SOME OBSERVATIONS ON THE TIMING OF FO-EVENTS

Bertil Lyberg

Swedish Telecommunications Administration
Technology Department, S-123 86 Farsta, Sweden.

ABSTRACT

The present study examines the effects which final consonants have upon the timing of the fundamental frequency contour in words carrying the sentence accent in Swedish. Monosyllabic test words containing both phonologically long and short vowel segments are placed in initial and final utterance positions. Results show that the timing of the FO-events that signal the sentence accent is dependent on whether the consonant following the vowel is voiced or not, especially when the vowel is phonologically short. The fundamental frequency fall in the case of a short vowel followed by an unvoiced consonant has to occur earlier than in the other cases in order to get the frequency fall within the vowel segment, otherwise the prosodic information will get lost.

INTRODUCTION

The fundamental frequency contour of an utterance is heavily influenced by the segmental composition. In investigations about the fundamental frequency contour utterances built up of only sonorants are often used [1] and thereby the influence of constrictions in the vocal tract is avoided or at least diminished. It is assumed that the fundamental frequency contour obtained in such a way will reflect some basic pattern that is perturbed by the segmental composition in ordinary utterances. Most studies of the fundamental frequency contour are dealing with overall patterns i.e. in which syllables maxima and minima will occur in the segmental flow [1] [2]. In models for generating an accurate fundamental frequency contour it is also necessary to take into account how the location of the maxima and minima is affected by the segmental composition. The exploration of such effects in greater detail is hopefully of great importance for the generation of synthetic speech with a naturalness and intelligibility that is acceptable in different types of communication systems. A systematic mapping of the variations of the locations of the extremes owing to the segmental setup can also be expected to provide some insight into certain aspects of the mental processes underlying the temporal organization of spoken language.

In the present investigation the fundamental frequency contour associated with sentence accent is studied in greater detail. The syllable structure and word position are systematically varied and the effects on the location of the extremes are examined.

Some fundamentals of Swedish prosody

The fundamental frequency contour of a monosyllabic word carrying sentence accent will in phrase final position have a maximum point in the vowel followed by a minimum point in the vowel or the following consonants. For a more detailed description of the fundamental frequency contour in different positions of Swedish utterances, see e.g. Bruce [1] and Lyberg [3]. The fundamental frequency manifestation associated with the signalling of sentence accent in Swedish seems to be very similar to the corresponding frequency manifestation in American English according to e.g. Pierrehumbert [2]. There are, however, discrepancies in the interpretation of the underlying parameters, and in the terminology used by the two authors Bruce and Pierrehumbert.

Two degrees of quantity are distinctive in Swedish, the short/long distinction. There is also a complementary distribution of phonological length between vowels and consonants in stressed syllables. A long vowel is in stressed syllables followed by a short consonant and a short vowel by a long consonant [4].

EXPERIMENTAL DESIGN

Speech material

A set of utterances containing one and three lexical main stresses was constructed. The test word was in the case of three-word sentences placed in both initial and final positions and the sentences were pronounced either with the test word in focus or with a "neutral" stress pattern i.e. with a conscious effort of the speaker to avoid junctures and contrastive stresses.

The test word was build up of both phonologically short and long vowel segments in order to elucidate the interaction between the signalling of the quantity distinction and the fundamental frequency contour. In addition to that the surrounding consonants were varied in a systematic way so that both voiced and unvoiced consonants occurred in postvocalic position. The test words were always monosyllables and may be considered as nonsense words. The seminonsense three-word utterances where built up of both nonsense (test words) and semantically non-anomalous words (always /såg/ saw in English).

The inventory and syntactic structure of the test sentences are presented in Table I. The sentences were read in ten randomly ordered sequences by a trained phonetician.

TABLE I

Sentences	Syntactic structure
{Dad, Dadd, Dat, Datt}	S – NP – N
{Dad, Dadd, Dat, Datt} säg {dad, dadd, dat, datt}	S [NP [N] VP [V NP [N]]]

+ A phonologically short vowel is in the orthography followed by two consonants.

Measurements

The duration of the vowel segment in the test words was measured. The duration of the vowel segment is defined as the interval between the release of the consonant preceding the vowel (always /d/; a rapid increase of intensity) and the occlusion of the following consonant (always /d/ or /t/; a rapid decrease of intensity).

The fundamental frequency was measured in nine equally spaced points of the vowel segment in the test words, at the beginning, at the end, and at seven points within the vowel segment.

OBSERVATIONS

The fundamental frequency contour is in figs. 1 and 2 shown for the final position of the three-word utterances. The diagram in fig.1 shows the fundamental frequency contour when the utterance is pronounced with a "neutral" intonation pattern and the diagram in fig.2 the frequency contour when focus is assigned to the final position of the utterance. Every point in the diagrams represents a mean value of ten recordings ot the same utterance. The maximum point and the following minimum point of the fundamental frequency contour will occur within the vowel segment no matter whether the following consonant is voiced or not and whether the vowel in question is phonologically long or not. The fundamental frequency contour after the minimum point is, in the case of a phonologically long vowel followed by an unvoiced consonant or a short vowel followed by a voiced consonant, more or less truncated in comparison to the frequency curve in the case of a long vowel followed by a voiced consonant. When a short vowel is followed by an unvoiced consonant the fundamental frequency fall will occur about 20 to 30 msec. earlier in the vowel segment in comparison to the other cases.

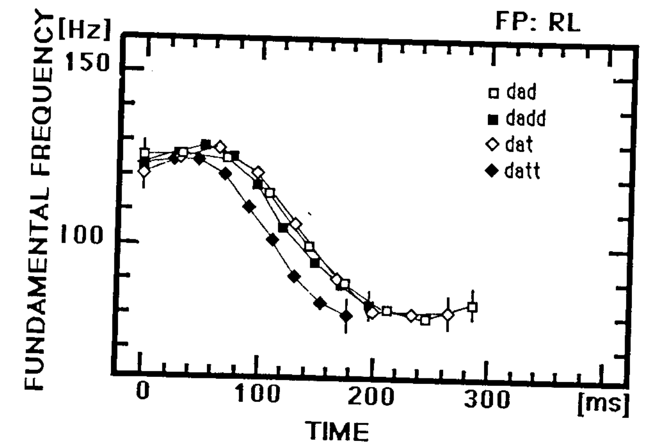

Fig. 1 The fundamental frequency contour of the vowel for different test words in final position of three-word utterances. The utterance is pronounced with a "neutral" intonation pattern.

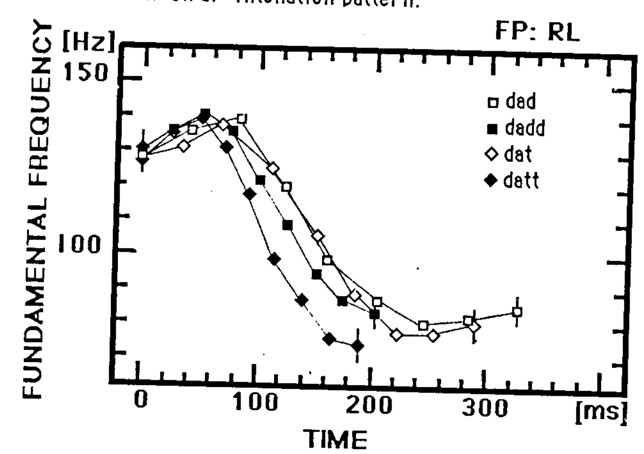

Fig. 2 The fundamental frequency contour of the vowel for different test words in final position of three-word utterances. Focus is assigned to the final position.

In the diagram in fig.3 the fundamental frequency contour is shown for the one-word utterances. The diagram shows that mainly the same timing difference of the fundamental frequency fall is apparent in these utterances. When a short vowel is followed by an unvoiced consonant the frequency fall will happen earlier in comparison to the other cases. A comparison between the fundamental frequency fall of a final test word carrying sentence accent in a three word utterance and a comparative test word in a one-word utterance shows that the fundamental frequency fall in the one-word utterance will happen later than in the three-word utterance.

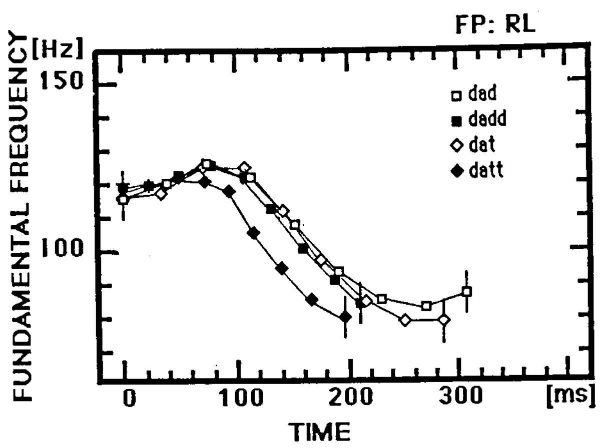

Fig. 3 The fundamental frequency contour of the vowel for different test words in one-word utterances

When focus is assigned to the initial word position the acoustic manifestation of the sentence intonation is in the studied utterances a maximum point of the fundamental frequency contour in the vowel segment of the test word followed by a minimum point, but in this case the minimum point seems to be located outside the vowel segment (fig.4). Most of the frequency fall is, however, still within the vowel segment. Some limited data from another speaker show a somewhat another strategy. For that speaker the fundamental frequency fall is more or less outside the vowel segment. The timing difference of the fundamental frequency fall for the different test words in this position is nevertheless the same as in the other word positions.

The duration of the vowel segment in the different test words is in fig.5. shown for different focus assignments of the three word utterances. The diagram shows that final lengthening is to a great extent dependent on the location of the focus position. When focus is assigned to the initial utterance position the duration of the vowel in the final test word is shortened. The speaker seems to have a prolongation of the vowel segment in the initial word position when it is in focus position that is more or less of the same magnitude as the lengthening process in utterance final position.

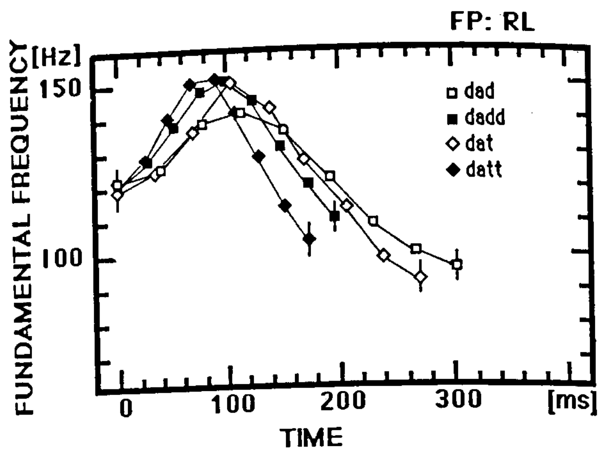

Fig.4 The fundamental frequency contour of the vowel for different test words in initial position of three-word utterances. Focus is assigned to the initial position.

dad

	-foc	+foc
-fin	216	304
+fin	245	330

dadd

	-foc	+foc
-fin	156	198
+fin	162	202

dat

	-foc	+foc
-fin	199	273
+fin	235	294

datt

	-foc	+foc
-fin	132	174
+fin	157	186

Fig. 5 The duration of the vowel segment in msec. is shown for different combinations of focus and utterance positions.

DISCUSSION

The main observation on the timing of the frequency fall connected with the signalling of sentence accent is that the frequency fall will occur earlier in a phonologically short vowel followed by an unvoiced consonant in relation to the CV-boundary than in the other cases. The duration of the vowel is in this case extremely short and the unvoiced consonant cannot convey any information about the fundamental frequency fall. When focus is assigned to the utterance final position and the final word consists of a monosyllable, the frequency fall has to occur within the vowel or within the vowel and the following consonants. When the following consonants are unvoiced, it seems to be necessary to move the frequency fall to an earlier point in relation to the CV-boundary in order not to lose the prosodic information.

It is in non-final position possible to partly locate the fundamental frequency fall to following syllables and words. The speaker in this study locates most of the frequency fall in the vowel but a limited study of another speaker seems to support the idea that it is a possible strategy for some speakers to locate the minimum point of the frequency fall in a following syllable. This is sometimes the case when a monosyllabic word is built up of unvoiced consonants after the vowel. The prosodic information is then signalled by the change of fundamental frequency level in the successive syllables. Similar data can be observed in American English [5].

CONCLUDING REMARKS

It seems possible to assume that an underlying intonation scheme is similar for sentences with the same prosodic pattern but built up of different segments and words. The timing pertubations observed seem to be possible to handle by means of adjustment rules on a more peripheral level. A complete intonation model of a language must at least include the following parts.

- An underlying intonation scheme.
- Timing perturbations owing to the syllable composition.
- Frequency perturbations owing to physiological factors such as e.g. inherent pitch.

The importance of accounting for the different types of perturbations of the fundamental frequency contour in the generation of synthetic speech to obtain a higher degree of naturalness and intelligibility must be determined by perceptual tests.

REFERENCES

[1] Bruce, G. (1977): "Swdish Word Accent in Sentence Perspective". Travaux de l'Institut de Linguistique de Lund XII. B Malmberg and K. Hadding (eds.), Lund : Gleerup.

[2] Pierrehumbert, J. (1980): "The Phonology and Phonetics of English Intonation". Ph.D. Thesis, MIT.

[3] Lyberg, B (1981): "Some observations on the vowel duration and fundamental frequency contour in Swedish". Journal of Phonetics 9, pp. 261-272.

[4] Elert, C.-C. (1964): "Phonological Studies of Quantity in Sweidsh", Almqvist & Wiksell, Uppsala, Sweden.

[5] Lyberg, B. (1984): "Some fundamental frequency perturbations in a sentence context". Journal of Phonetics 12, pp. 307-317.