PRODUCTION AND PERCEPTION OF PROSODIC PROMINENCE

Jacques Terken

Institute for Perception Research, Eindhoven,
The Netherlands

ABSTRACT

The variation of prosodic prominence in human speech is ascribed both to pragmatic and lexical/metrical factors. In order to account for the influence of pragmatic factors, the traditional Given/New distinction must be replaced by a hierarchical ordering, reflecting the relative importance of expressions. In addition, prominence differences serve a demarcative function. On the perceptual side, the demarcative function seems more relevant than the pragmatic function: it seems unlikely that listeners can use prominence differences to recover fine gradations in relative improtance.

1 INTRODUCTION

Variation of prosodic prominence is an important feature of natural speech. By this we do not only mean the presence or absence of prominence such as established by the distribution of pitch accents, but also the relative differences in prominence for accented syllables. Synthetic speech which does not contain such relative differences in prominence sounds rather dull.

This observation immediately raises a number of questions, both with respect to the production and perception of prosodic prominence. This paper will briefly discuss some of these questions. Before doing so, we will first indicate how the present discussion relates to the traditional classifications of prosodic prominence.

Prosodic prominence is usually defined in terms of variation in duration, F0 and amplitude. In the frequency domain, prominent speech units coincide with appropriately timed F0 changes (or local F0 maxima or minima). In the temporal domain, prominent speech units are lengthened in comparison with non-prominent units: unit duration exceeds the duration that would be expected on the basis of speech rate, phonological class and phonemic context if the unit were non-prominent. The perceptual tolerance for temporal variation is quite large (Nishinuma & Duez, 1989), and F0 variation appears to be the most reliable acoustic correlate for prosodic prominence. For that reason, prominence is usually discussed in terms of the distribution of pitch accents: speech units can be either prominent by virtue of the presence of a pitch accent or non-prominent if there is no pitch accent. This makes prominence a binary feature. However, a finer differentiation can be made for prominent speech units. Liberman & Pierrehumbert (1984) present evidence that speakers can very reliably comply with the instruction to make a word more or less prominent. The effect is that the F0 maximum increases if the speaker is asked to pronounce the word with a greater "degree of overall emphasis or excitement".

Our purpose is to incorporate such quantitative differences between accented speech units into the treatment of prosodic prominence.

2 THE PRODUCTION OF PROSODIC PROMINENCE

Traditionally, prosodic prominence has been related to information structure (e.g. Halliday, 1967): Given information is expressed by unaccented, i.e. non-prominent expressions, and New information by accented, i.e prominent expressions. In this treatment, no satisfactory account was given as to the location of accents (i.e. the location of prosodic prominence) within the expressions conveying New information. Later treatments, building on this framework, have related prosodic prominence to the focus structure of the discourse. Expressions can be $[+focus]$ or $[-focus]$. The assignment of $[\pm focus]$ is driven by pragmatic factors such as the Given/New status. Within a $[+focus]$ expression, metrical rules determine the position of the accent. These metrical rules are sensitive to syntactic properties of the sentence, such as functor-argument relations. Stylistic considerations determine whether additional words or syllables will be accented in $[+focus]$ expressions. In this way, the search for factors determining the assignment of focus can be separated from the search for rules determining the assignment of accents within focal expressions.

Now, the question is to which level relative differences in prominence must be ascribed. There are two broad classes of models:

1. according to one class of models, the speaker may decide that not all $[+focus]$ expressions are to be focussed upon to the same extent, for reasons which have to do with the pragmatic context and/or the thematic structuring of the sentence. The consequence is that we replace the binary feature $[\pm focus]$ by a n-ary valued feature $[\alpha focus]$. The mapping of $[\alpha focus]$ onto F0 values is done by means of a grid. We will call this the VARIABLE FOCUS view;

2. according to a different class of models, the speaker assigns $[\pm focus]$ to expressions on the basis of pragmatic considerations, and prominence differences originate from the mapping of $[+focus]$ onto F0 values. Particular models may differ in the way this mapping is conceived of. Some models attribute prominence differences to the outcome of the metrical rules which determine the location of accents in $[+focus]$ expressions. Essentially, this boils down to replacing the binary feature $[\pm accent]$ by a n-ary feature $[\alpha accent]$. Other models attribute prominence differences to lexical factors: each part of speech has associated with it a fixed prominence (cf. Allen, Hunnicutt & Klatt, 1987). Still other models attribute prominence differences to the operation of prosodic rules such as downstep. We will call this the MAPPING view, since it ascribes prominence differences to factors which come into play when $[+focus]$ is mapped onto F0 values.

Before we can discuss the different options, the status of α must be considered. From a linguistic point of view, it is implied that α is a nominal variable, i.e. the different values that α can take are qualitatively different and distinctive. We want to be more lenient and to avoid these implications. Instead, we consider α as an ordinal variable: the values of alpha are defined as the set {more, less, equal}. Also, the requirement of distinctiveness must be replaced by a requirement that different values of α are associated with different felicity conditions.

The variable focus view accounts for prominence differences between focal expressions, since it assigns to each focal expression a particular value for α. Support for this view can be found in different sources. In the introductory section, we have already referred to Lieberman & Pierrehumbert (1984). Kruijt (1985) shows that F0 maxima are lower in focal expressions referring to Given referents than in focal expressions referring to New referents (Given is used in the sense of "mentioned in the immediately preceding context",

cf. Chafe, 1974; Brown, 1983; Terken, 1985). Wells (1986) describes an experiment in which listeners were presented with utterances isolated from their context. He found that rankings of the relative importance of information conveyed by expressions in the utterances were systematically related to the prosodic characteristics of the expressions. Thus, it appears that prosodic features can be used by speakers to convey gradations in focus. Needham (1990) shows that focal expressions referring to non-typical parts of a previously mentioned whole are associated with higher F0 maxima than the same focal expressions referring to typical parts (we assume that this is a priming effect).

These findings can be accounted for in a natural way, if we take inspiration from proposals within the area of computational linguistics, which are intended primarily for anaphora resolution (e.g. Asher & Wada, 1988; Hajicova, Kubon & Kubon, 1990). They elaborate on the intuition that a hierarchical ordering can be established on the items which have been made accessible by the discourse. Anaphora resolution is guided by the hierarchical ordering of the set of candidate antecedents for a given anaphoric expression. For instance, Hajicova *et al.* describe an algorithm by which a salience index can be computed for the set of accessible items (i.e. the stock of shared knowledge).

In order to relate these proposals to relative differences in prosodic prominence, we must extend the idea of a hierarchical ordering to the information which is to be transmitted by the speaker:

1. for items in the stock of shared knowledge, prosodic prominence is directly related to their hierarchical ordering: the most accessible items are least prominent;

2. for items which are to be transmitted, a hierarchical ordering is established by the relative weight which is assigned to them by the speaker; this relative weight is affected by thematic roles (more central roles carry more weight than less central ones) and priming effects (information which can be more easily activated from the information in the stock of shared knowledge carries less weight than information which can be less easily activated); prosodic prominence is highest for items which are highest in the hierarchy.

If this account of prominence differences between focal expressions in terms of pragmatic factors is appropriate, prominence differences within focal expressions require a different explanation; such within-expression differences cannot be accounted for in terms of different values of $[\alpha focus]$, since we have assumed that $[\alpha focus]$ is assigned to each focal expression as a whole. It can be argued, however, that such within-expression differences do not emerge from pragmatic influences but from lexical and/or phonological factors. This means that the mapping view would be appropriate for relative differences in prominence within focal expressions.

Terken (1991b) describes findings related to the issue of prominence variation in read aloud text. The materials consisted of referring expressions embedded in texts read aloud by a professional speaker. Although there were clear differences in prominence within these expressions, no general pattern emerged. In order to determine the perceptual tolerance for different prominence patterns, listeners were presented with manipulated versions of expressions containing three accented words, embedded in their sentential context. The expressions constituted the maximal projections of Noun Phrases or Preprositional Phrases.

In general, listeners had equal preference for two different prominence patterns. In one pattern, which would be predicted on the basis of rhythmic alternation (Monaghan, 1988), maximal prominence was associated with the "edges" of the expression. In the other pattern, maximal prominence was associated with the left edge of the expression, and prominence decreased as a function of serial position. This pattern more resembles a downstep pattern. Both patterns were strongly preferred over a pattern in which maximal prominence was on the middle of three accented words. In both cases, prominence differences can be said to have a demarcative function: words with major prominence are at both edges or at the left edge of a syntactic constituent.

In addition, preferences for a particular prominence pattern were modified by lexical factors: if the word at the left edge was a semantically weak adverb such as "rather", there was a preference for reduced prominence at the left egde, and major prominence was shifted to the middle word.

These findings are supported by unpublished data from read aloud isolated utterances. In these data, a typical decrease in prominence appears to be associated with serial order within referring expressions containing several accented words: each following accent is less prominent than the preceding one.

As a consequence, we have two factors which contribute to differences in prosodic prominence: pragmatic factors govern the relative prominence of $[+focus]$ expressions, and lexical and phonological factors affect the relative prominence of accented words within $[+focus]$ expressions. The phonological factors concern the demarcation of syntactic constituents.

3 THE PERCEPTION OF PROSODIC PROMINENCE

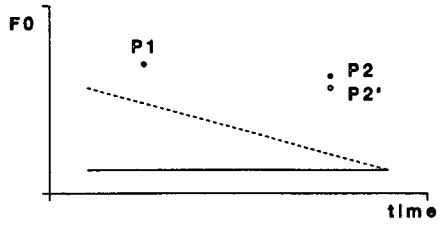
If listeners should be able to recover the value of $[\alpha focus]$ from relative differences in prominence, we must assume that they can establish prominence differences between non-adjacent pitch accents. Due to the transient character of the speech signal, it seems unlikely that this can be done on the basis of F0 values directly. Therefore, it must be assumed that relative prominence is recoded in terms of a grid-like structure, where the grid defines a set if "iso-prominence" curves. However, as Lieberman demonstrated already in 1965, even expert listeners cannot do so reliably on the basis of acoustic information.

This implies that, although there may be reliable relations between $[\alpha focus]$ and relative differences in prominence on the part of the speaker, it is unlikely that a one-to-one mapping of $[\alpha focus]$ can be recovered by the listener. Instead, we assume that listeners can give rather accurate judgments of relative differences in prominence between pairs of prominent syllables. For two successive accents A and B, A can be more or less prominent than B, or they can be equally prominent. If A and B are equally prominent, or B is less prominent than A, and there is a third accent C which is less prominent than B, from this it follows that C is also less prominent than A. However, if B is more prominent than A, the relation between A and C cannot be established.

If this is valid, it would imply that relative differences in prominence have primarily a demarcative function by telling the listener when a new constituent starts, and that the signalling of focus strength is of secondary importance. Only if successive constituents each contain just one accented word, can the speaker signal focus strength by means of relative differences in prominence.

Further experiments are needed to determine whether this picture is valid. Before conducting these tests, we need more insight into the perception of prominence differences. The primary acoustic correlates of prosodic prominence are well-known, but it is not fully clear how they contribute to the perception of prominence differences. In particular, there is no fixed procedure to determine which one of two accented words will be perceived as more prominent. Since pitch information will play an important part in such a procedure, an experiment was done addressing the question of how F0 variation contributes to prominence (Terken, 1991a).

In the experiment, utterances containing two accented syllables were presented to listeners. The two accented syllables were each associated with a frequency maximum, P1 and P2, respectively. The temporal distance between the two accented syllables was kept constant. The listeners were asked to adjust P2 so that the second accented syllable was judged to have the same prominence as the first accented syllable, for different values of P1. There were two different conditions. In one condition there was no baseline declination and different values of P1 were associated with variations in the distance between the topline and the baseline. In the other condition, the slope of the baseline was varied, so that the distance between topline and baseline within the utterance remained constant, but the scaling of the contour within the overall frequency range varied. A schematic representation of the results is shown in Figure 1. Here, for a given P1, the corresponding P2 which gives the same prominence as P1 is shown for the conditions with baseline declination (P2', the dashed line indicates the baseline, the open circle indicates the position of P2' giving the same prominence as P1) and without baseline declination (solid line, filled circle). P2 is adjusted to lower values in utterances with baseline declination than in utterances without baseline declination.



As can be seen from the schematic representation, a steeper slope of the baseline was associated with an upward shift of the initial part of the contour within the overall frequency range. Now, we assume that the speaker, when he is higher up in the overall range, has less room to bring about prominence variations by means of F0 variation due to a ceiling effect. In general, the range of F0 values employed and the position of the latter part of the contour within the overall range vary little for a given speaker. From these considerations, it may be concluded that the prominence associated with a given F0 maximum is affected by the actual frequency range employed by the speaker: if he can employ a small frequency range in the beginning of the utterance due to an upward shift of the baseline, it appears that the listener also expects a small frequency range near the end of the utterance. Further experiments are required to find out how these results generalize to utterances with three accents and to varying time intervals between the accented syllables. On the basis of these additional experiments, perceptual tolerances can be established for F0 variation, and questions can be anwered with respect to the communcative function of relative differences in prosodic prominence.

4 CONCLUSION

On the production side, prominence differences appear to be associated both with pragmatic factors affecting the information status of focal expressions, and with lexical and phonological factors affecting the relative prominence of accented words within focal expressions. In order to account for the pragmatic factors, the notion "information structure" must be defined in terms of a hierarchical ordering instead of the binary "Given/New" distinction.

On the side of the listener, it seems unlikely that this hierarchical ordering can be recovered from the speech signal on the basis of differences in prominence. Here, prominence differences appear to have primarily a demarcative function. Only in relatively simple sentences, may prominence differences help the listener to recover intended differences in relative importance.

References

Allen, J., Hunnicutt, M.S. & Klatt, D. (1987). *From text to speech: the MITalk system*, Cambridge U.P.

Asher, N. & Wada, H. (1988). A computational account of syntactic, semantic and discourse principles for anaphora resolution. *Journ. of Semantics*, **6**, 309-344.

Brown, G. (1983). Prosodic structure and the Given/New distinction. *Prosody: Models and Measurements*, eds. A. Cutler & R.D. Ladd, Springer Verlag, Berlin, 67-77.

Chafe, W.L. (1974). Language and consciousness. *Language* **50**, 111-133.

Hajicova, E., Kubon, P. & Kubon, V. (1990). Hierarchy of salience and discourse analysis and production. *Coling 90, vol. III*, ed. H. Karlgren, 144-148.

Halliday, M.A.K. (1967). Notes on contrastivity and theme II. *Journal of Linguistics* **3**, 199-244.

Kruyt, J.G. (1985). *Accents from speakers to listeners. An experimental study of the production and perception of accent patterns in Dutch*, Unpubl. Doct. Diss, Leyden.

Liberman, M. and Pierrehumbert, J. (1984). Intonational invariance under changes in pitch range and length. *Language sound and Structure*, eds. M. Aronoff and R. Oehrle, MIT Press, Cambridge, 157 - 233.

Lieberman, P. (1965). On the acoustic basis of the perception of intonation by linguists, *Word* **21**, 40-54.

Monaghan, A.I.C. (1988). Generating intonation in the absence of essential information. *Proceeding of Speech '88, Seventh FASE Symposium* eds. W.A. Ainsworth and J.N. Holmes, 1249-1256.

Needham, W.P. (1990). Semantic structure, information structure and intonation in discourse production. *Journ. of Memory and Language* **29**, 455-468.

Nishinuma, Y. & Duez, D. (1989). Perceptual optimization of syllable duration in short French sentences. *Proceedings of European Conference on Speech Communication and Technology, Eurospeech 89*, eds. J.P. Tubach & J.J. Mariani, Vol. II, 694-697.

Terken, J. (1985). *Use and Function of Accentuation. Some Experiments.* Unpublished doctoral dissertation, University of Leyden.

Terken, J. (1991a). Fundamental frequency and perceived prominence of accented syllables. *JASA*, in press.

Terken, J. (1991b). Synthesizing natural sounding intonation for Dutch: rules and perceptual evaluation. *Computer Speech and Language*, submitted.

Wells, W.G.H. (1986). An experimental approach to the interpretation of focus in spoken English. *Intonation in Discourse*, ed. C. Johns-Lewis, Croom Helm, London, 53-75.