language learning'. In N. Waterson & C. Snow (eds.), *The development of communication*, 173-188. New York: Wiley & Sons.

Linell, P. & L. Gustavsson. 1987. *Initiativ och respons. Om dialogens dynamik, dominans och koherens*. Linköping Studies in Communication 15.

McTear, M. 1985. *Children's Conversation*. Oxford: Blackwell.

Newport E., H. Gleitman & L. Gleitman. 1977. 'Mother, I'd rather do it myself: Some effects and non-effects of maternal speech style'. In C. Snow & C. Ferguson (eds.), *Talking to children. Language input and acquisition*, 109-49. Cambridge: Cambridge University Press.

Söderbergh, R. 1984. 'Linguistic effects by three years of age of extra contact during the first hour post partum'. Paper presented at the Third International Congress for the Study of Child Language, Austin Texas.

Wells, G. 1980. 'Apprenticeship in Meaning'. In K. Nelson (ed.), *Children's language 2*, 45-126. New York: Gardner Press.

Lund University, Dept. of Linguistics
Working Papers 38 (1991), 45-70

On the Perception of Prosodic Phrase Patterns

Eva Gårding and Lars Eriksson

In our search for perceptual correlates of intonation and accentuation we have conducted a series of experiments with some Swedish prosodic phrase patterns: Prototypical productions have been digitized and subjected to manipulations which have served as stimuli in categorization tests. The stimuli are obtained in different ways, (1) by shifting a fundamental frequency (Fo) peak in the time domain of two different carriers, (2) by shifting Fo values over neighbouring vowels in the frequency domain of one carrier and (3) by stripping the prototypical signals of their various acoustic components in four carriers. The results, displayed as categorization functions (1,2) and confusion matrices (3), indicate that the pitch movements over the vowels, their relational pitch levels and the temporal and spectral properties of the carrier are important cues to a prosodic phrase pattern. The importance of acoustic correlates varies from one prosodic pattern to another in such a way that an absolute rank order between them does not seem meaningful. The notion of markedness may be used to explain the asymmetry of confusion patterns.

Introduction

The phrase and its place in a linear or hierarchical structure of speech has gained increasing importance in phonetic and phonological analyses.

In Gårding and House 1987 production and perception of phrases in Scandinavian dialects and Finnish were studied in utterances consisting of different groupings of similarly accentuated numbers and the results supported the following phonetic definition of a prosodic phrase: A prosodic phrase is a part of an utterance which is connected by a special rhythmic and tonal pattern and demarcated by discontinuities in the range or general direction of the pitch contour (pivots).[1]

In the experiments to be reported here, we have used complex accentuation patterns as our basic material, namely a segmentally invariant sequence *långa män* which on account of its prosodic pattern may be tied to distinctive syntactic structures and semantic meanings (Table 1 and Fig. 4).

[1] Similarity of the elements of a group and recurrent special patterns ('hats' or 'troughs' depending on the dialect) could be interpreted as connective cues, breaks of similarity and special markers at the beginning or end of a group as demarcative ones (Gårding & House 1987). For a conceptual framework see Gårding 1985.

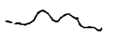
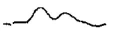
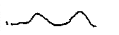
Our present goal is to define such patterns of accentuation in terms which are naturally related to perception. The study is here restricted to declarative intonation. Other intonations will be treated later.

Material and method

Some segmentally equivalent but prosodically and semantically different sentences were elicited in declarative intonation from a trained phonetician representing a modified Stockholm dialect. The sentences, in which the test phrase was preceded by *dom e* (they are), were uttered three times each in well-defined situational contexts.

The table below shows various representations of the material used in our experiments, E0, E1, E2. The segmental composition is /lɔŋːa mɛnː/ which may carry several prosodic patterns differing in accentuation and juncture giving them specific meanings. The examples are borrowed from Bruce. For other speakers of the same dialect see Bruce 1977.

Table 1. The prototypes of E0, E1 and E2

	Written form	Syntax	Meaning	Prosodic form	Accentuation	F0 contour
a	lánga män	2-word phrase	tall men tall men	ˋlɔŋːa ˊmɛnː	accented, unacc, accented	
b	lánga män lánga, men	2-word phrase	tall men tall but	ˋlɔŋːa ˌmɛnː	accented, unacc, deaccented	
c	Långamän	compound	men from Långa	ˋlɔŋːaˊmɛnː	accented, unacc, accented	

Primary accents are marked ˊ for A1 and ˋ for A2 , deaccentuation is marked by ˌ , word boundary by spacing and focus by underlining. (a,b) are used in E0, (a,b,c) in E1 and (a,c) in E2. For E3, see Fig.4.

In the following presentation the two-word phrases will be called the accented and the deaccented case respectively, referring to the status of the accentuation of *män*, and the compound phrase will be called the compound.

One production of each test sentence was chosen as a base for our experiments. It is the Fo curves of these prototypes that are shown in the last column of Table 1.

The productions were analysed digitally and manipulated in various ways by means of ILS and connected programs developed by one of the authors (LE). In this way we obtained series of stimuli which were judged in forced-choice categorization tests.

The tests followed routine procedures with the stimuli appearing three times each in random order in groups containing 8-11 test items and with a presentation of typical stimuli introducing each test. The test numbered 0 was administered only in the Lund speech laboratory, the other tests were given at KTH, Stockholm and in the perception laboratories of the phonetics departments in Lund and Stockholm. The listeners represented prosodically different dialects with southern dialects predominant in Lund and non-southern dialects in Stockholm. For details about listeners and task see the respective section below.

Outline

Apart from a summary desciption of our preliminary testing (0), three sets of experiments will be reported.

(1) Experiments in which a pitch peak has been shifted in time
(2) Experiments in which a pitch peak has been shifted in frequency
(3) Experiments in which the signals have been stripped systematically of their different acoustic correlates, Fo contour, intensity, spectral characteristics, etc.

Experiments 1 and 2 have been published in a condensed form in the proceedings from conferences in Stockholm and Budapest.

E0. Preliminary testing. The contrast accented/deaccented

The strong cueing power of Fo to stress has been documented in a number of well-known experiments ever since technical advances made manipulation with synthetic speech possible. In principle these early tests were designed to find some rank order of various acoustic correlates to stress, which proved to be in the following order: Fo, duration, intensity and spectral characteristics (Fry 1958). The importance of Fo was also evidenced for Swedish by Zetterlund, Nordstrand & Engstrand 1978.

Our first experiment, E0, concerns the contrast between accented and deaccented. It is generally assumed that in connected speech there are four relevant levels of stress (here called accentuation), apart from accented and unaccented also deaccented and strongly accented (used for focus, emphasis etc.). It should be noted, however, that a manifested stress level has no unique correspondence to a particular function (sentence accent, phrase accent etc.) but follows a principle of economy relative to the message (Gårding 1989).

Since we believed that it would be possible to bring out the contrast by means of the Fo curve alone, we chose as prototypes for our manipulations the two productions which had the most similar temporal patterns (a and b in Table 1). These prototypes were digitized and appeared in a listening test together with 9 edited variants.

Among the edited variants were simplifications of the Fo curves of the prototypes by the use of straight interpolation between conspicuous turning points and by reduction of the number of turning points; 'complifications' by the introduction of a new turning point in the deaccented prototype (to create a hat pattern) and by the exchange of the Fo contour and the carrier of the two prototypes. Conspicuous turning points are defined as such turning points which cannot be bypassed in straight interpolation without a loss of prosodic information. This definition is based on the outcome of earlier analysis-by-synthesis experiments (see e.g. Gårding & Stenberg 1990).

Some of these edited contours together with the two digitized prototypes, altogether 11 items, were triplicated and ordered in groups of 11 for a test which was given to 11 listeners from the Lund department. The answer sheets had two semantic alternatives: 'män' (men) or 'men' (but).

Results and discussion

Judging from the response scores for the digitized versions of the two prototypes, both patterns are recognized but the accented pattern has a better score than the deaccented one (85%/70%). Some of the listeners objected that the deaccented pattern was inherently ambiguous in that both the meaning 'but' conjunction, as given in the answer sheets, and the meaning 'män' following focussed *långa* could be attributed to it.

Contrary to our expectation, the scores of the manipulated stimuli showed that it is difficult to turn a pattern into its counterpart by means of Fo manipulations only. It is only when a new turning point is introduced over the deaccented vowel (thereby creating a hat pattern) that the accented alternative is given some but not overwhelming predominance (63%).

Conclusion

In the prototypes, chosen on account of their similar temporal characteristics, the Fo curve cannot override the cues provided by their spectral and intensity properties.

E1. Peak shifts in the time domain. Coordination of pitch and segments

To avoid the somewhat arbitrary character of our preliminary manipulations we turned to a new experiment which involved small systematic changes in an intonation curve: We decided to let a pitch peak of a constant triangular shape 'wander' across a segmental carrier in fixed small steps, thus determining its Fo contour.

The method of shifting the time location of a pitch peak in synthetic speech to study the effect on listeners is not unusual. Typically the aim has been to determine boundaries between distinctive prosodic categories, e.g. Accent 1 and Accent 2 in Swedish dialects (Malmberg 1955, Bruce 1977 and 1983), Serbo-Croatian accents (Purcell 1976), sentence accent in American English (Gårding & Gerstman 1960) and accents with different pragmatic values (Kohler 1987). When we use this method here, our main concern is the perceptual aspects of such a shift on more complex accentuation patterns. More precisely we ask the question: For a certain phrase contour, what are the perceptually relevant combinations of pitch movement and segment?

Since the deaccented contour (b) had received a lower recognition score than the accented one, a new prototype was chosen which was more clearly distinct from (a) on account of its temporal pattern. As can be seen in Figure 1, the *ŋ-a* segments have durations characteristic of a marked pre-boundary lengthening. The Fo pattern was simplified as before by using straight interpolation between the conspicuous turning points (see Fig. 1). Another simplification was that the slight declination of the topline was made level. In this way the pitch movement over the accented word *långa* becomes a two-level contour, i.e. maximal and minimal pitch levels are limited to two values.

The peak shift experiment was based on two of these simplified contours, each superposed on its spectral carrier, one from prototype (a) ending in accented /mɛn:/, the other from the new prototype (b), ending in deaccented /mɛn:/. As can be seen in Figure 1, the first peak over the syllable /lɔŋ/ has been kept fixed, the second peak has been shifted in steps of 20 milliseconds. The base of the peak was made proportional to the duration of the /ɛ/ of the /mɛn:/-syllable.

On listening to the two series of stimuli obtained we discovered that apart from the intended categories (a) and (b) they also included a third, a compound phrase. This called for new recordings with our speaker and

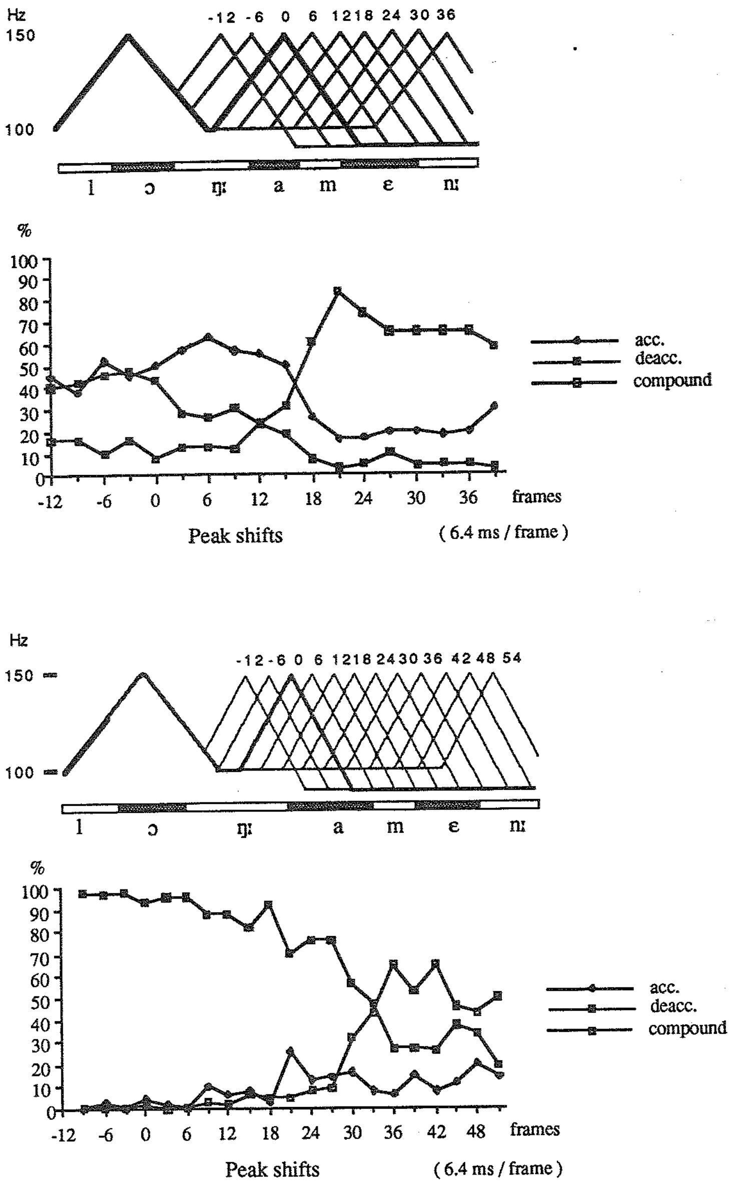


Figure 1. Response functions for stimuli with shifted peaks. (a) Carrier with accented /mɛn:/. (b) Carrier with deaccented /mɛn:/.

among the new versions, (c) was chosen as a prototype for the compound. Since the compound productions resembled the accented (a) contour in their temporal characteristics we did not find it necessary to use (c) as a carrier for a peak shift. In this way we could keep the number of test items manageable.

In the subsequent tests there were 3·40 (120) stimuli (ordered in groups of ten) which were judged by two groups of sixteen listeners each from Lund and Stockholm. The listeners, members of the linguistics departments, heard the stimuli over loudspeakers in the perception laboratory of their respective department. The task was to place each item heard in one of the three categories (a), (b) or (c), as shown in Table 1. In the response sheets, category (b) was now given two semantic labels (see Table 1), a change motivated by the reactions of the listeners in our preliminary test.

Results and discussion

Figure 1 shows the results of Test 1. With the carrier derived from the accented prototype, (a), there is no cross-over between the two-word phrases. The response functions follow each other rather closely with some, but less than expected dominance for the accented /mɛn:/ responses. When the votes cross over from two-word phrase to compound, the time-shifted pitch peak is at the beginning of the vowel /ɛ/, the rising ramp over the inital consonant /m/ leaving the preceding vowel /a/ in a low and level portion of the pitch curve.

In the deaccented carrier, (b), there is a strong preference for deaccented as long as the peak leaves the latter part of the phrase /a-mɛn:/ in low pitch, or in low preceded by a fall over /a/. The accented alternative is almost rejected. The top cross-over from the two-word phrase to the compound, which is well accepted, occurs as before when the peak is at the beginning of /ɛ/ and the /m/ in the preceding rising ramp.

There is, then, a difference in scores, obtained with stimuli which have a similar location of the time-shifted peak: The deaccented pattern reaches 95% in its own carrier and is ambiguous in the other one. The accented pattern, which gets scores around 60% in its own carrier is practically rejected in the deaccented case. The compound, finally, has a better score in the accented carrier (82%) than in the deaccented one (60%). It is evident that the temporal pattern, together with the intensity and spectral characteristics of the carrier plays an important role in the identification of our three prosodic phrase patterns.

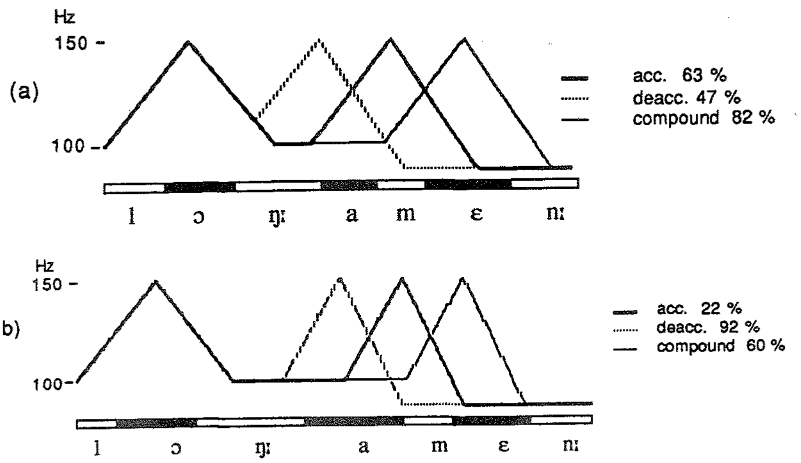


Figure 2. Contours with highest scores. (a) Carrier with accented /mɛn:/. (b) Carrier with deaccented /mɛn:/.

In spite of the strong influence of the carriers, our peak-shift experiment throws some light on the nature of the Fo contours of the prototypes. Figure 2 shows the contours with the highest scores for both carriers. With the exception of the accented case, the prosodic patterns can be given carrier-independent contour descriptions. These descriptions make use of the pitch movements over the vowels only.

	Carrier-independent description		Prototypical description cf. Table 1	
	over /a/	over /ɛ/	over /a/	over /ɛ/
Deaccented	fall	low	fall	low
Compound	low	fall	low	rise-fall
Description valid in (a) carrier				
Accented	rise	fall	rise	fall-low

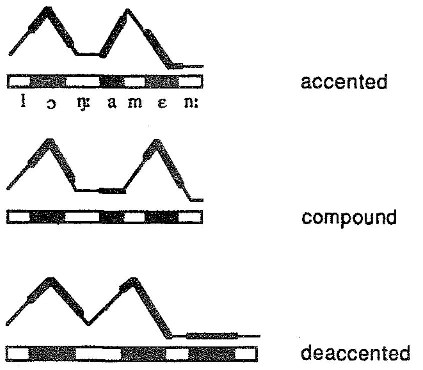
The close fit between the carrier-independent description and the prototypical one supports the contention that the pitch movements over the vowels are prime perceptual cues to a prosodic pattern. Further support comes from results of analyses of Chinese and Swedish which showed that the tone (accent) carrying part of the syllable is constituted by the sonorant segments, for Chinese the rhyme (Howie 1976), for Swedish the vowel (House et al. 1988). This may explain the rejection of the accented pattern in the (b) carrier. Due to the narrow base of the pitch peak no stimulus in this series has the prototypical movements over the vowels. It may also explain the rather low percentage of the accented pattern in its own carrier. With the pointed peak used in the experiment, the pitch of the accented /ɛ/ vowel, which has if not a hat at least a bowler pattern in the original, is poorly approximated.

The compound seems to have a special status among listeners as compared to the two-word phrases. Post-test reactions suggested that the compound could be separated from the other two by a simple binary decision between compound vs. non-compound, here materialized as low /a/ vs. non-low /a/. This binary choice might then come before the choice between the two-word phrases, thereby favouring votes for the compound. (The low distinctiveness of the two-word phrases in the accented carrier will receive a plausible explanation in the experiments of section 3.)

An inspection of individual scores did not show any dialectal bias in the listeners' responses but they did reveal that in contradistinction to the native speakers a group of five non-native subjects had difficulty with the compound.

Conclusions

1. The perceptually important segmental pitch carriers are the vowels.
2. The most successful combinations of pitch movement and vowel are found in carrier (a) for accented and compound and in carrier (b) for deaccented. They can be schematized as follows.



3. The temporal, intensity and spectral characteristics of the carrier phrase play an important role which cannot be clearly evaluated in this experiment.

Remarks

1. Follow-up experiments

Our difficulties in having the accented pitch pattern accepted in the deaccented carrier were overcome with a stimulus in which the two-level contour was replaced by a three-level one: the last peak was raised from 150 to 180 Hz. The same experiment, which was formally tested but not reported here, also showed that the time location of the peak in the 3-level contour is crucial. Part of the rising ramp has to occur in the ε-vowel. When the raised peak is at the C/V boundary with the initial consonant involved in the rise there is a sudden shift in the test responses to deaccented. Hence a large dose of high frequencies in the vowel are needed to make listeners accept the pattern as accented in the deaccented carrier.

2. Peak location

In the peak shift experiment the location of the movable peak determines the neighbouring pitch movement and the pitch movement determines the peak location so one feature may seem as good as the other for the description of the curve. However, we are interested in perceptual correlates and therefore the pitch movement has been chosen as the decisive factor. The use of peak location as a descriptive or distinctive feature in intonation analysis (Ladd 1983) has only indirect perceptual relevance.

E2. Peak shifts in the frequency domain: Intersyllabic pitch relations

Our preliminary experiments had shown that the contrast between accented and deaccented was not strong enough to be simulated by the Fo contour in a carrier derived from the opposite category. The contrast between compound and two-word phrase, on the other hand, was well established in the accented carrier used in the preceding peak-shift experiment and the results suggested that the distinction was carried by the pitch of the unaccented *a*-vowel, with low pitches signalling compound and high pitches a two-word phrase.

These results motivated our choice of the compound/two-word phrase distinction for our continued experimentation. We shall report on two of these experiments which concern the perceptual importance of 1) the pitch value of the first, accented vowel, 2) the pitch value of the second, unaccented vowel. Taken together the results will show the significance of the pitch relations between successive vowels for the identification of a particular pattern. In these experiments the third vowel was kept fixed since it was clear from our production data that this vowel carries a similar pitch peak in both patterns as a reflection of the sentence accent. Also, in a similar experiment exploring the same perceptual contrast, Bruce had shown that the pitch value over the sentence-accent carrying vowel did not have any influence on the compound/two-word distinction which was clearly manifested in the preceding syllables (Bruce 1977).

As a carrier for our pitch manipulations we chose the compound prototype (c) of Table 1. Listening tests were composed as before and the tests were administered to 14 listeners from Lund and 13 from Stockholm. The listeners were asked to label each signal as a compound or a two-word phrase without further specification.

Results and discussion

Figure 3 shows the design of the edited curves and the listeners' responses.

1. To begin with, the first peak was lowered gradually from 170 to 110 Herz in steps of 10 Hz. The response function shows a sharp cross-over from compound to two-word phrase between 140 and 130 Herz.

Obviously the height of the first peak is important for the categorization of the stimuli. A low first peak makes on the listeners (and the experimenters) the impression of deaccented *långa* appropriate for the two-word phrase *långa MÄN* with *män* in focus.

2. In the second experiment belonging to this group, the pitch portion over the mid vowel was raised in a series of ramps starting from the low end point of the first peak and ending at the second peak in 10 Hz-steps from 90 to 170. The response function for this set of stimuli shows that there is a sharp cross-over from compound to two-word phrase between 120 and 130 Hz.

The results of experiments (1) and (2) point to a critical relation between the average pitch values of the first and second vowels. When the average pitch of the second vowel is at least 35 Hz below the peak of the first one, the contour is judged as a compound, otherwise it is taken for a two-word phrase. In relative terms, the critical quotient of the first pitch to the second is 4/3. Above this value the contour belongs to the compound category,

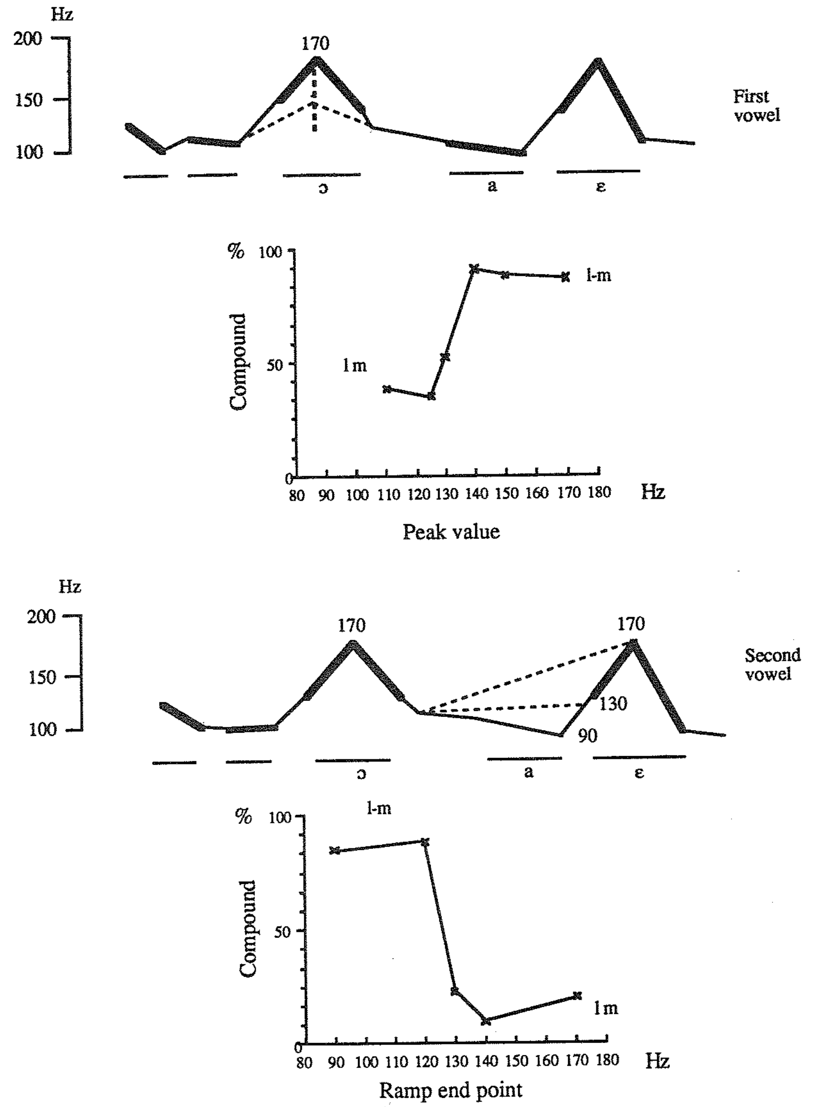


Figure 3. Response functions for intersyllabic pitch variations.

below to the two-word phrase category. In Bruce's experiment the quotient was 5/3 (Bruce 1977). The design was slightly different, however.

It is very probable that such quotients are the invariant tonal features of accentuation patterns which make them recognizable in different kinds of intonation and different voice registers.[2]

Conclusion

Different pitch relations over neighbouring vowels may separate two accentuation patterns in the same carrier.

E3. Stripped stimuli. The complexity of recognition

In the following experiments we are trying a more radical approach to our problems. Instead of making local changes in the curves of which the global effects are difficult to judge and instead of having to take different carriers into consideration, we now make global changes by stripping off layer by layer of acoustic properties from the signals.

New recordings were made with the same speaker. Segmentally the same phrase was used but this time the speaker was asked to give answers to questions which called for five different accentuation patterns, namely even accentuation (in the literature also called broad focus) (1), focus on either of the two words: to the left (2), to the right (3), a neutral compound (4), and a compound in which the first part, a place-name, was expected to be in contrast with another place-name mentioned in the eliciting question (5).

The prototypes of the first four categories will be used in our stripping experiments. They are represented in Figure 4 by their digitized wave forms and pitch curves and have the following labels. Small letters, l m, are used for even accentuation, capital letters denote the focus location, L m versus l M, and hyphen denotes the compound, l-m. The intended contrastive compound (5) turned out not to be distinct from the neutral one and was therefore discarded. According to the speaker, the compound 'being already contrastive' could only be pronounced in one way. (This may not be true for all speakers and all dialects.) Of the four remaining patterns

[2] The importance of the pitch relation between the second and third vowel is in our case not dependent on the accentuation pattern, but a reflexion of the sentence accent which hits the test phrases in a similar way. Without the sentence accent, for instance in a sentence with an earlier focus as in *JA, jag vill ha långa män (Långamän)* the pitch peak over *män* is lowered or flattened out in both cases. However, as is evident from production data and informal testing the two prosodic patterns may still be distinct on account of the different pitch relations of the first two vowels.

the compound *långamän* 'men from Långa' (a not very well-known place) may seem semantically far-fetched but in other respects it represents a regular and productive pattern.

The stimuli were made and labelled in the way displayed by Figure 5. The order is determined by the natural assumption that the meter (as defined below) represents the most basic (naked) part of the signal. The figures on top of the panel displaying meter refer to relative durations of segments measured from vowel onset to vowel onset. The same measure has been used with some success in an automatic prosodic parser developed in Lund (House et al.1988) and for cross-language comparisons of rhythm (Fant & Kruckenberg 1989, Strangert 1985). In the panel representing intensity, the figures refer to relative impulse areas over the vowels obtained by the prosodic parser.

1. Meter: A constant 100 Hz sine wave of the four prototypes with intensity neutralized to a certain level. The consonants are devoiced, which is marked by gaps in the figure.

2. Rhythm: The same signal but with the original intensity of the prototypes added.

3. Rhythmic speech: The speech wave is added.

4. Melodic speech: The melody is added. The melody is obtained by making Fo level with each level tone represented by the average Fo value of the pitch movements over the vowels.

5. Accented speech: The accentual movements are added by making piecewise linear approximations of the pitch movements over the vowels.

We can conclude from Figure 5 that, according to the five different parameters used in the experiment, the speaker makes use of three temporal patterns (the same is shared by l m and l-m, as shown by parameters 1 and 3), three intensity patterns (again, a very similar one is shared by l m and l-m, parameter 2), two melodic patterns, schematically rising over *långa* in l m and L m and falling over the corresponding word in l M and l-m (parameter 4). Parameter 5 adds the pitch movements of the melodic patterns, thus emphasizing their different ranges, corresponding to a two-level contour for l m and l-m and three-level contours, expressing focus or phrase accent for the other two patterns.

Eleven listeners from Lund and 10 from Stockholm took part in the test which included 3·20 stimuli. The test items were presented through head phones. They were ordered in two main groups, each with 7 subgroups consisting of 8 stimuli. The speech-like stimuli belonged to the first main group.

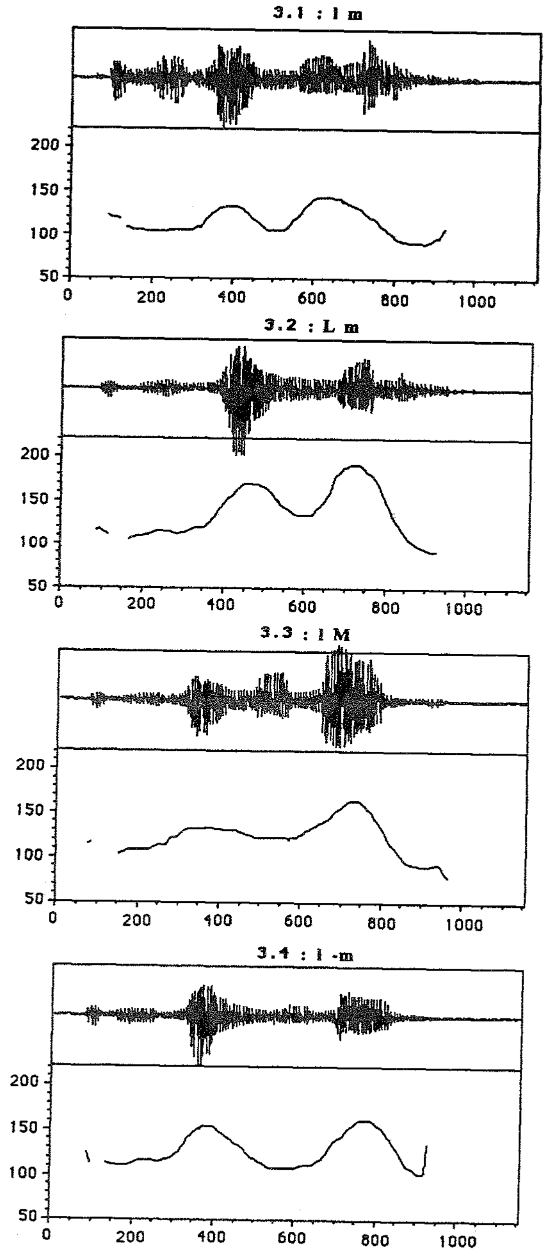


Figure 4. Waveforms and F0 contours of productions used in E3.

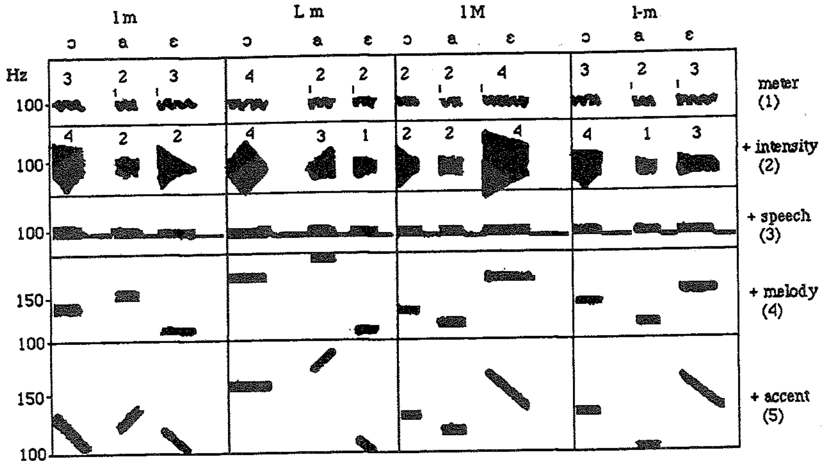


Figure 5. Experiments with stripped patterns. The chosen parameters are added from top to bottom. See text.

Results and discussion

Our results are presented as confusion matrices (Fig. 6). The intended patterns correspond to the rows and the heard ones to the columns. The figures refer to percentage votes for a certain category and add up to 100 for every row. The number 32 in the left-hand upper corner of Matrix 1 (Meter) for instance, means that 632% of the votes are for l m, the pattern with even accents, when only the metrical pattern is presented.

Ideal identification would be 100% in the diagonal and 0 outside. Non-zero numbers outside the diagonal indicate confusion. Two stimuli, a and b, may be confused in two ways. An intended pattern (a) may be heard as a pattern (b) which in turn, when presented, may be heard as pattern (a). When the two percentages are nearly the same, we call the confusion symmetrical, otherwise asymmetrical. The pair l m/Lm in Matrix 4 is an example of symmetrical confusion and the confusion of the pair l m/l-m in Matrix 3 is asymmetrical.

One case of asymmetrical confusion of a pair (a,b) is a high identification score (in the diagonal) of one member, for instance (a). This prevents (a) from being confused with other patterns but it does not prevent (b), when presented, from being confused with (a). Asymmetrical confusion

is frequently connected with a situation in which one member of a confused pair has some special acoustic mark lacking in the other one. When there is no such mark present, listeners may still have some reason, as we shall see below, to more or less identify one but hesitate about the other.

On the whole the sine-wave stimuli (Matrices 1 and 2) were judged as very difficult by the listeners and the difficulty is borne out by the low identification scores and the large, often symmetrical confusions of Matrix 1 in particular. The difficulty is understandable if we consider that in the test situation a listener is asked to give a sine-wave stimulus a semantic label, an associative task which is by no means a natural one. In our further discussion we shall not give much weight to Matrix 1.

In Matrix 2, representing sine-wave stimuli with added intensity, the identification score of focus right, l M, jumps to 58 from the figure 5 in the previous matrix. Another observation in Matrix 2 is the similar scores in the rows of l m and l M with the majority of votes for l M. This may be due to a common feature, not covered by the impulse area parameter, namely the fast decrease of intensity from a high level on the last vowel.

The stimuli behind matrices 3-5 are speech-like and the scores in the diagonals have improved considerably. Matrix 3 with responses to the 'rhythmic speech' patterns in monotonic Fo contours, shows that all the three two-word phrases are distinguishable without any help from Fo. The compound, on the other hand, is confused with even accentuation, l m. It seems, then, that the speech-like character of the stimuli has brought out the similar temporal characteristic (3-2-3) which was not so clear earlier. For the asymmetry of the pair l-m/l m, see below. When the melody is superimposed, Matrix 4, the compound is well identified but the even accent pattern is confused with focus left. It is as if the similarity of the melodic patterns with *långa* in a fall-rise masks the difference in rhythm which kept them apart in the preceding acoustic outfit.

With added accentual movements, Matrix 5, the identification scores are very close to those of the prototypes in Matrix 6, including the confusion between l m and L m.

This confusion needs a special comment. We believe that the two-level contour used by our speaker over *långa* in even-accented (focus-free) l m is the unmarked, ambiguous case which may be associated with the same meaning and function as focus left. A special feature is needed for successful focus identification, and this special feature is a pattern which makes use of three Fo levels. (Three levels are typically used by the heard

intended

	lm	Lm	lM	l-m
lm	32	42	18	8
Lm	30	38	20	12
lM	12	62	5	22
l-m	50	27	15	8

1. Meter

	lm	Lm	lM	l-m
lm	24	5	62	9
Lm	40	33	13	13
lM	22	8	58	12
l-m	10	59	3	28

2. + Intensity

	lm	Lm	lM	l-m
lm	75	11	6	8
Lm	11	75	6	8
lM	31	2	64	3
l-m	53	9	0	38

3. + Speechwave

	lm	Lm	lM	l-m
lm	43	49	3	5
Lm	53	46	0	1
lM	0	0	99	1
l-m	0	4	6	90

4. + Melody
(piecewise level)

	lm	Lm	lM	l-m
lm	51	46	0	3
Lm	0	98	2	0
lM	0	3	95	2
l-m	8	0	5	87

5. + Accentual
movements
(piecewise linear)

	lm	Lm	lM	l-m
lm	36	64	0	0
Lm	2	94	0	0
lM	2	1	97	0
l-m	6	0	0	94

6. Prototypes

Figure 6. Confusion matrices for stripped stimuli and prototypes.

Stockholm speakers analysed by Bruce 1977. See also E1R above.)[3] The unmarked character of l m may also explain the large number of votes for l m in Matrix 3 and the asymmetric confusion of l m/l-m in the same matrix. When in doubt about the l-m stimulus, which at this stage has not received its special mark, half of the votes go to l m, which for reasons of semantic probability represents the more acceptable alternative.

Figure 7 gives an overview of how the the four prosodic patterns are recognized in their different acoustic outfits. The identification scores are presented as functions of our six parameters, starting with the most undressed one and ending with the fully dressed prototypes. The diagonals indicate what the functions would have looked like if every parameter had contributed to identification in the linear fashion that we expected. The zigzagging curves emphasize the lesson taught by the confusion matrices, namely that the relation between the parameters and the recognition of a pattern is far from this ideal. We shall summarize the most striking features.

The pattern l M is set apart from the others already by intensity but the compound l-m needs melody for its identification.The patterns l m and L m are identified in rhythmic speech but the identification is lost when melody is introduced. The identity of L m is fully restored in accented speech.

The curves of Figure 7 show that a parameter which is important in one acoustic outfit of the signal may not be important in another. It is therefore difficult and perhaps not meaningful to establish an overall rank order of the features. Instead, our results strongly suggest that for the identification of a pattern in this set of four alternatives, listeners rely on some special feature or a combination of features which distinguishes it from the others.

Table 2 focuses on the importance of the chosen parameters for pattern recognition, as reflected by the confusion matrices: Intensity (M2), speech rhythm (M3), melody, rise-fall or fall-rise (M4), accentual pitch movements (M5). 3 degrees have been considered in the identification process, 'set apart' marked | (around 50% and above), 'identified' + (around 70% and above), and 'recognized' ++ (around 90% and above). lm/Lm means confusion between lm and Lm.

[3] The three-level focus pattern is very typical of present-day Stockholmian. According to J. J. who coaches news reporters for Radio Sweden, this style of pronunciation is not recommended (personal communication).

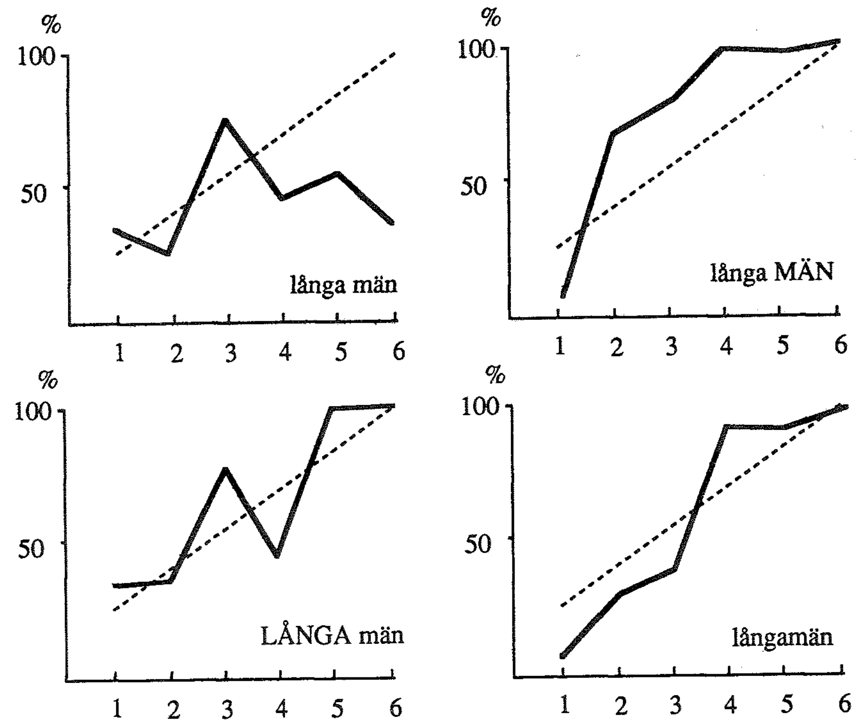


Figure 7. Recognition of four prosodic patterns as a function of acoustic correlates: 1 (Meter), 2 (+Intensity), 3 (+Speechwave), 4 (+Melody), 5 (+Accent), 6 (Prototypes).

Conclusions

1. For the identification of a pattern, listeners rely on some special feature or a combination of features which distinguishes it from the others.
2. There seems to be no perceptual rank order of acoustic features. Instead, the importance of an acoustic feature for the recognition of a prosodic pattern may depend on the set of features available in the signal.
3. Asymmetric confusions may indicate that one member of a confused pair has a special acoustic mark lacking in the other member. The results of our stripping experiments, as presented in Table 2, suggest that this special mark may be due to intensity (M2), speech rhythm (M3), melody (M4: Rise-Fall or Fall-Rise) or accentual pitch movements (M5).

Table 2. Pattern distinguishing features

	M2	M3	M4	M5
l m		+	lm/Lm	lm/Lm
L m		+	lm/Lm	++
l M	l	l	++	++
l-m			++	++

Summary of results and general discussion

Although our experiments have not followed a strict plan we shall attempt to summarize the results in a coherent form.

The temporal, intensity and spectral properties of a prosodic pattern have to be in phase with the Fo contour to secure good identification (E0, E1). Although Fo is known to have a strong perceptual cue value, a prosodic pattern appears as a sequence of events in which each event is composed of a tight bundle of acoustic features which cannot be separated from each other without impairing the identification of the pattern (E3). The time structure of these events is important. In so far as a prosodic phrase has a specific rhythm (here defined by the rhythmic speech wave parameter), the rhythmic pattern alone makes it recognizable in a monotonic Fo contour (E3). The importance of the rhythmic pattern is also noticeable in the results of the peak shift experiment which show how a given carrier favours the recognition of the patterns which have a rhythm compatible with that of the carrier phrase (E1).

An overall impression, then, is that pitch does not have the expected domineering role in the identification process. Rather different acoustic correlates have different importance for different phrase patterns (E3). Intensity plays an important role for at least one of the patterns (l M) and the rhythmic speech properties are sufficient for listeners to separate l m, L m, l M in a monotonic Fo contour (E3).

As has been shown in early perceptual experiments with Norwegian accents (Efremova, Fintoft and Ormestad 1963 and Fintoft and Martony 1964), the important segmental carriers of a pitch pattern are the vowels. This is borne out by recent work by House 1990 and our own results which show that when the pitch movements over the vowels are 'right' (prototypical), a prosodic pattern is easily recognized (E1).

Rises and falls of the same pitch interval in two-level contours are sufficient to make listeners distinguish between the contours of the deaccented, accented and compound prototypes. For the distinction between even and focal accentuation, l m/ l M, on the other hand, two different pitch intervals seem necessary in a three-level contour (El R and E3). Hence the labels H and L are not sufficient for the representation of these patterns.

The pitch movements are accompanied by rather fixed positions of the Fo turning points relative to the vowels. These turning points, which have played an important role in our production-oriented model of intonation (Bruce & Gårding 1978) cannot be regarded as perceptual units even if they are likely to have an important function in the auditive process.

The 'timing of a peak or a valley' can only be observed in the acoustic record and notions like 'delayed timing' of turning points (Ladd 1983) do not fit in the perceptually oriented model of intonation that we are aiming at. From a listener's standpoint it is the shape (fall, rise, level) of the Fo movement in the vowel that counts and for a prosodic phrase it is the combination of such shapes in a proper rhythm which is crucial.

Considering the general nature of perception it is desirable that all basic units be stated in relative terms, also the combination of the pitch shapes that constitute a prosodic phrase pattern. Strong candidates for such basic units are the quotients between the pitch values over neighbouring vowels (E2). This finding is worthy of further investigation.

The occurrences of asymmetric confusions (E3) remind us of the notion of markedness. In the present context we take it to mean that one member of a pair bears a characteristic feature which prevents confusion with the other member but not confusion in the other direction.[4]

The confused member can be regarded as the unmarked (default) case capable of carrying more than one meaning, whereas the special feature marks the other member for a particular meaning.

We shall end with diagrams summarizing perceptually oriented features for the four prosodic phrases (Fig. 8). The row below the transcription

[4] The concept of markedness for prosodic patterns was found useful in earlier studies. 'Internal juncture' was defined as a marked syllable boundary in a phrase. The definition was based on perceptual tests which showed that a pronunciation following natural syllabification rules was ambiguous as to morpheme boundaries whereas a pronunciation going against these rules gave a unique solution (Gårding 1967).
The situation is similar for accentuation patterns: *Anders tankar på gården* with the same accentuation level on *Anders* and *gården* is unmarked and ambiguous. It may mean 'Anders gets gas (*tankar*) in the yard' and 'Anders's thoughts of the yard'. An added terminal juncture after the noun phrase *Anders*, manifested as rising phrase intonation and/or preboundary lengthening restricts the meanings of the sentence (Gårding 1989).

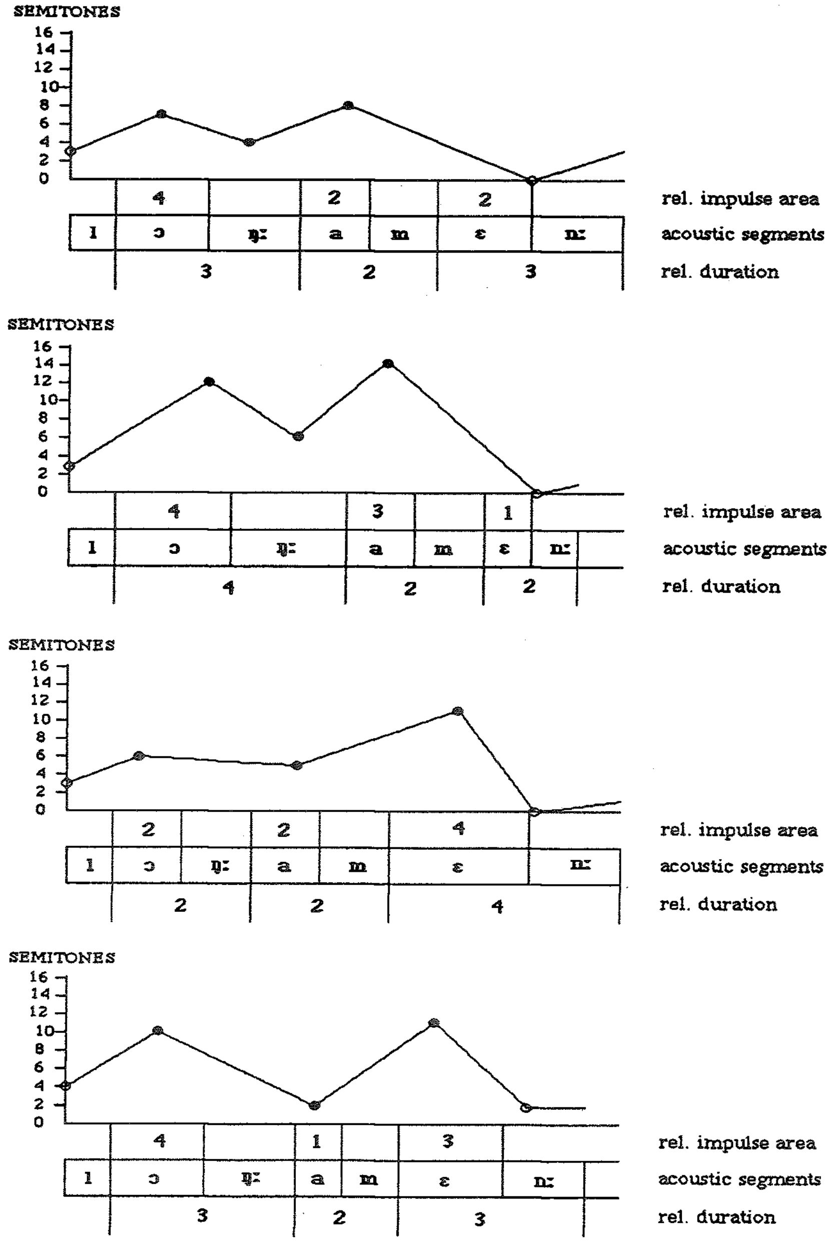


Figure 8. Perceptually oriented features of the accentuation patterns of E3.

gives relative durations of segments stretching from vowel onset to vowel onset and the row above the text presents relative impulse areas (dB envelopes) over the vowels. For an individual pattern the values of these parameters go in the same direction. The unfilled circles indicate the beginning and the end of the common phrase intonation. The different patterns of accentuation are represented by the three filled circles showing the pitch values of the main peaks and the main valley in a perceptual (semitone) scale. Note the different arrangement of these pitches in relation to the text depending on the location of focus. In (a,b) the highest pitch is over /a/, in (c) over /ɛ/ and in (d), the compound, the pitches over /ɔ/ and /ɛ/ are similar.

It is noteworthy that with the perception oriented description we are not far from the impressionistic notation used by the classic scholars of Scandinavian languages. For an overview of the treatment of Swedish stress-patterns 'tryckmönster' see Elert 1970.

Based on our results we conclude that a prosodic phrase has an accentuation pattern, i.e. a rhythmic structure (in our E3 experiments represented by meter, intensity and speechwave) to which a typical sequence of pitches is coordinated. This pattern is an essential part of the perceptual characteristics of a prosodic phrase.

Acknowledgement

Thanks are due to our listeners, to David House for comments on an earlier version, to Bob McAllister for help with the administration of some of the tests in Stockholm, to Mats Nyström for a discussion about statistics, and to Anders Kambrin for technical assistance.

References

Bruce, G. 1977. *Swedish word accents in sentence perspective.* Travaux de l'Institut de linguistique de Lund XII, Lund: Gleerup.

Bruce, G. 1983. 'Accentuation and timing in Swedish'. *Folia Linguistica* 17, 221-238.

Bruce, G. & E. Gårding. 1978. 'A prosodic typology for Swedish dialects'. In E. Gårding, G. Bruce & R. Bannert (eds.), *Nordic Prosody,* 219-228. Travaux de l'Institut de linguistique de Lund XIII, Dept. of Linguistics. Lund University.

Efremova, I. B., K. Fintoft & H. Ormestad. 1963. 'Intelligibility of tonic accents'. *Phonetica* 10, 203-212.

Elert, C.-C. 1970. *Ljud och ord i svenskan.* Stockholm: Almqvist & Wiksell.

Fant, G. & A. Kruckenberg. 1989. 'Preliminaries to the study of Swedish prose reading and reading style'. *STL-QPSR* 2, 1-83.

Fintoft, K. & J. Martony. 1964. 'Word Accent in East Norwegian. *STL-QPSR* 3, 8-15.

Fry, D. B. 1958. 'Experiments in the perception of stress'. *Language and Speech* 1, 126-152.

Gårding, E. 1967. *Internal juncture in Swedish.* Travaux de l'Institut de phonétique de Lund VI, Lund: Gleerup.

Gårding, E. 1985. 'In defence of a phrase-based model of intonation'. *Working Papers* 28, 1-18. Dept. of Linguistics and Phonetics, Lund University.

Gårding, E. 1989. 'Intonation in Swedish'. *Working Papers* 35, 63-88. Dept. of Linguistics and Phonetics, Lund University.

Gårding, E. & L. J. Gerstman. 1960. 'The effect of changes in the location of an intonation peak on sentence stress'. *Studia Linguistica* 14, 58-60.

Gårding, E. & D. House. 1987. 'Production and perception of phrases in some Nordic dialects'. In P. Lilius & M. Saari (eds.), *The Nordic Languages and Modern Linguistics* 6, 163-175. Helsinki University Press.

Gårding, E. & M. Stenberg. 1990. 'West Swedish and East Norwegian intonation'. In K. Wiik & I. Raimo (eds.), *Nordic Prosody* V, 111-31. University of Turku.

House, D. 1990. *Tonal perception in speech.* Travaux de l'Institut de phonétique de Lund 24. Lund University Press.

House, D., G. Bruce, L. Eriksson & F. Lacerda. 1988. 'Recognition of prosodic categories in Swedish: Rule implementation'. *Working Papers* 33, 153-161. Dept. of Linguistics, Lund University.

Howie, J. M. 1976. *Acoustical studies of Mandarin vowels and tones.* Cambridge University Press.

Kohler, K.J. 1987. 'Categorical pitch perception'. In Ü. Viks (ed.), *Proc. of 11th Int. Congr. of Phonetic Sciences* 5, 331-333. Tallinn: Academy of Sciences of the Estonian S.S.R.

Ladd, D.R. 1983. 'Phonological features of intonational peaks'. *Language* 59, 721-59.

Malmberg, B. 1955. 'Observations on the Swedish word accent'. Haskins Laboratories Report. Mimeographed.

Purcell, E.T. 1976. 'Pitch peak location and the perception of Serbo-Croation word tone'. *Journal of Phonetics* 4, 265-270.

Strangert, E. 1985. *Swedish speech rhythm in a cross-language perspective.* Stockholm: Almqvist & Wiksell.

Zetterlund, S., L. Nordstrand & O. Engstrand. 1978. 'An experiment on the perceptual evaluation of prosodic parameters for phrase structure decision in Swedish'. In E. Gårding, G. Bruce & R. Bannert (eds.), *Nordic Prosody*, 15-23. Travaux de l'Institut de linguistique de Lund XIII.

Lund University, Dept. of Linguistics
Working Papers 38 (1991), 71-82

Choosing Aspect in Automatic Translation into Russian and Polish

Barbara Gawrońska

The paper presents an experimental procedure choosing the perfective/imperfective aspect in automatic translation from Swedish and English into typical aspect languages: Russian and Polish. The program described is based on the assumption that there are certain similarities between the (in)definiteness of Swedish/English NPs and the Slavic aspect. Both categories (aspect and definiteness) may be related to the conceptual distinction between unique referents and referents which are unmarked as to their uniqueness. The uniqueness-based approach takes into account both sentence-internal cues for aspect choice and the linguistic context of the sentence to be translated. A kind of knowledge representation is utilized as well.

Introduction

Russian and Polish – two of the five languages involved in the experimental MT-system SWETRA (Dept. of Linguistics, Lund University) are known as typical aspect languages. The lexical inventory of both Russian and Polish contains aspectually marked verb pairs, i.e. each verb (except a small group of biaspectual verbs) is inherently either perfective or imperfective. The distinction is usually marked by a prefix (Pol: *czytać/przeczytać*, R: *čitat'/pročitat'* 'to read imp/perf') or a change in the stem, e.g. Pol: *podpisać/podpisywać*, R: *podpisat'/podpisyvat'* 'to sign perf/imp', Pol: *brać/wziąć*, R: *brat'/vzjat'* 'to take imp/perf'. This means that a translator formulating a Polish/Russian equivalent of an English VP almost always has to choose between two members of a certain verb pair. Human translators, who are native speakers of Russian or Polish, normally perform this task without difficulty. What cues do they use when deciding which aspectual variant fits into the given context properly? Can the principles for aspect choice be formalized and used in an MT-system?

The aspect category as a linguistic problem

Do all languages express the category of aspect in some way? What exactly is expressed by this category? Questions like these have been discussed in an enormous number of works in general linguistics. Nevertheless, little agree-