making measurements, despite a tendency for the position of the scale with respect to the baseline to vary. The template was constructed on the basis of the calibration tone. Measurement of the first three formants of /l/ or /r/ and of its vowel environment was attempted, though in the discussion below the third formants of the vowels are omitted because of difficulties encountered in measuring them; in particular the low back vowels /ɒ/ and /ɔ·/ often presented a series of ill-defined peaks above the clear second formant. The actual values assigned to formants were estimates of their centre frequencies; if the harmonics dropped off in amplitude symmetrically on either side of the harmonic of highest amplitude in a particular region of the spectrum, then the frequency of this harmonic was taken as the centre frequency of the formant; in the majority of cases, however, peaks were asymmetrical, and in these cases the frequency chosen would lie between the harmonics of highest and second highest amplitude.

In the discussion of /r/ below data from 13 speakers is used; one speaker was left out because of an erratic tendency to use a tap, rather than the more usual approximant articulation of /r/; and another because of the high damping of his upper formants combined with a relatively high fundamental frequency made measurement of even the first three formants unreliable. In the case of /l/, the "third" formant presented is, following Lehiste (1964), the third spectral prominence detectable. The spectrum of a lateral usually contains an antiresonance in the 2-3 kHz region caused by the shunting cavity behind the central apical closure (Fant (1960:164)) which has the effect of reducing the level of or cancelling completely the third formant of the lateral (which is dependent on the mouth cavity in front of the apical constriction and will be centred on the 2.3 to 2.6 kHz region). In the majority of laterals in the present data the "real" third formant is more or less completely cancelled, and so the figures presented as the third formant (around 3kHz) derive from what is theoretically the fourth.

3. Results

Table I presents the mean frequencies over all speakers of the first three formants of /l/ in each vowel environment, together with the first two formants of the vowel environment; and then the equivalent information for /r/. The information from Table I is also displayed graphically in Figs. I and 2. These graphs have the ten vowel environments plotted along the x-axis in order round the vowel quadrilateral from close front through open to close back, so as to make the F_2 and F_1 trends optimally clear.

Table I

Overall means (all speakers) of formants of initial /l/ and /r/, and their vowel environments; below, the mean (across vowels) of the means, and for comparison equivalent figures from other sources (see text).

	lF_1	lF_2	lF_3	VF_1	VF_2	rF_1	rF_2	rF_3	VF_1	VF_2	
i·	340	1510	3010	310	2270	310	1160	1750	280	2120	
ɩ	350	1470	3040	460	1820	330	1140	1720	410	1670	
e	380	1410	3060	590	1710	340	1150	1710	550	1570	
æ	410	1300	3050	720	1440	340	1110	1670	630	1400	
ʌ	390	1290	3130	630	1230	330	1100	1670	580	1160	
ɑ·	370	1280	3080	620	1130	330	1090	1640	570	1080	
ɒ	370	1230	3030	570	990	320	1050	1630	520	970	
ɔ·	340	1290	3040	470	900	310	1000	1580	420	850	
ɷ	350	1320	3070	460	1200	320	1060	1660	400	1110	
u·	340	1410	2980	340	1500	300	1090	1640	320	1360	
MEAN	360	1350	3050	520	1420	320	1090	1670	468	1330	Nolan
MEAN	295	980	2600	580	1390	280	930	1360	380	1200	Lehiste
MEAN	365	1305	2780								Al-Bamerni

The first row of means in Table I are means of the formant frequencies of the present study, shown in the same table; the second row are for values of initial /l/ and initial /r/ given by Lehiste (1964: Tables 2-III, 3-III and 3-XII), with only the nine of her vowel environments with equivalents among the present study's ten included in the calculation (Lehiste's /i ɪ ɛ æ ə ɑ ɔ ʊ u/ equivalent to /i· ɩ e æ ʌ ɑ· ɒ ɔ· ɷ u·/); and the third row (/l/ only) are from Al-Bamerni (1975) and relate to the present ten British English vowel phonemes plus /ə·/. The present mean lF_1 (360 Hz) agrees well with that of Al-Bamerni (365 Hz) and also with Fant's (1960:164) measured

F_1 of Russian non-palatalised /l/ (350 Hz). Lehiste's lower value (295 Hz) prompts no obvious explanation, though in theory (see Fant (1960:164)) it should indicate a smaller cross sectional area of the lateral passage. It is difficult to guess without further (e.g. radiographic) evidence whether this might be a general trend in American English. lF_2 values from the two studies of British laterals agree quite well (1350 and 1305 Hz), and the generally held view (e.g. Jones (1972:359)) that American /l/ tends to be realised fairly dark (velarised/pharyngealised) even where British English has a clear allophone is borne out by the considerably lower mean lF_2 of Lehiste. The higher lF_3 (3050 Hz) of the present study as compared with Lehiste's (2600 Hz) and Al-Bamerni's (2780 Hz) is perhaps caused in part by the problems of locating formant peaks in what in many cases is a rather flat spectrum above 2500 Hz. Nevertheless Fant (1960:164-7) predicts an F_4 (equivalent to "F_3" here and in Lehiste) of 2900 Hz for Russian non-palatalised /l/ and of 3050 Hz for palatalised /l/ so that 3050 Hz may not be an unreasonable figure for the first peak above the major antiresonance.

Lehiste's mean values for the formants of initial /r/ are all lower than those of the present study - rF_3 considerably so (1360 versus 1670 Hz). This would seem to point to a generally "weaker" articulation of /r/ in British English, in particular with a less extreme degree of stricture (higher F_1) and less retroflexion (higher F_3). The implications of this are discussed in section 6 below.

Figs. 1 and 2 show clearly the relation between the formant frequencies of the vowel environments and the consonants preceding them. The clearest case of coarticulation is lF_2, which in almost every environment echoes the variation of VF_2. The range of variation in lF_2 is smaller than that of the equivalent initial allophone of Bladon and Al-Bamerni (1976: fig.2) - approximately 300 Hz as opposed to approximately 500 Hz - but this might be due to the preceding [ə] in the context used in the present study (the formant measurements having been made on the items preceded by the definite article, as this allowed for better visual tracking of the formants); and to the inclusion in the means of certain speakers who exhibited very little coarticulation in /l/. lF_1 can also be seen to vary with F_1 of the vowel environment, within a range of 70 Hz. For /r/, the range of variation of F_2 and F_1 is smaller than for /l/, but scrutiny of the values in Table I confirms that within the limited range the variation is generally in the same direction as that of the equivalent formant of the vowel environment. Also shown are the high mean F_3 of /l/ and the low mean F_3 of /r/, this latter showing some tendency to follow rF_2.

Fig. 1 Formant frequencies of /l/ and following vowel (mean of 15 speakers)

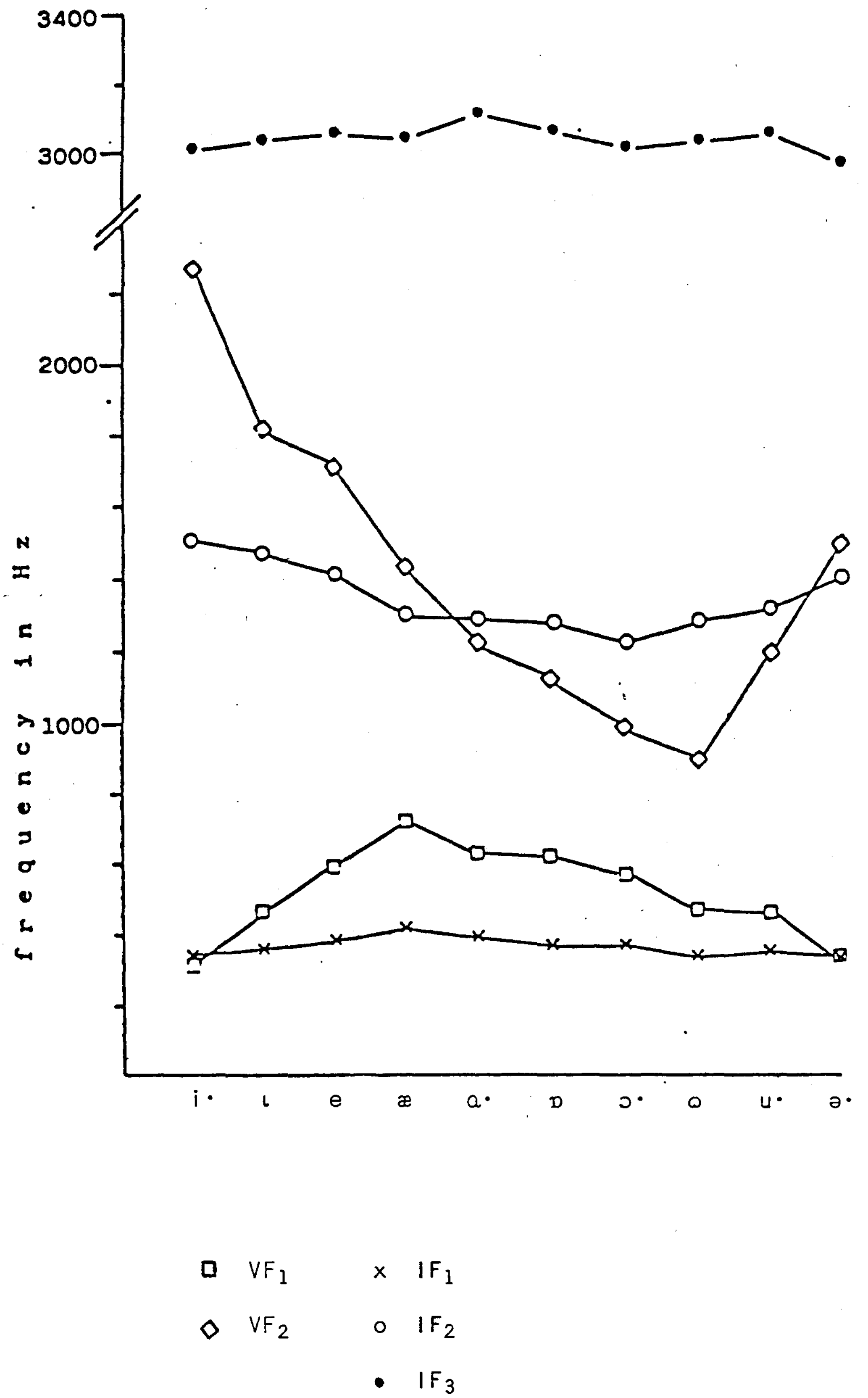
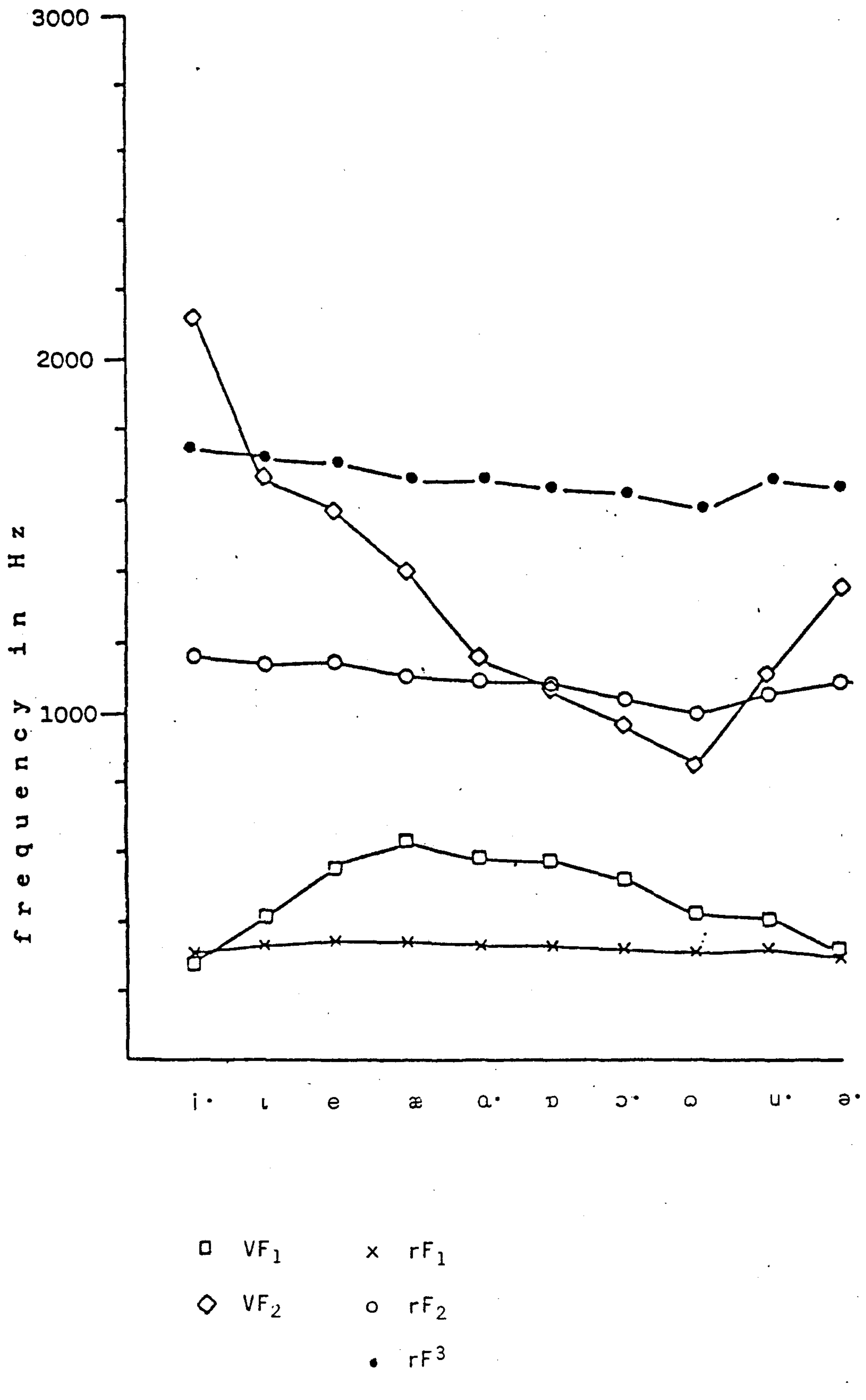


Fig.2 Formant frequencies of /r/ and following vowel (mean of 13 speakers)



4. Quantifying Degree of Coarticulation

So far only subjective statements, on the basis of the formant plots and the table of formant values, have been made regarding the degree of coarticulation of the two consonants. Clearly some quantitative expression of the degree of coarticulation is desirable, both for the present purpose of comparing the two consonants and for subsequent work aimed at discovering differences in the coarticulatory strategies of individual speakers.

A method considered was that of calculating a canonical formant value for /l/ or /r/, and finding the mean of the deviations from that value in each vowel environment. This is the method of Bladon and Al-Bamerni (1976:142):

"To quantify the degree of coarticulation in a given context, an appropriate measure would seem to be the mean deviation in F_2 values over the set of vowel environments. Specifically, we determined for a chosen context condition ... [such as syllable initial position] ... the mean value of F_2 over the set of vowel environments ... and called this for the context in question the canonical F_2 value. The mean distance in mels by which the [F-values of /l/ in the vowel environments] exceeded or fell short of this canonical value was called the degree of coarticulation or mean coarticulatory distance."

In practice, this measure provides for intuitively appealing comparisons of the degree of coarticulation of different extrinsic allophones of /l/ in their study. But it is prey to theoretical objection, since it is based purely on variation in F-values of /l/ (cf. the lF_2 line of Fig.1) and takes no account of whether this variation is related to variation in the equivalent formant of the vowel environment. For instance, a high value in mels (or Hz) of their "degree of coarticulation" measure for F_2 of /l/ could indicate an extreme case of lF_2 assimilating to that of the following vowel; but equally it could, in theory, indicate an extreme case of dissimilation in F2, or indeed an extreme case of random variation in F_2 of /l/.

This possible shortcoming in the above method for measuring degree of coarticulation in formants highlights the nature of the problem which is basically that of the statistician needing to know the strength of the relationship between two variables - in the present instance, the variables being for example lF_2 and VF_2. The strength of such a relationship can be summarised in the single value of one of the widely used correlation coefficients, such as Pearson's product-moment coefficient of correlation or Spearman's coefficient of rank correlation; Nie et al. (1975:276):

"Bivariate correlation provides a single number which summarizes the relationship between two variables. These correlation coefficients indicate the degree to which variation (or change) in one variable is related to variation (or change) in another. A correlation coefficient not only summarizes the strength of association between a pair of variables, but it also provides an easy means for comparing the strength of the relationship between one pair of variables and a different pair."

The value of either of these coefficients lies between +1 (perfect positive correlation) and -1 (perfect negative correlation), with 0 indicating absence of linear correlation. In calculating Spearman's rank correlation coefficient (r_s) the value of each case with respect to the variable in question is replaced by the rank number of that case within the variable; it is a non-parametric statistic, and does not assume that the variables in question have normal distributions (Nie et al. (1975:276)). It was decided to use Spearman's coefficient to measure the degree of correlation between /l/ and /r/ formants and those of the following vowel because the values of a given formant over the list of 30 test items are not likely to be normally distributed. (Taking VF_2 as a strong example, the peak of the distribution will be near the lower end of its range at around 1000-1200 Hz, since a number of vowel phonemes (/ɑ· ʌ ɷ, possibly u· ɒ/) have F_2 in this region, whilst the middle, and upper end of the range will apply to rather few vowels - /i· ɩ/ and perhaps /e/. Computation confirmed F_2 to have this "positive skewness".)

Using the SPSS NONPAR CORR subprogram, r_s was computed to evaluate the strength of the correlation of F_1 and F_2 of /l/ and /r/ with the corresponding formants of the vowel environment. The values of r_s for the complete corpus are shown in Table 2. It can be seen that the highest correlation is that of lF_2 with F_2 of the vowel environment ($r_s = 0.575$), and the second highest is that of lF_1 with F_1 of the vowel environment ($r_s = 0.533$). A similar pattern emerges with /r/: the correlation of rF_2 with F_2 of the vowel environment ($r_s = 0.405$) being the strongest, followed by rF with F_1 of the vowel environment ($r_s = 0.305$).

<u>Table 2</u>

Spearman coefficients of correlation between (across) formants of /l/ or /r/ and (down) of the following vowel, computed for data from 15 (/l/) or 13 (/r/) speakers.

	lF_1	lF_2
VF_1	0.533	-0.352
VF_2	0.032	0.575

	rF_1	rF_2
VF_1	0.305	0.036
VF_2	0.101	0.405

The other values in Table 2 are for cases where little correlation would be predicted, for instance of $1F_1$ with VF_2 (r_s= 0.032). But the moderate negative correlation of $1F_2$ with VF_1 (r_s = -0.352) warns against uncritical equation of "positive correlation" with "assimilation" or of "negative correlation" with "dissimilation" in the phonetic sense. This moderate negative correlation is arguably the indirect joint result of the phonetic quality of the vowels in the corpus and the relatively strong ASsimilation of /l/ to the following vowel. Suppose for a moment there existed a highly regular assimilation of /l/ and the following vowel. Crudely one might expect F_2 of /l/ to fall as F_1 of /V/ rises through vowels /i· ɩ e æ ʌ (ɑ·)/ (negative correlation), but for F_2 of /l/ to continue falling as F_1 of /V/ falls through vowels /(ɑ·) ɒ ɔ· ɷ u·/ (positive correlation), thus more or less cancelling out the negative correlation. (See Ladefoged (1971:73 Fig.11) for a schematic diagram of vowel formants conforming to the described pattern). But remember that in the present data (see Figs. 1 and 2) /ɷ/ and /u·/ are generally realised by centralised and somewhat unrounded close vowels, with F_2 correspondingly rather higher than in earlier published data on English (/ɷ/ around 1200 Hz, /u·/ around 1400 Hz); the result is that the second (positive) half of the correlation breaks down, leaving the negative half to predominate, and yielding a high negative r_s for $1F_2$ with VF_1. Now whilst in reality the assimilation of /l/ with the vowel environment is not perfectly reliable, it is suggested that the tendency may be strong enough to account in this manner for the moderate negative correlation of $1F_2$ with VF_1.

Given phonetically critical interpretation, however, the correlation coefficient r_s of the variation in a formant of one sound with the variation in the corresponding formant of an adjacent phonetic segment provides, it is claimed here, a valid measure of the (acoustic) assimilation of one to the other. The validity is supported by evidence from a second analysis of the data involving linear prediction, and a method of estimating coarticulation suggested by Su et al. (1974). For each speaker, smooth spectra of each example of /l/ and /r/ were estimated by linear prediction, and then (using a somewhat arbitrary "front-back" division whereby the first five vowels along the x-axis of Fig.1 were classified as "front") a mean /l/ or /r/ spectrum in "front", and in "back" environments computed. The difference between the mean front and mean back consonant spectrum of a speaker was taken as an estimate of his coarticulation in that consonant. In general, close agreement was found between the formant correlation method, applied now to individual speakers, and the spectral distance method, when used to rank speakers from highest to lowest degree of coarticulation. Figs. 3 and 4 show such a ranking for the speakers for whom complete analyses were possible; two rankings using the spectral difference method are presented, SA relating to the same recording session as the formant correlation rankings were calculated from, and SB to a subsequent recording of the same words. With the curious exception of PVE, the rankings by /l/ coarticulation show consistent trends (for example with GRP clearly favouring coarticulation, and JRC exhibiting rather little); the rankings by /r/ coarticulation are less consistent, but this may be because of the altogether smaller amount of coarticulation to which /r/ is susceptible (a finding confirmed, incidentally, by the spectral difference measure).

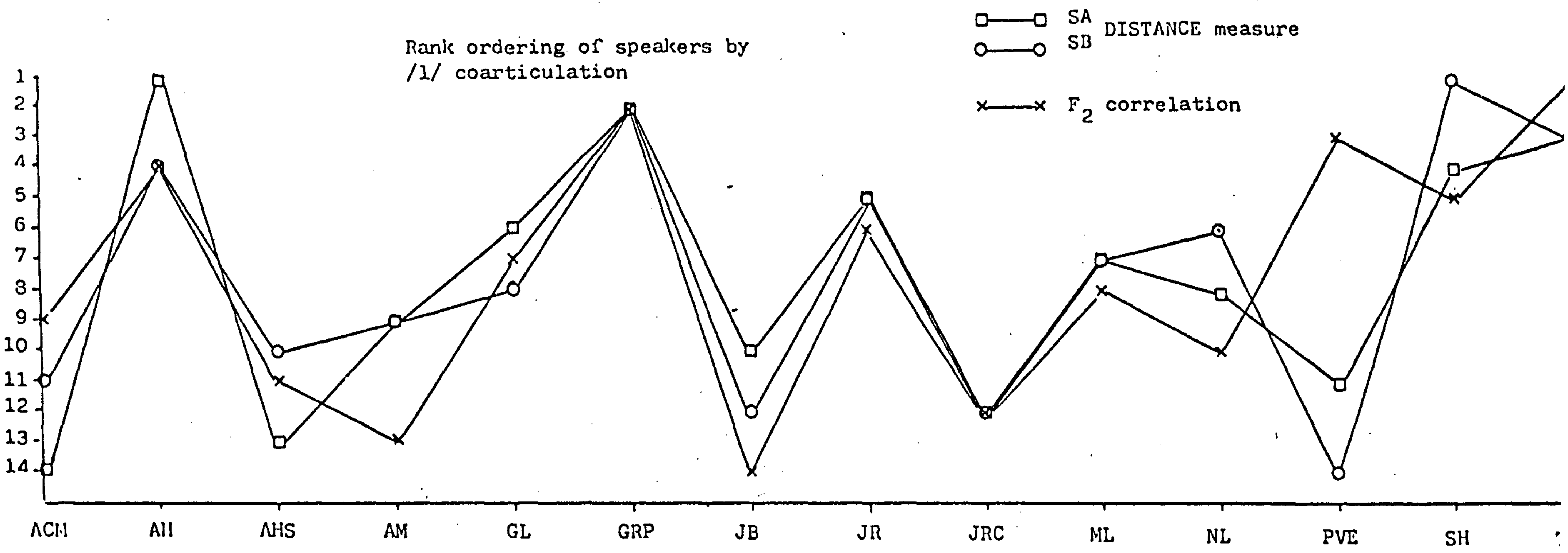

Rank ordering of speakers by
/l/ coarticulation
SA
SB
DISTANCE measure
F_2 correlation
1
2
3
4
5
6
7
8
9
10
11
12
13
14
ACM
AH
AHS
AM
GL
GRP
JB
JR
JRC
ML
NL
PVE
SH

Fig. 3

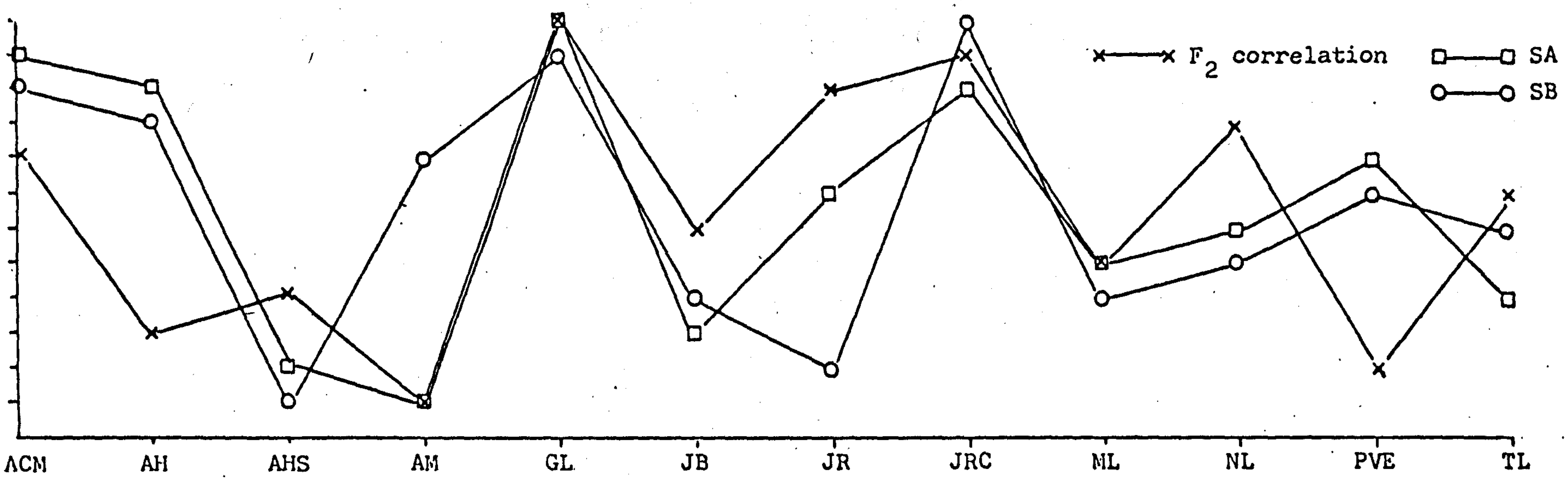


Fig. 4
Rank ordering
of speakers
by /r/
coarticulation

Fig. 4

The finding that /l/ assimilates more than /r/ to a following vowel is in accord with Lehiste's (1964) conclusions on her American English data. But her statement that

"The data suggest that the following vowel has but little influence on the initial allophone of /r/."

is not so applicable to the present data, since moderate correlation between the two main quality-determining formants of /r/ and those of the following vowel has been found; further, her rF_2 range across the nine relevant environments (see section 3 above) is only 110 Hz compared to 160 Hz in the present study, and her highest (990 Hz) and lowest (880 Hz) rF_2 values (happen) to be in the environment of /ə/ and /ɩ/ respectively - just about the reverse of the "coarticulatory" pattern exhibited in Table I and Fig.2 above. Although not absolutely clear cut, the evidence points strongly to an allophone of /r/ in initial position which coarticulates more readily in British than American English.

5. Coarticulation Resistance

The concept of Coarticulation Resistance was introduced by Bladon and Al-Bamerni (1976) to account for their finding that the three extrinsic allophoneds of /l/ coarticulated with adjacent vocoids to different degrees - [l] more than [ɫ], and syllabic [ɫ̩] least of all. They explain the concept as follows (1976:149):

"Coarticulation on RP /l/ takes place freely from either direction, but the direction itself is unimportant. Antagonistic vocal tract adjustments apart, coarticulation is inhibited only by coarticulation resistance (CR) at some point in the succession of speech events. Each extrinsic allophone (and indeed each boundary condition) is assigned a value for CR by rules which in some instances may be language particular and in others quasi-universal. The CR value would be represented as a numerical coefficient attaching to a phonetic feature, say [3 CR], along the lines proposed by Chomsky and Halle (1968) for all other phonetic specifications in the phonological system."

Later, mentioning the work on speaker identification using nasal coarticulation by Su et al. (1974) they add that

"there seem to be cases where CR is assigned on a highly idiolectal basis."

The CR index of a segment will be subject to constraints which comprise physiological, aerodynamic and acoustic factors. Bladon and Nolan (1977:193) in a study using X-ray techniques claimed that

"the specification for coarticulation resistance which attaches to English /s/ and /z/ must, at least in the contexts studied here, be high - and higher than that of the other English alveolars."

Speakers using a laminal articulation for /s/ and /z/ in vocalic contexts but an apical articulation for the other alveolars were round to maintain their fricative articulation unchanged and adjust /t d l n/ towards more laminal articulations when the latter sounds occurred in clusters with /s z/. In this case the high CR index may be the result of the rather precise articulation required to produce the correct aerodynamic conditions for these "groove" fricatives (c.f. Amerman and Daniloff (1977:111)). From informal investigation it emerges that people who habitually use a laminal /s z/ articulation need practice in producing it with the apex if it is not to sound "lispy". The production of an auditorily adequate /t d n/ on the other hand is less demanding, requiring only a complete closure of the vocal tract in the region of the alveolar ridge, so that if coarticulation in a cluster containing /s/ or /z/ and a non-fricative alveolar is to take place, it will be the less resistant non-fricative that will adjust. (Although /l/ would appear to involve a more complex articulation than the stops, it was found to coarticulate in a way similar to /t d n/.) Here then is a case where a high CR index reflects strong aerodynamic-acoustic constraints on articulation. A high resistance to coarticulation would also be predicted for other sounds involving finely balanced aerodynamic effects, such as trills.

In addition, the CR index of a particular segment will also reflect constraints placed on that segment by the phonological system of the particular language. In a language such as Russian, for example, which contrasts a palatalised series of consonants with a velarised series, these segments would have to have relatively high CR indices, as coarticulation with front and back vowels could eliminate their phonetic distinctness. Possibly in Russian the high CR index of the consonants causes the articulatory control mechanism, in its efforts to smooth the transitions between segments, to take advantage of a (hypothetically) low index on the vowels, with the result that the vowels exhibit more coarticulatory variation than in a language such as English, where the consonants are free to coarticulate. A similar situation is evidence by Irish, which like Russian has a primary division of consonants into a palatalised a velarised set (the resonants being further subdivided by a "lenited-nonlenited" opposition, which provides Irish with four (or in dialects where the lenited phonemes have coalesced, three) lateral phonemes. Ní Chasaide (1977:44) found that

"... a rather striking lack of qualitative coarticulation [of the laterals] was observed ... the picture in Donegal Irish would seem to be of a rather stable lateral quality, whereas ... the vowel quality may vary considerably according to the adjacent consonant."

She goes on to note that whilst the communicative function of the English lateral is not impaired by its extrinsic allophonic and coarticulatory variation in quality

"In Irish, the fact that distinctions are made on relatively small qualitative differences ... necessitates a more compact range for each lateral if intelligibility is not to be impaired."

The exact status of the concept of Coarticulation Resistance is still in question; at first sight it may appear to be simply a terminological device for summarising statements about the degree of coarticulation occurring in different allophones; and this is similar to the view taken by Kent and Minifie (1977:120):

"the CR values do not in themselves generate predictions, but rather appear to be part of a numerical description of speech articulation."

It is however less than obvious what sort of "predictions" the values themselves could be expected to generate, for it is not normal to speak of the values of other phonetic features ([3 nasal] or whatever) "generating predictions". Rather these also are part of a numerical description of speech articulation; and neither kind is successfully dismissed by labelling it "descriptive".

The prediction which is in fact made by the hypothesis of Coarticulation Resistance is that coarticulation is limited on a segmental basis and is not determined by principles dependent on a higher articulatory unit such as Kozhevnikov and Chistovich's (1965) "articulatory syllable". The claim will be falsified if and when it can be demonstrated that coarticulatory effects are consistently limited by a particular suprasegmental unit of production; the evidence summarised by Kent and Minifie (1977: 118-9) shows that at present this goal is receding rather than coming nearer, the number of candidate articulatory units ($(C)^nV$; VC; VVC; V_1CV_2) apparently growing each time a new coarticulated property (lip-protrusion, nasality, etc.) is investigated. In this light it does not seem unreasonable to put forward and elaborate an alternative hypothesis which puts greater emphasis on the variety of segment-based factors limiting coarticulation.

A clear instance of the need for such a segmental emphasis is the finding by Bladon and Al-Bamerni (1976) that greater coarticulation took place between /l/ and the preceding vowel in the clear [l] of [Vlɩ#] sequences than in the dark [ɫ] of [Vɫ#] sequences (1976: Fig.4, contexts B and MI) - it cannot be the /ɩ/ which causes the greater coarticulation, since it functions as an invariant context to condition the occurrence of the clear allophone and remains constant for the various /V/ environments. A theory of coarticulation which does not allow different segments to have, or be assigned, different coarticulatory susceptibilities would presumably have to claim that the greater degree of coarticulation occurs because the following /ɩ/ somehow causes greater cohesion between /V/ and /l/. Such a solution does not seem attractive.

As a second example, the tendency of open vowels to nasalise more readily than others adjacent to nasals may be due to the low tongue dragging down the velum by means of the connecting palatoglossus muscle (cf. Hardcastle (1976:124)), in which case it might be classed as "unplanned" coarticulation and so require no statement at the level of articulatory planning; but Ohala (eg. 1974) has suggested this to be disconfirmed by electromyographic evidence, and that a more likely explanation lies in the lesser perturbation caused by nasalisation in the acoustic spectrum of open as opposed to non-open vowels. The speaker is as it were constrained by the more serious acoustic/perceptual consequences to limit, at the planning stage, the coarticulated nasality in non-open vowels - or in other words to assign these vowels a higher coarticulation resistance index.

The question arises of the nature of the mechanism by which the CR index might be assigned. Is it a feature present in the systematic phonemic representation of each segment, which then takes on particular numerical values through the operation of rules of the phonological component, as [+nasal] or [+round] might? This view is rejected, because unlike other (binary, phonological) features [CR] serves no classificatory function - and indeed could not unless certain feature complexes were to contain a specification [-CR], of which the implications are distinctly odd: in the view put forward here, a complete lack of coarticulation resistance must lead to the total assimilation of, and hence non-existence of a segment (though conceivably [CR] might replace the feature [segment] as a means of distinguishing between segments and boundaries). In addition, as an "articulatory control principle" coarticulation resistance does not belong, theoretically, in the abstract phonological system but rather on the "performance" side of the elusive and questionable boundary between performance and competence.

It is therefore proposed that CR indices are assigned to segments at the phonetic level, which is in accord with the findings of Bladon and Al-Bamerni (1976) that there need be no similarity in the coarticulatory behaviour of different allophones. Furthermore coarticulation resistance will not be an additional feature added on to the feature complex of a phonetic segment, but will instead take the form of indices relating to subsets of those features. That this is necessary can be seen from the fact that a given segment may be resistant to coarticulation in one respect, yet free to coarticulate in another. [ɫ], whilst resisting tongue-body coarticulation with preceding vowels, may nevertheless change its tongue-tip articulation from alveolar to dental under the influence of a following dental, or to post-alveolar under the influence of a following post-alveolar - eg. [hɛɫ̪θ] (health), [wɛɫ̠ɹɛd] (well-read) (cf. Gimson (1970:203)). In this case a high CR index would be assigned to "tongue body features" (cf. Chomsky and Halle (1968:299)), and a relatively lower one to those features applying to the detailed positioning of the tip of the tongue. Discussing alveolar stops and fricatives, Amerman and Daniloff (1977:111) suggest that with the exception of the region of constriction, tongue shape may not be any more precisely specified for the fricatives than for the stops - the vowel positions in see-saw might well be reflected in the tongue body position during the preceding /s/. Here tongue body and tip CR values for /s/ are presumably the reverse of those for the dark allophone of /l/.

Fig. 5
CR Assignment

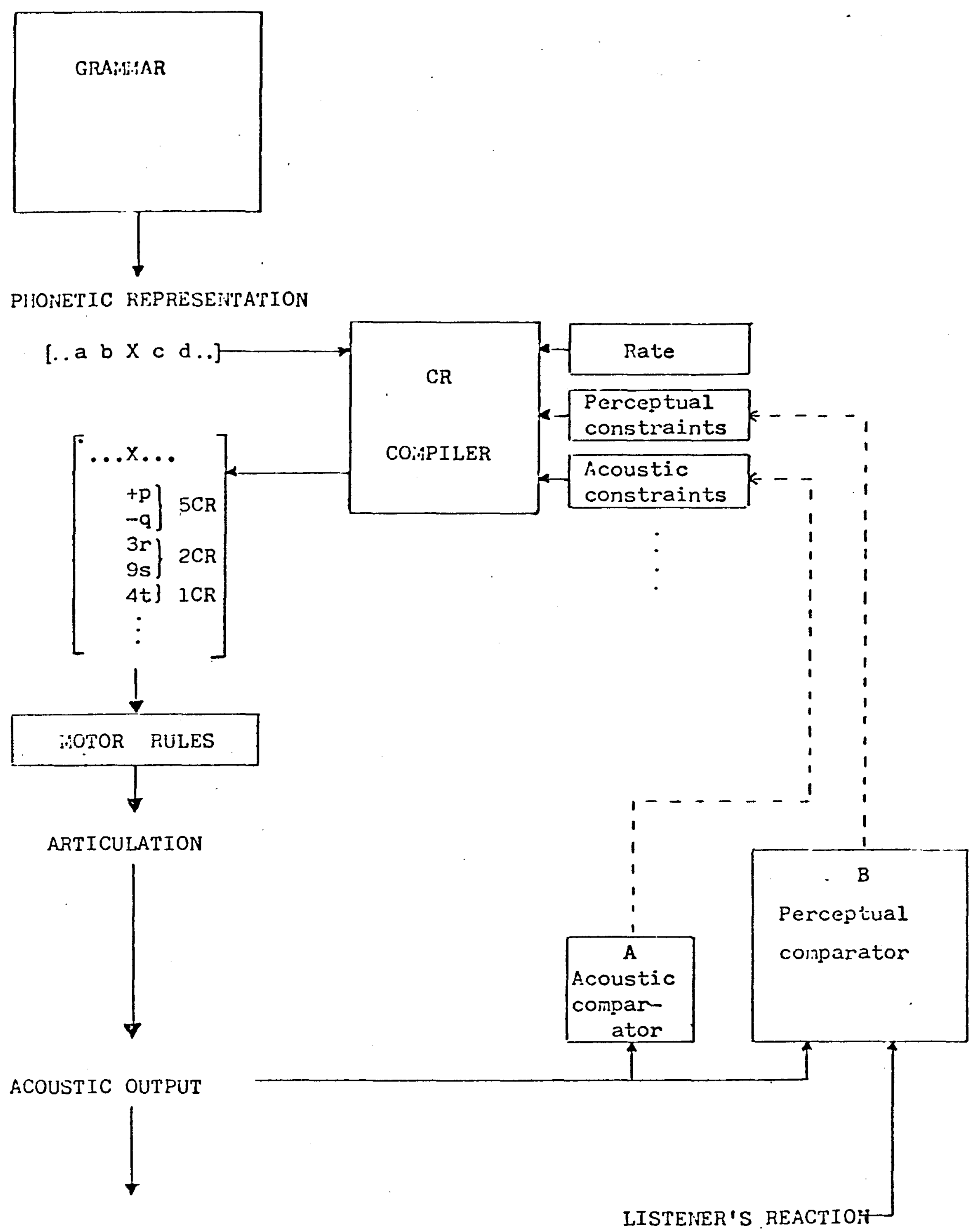


Fig.5 presents a suggested mechanism for the assignment of the CR indices to a segment. Starting at the top left of the diagram, it can be seen that the mechanism, although concerned with speech production, assumes the incorporation of some kind of competence model, or grammar. The exact nature of this part of the model is not important here, provided that it yields as output phonetic representations in the form of feature matrices, the features being in many cases multivalued, and the representations being such as to account for extrinsic allophones, but not to contain information on those (intrinsic) allophones resulting from coarticulation; this view of the interface between the phonological component and the strategies of speech production is in accord with, for instance, that of Fant (1971:221-2):

"Before we can accomplish the happy marriage between phonology and phonetics we have to work out the rules for predicting the speech event given the output of the phonological component of the grammar ... The starting point is the feature matrix of a message as successive phonological segments, ie. columns, each with its specific bundle of features, ie. rows. The particular choice of classificatory features at this stage is not very important providing the conventions relating phonemes to alternative feature systems are known.

The derivation of the rules of this 'phonetic component' of language aims at describing the speech production, speech wave, or perception correlates of each feature given the 'context' in a very general sense of co-occurring features within the phonological segment as well as those of following and preceding segments. One set of sequential constraints are expressible as coarticulation rules which may be both universal and language specific."

A copy of the feature matrix for the segment X together with the segments and boundaries included in its coarticulatory environment (which may turn out to be as extensive a unit as the phonological phrase in which the segment in question occurs) is read into the CR compiler. This contains information on that segment - including language-specific information of an arbitrary kind (such as the high resistance of [ɫ] in English); information about the idiosyncratic coarticulatory behaviour of the individual speaker; and it also contains as an operating principle the goal of assigning the minimum value for each CR index, in keeping with the presumed tendency of the speech control mechanism to minimise articulatory effort. Such a principle of least articulatory effort is not universally accepted by phonologists, and has for example been explicitly rejected by Jakobson (1971:8), though he rejects it as an explanation of the development of a child's phonological system, and may therefore not be denying its existence among the strategies of speech production. Lindblom (1972) on the other hand incorporates the principle of "least effort", together with the principle of "maximal perceptual contrast", into a model which accounts for preferred vowel systems in child acquisition of speech, and by extension in the phonological systems of mature speakers (1972:69):

"It does not seem controversial to suggest that contrasts and structures that the child discovers, produces, and uses easily and at an early stage, should stand a better chance of 'slipping through' and remaining in the language."

Lindblom suggests that children acquire the syllables [pæ bæ mæ] first, and that this is because these sounds are the ones which will result from a simple up and down jaw movement combined with minimum "deformation", and hence minimum physiological effort, of the tongue and lips. The principle of least effort applies similarly to mature speakers, who are constrained to produce the complete range of distinctions required by the phonological system of their language, but to do so in the most economical way; for instance Lindblom (1972:77-9) and Lindblom and Sundberg (1971) show that compensatory articulations permit a speaker to produce an [i] with lowered jaw, but that this involves a "supershape" of the tongue - the tongue is deformed considerably from its neutral position, and there is considerable antagonism between the tongue muscles (eg. the genioglossus) and the depressor muscles of the jaw. The effort involved rules out such a configuration from being preferred in normal speech production. Now whilst Lindblom's work deals with static postures, it is reasonable to extend the principle of least effort to the dynamic behaviour of the vocal organs, and to assume that gross transitions between antagonistic configurations, or trajectories, will be avoided if possible. It is this general preference for reducing dynamic articulatory effort which underlies the specific principle of minimising CR indices by which the operation of the CR compiler is guided.

In its computation of CR indices the compiler also draws on information external to itself, represented by the (no doubt incomplete) set of inputs to its right on Fig.5.

The compiler will certainly require information on the rate of the utterance in order to make its assessments of the CR indices. If the articulators are to fulfil the demands made on them by an increase in speaking rate, an increasing compromise will have to be made between the full phonetic specification of the utterance and the time available to realise that specification. It will be up to the CR compiler to loosen the coarticulation resistance of the utterance sufficiently for the articulators to be physically capable of performing it in the time available; and to compute this relaxation of the CR indices in a way minimally damaging to the intelligibility of the utterance.

The third input, labelled "acoustic constraints", forwards values to the CR computation which try to ensure that the articulation performed for a given segment results in an adequate acoustic realisation of the feature complex of that segment. This assumes that at least some features will be in acoustic or auditory terms, or at least have readily derivable acoustic/auditory interpretations. For example [s] might be specified [+continuant, 4 strident]. The high degree of stridency involved in such a specification puts a strong requirement for precision on the configuration of the forward part of the tongue and it must therefore be prevented from adjusting to any great degree to the configuration of the adjacent segments (to the extent that these are not strident; [naɩʃʃɒp] nice shop is acceptable whilst [naɩθθɩmbɫ] nice thimble is not) - hence the "acoustic constraints" component will specify a high coarticulation resistance index for the part of the tongue involved in a strident articulation.

The second input is more problematical, but constitutes the principles which try to ensure that the phonetic distinctions specified in the phonetic representation are preserved in the resultant speech. The main perceptual constraint states that, at a given rate and in the same coarticulatory environment, two segments may not have perceptually indistinguishable realisations unless these segments are identically specified in the phonetic representation. This "perceptual distinguishability" constraint is a direct reflection of the phonological system's limiting influence on coarticulation. For example in Irish, if tongue body coarticulation were to threaten the perceptual distinguishability of the realisations of (nonlenited) /l/ and /ɫ/, the perceptual constraints values forwarded to the CR computation would be increased so as to restore the CR indices of these segments to a point where their distinguishability was maintained. This would be prompted, as discussed below, by some kind of internal or external feedback - and would of course be subject to a variety of pressures capable of inhibiting the corrective loop, as happens in the case of language change where two phonemes, or extrinsic allophones, coalesce.

On the basis of the available information the CR compiler then computes the CR indices for the segment in question and returns them to the phonetic representation, which forms the input to articulatory control rules, the eventual output of which is the speech movements of the vocal tract. The articulatory movements define the acoustic output, which is monitored (neglecting for the present monitoring of earlier stages through kinaesthetic or tactile feedback, or of the motor commands themselves) by components A and B.

The first, A, compares the acoustic effects produced (or strictly, the auditory impressions thereof) with reference samples (either stored or generated) which indicate the expected relationship between particular combinations of feature values and acoustic output. If the output falls outside the tolerance limits associated with the reference sample - for example if [s] lacks aperiodic energy above 4 kHz - a command is sent to the "acoustic constraints" component so that its contribution towards the appropriate CR index will be increased. This correction might become necessary in cases where a fast rate of speech production had caused an over-reduction of the CR indices.

The perceptual comparator, B, has two inputs. The first is the acoustic feedback from the speaker's own speech, which, from the fact that speakers detect and correct their own errors, may be presumed to undergo monitoring (at various levels of analysis) by a continuous perception process. If this monitoring process detects a discrepancy between intended and produced utterance, then (depending on the seriousness of the discrepancy) a corrective message is forwarded to the appropriate level. Fig.5 shows only the output relevant to coarticulation resistance. If the realisation of a particular phonetic feature complex is found to be perceptually indistinguishable from the realisation of another feature complex - for example if the realisation of Irish /l/ before certain vowels is no longer distinguishable from that of /ɫ/ - the perceptual component may update the perceptual constraints value in an attempt to prevent coarticulatorily induced loss of distinguishability in future utterances. The likelihood of this updating taking place will be affected by the second input to the perceptual comparator.

This second input is the reaction of any listener(s) present. In the extreme case - when unintelligibility results - it will almost certainly trigger the corrective procedure into action, the discrepancy between (in the present instance) the linguistically specified, "intended" phonetic characteristics and the actual acoustic output being "serious" enough. This input may also, if the listener gives an indication (most straightforwardly by repeating the context of the word he has difficulty with) of the location of the communicative failure, help focus the corrective action on the part of the utterance where the discrepancy is most damaging.

If no corrective updating takes place, then the CR indices which have led to the lack of distinguishability in the present utterance will be allowed to be computed for future utterance tokens - perhaps just at a fast rate, or perhaps spreading to all rates and registers; and if the latter, then the phonetic representation (for new learners at least) will no longer maintain the distinction between the two phones, necessitating a change in the phonological rules. Note that it is not "phonological neutralisation" which is being discussed here - the speech production mechanism is considered here to be oblivious of the possibility that identical phonetic segments may be the result of mapping by the phonological rules of different phonemes, and takes into account only the phonetic representation. Nevertheless consistent realisation of distinct phonetic segments derived from separate phonemes as indistinguishable outputs will, if left uncorrected, lead to partial or total neutralisation of those phonemes; the evidence for distinct phonetic segments available to the language learner will have been lost, and he will structure his phonological rules on the assumption that the phonetic representation is not to distinguish the two segments. This process could lead to the loss from the language of the opposition of two phonemes; but equally to the loss of the distinctness of two extrinsic allophones of a phoneme.

6. Coarticulation and the loss of postvocalic /r/

The above outline of the concept of coarticulation resistance, though undoubtedly wanting in many respects, provides a framework within which to examine the finding that the data in the present study show initial /r/ to coarticulate more freely in Southern British English (SBE) than in the American English of Lehiste (1964). If this does represent a valid generalisation to the effect that initial /r/ in SBE coarticulates more than its American counterpart, it reveals an interesting parallel with the behaviour of postvocalic /r/.

SBE, of course, now lacks any post-alveolar final allophone of /r/; the interesting point is that American /r/ also seems "weaker" in the sense of "less resistant to coarticulation" in syllable final position; Lehiste (1964:61):

"The final allophones, on the contrary, appear to depend to a considerable extent upon the vowels associated with them."

Furthermore, the average F_1 values given by Lehiste (1964: Tables 3-IV and 3-V) for initial and final /r/ are 280 Hz and 455 Hz respectively. The higher F_1 indicates a lesser degree of stricture in the vocal tract (cf. discussion by Fant (1960:168) on the changing F_1 of a trilled [r] in its different phases) and therefore, arguably, an articulation more assimilated to the adjacent vocalic, strictureless articulation.

On the basis of the evidence, the following values of coarticulation resistance are posited for tongue blade and tongue body at the time of a hypothetical divergence of the two in their /r/ behaviour:

		American English		Southern British English	
		blade	body	blade	body
/r/	initial	6	6	4	4
	final	2	4	0	2

This may be interpreted as follows. American English initial /r/ has high CR of both tongue body, which prevents the resonance of the sound assimilating to that of the following vowel, and of tongue blade, which prevents the blade assimilating to the following vocalic articulation and so ensures a close stricutre and low F_1. At some stage in the development of SBE (probably at the end of the eighteenth century, cf. Gimson (1970:210)) factors conspired in this dialect to lower the CR indices of the initial allophone. The tongue body CR index was then low enough to allow some coarticulation with the resonances of the following vowels, and this is reflected in the above findings on present day SBE that initial /r/ does coarticulate with following vowels.

Under pressure from the same (putative) factors, the CR index of the tongue blade (here taken to include all the foremost region of the tongue, including tip) was lowered. Whilst not yet low enough to allow the post-alveolar constriction to disappear (by assimilation with the following vocalic segment), the lower CR makes it less likely that the tongue blade will attain as extreme a deformation from its normal non-retroflex or non-bunched shape. This trend would cause in particular a change in F_2 frequency, since it is probable that increasing retroflexion, from slight retroflexion at the alveolar region to strong retroflexion at a post-alveolar or palatal region will cause a progressive lowering of F_3 (cf. Fant (1968:239)). Lehiste's average F_3 for initial /r/ of 1350 Hz (1964: Table 3-IV) compares with a mean F_3 for initial /r/ over the (twelve measurable) speakers of the present study of 1670 Hz (Table 4 above), which seems to confirm that the American initial /r/ is the more retroflex. The mean of 1090 Hz for F_2 of /r/ in Table 4 is also higher than the equivalent figure of 920 Hz in Lehiste's Table 3-IV.

This may in part be the result of coarticulation in the English data, where front vowels with high F_2 will tend to pull up the mean F_2 of preceding /r/. At the recording session three speakers - ACM, GL and JRC - were noted (auditorily) to use fairly strongly labial realisations of /r/, and so might be considered unfairly included in the means. If these speakers are excluded, the present F_3 and F_2 values remain well above those of Lehiste, at 1570 and 1060 Hz.

As noted above, the formant data for American final realisations of /r/ suggest greater assimilation to an adjacent vowel than in the case of the initial ones - F_2 undergoes more perturbation, the average F_1 is higher and nearer to an expected figure for a mean F_1 for a set of vowels, and at 1560 Hz F_3 suggests less severe retroflexion than does the 1350 Hz mean F_3 of the initial allophone (Lehiste's Tables 3-IV and 3-V). Now, although in the production of an utterance the CR compiler operates on a phonetic representation and is ignorant of the phonemic identity of the initial and final allophones of /r/, it is nevertheless not surprising that whatever the influencing factors are (eg. rate) causing a change in the CR index of one allophone of a phoneme, they may also have a parallel effect on other allophones, by virtue of their phonetic similarity. Hence, reductions of blade and body CR indices parallel to those of the initial allophone can not unreasonably be postulated for the final allophone, as part of the development of SBE. This leaves the SBE final /r/ with no resistance to blade coarticulation - hence no retroflexion (in the absence in English of retroflex vowels to cause coarticulated retroflexion) - and a small resistance to tongue body coarticulation, yielding [V+ neutral vowel] sequences where previously there were [V+ɹ] sequences. Since the original loss of postvocalic /r/ as a retroflex articulation the remaining coarticulation resistance of the segment has continued to weaken, with the resulting total (except perhaps durational) assimilation of /r/ after vowels other than close vowels - [aə] → [a·] (heart), and more recently a tendency towards [ɛə] → [ɛ·] (air) and [ɔə] → [ɔ·] (port).

The chronology inherent in the changes outlined above needs to be clarified; it is envisaged that a given generation of speakers allowed their production of [Vɹ] sequences, at least under certain conditions (or rate, informality and the like), to admit greater coarticulation, and this resulted in /r/ being produced in those circumstances as a neutral vowel. At a time when, in that or subsequent generations, this production of /r/ became unconditional, new learners of the language would have no evidence to support the phonetic existence of postvocalic retroflex sounds, and so they would acquire a modified set of phonological rules to derive appropriate phonetic representations. Furthermore, as evidence for postvocalic /r/ is reduced, there would be pressure for change at the systematic phonemic level, although the underlying representations may justifiably be considered still to contain postvocalic /r/ if this is realised as a "linking" [ɹ]; less justifiably so in the case of those who have a generalised process of inserting an "intrusive" [ɹ] whenever a non-close vowel precedes another vowel with a word-boundary intervening.

In the dialects of South East England, or at least in certain restricted subdialects, it appears that the tongue CR indices have weakened still further, causing initial /r/ to be realised as a bilabial or labiodental approximant (a labial contribution is a common adjunct to an alveolar [ɹ]; this further development may well be a compromise articulation accepted by the acoustic comparator as it preserves the essentially grave nature of [ɹ], while preferred by the articulators as requiring less articulatory effort than retroflexion of the tongue tip - cf. Lindblom (1972) who claims that labials result from an essentially neutral (minimum effort) vocal tract when the jaw is raised). Although often felt to be "defective", labial /r/ realisations are now common enough to be widely accepted in the South of England.

7. Conclusions

Using evidence from the /r/ phoneme of British and American English, an attempt has been made to extend the scope of the "Coarticulation Resistance" model of articulatory control. It is suggested that certain regularities in the historical development of /r/ in Southern British English, namely a general "weakening" of its realisations leading to loss of its postvocalic allophone and a less retroflex initial allophone, can be regarded as the result of a lowering of the CR indices assigned to these allophones. The evidence from American English points to that dialect being at a less advanced stage in the same process, having an initial allophone highly resistant to coarticulation and a final allophone with a smaller resistance. In this way a possible description of this diachronic process results from the coarticulation resistance model originally introduced to account for synchronic coarticulatory behaviour.

In addition, support is given to the notion that coarticulation is a "spreading" of phonetic properties limited only by inhibiting factors associated with segments (extrinsic allophones): Present day American English demonstrates contextually determined allophones of a phoneme, /r/, which differ in their readiness to coarticulate with a vowel environment; this appears at first sight no more than was demonstrated for /l/ in British English by Bladon and Al-Bamerni, and shown to hold also for American English by Lehiste. But then note the implications of the combination of evidence from /r/ and /l/ in American English. Whilst Bladon and Al-Bamerni admit (1976:148) that their evidence in general does not conflict with the theory of a $(C)^nV$ "articulatory syllable" within which (anticipatory, and they found in their case also retentive) coarticulation is greater than across articulatory syllable boundaries, and whilst the same seems to hold true for Lehiste's data for /l/ in American English, the problem now arises - since post - not pre-vocalic allophones of /r/ coarticulate most with the vowel environment - that /r/ requires a different articulatory syllable to match up with its coarticulatory behaviour. Must two contradictory articulatory syllables - CV for /l/ and VC for /r/ - be recognised even for the phonetically similar liquids? The evidence tells against the articulatory syllable theory, and in fact indicates the probability of coarticulation being limited by values associated with segments.

References

Al-Bamerni, A (1975): *An Instrumental Study of the Allophonic Variation of /l/ in RP*. Unpublished MA Dissertation, University College of North Wales, Bangor.

Amerman, JD and Daniloff, RG (1974): Aspects of lingual coarticulation. *Journal of Phonetics* 5 107-113.

Bladon, RAW and Al-Bamerni, A (1976): Coarticulation resistance in English /l/. *Journal of Phonetics* 4 137-150.

Bladon, RAW and Nolan, FJ (1977): A video-fluorographic investigation of tip and blade alveolars in English. *Journal of Phonetics* 5 185-193.

Chomsky, N and Halle, M (1968): The Sound Pattern of English. New York: Harper and Row.

Fant, G (1960): *Acoustic Theory of Speech Production.* The Hague: Mouton.

Fant, G (1968): Analysis and synthesis of speech processes. In Malmberg, B (Ed), *Manual of Phonetics.* Amsterdam: North Holland.

Fant, G (1971): Distinctive features and phonetic dimensions. In Perren, GE and Trim, JLM (Eds), *Applications of Linguistics.* Cambridge: CUP.

Gimson, AC (1970): *An Introduction to the Pronunciation of English.* London: Arnold.

Hardcastle, WJ (1976): *Physiology of Speech Production.* London: Academic.

Jakobson, R (1971): *Studies in Child Language and Aphasia.* The Hague: Mouton.

Jones, D (1972): *An Outline of English Phonetics.* Cambridge: Heffer.

Kent, RD and Minifie, FD (1977): Coarticulation in recent speech production models. *Journal of Phonetics* 5, 115-133.

Kozhevnikov, VA and Chistovich, LA (1965): *Speech: Articulation and Perception.* Translation: Joint Publications Research Service, 30.543. US Department of Commerce.

Ladefoged, P (1971): *Preliminaries to Linguistic Phonetics.* Chicago: University of Chicago Press.

Lehiste, I (1964): Acoustical characteristics of selected English consonants. *International Journal of American Linguistics* 30, No.3.

Lindblom, B (1972): Phonetics and the description of language. *Proceedings of the 7th International Congress of Phonetic Sciences,* 63-97. The Hague: Mouton.

Lindblom, B and Sundberg, J (1971): Acoustical consequences of lip, tongue, jaw and larynx movement. *Journal of the Acoustical Society of America* 50, 1166-79.

Ní Chasaide, A (1977): *The Laterals of Donegal Irish and Hiberno-English: An Acoustic Study.* Unpublished MA Dissertation, University College of North Wales, Bangor.

Nie, NH, Hull, CH, Jenkins, JG, Steinbrenner, K and Bent, DH (1975): *Statistical Package for the Social Sciences* (2nd Ed). New York: McGraw-Hill.

Su, L-S, Li, K-P and Fu, KS (1974): Identification of speakers by use of nasal coarticulation. *Journal of the Acoustical Society of America* 56, 1876-1882.

REVIEW of Gillian Brown, *Listening to Spoken English*, London, Longman, 1977 pp. xv & 175

Those of us who number even advanced foreign students of English among our charges are all too used to the dismayed protestation that, 'they don't speak English here like they did in our classes at home'. It is indeed high time that teachers of English as a foreign language paid more attention to bridging the gulf between the slow, careful pronunciation to which students are usually exposed in courses, and the extremely messy signal they will have to contend with in interacting with native speakers in everyday situations. Gillian Brown's book is an attempt to provide teachers with a tool which will allow them to prepare their students for this acoustic onslaught. Published as part of the Longman 'Applied Linguistics and Language Study' series, it will, I am sure, quickly find its way on to the bookshelves of EFL teachers, beside other volumes of the series, such as Crystal and Davy's Advanced Conversational English and Widdowson's Stylistics and the Teaching of Literature.

The author sets herself the task of describing the differences in pronunciation between the kind of English which students will need to understand when following advanced courses in Britain, and the much slower, more explicit style to which they will no doubt have been exposed in their courses at home. There is clearly a problem in deciding what sort of material to use as a model: the choice made here is a sensible one, observations being based on recordings of BBC newscasts and interviews, which show considerable stability of patterning, and yet are informal enough to demonstrate a wide variety of modifications to the 'idealised' pronunciation students might have been led to expect. The author is at pains to stress that she is not advocating the use of such a naturalistic model for the teaching of pronunciation, but believes that 'slow colloquial' is the only workable model for this, with all but the most advanced students. Indeed, I think she goes a little too far in saying, "I do not approve of teaching students to produce 'assimilated' forms and elided forms. Sophisticated students who have been taught to be aware of these forms will introduce them into their own speech in a natural context when they feel able to control them." (pp.156-7). In my experience, such desirable changes in students' English very rarely come about in this spontaneous way, except perhaps after very long periods of exposure.

The book is organised into eight chapters, plus an introduction, appendix, bibliography and index. The first and last chapters are of a general nature, the first discussing the whole raison d'être of the book and underlining the need to teach listening comprehension on a par with other skills, the last drawing together the various strands and making some useful, if at times rather obvious, suggestions for the inculcation of listening comprehension skills, using materials of the type analysed in the central chapters. It is interesting, and significant, that of the six chapters which form the core of the work, only two (numbers 2 and 4) are concerned with segmental phenomena. Indeed, one of the main features of the book is its emphasis on stress and rhythm (chapter 3), intonation (chapter 5), verbal 'fillers' (cahpter 6) and paralinguistic features (chapter 7).

In general, the author handles her extremely difficult task with enviable skill, combining a strong theoretical interest with the down-to-earth common sense of one who has had first hand experience of dealing with students' comprehension problems. Particularly valuable, for instance, is the demonstration of how useful visual clues can be in our attempt to decode the noisy messages we receive from our interlocutors. Muscular movements made by the speakers are in time with stressed syllables and so give a clue to rhythmic structure. Movements such as that of the jaw and lips in the formation of labial consonants, and the lip pouting which often accompanies formation of [ʃ,ʒ,tʃ,dʒ,r], although much reduced in informal speech, may yet serve as pointers to these particular sets of sounds for students who have been trained to watch as well as to listen. The book contains many practical hints of this kind, encouraging us to inspire in our students the confidence to use all the many types of information which can aid in the interpretation of the spoken message, rather than relying on the clear perception and understanding of all (and only) the segmental features of utterances, as so many foreign learners do.

Few works, however competent, clear and useful they may be, are successful in every respect. Gillian Brown's book is no exception and it seems worthwhile to comment on the balance of strengths and weaknesses in each of the areas covered.

In chapter 2, the 'ideal' forms of sound segments are discussed, as a preliminary to an analysis, in chapter 4, of the modification of these forms in normal informal speech. The interpolation, between these related chapters, of a section on stress and rhythm is not explicitly justified in the book - it is clear, however, that rhythmic factors are among the most important influences on the modification of idealised segments in normal continuous speech. In presenting an overview of the phonemic make-up of English, the author very commendably attempts to "group phonemes together into classes which are determined by the way in which they function in English" (p17), rather than relying on phonetic criteria. Unfortunately, however, this attempt does not entirely come off. In a discussion of the vowel system, for instance, the following classes are proposed (Table 3, p33):

<u>A</u>	<u>B</u>	<u>C</u>	<u>D</u>
ɪ	i	jʊ	iə
e	eɪ		eə
æ	aɪ	aʊ	ɑ
ɒ	ɔɪ	əʊ	ɔ
ʊ		u	ʊə
ʌ			
ə			ɜ

It is claimed that these classes are distributional, and class A is, in fact, so defined, in that none of its members can occur in an open, stressed monosyllable. Classes B and C, however, are defined in phonetic terms (as 'front closing' and 'back closing' vowels, respectively), while the criterion for class D appears to be primarily orthographic, viz. that most words containing these vowels have an 'r' in the spelling, which appears as an [r] glide when such vowels are followed immediately by a second vowel in a separate syllable.

A further problem in this chapter on idealised sound segments, which becomes even more acute in the later chapter on patterns of simplification in informal speech, arises out of the necessarily rather detailed descriptions given. The phonetic knowledge required for an understanding of such descriptions, while by no means out of the way for the professional linguist, may well prove too daunting for many EFL teachers. Although the author claims (perhaps significantly on the last page of the book) that "there is certainly no necessity to introduce all the technical terms", it is difficult to see how students could be taught to recognise and understand some of the subtler distinctions made, unless they possessed the relevant technical armoury. This problem is reinforced by the author's liberal use of what she admits to be somewhat unorthodox phonetic transcription conventions, in an attempt to put over the relevant distinctions. On page 30, for example, we find the items <u>knack</u> and <u>nag</u> distinguished phonetically as [næ?k^{ə}̊] and [næ̣: ^{g}k^{ə}̊] respectively.

The short chapter 3 on stress and rhythm contains a valuable discussion of the problems of teaching English rhythm to foreign learners, and rightly criticises the tendency, in many commercial courses, to use a model in which little distinction is made between the pronunciation of otherwise identical stressed and unstressed syllables. Visual and phonetic clues to stress are discussed. The theoretical framework used for the description of rhythm is based on the 'foot', as the term is used by Abercrombie (1965) and taken over by Halliday (see eg. Halliday 1970), which certainly has merits for the teaching of rhythm, but is perhaps of somewhat doubtful theoretical validity.

Chapter 4 presents a very thorough account of the simplification of sound segments in continuous, informal speech, by the processes of assimilation and elision. The points made are amply exemplified from the taped BBC data. There is an extremely valuable section on visual clues, of which I made mention earlier. The one criticism which could be levelled at the exposition in this chapter is that of at times excessive detail, leading to the problems of transcription and analysis discussed above.

Chapter 5, on intonation and its functions, is one of the least successful, because here the reader comes up against the difficulty of hearing in his head the exact pitch patterns suggested by even a fairly pictorial representation of intonation contours. This problem could be overcome by making available a tape of the relevant BBC material, assuming that permission could be obtained. One important observation of which I, at least, should like documentary proof, is that all tone groups (ie. stretches of language containing one major pitch change) are delimited by pauses in the news broadcasts analysed. Not surprisingly, this was not always the case in the unscripted radio interviews analysed later in the chapter. A further claim, again a little surprising to anyone brought up on the well-known accounts of intonation in English, such as O'Connor and Arnold (1961/1973), Halliday (1970), is that a large proportion of even non-final tone groups in news broadcasts have falling intonation. Indeed, early in the chapter (p86) the reader is misled into thinking that all tone groups in these extracts have falling tone: only later is it admitted that non-final tone groups may have rising or falling-rising patterns. Curiously, the over-short discussion of pitch direction in chapter 5 is continued in chapter 7, where it is (surely wrongly) presented as a paralinguistic feature. Here, the interesting claim is made that high rising intonation is quite rare, and is not the usual pattern for polar questions, which in fact normally take a fall. Despite such observations, which clearly point to the need for more detailed investigation of this area, the discussion of the function of tone choice is disappointing. In particular, very little is said about the function of the low-rising and falling-rising tones which are readily apparent in every news broadcast I have ever heard. The discussion of the information-structuring functions of intonational choices, though accurate, is scanty, despite the fact that even advanced foreign learners find especial difficulty in maintaining the cohesion of spoken text by appropriate placement of tonic elements.

Chapter 6, on verbal 'fillers', contains, for the most part, fairly commonsensical observations - it is nevertheless good that teachers should be reminded of the importance, in conversational interaction, of feedback signals indicating the listener's continued attention, and of repetitions which have a cohesive function in the discourse, or serve as time-fillers while the rest of the speaker's utterance is being planned. The author differentiates between repetition of 'lexical' words, which occurs mainly in more formal speech situations, and can have powerful rhetorical effects, and repetition of lexically weak items, which is characteristic of informal speech. The former type of repetition, it is argued, may aid rather than inhibit comprehension, while the second type, together with the occurrence of false starts, hesitations and the like, contributes substantially to the difficulties foreign learners find in following conversations. The chapter ends with a useful piece of practical advice to students - concentrate on the ends of utterances, since the beginnings, often loaded with introductory fillers, usually contain little of the meat of the message.

Chapter 7, on paralinguistic features, is somewhat marred, as mentioned earlier, by the unwarranted inclusion of pitch direction as a paralinguistic phenomenon. Another doubtful area included here is the degree of precision of articulation (eg. the use of 'strong' forms where 'weak' forms might be expected). The chapter nevertheless contains useful discussion of pitch range, placing within the range, tempo, loudness, voice setting ('breathy' and 'creaky' voice, etc), articulatory setting (tenseness of articulation), lip setting and pause. In discussing these areas, the author makes use of quotations from literary works, in which the manner of production of direct speech is indicated in the linguistic context, often by means of adverbials such as sharply, indignantly, harshly, etc. The device works well, aiding the reader to imagine more accurately the voice qualities intended, and to some extent making up for the lack of a taped version of the BBC extract discussed later in the chapter. One aspect of the attempted analysis of paralinguistic features is perhaps not entirely satisfactory. The author refuses to be led into too facile an association of paralinguistic features with particular attitudes, preferring, in most cases, to claim that deviations from the norm simply indicate that some attitude or other is being expressed. The crucial question from the point of view of the hearer, however, is which attitude is being conveyed. This question the author fails to tackle.

Typographically, the book is of a high standard - I found only one error, viz. the omission of an underlining for may in Extract I, p90. The references given are perhaps rather too few, though well chosen. The index, too, could profitably have been expanded.

Taken as a whole, then, this is a book to be recommended, if with some reservations. It will certainly repay the perhaps not inconsiderable labours of those teachers who are prepared to get to grips with it.

References

Abercrombie, D (1965) 'A phonetician's view of verse structure' and 'Syllable quantity and enclitics in English' in Studies in Phonetics and Linguistics, London, Oxford UP.

Halliday, M.A.K. (1970) A course in spoken English: intonation London, Oxford UP.

O'Connor, J.D. and Arnold, G.F. (1961/1973) Intonation of colloquial English London, Longman

C.S. Butler, University of Nottingham

REVIEW of Eli Fischer-Jørgensen, *Trends in phonological theory: a historical introduction.* Akademisk Forlag, Copenhagen, 1975. pp. xxiii & 474. £19.45.

This book offers a good deal more than its title promises. It is in fact a historical and critical survey of the work of all the individuals and schools that have made significant contributions to phonological theory in the twentieth century, and of the interrelations between them. Furthermore, since most phonological theories are embedded in some more or less precisely articulated general linguistic theory or conceptual framework, the book will also be of use to those who wish to learn, at not too great length and in an intelligible form, about the general principles of various linguistic schools - for example the key concepts of Saussure's general linguistics, the outlines (as far as possible, in plain English) of Hjelmslev's glossematics, the philosophical-psychological assumptions that lie behind the transformational-generative theoretical paradigm.

No work such as this, covering the whole discipline of phonology, had been seriously attempted before, and Fischer-Jørgensen's achievement is superb in content, in presentation, and not least in the impartiality and courtesy with which she avoids polemic and gives a hearing to every point of view. But her impartiality is not that of a person who can report the views of others because she has none of her own. Though she finds that almost every school has valid insights to offer in one area or another, she clearly considers some theories and some styles of analysis to be at any rate more useful than others; in particular, she is convinced of the validity of some form of classical phonemic analysis, of the "basic concepts of Prague phonology" and of the need for a distinctive-feature theory. She is inclined to doubt the psychological reality of generative phonologists' underlying forms, particularly when these are "abstract", yet insists (296) that the generativists' explicit and precise descriptions are "of great value, irrespective of [their] psychological adequacy"; this may sound like having one's cake and eating it, but I suspect it means (cf. 401-2) that even should they prove to be bad phonology, generative analyses are good morphophonemics.

Any account of twentieth-century phonology must give a central place to the Prague School, the post-Bloomfieldians and generative phonology; and Fischer-Jørgensen discusses all three, despite their great differences, fully and lucidly. The treatment of other schools is rather uneven in quantity (though not in quality), perhaps reflecting the book's origin as a manual for Copenhagen phonetics students. Thirty pages are devoted to glossematics without the author (never mind the reader) being convinced that it has anything worthwhile to contribute to our understanding of languages; stratificational theory also is perhaps given more attention than it deserves (though Fischer-Jørgensen is not the only recent author of whom this could be said); while for Firthian prosodic phonology, one hopes it is not a proof of insularity to say that four and a half pages is hopelessly inadequate. The most striking imbalance in the book, however, is a wholly welcome one. The eleventh chapter, the third longest in the book, is devoted to a topic of which most readers can learn all too little from other sources, phonological theory in the Soviet Union: if for about thirty years from 1920 Soviet phonology developed (or for much of that time decayed) almost independently of the West, Fischer-Jørgensen's account makes it clear that since 1952 and especially since 1960 Soviet phonologists have been working more and more in the same tradition as their Western colleagues, who, however, owing to the language barrier have been less than ideally placed to take advantage of fresh insights provided by Soviet work. Fischer-Jørgensen has performed a notable service in bringing this rich and multifarious tradition of phonological study to Western scholars' attention: a service that evidently (320 fn) involved a disproportionate amount of uncongenial toil.

The chapter on generative phonology is at least as good an introduction to the subject as most that have been written by generative phonologists themselves. Though not herself an initiate, the author is extremely well acquainted with the development and ideas of the school, describes its theories with a sympathetic understanding not very common among either generativists or anti-generativists, and in criticism can be acute and to the point: see her remarks on the inadequate treatment of morphology in generative grammar (202 f), the notion of a "neutral position" of the vocal tract (226), the impossibility of accounting for external sandhi in the strict Sound Pattern of English model (249) and the evidence for the linguistic reality of a classical phonemic analysis (287-296). As generative phonology has for some years been in a state of Heraclitean flux, it should be mentioned that Fischer-Jørgensen's treatment goes down to 1973 but not beyond.

In addition to the schools already mentioned, chapters are devoted to "forerunners of phonological theory" (emphasis on the Icelandic "First Grammatical Treatise", on Baudouin de Courtenay, Saussure and Sapir), to Daniel Jones, to Jakobson's distinctive feature theory (the Chomsky-Halle and Ladefoged feature theories are dealt with elsewhere in the book), and to "contributions from outside the schools", a chapter which brings together, and compares with each other and with the major schools, the work of linguists as diverse as Malmberg, Kuryłowicz, Haas, Labov, Bazell, and the author herself. The book ends with a rich but intelligently selective bibliography and indexes of names and subjects.

To the serious student of phonology, at whatever level, this book will be essential for many years; and I would not be surprised were it to take rank with the two classic works that the science of phonology has so far produced, Trubetzkoy's Grundzüg der Phonologie and Chomsky and Halle's Sound Pattern of English.

A.H. Sommerstein, University of Nottingham