Dept. for Speech, Music and Hearing
Quarterly Progress and Status Report

Word accent, emphatic stress, and syntax in a synthesis rule scheme for Swedish

Carlson, R. and Granström, B.

KTH Computer Science and Communication

journal: STL-QPSR
volume: 14
number: 2-3
year: 1973
pages: 031-036

http://www.speech.kth.se/qpsr

A mixture of the acute and grave rule systems will give a tentative system predicting the fundamental frequency pattern of Swedish nonsense word sentences without limitations concerning word accent. It should be stressed that the present rules describe only a certain subject's behavior and that dialectal and ideolectal differences have not yet been considered[4].

One important presupposition underlying our rule system is that the intonation contour can be split into two parts: sentence intonation and word stress marking. In statements pronounced by our subject the sentence intonation could be described simply as a linear fall in F_o with constant starting and end values. The stress marking contour is characterized by the location of minima and by a positive F_o wave that connects these minima. The amplitude of this wave was found to be related to the duration of the preceding stressed vowel (Fig. III-B-1). This system will predict the intonational pattern of sentences consisting only of acute accent words (Fig. III-B-2).

With respect to F_o there is a close similarity between the syllable with secondary stress in grave words and the stressed syllable in acute words. However, the durational pattern is different. Typically a vowel carrying secondary stress was found to be about 23 % shorter than a main stress vowel, everything else being equal[2]. This will influence the peak height of the F_o wave (Fig. III-B-1).

The primary stress syllable of a grave word has a different stress marking correlate. The pattern is the inverse of that of the acute accent, and the reference point is a maximum rather than a minimum. The value of the maximum seems to be a function of the duration of the word preceding the grave syllable. Now there is a contradiction between two hypotheses viz., the truncation hypothesis and the rate adjustment hypothesis [3]. The contradiction leads to the question, should the word intonation contour of primary stress syllables in grave accent words have a fixed shape which is truncated in accordance with the duration of the vowel or should the minimum govern the pattern so as to cause an adjustment of rate? In our system we have chosen the second hypothesis in order to make the rules as simple as possible. However, this is at variance with the earlier results[3]. The adopted rule might or might not be of perceptual importance.

The F_0 rules

The rules have as input the stress and tonal pattern and the duration of successive segments in a given sentence.

(A) Temporal location of

1. F_0 minima

i) Locate one minimum in the middle of each stressed vowel segment except primary stress vowels in grave accent words.

ii) Move the minimum to the vowel onset in sentence boundary words.

iii) Two additional minima are located at the beginning and end of the sentence.

2. F_0 maxima

Locate a maximum in the initial part of stressed vowels carrying primary stress in grave accent words.

(B) F_0 values of maxima

Compute the value of maxima (dealt with in stage A2) using the durational coefficient for the preceding word.

(C) Connect maxima and minima

1. Connect maxima and minima by half a period of a cosine wave.
2. Connect minima with no maxima inbetween by a complete, inverted period of a cosine wave. The period length is determined by the duration inbetween the two minima and the amplitude is a function of the duration of the stressed vowel containing the first minimum (Fig. III-B-1).

(D) Add the sentence intonation component

For our subject: a linear fall from 120 Hz to 90 Hz.

An illustration of how these rules function is shown in Fig. III-B-3. It should be noted that the duration has been normalized for the sentence carrying the grave accent word in order to get synchrony between the two sentences.

Possible expansion of the model

Within the present descriptive framework it appears possible to make some generalizations aiming at a synthesis of a wider class of sentence realizations. Alterations of the phrase contour could obviously give such effects as expression of doubt (question tone) or signalling that something will follow (continuation tone).

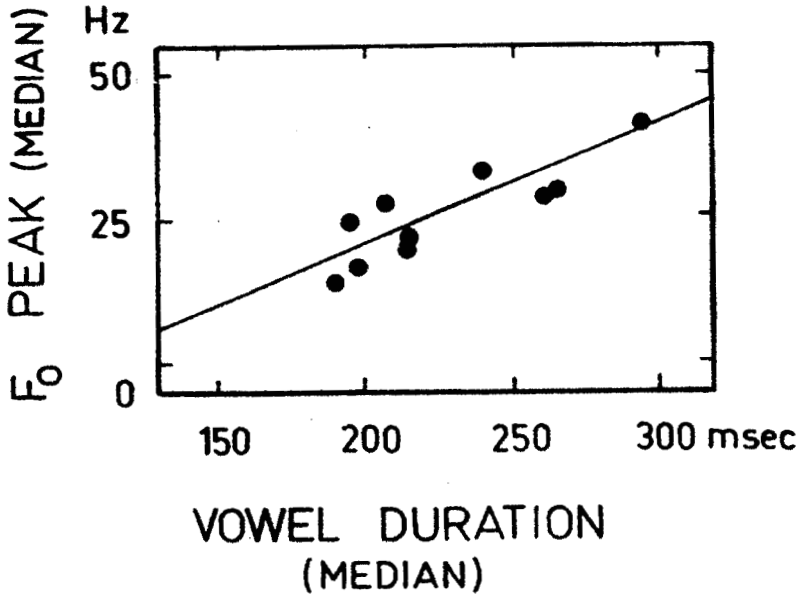

Fig. III-B-1. Deviation of F_o peak from the underlying sentence contour as a function of the duration of the preceding stressed vowel.

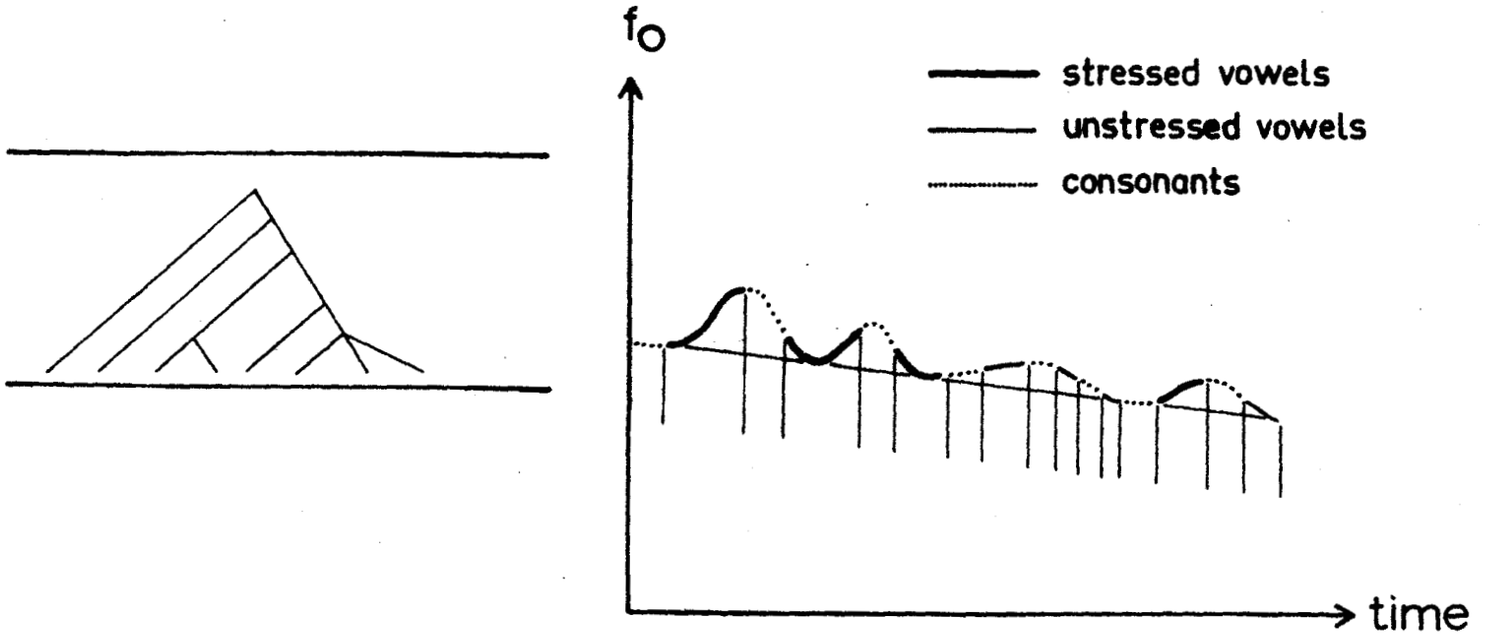

Fig. III-B-2. A fundamental frequency contour synthesized with the aid of the timing and F_o rules described in the text. The structural analysis of the sentence is shown to the left.

Depending on the structural analysis of the sentence some variation in the timing and intonation might be predicted. In this section we will demonstrate how the model handles junctural phenomena and emphasis within a single sentence. The sentence chosen is the same as the one used in the previous sections. One procedure that the model suggests for creating emphatic stress is briefly touched upon in ref.(2). The κ coefficient in the timing rule contains a certain factor greater than the one for the emphatic segments and the inverse of that factor for the surrounding syllables. Through this factor the absolute value of the κ coefficient may be made continuously variable so as to provide the appropriate degree of emphasis. The resulting modification of the time structure will cause a desirable modification of the intonational pattern without altering the F_o-rules. The pattern can be seen in Fig. III-B-4. Only one outstanding peak in F_o associated with the emphasis appears as also noted for other languages (Svetozarova)(5).

In Fig. III-B-4 two alternatives are presented; emphasis on the stressed syllable vs on the whole word. These realizations can be in free variation but with a definitive preference for the syllable emphasis.

The next example illustrates how in this model a difference in syntactic analysis of the same sentence will be reflected in different prosodic patterns. Dividing the reference sentence in two phrases at two different places will convey a shift of meaning as:

"John tog bilen i garaget" "John took the car in the garage"
"John tog bilen i garaget" "John took the car in the garage"

In the model this has been handled by letting the phrases pass through the rule system independently but deleting the final and initial specification of F_o at the juncture. The result can be seen in Fig. III-B-5. In this example informal listening tests indicate that no impression of subordination of the second part of the sentence is conveyed but a clear prejunctural increase in perceived stress is apparent. The subordination could easily be taken care of by postulating a higher node with appropriate α , β coefficients as discussed extensively in ref.(2).

Conclusion: At least two classes of effects are omitted in the present model of prosody. One is the variation in intensity depending on the amplitude and spectral shape of the sound sources, the other is the

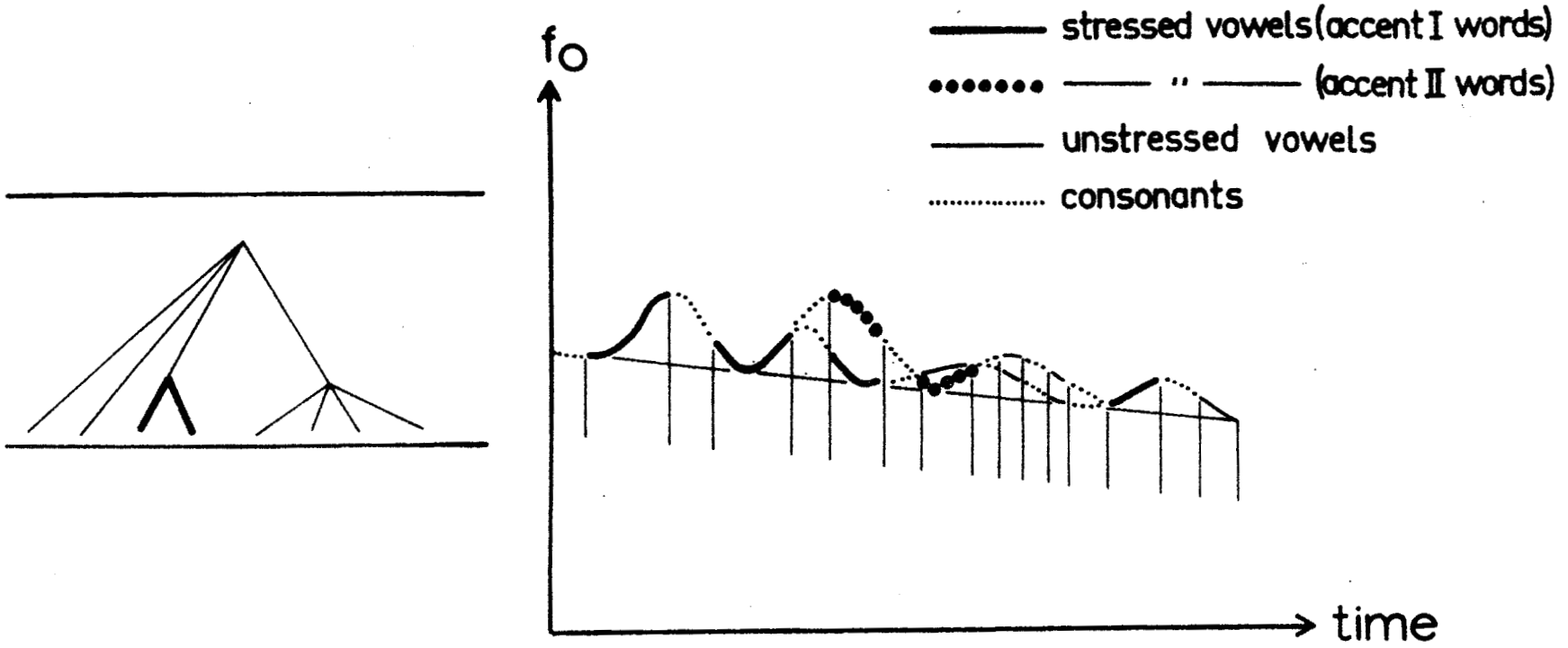

Fig. III-B-3. Synthesized fundamental frequency contours showing the difference between accent I and accent II.

systematic distribution of positional allophones. This does not mean that we deny the importance of these effects, only that we want to concentrate on what we believe to be the two most important prosodic factors that are not segment-specific namely timing and F_0 intonation.

The relevance of our rule system has been demonstrated by synthesizing nonsense sentences including the given examples and evaluating them in an informal listening situation. The present rule system is only tentative and obviously insufficient in many respects. However, it is intended as a basis for a description that could be tested against new classes of analytical data and it could also be subject to more formal perceptual evaluations. This will surely urge modifications and expansions but still we feel that at this rather preliminary stage some important generalizations have all the same been captured.

References

(1) R. Carlson, B. Granström, B. Lindblom, and K. Rapp: "Some timing and fundamental frequency characteristics of Swedish sentences: data, rules, and a perceptual evaluation", STL-QPSR 4/1972, pp. 11-19.

(2) B. Lindblom and K. Rapp: "Some temporal regularities of spoken Swedish", paper presented at the Symposium on Auditory Analysis and Perception of Speech, Leningrad, Aug. 21-23, 1973.

(3) Y. Erikson and M. Alstermark: "Fundamental frequency correlates of the grave word accent in Swedish: the effect of vowel duration", STL-QPSR 2-3/1972, pp. 53-60.

(4) E. Gårding and P. Lindblad: "Constancy and variation in Swedish word accent patterns", Working Papers 7, 1973. Phonetics Laboratory, University of Lund.

(5) N.D. Svetozarova: "The inner structure of intonation contours in Russian", paper presented at the Symposium on Auditory Analysis and Perception of Speech, Leningrad, Aug. 21-23, 1973.

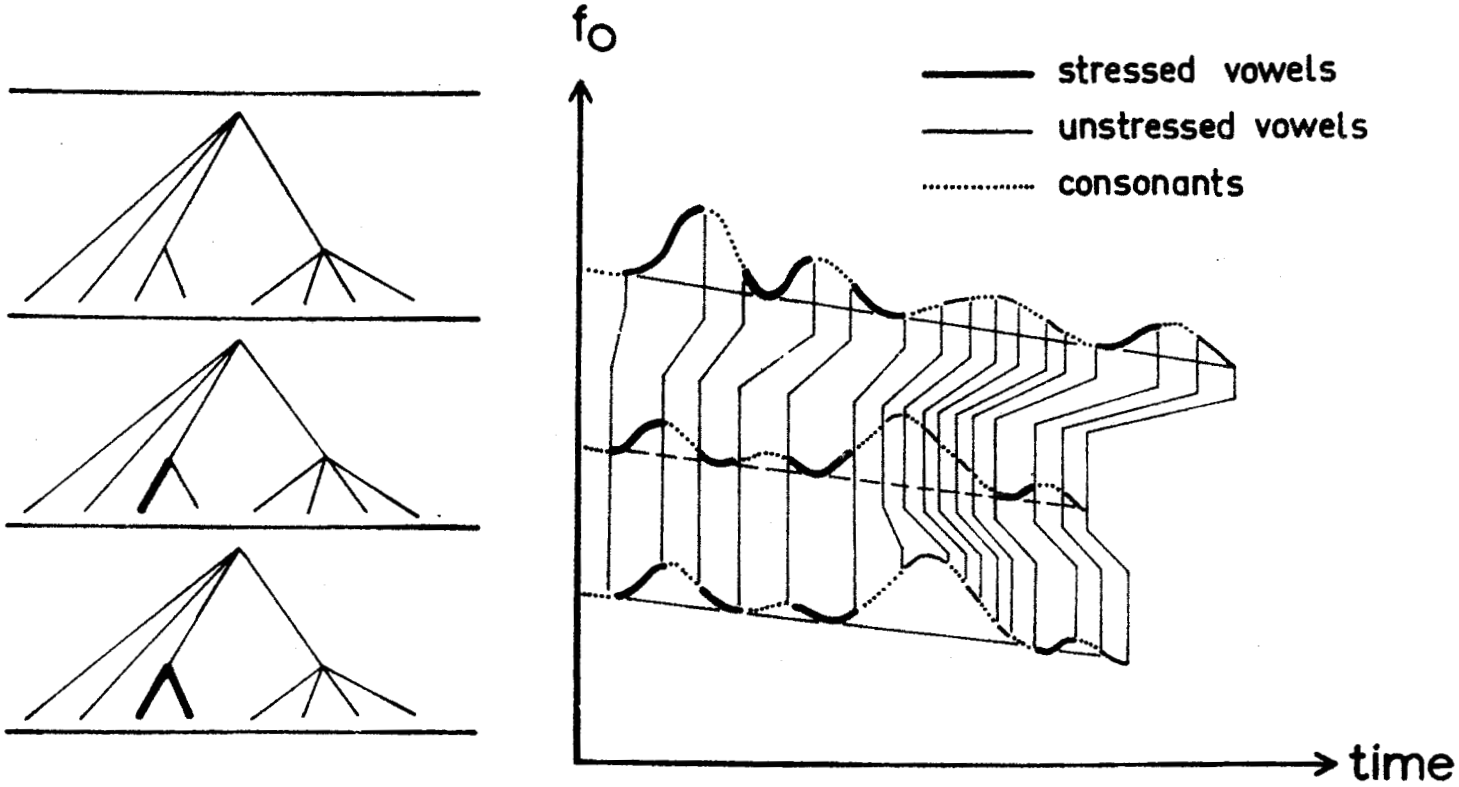

Fig. III-B-4. Synthesized fundamental frequency contours showing two examples of emphasis. Emphasized syllables are shown in the structural analysis by thick lines.

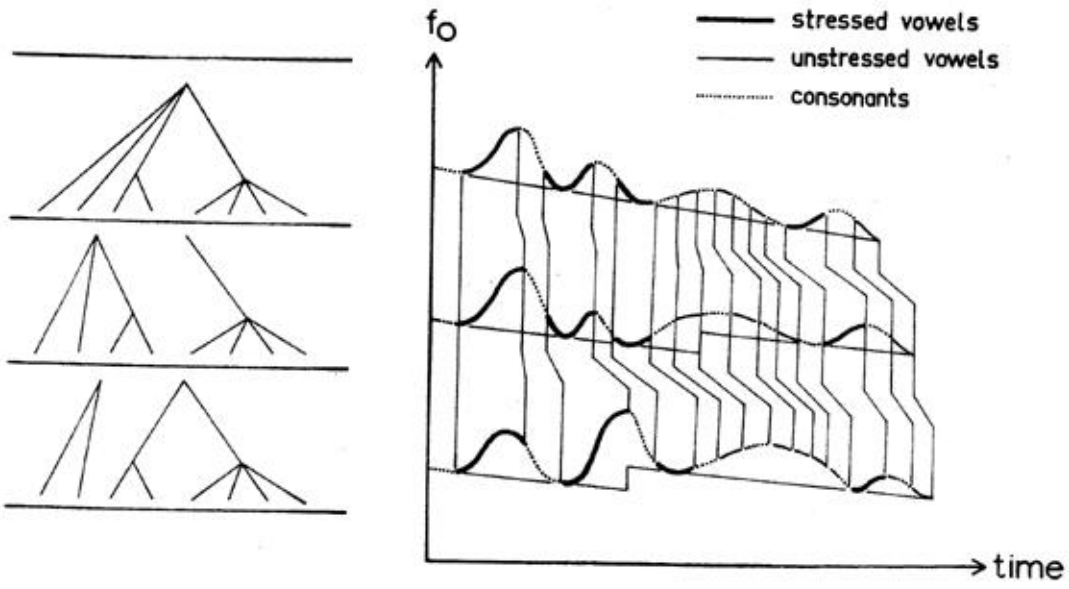

Fig. III-B-5. Synthesized fundamental frequency contours showing two juncture positions.

APPENDIX I

The duration assigned to a given segment is given by

$$S = D^{*} \cdot \varkappa \cdot \lambda \prod_{i=1}^{k} \frac{1}{(a_i+1)^{\alpha_i}(b_i+1)^{\beta_i}}$$

The symbols have the following meanings:

D^{*} = context-perturbed intrinsic duration;

$\varkappa$ = stress parameter; $0 \leq \varkappa \leq 1$ (cf. discussion of emphasis);

λ = phonological length parameter; $0 \leq \lambda \leq 1$;

a_i = size of portion that remains to be processed at a given point within the constituent at the i:th level;

b_i = size of portion that has already been produced at a given point within the constituent at the i:th level;

α_i = exponent controlling the degree of which the portion to be processed at the i:th level is allowed to influence segment duration, degree of anticipatory adjustment. $0 \leq \alpha_i \leq 1$;

β_i = exponent controlling the degree to which the portion already processed at the i:th level is allowed to influence segment duration, degree of backward adjustment. $0 \leq \beta_i \leq 1$;

k = upper limit on number of levels at which fraction is to be calculated. For any given segment k is equal to the number of nodes that are specified in the tree for that particular segment.

(From B. Lindblom and K. Rapp)

IV. PSYCHOACOUSTICS

A. DETERMINATION OF DIFFERENCE LIMEN AT LOW FREQUENCIES
Summary of thesis work for "civilingenjörsexamen"

A. Askenfelt

Abstract

Difference limen for frequency was determined in the region 40-200 Hz. Three types of stimuli were used; double-bass synthesis, low-pass filtered pulse train, and pure tones. Subjects listened to pairs of tones and adjusted the frequency of the second tone until they had matched the pitches to their own satisfaction. The standard deviation of the settings was used as measure of DL.

The obtained DL's are a tenth of the often quoted data of Shower & Biddulph and Zwicker & Feldtkeller but in close agreement with the data of J. Nordmark (1968). Pure tones gave nearly twice as high values of DL's as the complex tones.

1. Introduction

It has been realized for a long time that data on pitch discrimination are of relevance not only to psychoacoustics but also to musicology[1]. Since the end of the 19th century a number of investigations of difference limen (DL) for frequency has been carried out [2,3,4,5,6,7]. The results of some of the more important works are shown in Fig. IV-A-1. As seen in the figure there is a considerable degree of disagreement, which in part is probably due to differing methods of measurement. Also, it appears to a musician that some of the values reported are strikingly high. For example, the classical work of Shower & Biddulph[3] gives a DL of 3 Hz at a frequency of 60 Hz. This corresponds to 85 cents, almost a semitone (1 semitone = 100 cents). All but one of these results are derived from stimuli consisting of pure sinusoids. No attempt has earlier been made to measure DL for frequency with more "natural" sounds. The purpose of this work is to determine DL for low frequencies using stimuli similar to the tones from a musical bass instrument.

2. Stimuli

The double-bass was chosen to represent the bass instruments. Six tones in the frequency region 40-200 Hz were recorded in an anechoic room and analyzed by an audio-frequency spectrograph. A typical spectrum is shown in Fig. IV-A-2. The tones were synthesized according to