Swedish intonation contours in text-to-speech synthesis

Dieter Huber

ABSTRACT

The purpose of this study is to analyse the intonation contours of ten Swedish sentences pronounced by three different native speakers, in order to define a set of generative rules that can be useful in text-to-speech synthesis. Intonation in this context is understood in terms of interactions between word accents, sentence accents, initial juncture and terminal juncture, but without any reference to intrinsic qualities, segmental conditioning factors or various emotional and idiolectal features.

Out of a number of already existing intonation models for Swedish, the procedure developed by Eva Gårding and her collaborators at Lund University is adopted in a simplified version to synthesize 'idealized' F_o-contours for each of the test sentences in the pronunciation characteristics of each speaker (Bannert, 1984; Bruce, 1977; Gårding, 1977, 1979, 1981, 1983, 1984). The obtained results are compared with the original pitch sequences, mapping matches and mismatches. Some minor adjustments and modifications are proposed to improve the model. Finally, validity testing demonstrates the efficiency of the applied procedure and its underlying concepts with regard to the analysed text material, at the same time indicating areas for further research.

1. PRESENTATION

1.1 Text Material

The following ten Swedish sentences are analysed:

- Bussens förare fick körkortet indraget.
- Isen kan omöjligt bära en vuxen.
- Torpet hade blommor och gräs på taket.
- Många trivs med att vandra i fjällen.

- Sikten är ganska skymd i kurvan.
- Skorna var nya och alldeles för trånga.
- Lingonen brukar mogna i september.
- Pumpen på gården hade rostat på vintern.
- Vågorna slog högt över bryggan i stormen.
- Dagen firades med klang och jubel.

These sentences were chosen from lists compiled by Margareta Korsan-Bengtsen (Distorted Speech Audiometry, Acta Ota-Laryngologica, Suppl. 310, Göteborg 1973) and constitute statements without any prominent contextual features. They were recorded in an echoless sound studio at the Acoustics Laboratory of the Swedish Telephone Company in Stockholm under equally controlled conditions for each speaker.

1.2 Test Speakers

The three speakers comprise a 37-year old man, a 32-year old woman and a 11-year old boy, all of them living in the Stockholm area and speaking Swedish without noticeable particularities. They were instructed to read the ten statements one after another, with short intervals, in a normal and colloquial fashion as if talking to another person. As a matter of convenience I have chosen to apply the labels DIAMAL, DIAFEM and DIACHI throughout the course of this study whenever I refer to the male, the female or the child's voice respectively.

1.3 Vocoded Pitch Contours

A copy of the original tape recording was analysed at Chalmers University of Technology in Göteborg with respect to voicing determination, pitch value extraction and allophonic segmentation, using the vocoder system developed by and available at the Department of Information Theory (Hedelin, 1981). Pitch extraction was performed by the autocorrelation method (Hess, 1983; Rabiner & Schafer, 1978). Allophone labelling was carried out manually with judgement based on both visual and auditory evidence. Word accents are marked only for syllables receiving primary stress. Word boundaries and lower levels of stress are disregarded, as well as phrase accents and phrase boundaries. This analysis resulted in a set of thirty vocoded intonation contours, the first three of which are demonstrated in figure 1.

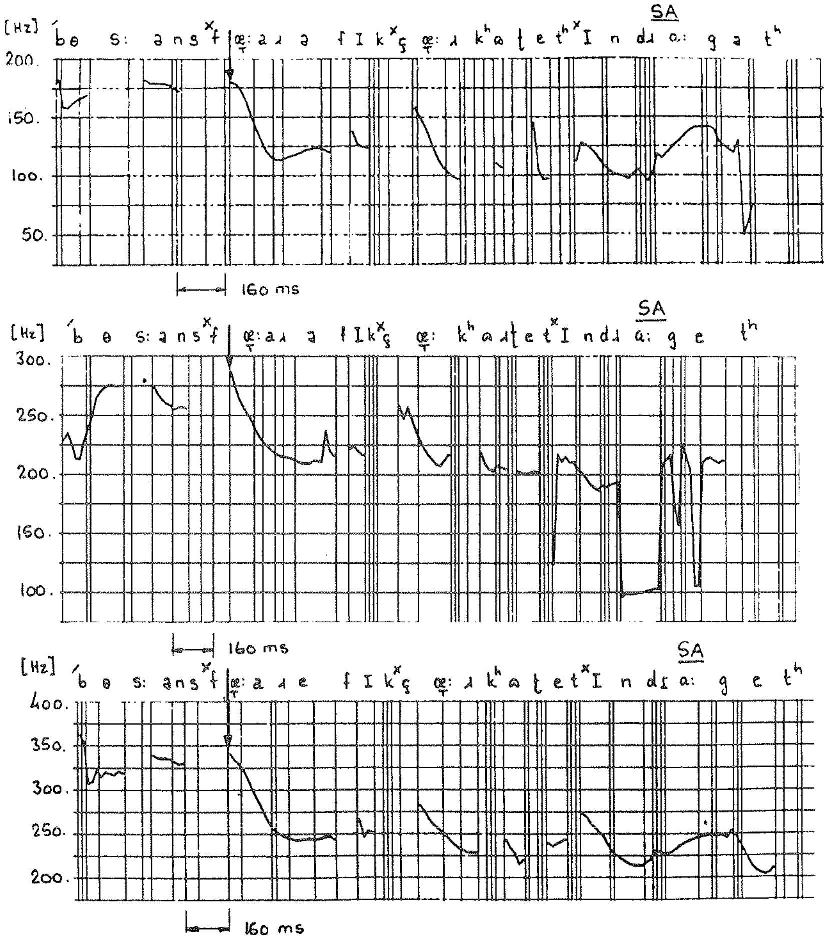


Figure 1. 'BUSSENS FÖRARE FICK KÖRKORTET INDRAGET'
Vocoded F_0-contours for DIAMAL, DIAFEM and DIACHI.
The line-up point is at the CV-boundary of the
stressed syllable in 'FÖRARE'.

1.4 Generated Pitch Contours

A first superficial glance at the lined-up text material reveals both similarities and dissimilarities in the pitch fluctuations of the various F_o-contours. Some of the irregularities can be easily accounted for in terms of individual variation. Others might prove more difficult to explain. One first point of interest is that we obviously have to consider large-scale pitch movements over relatively wide ranges of frequency and duration as well as comparatively minute vacillations that are added to or superimposed upon the larger strctures without, hcwever, changing their overall course of direction.

One possible approach to the investigation of the problems involved would be by way of a detailed description of the pitch sequences in the presented text material. Thus correlating the data obtained for DIAMAL, DIAFEM and DIACHI and viewing them in the broader light of established phonetic theory, we might hope to unearth recurrent regularities that can be formulated into a set of generative rules and corroborated by further systematic research. Such an analysis method requires, however, large and variable text material from many different sources if it is to yield reliable results, and does therefore not seem feasible within the limited scope of this study. Instead I have chosen an analysis-by-synthesis procedure, applying a simplified version of the Lund model of sentence intonation to synthesize 'idealized' pitch contours of all ten sentences in the pronunciation characteristics of the three test speakers, which I will then compare with the 'real' (vocoded) F_o-contours.

The Lund model of sentence intonation comprises in its most comprehensive application nine consecutive steps which gradually transform a given string of phonemic symbols representing the utterance in common alphabetic writing (INPUT) into its concomitant F_o-contour (OUTPUT). For a more detailed description of the entire procedure see Gårding, 1984.

Applying this model to the text material and using a tonal grid defined by four parallel lines at a distance of 20/40/20 Hz from each other, with a overall fall of 50 Hz, I obtained thirty generated intonation contours, one of which is shown in figure 2.

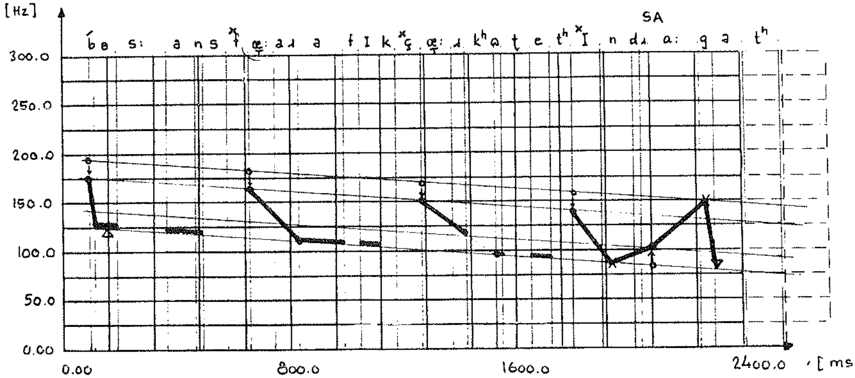
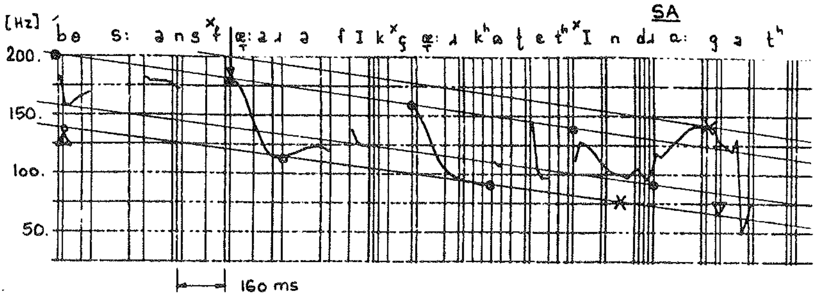


Figure 2. 'BUSSENS FÖRARE FICK KÖRKORTET INDRAGET'
Generated F_0-contour for DIAMAL.

2. COMPARISON

In order to facilitate comparison of the 'idealized' with the 'real' sequences I have transferred the generated HIGHs and LOWs directly into the vocoded contours, using prominent turning points in the latter to establish the grid. This was not always possible under the conventions suggested in paragraph 1.4. Compromise was sometimes necessary with respect to both parallelism and internal stratification. It should be noted, however, that construction of a tonal grid obeying some kind of regularity was possible without major difficulties in all the analysed sentences.

Figure 3. 'BUSSENS FÖRARE FICK KÖRKORTET INDRAGET'
Vocoded F_0-contours for DIAMAL, DIAFEM and DIACHI with superimposed grid and generated HIGHs and LOWs.



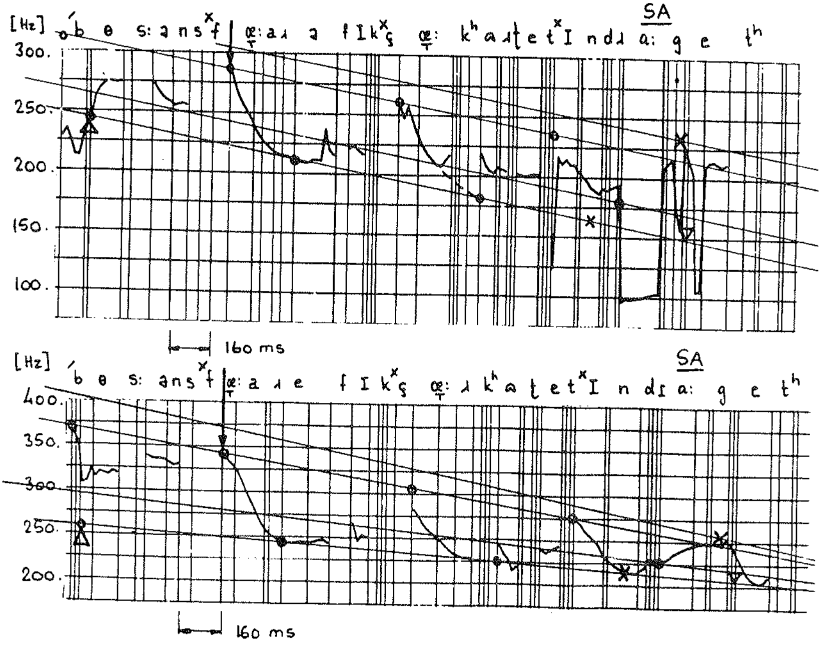


Figure 3 cont'd.

2.1 Matches and Mismatches

Comparing a total of thirty intonation contours representing ten Swedish sentences in the pronunciations of one male, one female and one child speaker, we are confronted with a bewildering number of both striking matches and blatant mismatches. Before entering a detailed discussion of the accumulated material I have tried to condense the results into a systematic arrangement which I hope will contribute to reveal some of the underlying regularities.

Table I. Matches and mismatches between vocoded and generated contours (Explanatory notes on page 8).

		GRID					INITIAL JUNCT.								WORD ACCENT							UNSTRESSED WORDS AND SYLLABLES							SENTENCE ACCENT						TERMINAL JUNCT.							
		o	▲	▼	↕	m	o	↑	↓	←	→	↗	↘	m	o	↑	↓	←	→	m	t	o	↑	↓	↗	↘	~	m	o	↑	↓	←	→	m	o	↑	↓	←	→	↗	↘	m
BUSSENS...	M	+	−	−	−	−	−	−	+	−	−	+	−	−	7	1	−	−	−	−	8	−	−	+	−	+	+	−	1	−	1	−	−	−	−	+	−	+	−	+	−	−
	F	+	−	−	−	+	+	−	−	−	−	+	−	−	6	2	−	−	−	−	8	−	−	+	−	+	+	−	1	−	1	−	−	−	−	+	−	+	−	+	−	−
	C	−	−	+	−	−	−	−	+	−	−	+	−	−	6	1	1	−	−	−	8	−	−	+	−	+	+	−	2	−	−	−	−	−	−	−	−	+	−	+	−	−
ISEN...	M	+	−	−	−	−	+	−	−	−	−	+	−	−	6	1	1	−	−	−	8	−	−	+	+	+	+	−	−	−	−	1	−	−	−	−	−	+	−	+	−	−
	F	−	−	+	−	−	−	−	+	−	−	−	+	−	1	4	3	−	−	−	8	+	−	+	+	+	+	−	2	−	−	−	−	−	−	−	+	−	−	+	−	−
	C	+	−	−	−	−	−	−	+	−	−	+	−	−	4	2	2	−	−	−	8	−	−	+	+	+	+	−	−	−	−	1	−	−	−	−	+	+	−	+	−	−
TORPET...	M	+	−	−	−	−	−	−	+	−	−	+	−	−	4	3	1	−	−	−	8	−	−	+	−	−	+	−	1	1	−	−	−	−	+	−	−	−	−	+	−	−
	F	+	−	−	−	−	−	−	+	−	−	−	+	−	3	2	3	−	−	−	8	−	−	+	+	+	−	−	2	−	−	−	−	−	−	−	−	−	−	−	+	−
	C	−	+	−	−	−	−	−	−	+	−	−	−	−	1	1	4	2	−	−	8	−	−	+	+	+	+	−	1	−	−	−	−	−	−	−	+	−	−	−	+	−
MÅNGA...	M	+	−	−	−	−	+	−	−	−	−	+	−	−	4	−	−	−	2	−	6	−	−	+	+	+	+	−	1	−	−	−	−	−	−	−	−	+	−	+	−	−
	F	−	−	−	+	+	−	−	+	−	−	+	−	−	1	2	3	2	−	−	8	−	−	+	−	+	+	−	−	1	−	−	−	−	−	+	−	−	−	+	+	−
	C	+	−	−	−	−	−	−	+	−	−	+	−	−	5	1	2	−	−	−	8	+	−	+	+	−	+	−	1	−	−	−	−	−	+	−	−	−	−	+	−	−
SIKTEN...	M	−	−	+	−	−	−	+	−	−	−	+	−	−	2	3	3	−	−	−	8	−	−	+	−	−	+	−	1	1	−	−	−	−	+	−	−	−	−	−	−	−
	F	+	−	−	−	+	−	−	+	−	−	−	−	−	3	3	2	−	−	−	8	+	−	+	+	+	+	−	1	1	−	−	−	−	−	−	+	+	−	+	−	−
	C	−	+	−	−	−	−	−	+	−	−	−	−	−	2	1	5	−	−	−	8	−	−	+	+	+	+	−	−	1	−	−	−	−	−	−	+	−	−	−	−	−
SKORNA...	M	+	−	−	−	+	+	−	−	−	−	+	−	−	6	1	1	−	−	−	8	−	−	+	+	−	+	−	1	−	−	−	−	−	−	−	+	+	−	+	−	−
	F	+	−	−	−	+	−	−	+	−	−	−	−	−	5	−	3	−	−	−	8	−	−	+	+	+	−	−	0	−	−	−	−	−	−	+	−	−	−	+	−	−
	C	−	−	+	−	−	−	−	+	−	−	−	−	−	5	−	3	−	−	−	8	−	−	+	−	−	+	−	0	−	−	−	−	−	−	−	+	−	−	−	−	−
LINGONEN...	M	−	+	−	−	+	+	−	−	−	−	−	−	−	6	1	1	−	−	−	8	+	−	−	−	−	−	−	−	1	−	−	−	−	+	−	−	−	−	+	−	−
	F	+	−	−	−	+	+	−	−	−	−	−	−	−	4	1	2	1	−	−	8	+	−	−	−	−	−	−	0	−	−	−	−	−	−	+	−	−	−	+	−	−
	C	−	+	−	−	−	+	−	−	−	−	+	−	−	4	−	4	−	−	−	8	+	−	−	−	−	−	−	0	−	−	−	−	−	+	−	−	−	−	+	−	−
PUMPEN...	M	+	−	−	−	−	+	−	−	−	−	+	−	−	5	1	2	−	−	−	8	−	−	+	−	−	+	−	1	−	−	−	−	−	−	−	−	+	−	+	+	−
	F	+	−	−	−	+	−	−	+	−	−	−	−	−	4	3	1	−	−	−	8	+	−	−	−	−	−	−	0	−	−	−	−	−	−	+	−	−	−	+	+	−
	C	+	−	−	−	−	+	−	−	−	−	−	−	−	2	2	4	−	−	−	8	+	−	−	−	−	−	−	1	−	−	−	−	−	−	−	+	−	−	−	−	−
VÅGORNA...	M	−	−	+	−	−	+	−	−	−	−	+	−	−	2	2	3	1	−	−	8	−	−	+	−	−	+	−	1	−	−	−	−	−	−	−	+	−	−	+	−	−
	F	+	−	−	−	+	+	−	−	−	−	+	−	−	5	1	2	−	−	−	8	+	−	−	−	−	−	−	0	−	−	−	−	−	+	−	−	−	−	+	−	−
	C	+	−	−	−	−	+	−	−	−	−	+	−	−	3	2	3	−	−	−	8	−	−	+	+	+	+	−	1	−	−	−	−	−	−	−	+	−	−	+	−	−
DAGEN...	M	+	−	−	−	−	+	−	−	−	−	+	−	−	3	3	2	−	−	−	8	−	−	+	−	−	+	−	1	−	−	−	−	−	−	−	−	+	−	+	−	−
	F	+	−	−	−	+	−	−	−	−	+	+	−	−	3	3	2	−	−	−	8	−	−	+	−	−	+	−	−	1	−	−	−	−	+	−	−	−	−	+	+	−
	C	−	−	+	−	−	−	−	+	−	−	−	+	−	4	1	3	−	−	−	8	−	−	+	+	+	+	−	−	−	−	1	−	−	−	−	+	+	−	+	+	−
Total		19	4	6	1	10	13	1	14	1	1	18	3	0	116	48	66	6	2	0	238	9	0	24	13	15	22	0	20(6)	7	2	3	0	0	7	6	11	11	0	24	7	0

For code, see next page.

M	=	DIAMAL
F	=	DIAFEM
C	=	DIACHI
o	=	complete correspondence between vocoded and generated aspect
▲	=	diverging grid
▼	=	converging grid
↕	=	misfit with respect to internal stratification
↑	=	generated accent/juncture mark to high
↓	=	generated accent/juncture mark to low
←	=	generated accent/juncture mark to early
→	=	generated accent/juncture mark to late
↗	=	F_o-rise missing in generated version
↘	=	F_o-fall missing in generated version
t	=	total number of accent HIGHs and LOWs
m	=	other misfits
∼	=	F_o-level in vocoded version

Code to preceding table

2.2 Discussion

2.2.1 The Tonal Grid

Permitting individual adjustments with respect to width, range and rate of fall, the model predicts correct grids in nineteen out of thirty occurrences. This figure may need some modification, taking into account those incidences (marked in the m-column) where considerable parts of the intonation contour fall one octave below the otherwise regular tonal surroundings. This happens eight times for DIAFEM and twice for DIAMAL. A closer review reveals 'higher-harmonic detection' or 'higher-harmonic tracking' (if continued for a longer time) of the

subharmonic at $F_0/2$ as the probable cause. Errors of this kind are also referred to as octave errors (Hess 1983) and can be treated in terms of individual aberrations on the phonation level with negligible impact on the intonation sequence as a whole. Therefore they do not need to be taken into consideration when drawing the grid.
Of the remaining 'non-correct' patterns in the presented text material, four can be described as diverging and six as converging tonal grids, all ten of them displaying accurate internal stratification. I am hesitant to dismiss those cases simply as not conformable, or rather, I expect that further advances in the research of the general concept of the grid and its applications to different languages and sentence types, including emotional as well as attitude features, will sooner or later provide us with the tools to ascribe correctness even to these kinds of deviations. One possible lead in this direction might be that non-parallelism in our material occurs nine out of ten times for male speakers (three times for DIAMAL, six times for DIACHI) but only once in the case of DIAFEM. Much larger text material has to be analysed, however, before these observations may be established as reliable facts.

One example, test sentence number four in the DIAFEM-version, displays a grid which clearly deviates from the internal 1-2-1 stratification principle, at the same time differing markedly even with respect to the overall intonation slope. Far from proffering any explanation I wish to stress that the vocoded intonation contour even in this case, however, provided enough clues to establish some kind of grid pattern without any greater difficulties.

2.2.2 Initial Juncture

The placement of the initial juncture LOWs has been predicted correctly in thirteen out of thirty cases in the presented text material. Faulty judgements are confined almost exclusively to the frequency scale with an overwhelming tendency towards underrating (14:1 ratio too low versus too high). Temporal defects are comparatively rare (once too early and once too late).

If the model does very nicely with respect to at least the temporal prediction of the initial juncture LOWs, it does not succeed very well in generating the F_o-rises that typically accompany those LOWs. Our text material displays eighteen such F_o-rises (versus three F_o-falls and nine F_o-levels), none of which is predicted by the generated sequences.

It will be discussed later, if theses F_o-rises are to be treated as intrinsic juncture features or rather as belonging to concatenation.

2.2.3 Word Accents

The thirty Swedish sentences constituting the text material for this study contain 238 word accent markers (HIGHs and LOWs), which are distributed in pairs over 119 words. 61 of these words receive acute accent (ACCENT I). The remaining 58 words have grave accent (ACCENT II).

116 of the 238 word accent markers, which means roughly half of them (48,7 %), are synthesized at their proper locations in the vocoded contours. 48 (20,2 %) are placed too high and 66 (27,7 %) too low, which adds up to a total of 114 word accent markers (47,9 %) which are positioned correctly on the temporal scale with deviation in pitch determination only. The remaining 8 word accent markers (3,4 %) are located inaccurately on the temporal scale. Half of them (1,7 %) are likewise faulty as to their frequency level and can thus be regarded as totally misplaced.

To sum up the results so far it can be stated that with regard to the presented text material, the model proves

- highly successful in the temporal prediction of word accent markers (96,6 %)
- considerably less reliable in frequency ranging (50,4 %)

with a total failure rate as low as 1,7 %.

Defects in correct F_o-labelling are distributed rather evenly between overshooting and undershooting values, which excludes inaccurate grid placement as possible cause. Improved results might be achieved by introducing phrase structure, by adjusting the algorithm for pitch generation to variable sentence types or by altering the concatenation rule.

2.2.4 Unstressed Words and Syllables

The Lund model of sentence intonation defines the F_o-sequences of unstressed words and syllables by way of concatenation, using copy and join rules to establish both levels and directions. Applying these conventions to our text material we usually obtain F_o-levels which are about half an octave below the vocoded ones. The table in paragraph 2.1 reveals nine correct occurrences versus 24 where the contour is placed too low. There is no incidence where the generated conour is located too high.

One further discrepancy between the synthesized and the vocoded contours is found in the realization of small scale F_o-rises (in 13 cases) and F_o-falls (in 15 cases) in unstressed words and syllables, which the model does not distinguish at all. Even though most of them may simply be 'ripples on waves on swells on tides' in Dwight Bolingers analogy (1964) and thus irrelevant to our perception of sentence intonation, I am still hesitant if it is wise to neglect them as is generally done in speech synthesis.

2.2.5 Sentence Accent

One first observation has to deal with the absence of sentence accent in six out of thirty analysed sentences, four of them being pronounced by DIAFEM, two by DIACHI. The appropriate contours were thus generated totally disregarding any kind of sentence accent commands, producing correct results in all six cases.

Including these six occurrences in the o-column, there is an allout number of 26 accurate sentence accent markers (68,4 %) versus 7 too high (18,4 %), 2 too low (5,3 %) and 3 too early (7,9 %).

Summing up these results under the same prerogative as in paragraph 2.2.3, it can again be noted that the conventions -

here for sentence accent marking - prove

- highly successful in the temporal prediction (92,1 %)
- less reliable in frequency ranging (76,3 %)

with a total failure rate all the way down at 0 %.

Two of the three markings which are placed to early, are specially interesting, as they seem to reveal a plausible explanation for the delays. The sentence accent HIGHs in both the DIAMAL and DIACHI version of 'ISEN KAN OMÖJLIGT BÄRA EN VUXEN' together with the adjoining terminal juncture LOWs determine the ordinates of two large-scale intonation contour falls, which according to the rules would have to be implemented entirely within the limits of voiceless speech segments. What the vocoded text material displays, however, is that the F_0-falls are placed not where the model predicts but in the voiced speech areas following immediately after. In other words, the whole F_0-contour confined by the sentence accent plus terminal juncture markers is in both cases transferred into the adjoining voiced section without changing neither the pitch levels nor the falling rate.

Similar occurrences of pitch contour reorganisation in connection with voiceless segments have been described by Rapp (1971), Eriksson (1973), Bannert & Bredvad-Jensen (1975) and others. The procedure seems both reasonable and logical. It would be difficult to imagine how large-scale pitch movements within the limits of unvoiced speech could be performed otherwise by the human voicing mechanism without simply truncating them. Two incidences out of thirty are, however, not enough to establish reliable evidence and have to serve in the context of this limited study as mere observations.

2.2.6 Terminal Juncture

The placement of the terminal juncture LOWs has been predicted correctly in seven out of thirty cases in the presented text material. Contrary to the results obtained for the initial juncture, inaccurate predictions occur both with respect to frequency and time.

F_o-rises after terminal juncture LOWs are even more common here (24 occurrences) than they were for initial junctures. The model does not predict any one of them. Again the question arises if these features are to be treated as an outcome of concatenation or rather as intrinsic juncture quality.

2.3 Modifications

It can reasonably be assumed that many of the shortcomings of the synthesized contours will be avoided once there are more adjustable rules and conventions on how to establish the grid to accommodate different contextual surroundings. This kind of improvement is specially to be expected with regard to frequency level determination in both stressed and unstressed parts of the speech utterance, but also when it comes to include phrase structure into the general model. I will therefore not deal with the problems of frequency ranging in this paragraph, leaving possible solutions to future research.

Some minor adjustments, however, I would like to suggest already here, in order to ameliorate the results obtained for the text material analysed in this study. One of these supplementary rules deals with the construction of the grid, two with initial juncture and terminal juncture respectively, and the two remaining ones are amendments to the concatenation process.

1. Permit diverging and converging patterns as well as parallelism when drawing the grid.
2. Connect initial juncture LOW with the next following word accent HIGH by direct interpolation.
3. Add F_o-rise to terminal juncture LOW.
4. Move large-scale F_o-movements into adjacent voiced areas without changing their properties, if otherwise they would be located entirely within voiceless sections.
5. Connect F_o-turning-points representing

different pitch levels by cosine interpolation.

Application of these amendments to our text material of thirty sentences would score the following improvements:

1. Ten more 'correct' grids, rendering a total of 29 out of 30 accurate predictions.
2. Considerable improvements both with respect to initial juncture rises (18 incidences) and prediction of unstressed words and syllables (27 cases).
3. 24 correct predictions out of 30.
4. Refers to sentence 'ISEN KAN OMÖJLIGT BÄRA EN VUXEN' (see paragraph 2.2.5).
5. Reflects the general impression of the F_o-contours in all thirty text sentences.

3. VALIDITY TESTING

One question left unanswered in this study so far concerns our perception of intonation contours. Do the simplified contours generated by the Lund model of sentence intonation actually produce acceptable human intonation when used in speech synthesis?

Using again the digital equipment available at the Department of Information Theory at Chalmers University of Technology in Göteborg, I replaced the vocoded contours of the first three sentences gradually by their respective generated ones, first without the modifications suggested in paragraph 2.3, later with them included. The final result for sentence 1 spoken by DIAMAL is demonstrated in figure 4.

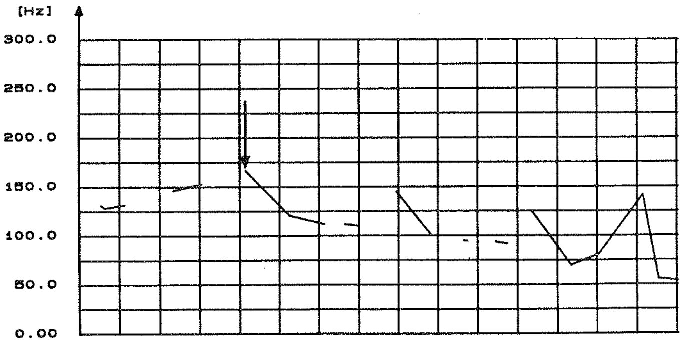


Figure 4. 'BUSSENS FÖRARE FICK KÖRKORTET INDRAGET'
Synthesized F_0-contour for DIAMAL in the modified final version. The arrow marks the CV-boundary of the stressed syllable in 'FÖRARE'.

The concomitant sound tracks were played to a group of listeners after every introduced change and were continually critisized by them as to the naturalness of the perceived intonation.

3.1 Results

Speech synthesis based on the unmodified contours only produced curiously base pitch level impressions in the first part of all sentences, whereas the later sections sounded quite acceptable. Sentence number two (ISEN KAN OMÖJLIGT BÄRA EN VUXEN) was beyond that characterized by the total absence of emphasis, which did not coincide with the impression from the original recording.

The first defect could be completely remedied in all incidences by applying modification number two, which means by linear interpolation between the initial juncture LOWs and the next following word accent HIGHs without clinging to the baseline.

Application of modification number four reestablished sentence accent in the voiced part of the last word in sentence two, which thus replicated the tonal contour of the original vocoded version.

Both rules number two and number four proved thus highly productive in the limited context of the synthesized material.

The equally proposed modification number three (adding F_0-rises after terminal juncture LOWs) did on the other hand not produce any easily discernable improvements and was judged negligible for the purpose of this study.

The remaining two modification rules number one and number five were not included in the test procedure at all, number one out of lack of appropriate features in the synthesized material (all three sentences show parallel tonal grids in the vocoded versions), number five because linear interpolation already produced highly satisfactory results.

BIBLIOGRAPHY

Bannert R. 1984. Towards a model for German prosody. Working Papers 27, 1-36. Department of Linguistics, Lund University.

Bannert R. and A.-C. Bredvad-Jensen. 1975. Temporal organization of Swedish tonal accents: The effects of vowel duration. Working Papers 10, 1-36 and 15 (1977), 133-138. Phonetics Laboratory, Lund University.

Bolinger D.L. 1964. Around the edge of language: Intonation. Harvard Educational Review 34, 282-293.

Bruce G. 1977. Swedish word accents in sentence perspective. Dissertation Lund University.

Erikson Y. 1973. Preliminary evidence of syllable locked temporal control of F_0. STL-QPSR 2-3, 23-30.

Gårding E. 1977. The Scandinavian word accents. CWK Gleerup, Lund.

Gårding E. 1979. Sentence intonation in Swedish. Phonetica 39, 288-301.

Gårding E. 1981. Contrastive prosody: A model and its application. Studia Linguistica vol 35, no 1-2, 146-166.

Gårding E. 1983. A generative model of intonation. In Prosody: Models and Measurements, Cutler&Ladd (eds.), 11-25, Springer Verlag, Berlin.

Gårding E. 1984. Comparing intonation. Working Papers 27, 75-99. Phonetics Laboratory, Lund University.

Hedelin P. 1981. A tone-oriented voice excited vocoder. Proceedings of the 1981 IEEE International Conference on Acoustics, Speech and Signal Processing, Atlanta GA, 205-208.

Hess W. 1983. Pitch determination of speech signals. Springer Verlag, Berlin.

Korsan-Bengtsen M. 1973. Distorted speech audiometry. Acta Ota-Laryngologica, Suppl. 310, Göteborg.

Rabiner L.R. & R.W. Schafer. 1978. Digital processing of speech signals. Prentice Hall, Englewood Cliffs NJ.

Rapp K. 1971. A study of syllable timing. STL-QPSR 1, 14-20.