Some Observations on Different Strategies for the Timing of Fundamental Frequency Events.

Bertil Lyberg† and Sonia Sangarig*

††Linköping University, Linköping, Sweden
†Telia Research AB, Farsta, Sweden
*Ericsson Mobile Communications AB, Linköping, Sweden

1. INTRODUCTION

The acoustic manifestations of the prosodic features are heavily influenced by the segmental compositions of the utterances. In order to study the acoustic correlates of the prosodic features it is necessary to diminish the influence of the segments. In investigations about the fundamental frequency contour utterances built up of only sonorants are often used and thereby the influence of constrictions in the vocal tract is avoided or at least diminished. In order to arrive at a detailed model of the fundamental frequency for e.g. speech synthesis it is necessary to exactly know what is happening in different segmental environments. This information can also shed light on the underlying mechanisms of the timing of the prosodic patterns. In the present study we are examining the effects which voiceless consonants have upon the fundamental frequency contour not only in words carrying the tonal word accents but also words carrying in sentence accent and terminal juncture. Three different hypotheses are considered the truncation, the timing adjustment and the rate adjustment hypothesis.

The results show that the truncation hypothesis seems to be of limited validity at least for a detailed description of what is happening in the fundamental frequency contour in Swedish.

2. SOME FUNDAMENTALS OF SWEDISH PROSODY

Duration

Two degrees of quantity are distinctive in Swedish, the short/long distinction. The duration of the short vowels is according to Elert (1964) about 65 per cent of the duration of the long vowels averaging over data from lists containing both single words and sentences. In stressed syllables the long vowels is followed by a short consonant and a short vowel is followed by a long consonant.

Pitch

In Swedish there are two kinds of tonal accents that a main stress syllable can have the acute accent (accent I) and the grave accent (accent II). The accent II pattern is typical of compound words and certain derivational forms. The acoustic differences between the two accents are found in the fundamental frequency contour, intensity and duration (Elert 1964).

The location of the extremes of the fundamental frequency in the segmental flow seems to be the most important parameter in signalling the distinction between accent I and accent II patterns (Bruce 1977).

3. HYPOTHESES

- The truncation hypothesis. There is no influence on the timing of the fundamental frequency events and the frequency contour is truncated in the unvoiced parts.
- The timing adjustment hypothesis. The timing of the fundamental frequency extremes are changed in such a way that the signalling part of the contour e.g. a fall will not be disturbed or truncated by the unvoiced part of the signal.
- The rate adjustment hypothesis. The rate of the fundamental frequency change is changed in

ABSTRACT

The acoustic manifestations of the prosodic features are heavily influenced by the segmental compositions of the utterances. In order to study the acoustic correlates of the prosodic features it is necessary to diminish the influence of the segments. In investigations about the fundamental frequency contour utterances built up of only sonorants are often used and thereby the influence of constrictions in the vocal tract is avoided or at least diminished. In order to arrive at a detailed model of the fundamental frequency for e.g. speech synthesis it is necessary to exactly know what is happening in different segmental environments. This information can also shed light on the underlying mechanisms of the timing of the prosodic patterns. In the present study we are examining the effects which voiceless consonants have upon the fundamental frequency contour not only in words carrying the tonal word accents but also words carrying in sentence accent and terminal juncture. Three different hypotheses are considered the truncation, the timing adjustment and the rate adjustment hypothesis.

The results show that the truncation hypothesis seems to be of limited validity at least for a detailed description of what is happening in the fundamental frequency contour in Swedish.
order to maintain the signalling part of the fundamental frequency contour within the voiced part.

4. EXPERIMENTAL DESIGN

A set of utterances containing three lexical main stresses was constructed in such a way that the main stress was either a monosyllabic test word or a disyllabic test word carrying the two different tonal word accents. These test words were systematically placed in all the possible word positions in the utterance. The segmental structure of the test words was varied in such a way that the vowels were followed by either a voiced or voiceless consonant. In the case of a following voiceless consonant, the consonant was always preceded by a short vowel in order to get as short vowels as possible in front of a voiceless consonant. If the subject need to do any modification of the fundamental frequency contour it ought to happen in this case.

In reading these sentences each speaker systematically assigned the focus to different positions of the utterances by imagining that he answered different key questions. For one of the sentences containing three main stresses the key questions and the focus assignments were:

Ve m vinner ålen?
Who wins the eel?

Mannen vinner ålen.
The man wins the eel.

Va d gjorde mannen med ålen?
What did the man do with the eel?

Mannen vinner ålen.
The man wins the eel.

Va d vinner mannen?
What did the man win?

Mannen vinner ålen.
The man wins the eel.

The inventory of utterances where the focus assignments are denoted by italics:

En man vann en ål.
(A man won an eel.)

Mannen vinner ålen.
(The man wins the eel.)

Mamman lånar nålar.
(The mother borrows the needles.)

Each sentence was read by the speaker substituting the words containing main stress one at a time by the nonsense words dad, datt, dadad, dattat (the disyllabic word pronounced with accent I). In the case of accent II it was very hard for the speakers to pronounce the nonsense test words and the test words were replaced by real words and in this case the baseline is a phonological short vowel followed by a long voiced consonant.

The sentences were read in five randomly ordered sequences by two male speakers substituting the words containing focus assignment by the nonsense words /dad/, /datt/, /dadad/ and /dattat/.

It was easier for the speaker to substitute only one word in the sentence by the nonsense word than to pronounce the whole sentence by means of reiterant speech (Nakatani and Schaffer 1978).

Measurements

The duration of the vowels in the test words was measured. The duration of the vowel segment is defined as the interval between the vowel onset (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 five equally spaced points in the stressed vowel segment in the test words, at the beginning, at the end and at three points within the vowel. The unstressed vowel in the disyllabic test words was measured in three equally spaced points, at the beginning, at the end and in the middle of the vowel.

5. OBSERVATIONS

In Figure 1 the fundamental frequency contour is shown for the test word /dad/ and /datt/ in focus position for different utterance positions. The maximum point will in the case of a short vowel followed by a unvoiced consonant (datt) occur about 30 msec. earlier in comparison to the test word with a long vowel followed by a voiced consonant (dad). It is also obvious that the fall of the fundamental frequency contour is to a certain degree truncated. To be noted is that the signalling of focus or sentence accent according to Bruce (1977) is a rise in the fundamental frequency
contour or a low followed by a high. The high point in the main
stress vowel is normally preceded by a low point in the vowel
preceding the focus position in the utterances.

The diagram in Figure 2 the fundamental frequency contour is
shown for the disyllabic test words in focus position. The rise of
the fundamental frequency is in the case of a short vowel fol-
lowed by a voiceless consonant earlier than in the utterance with
a long vowel followed by a short consonant. The fall in the short
main stress vowel is totally absent and is probably in this case
signalled by the following low point in the next syllable.

The diagram in Figure 2 the fundamental frequency contour is
shown for the disyllabic test words in focus position. The rise of
the fundamental frequency is in the case of a short vowel fol-
lowed by a voiceless consonant earlier than in the utterance with
a long vowel followed by a short consonant. The fall in the short
main stress vowel is totally absent and is probably in this case
signalled by the following low point in the next syllable.

6. CONCLUSIONS

In syllables built up of phonologically short vowels and followed
by voiceless consonants the subjects in this study are moving the
maximum point in the sentence accent to an earlier point and the
following fall has more or less the same slope as the fall in the
long vowel. The fundamental frequency will, however, not reach
the same low frequency value. That means that the results are
both in accordance with the truncation hypothesis and the timing
adjustment hypothesis. In the primary stressed syllable in the
accent II words there is a slight tendency for rate adjustment for
our speakers. This fact is in opposition to the results found by
Erikson and Alstermark (1972). Since the investigations, how-
ever, are based on a very limited material it seems possible to
assume that different speakers may use different strategies in
order to combine the prosodic signalling with the limited time
available to execute the fundamental frequency manifestation
owing to the phonologically short vowel and the following voice-
less consonant.

7. REFERENCES

1. Bruce, G. (1977): "Swedish Word Accent in Sentence Per-
spective". Travaux de l’Institut de Linguistique de Lund X.

2. Erikson, Y. and Alstrermark, M. (1972): "Fundamental fre-
quency correlates of the grave word accent in Swedish: The

3. Elert, C.-C. (1964): "Phonological Studies of Quantity in
Swedish", Almqvist & Wiksell, Uppsala, Sweden.

without words. Prosody cues for word perception". J.