THE INTONATIONAL MODEL USED FOR GERMAN TEXT-TO-SPEECH GENERATION *

Yuancheng Zheng, Harald Trost, Ernst Buchberger, and Johannes Matiasek

Austrian Research Institute for
Artificial Intelligence

ABSTRACT - In Text-To-Speech (TTS) systems the intonation quality depends mostly on the fundamental frequency (F0) contour. In this paper we propose a method of reconstructing the F0 contour in German TTS systems. The input to our system is an intonation hierarchy structure obtained by syntactic analysis. The nodes in our intonation hierarchy tree are scaled by relative numerical values which are normalized by pitch range. Assuming that the F0 tends to move continuously within a word, 11 patterns for reconstructing syllable F0 contour are presented according to the location of the syllable's stress.

I. INTRODUCTION

Although very good synthetic results have been obtained for single words in many different languages, the quality of SENTENCE synthesis is still dissatisfying. The synthesized sentences usually sound quite monotonous. It is commonly recognized that intonation makes a significant contribution to the quality of a synthetic sentence.

To synthesize with good-quality intonation, we must first have answers to the following questions:

a. What is the intonation structure of a sentence. This must provide gives information about the sentence tone pattern (i.e. falling or rising tone), the sentence focus, and the contribution of each word to the sentence pitch contour, i.e. the local pitch baseline of each word in a sentence. These word pitch baselines, which reflect the basic undulation of a sentence intonation, are relevant to the baseline of the whole sentence.

b. What is the pitch pattern of a word. This must provide information about the location of the stressed syllable , the syllable patterns, and the pitch height of each syllable relevant to the local word pitch baseline.

c. How to reconstruct the sentence pitch contour according to a sentence intonation structure and a word pitch contour pattern.

The main topic of the research presented in this paper is to provide a method to reconstruct the sentence pitch contour, the F0. How to obtain the sentence intonation structure and the morphological structure of a word is beyond the scope of this paper. In Section II of this paper we present a method to quantitatively represent a German intonational structure, assuming the sentence intonation structure has already been obtained by syntactic analysis. In Section III the F0 contour patterns of German syllables are discussed in detail. 11 rules for syllable F0 contour patterns are presented. In the last section, we discuss the synthesis results of sentence intonation in our German TTS system.

---

* This work has been supported by the Austrian "Fonds zur Förderung der Wissenschaftlichen Forschung" (FWF) under the grant M28-PHY.

II. THE FOCUS AND F0 STRUCTURES OF GERMAN SENTENCES

Like other intonation languages, in German a single sentence can be assigned many different intonational focus structures. Focus is a linguistic feature exhibited by part or all of a sentence and defined independently of its phonetic realization. Pragmatically, the focused constituents of a sentence contain relatively important information. In this paper we deal with only those cases of German sentences with a default ordering of constituents:

Subject + Adv3 + Adv2 + Obj + Adv1 + Predicate + V-Zusatz + Verb

Figure 1. The structure of a German sentence

In the default case, in a German sentence the predicate is focused. Based on the structure of German sentences (shown in Fig. 1), Pheby (1980) proposes the following Rhematic Hierarchy for German:

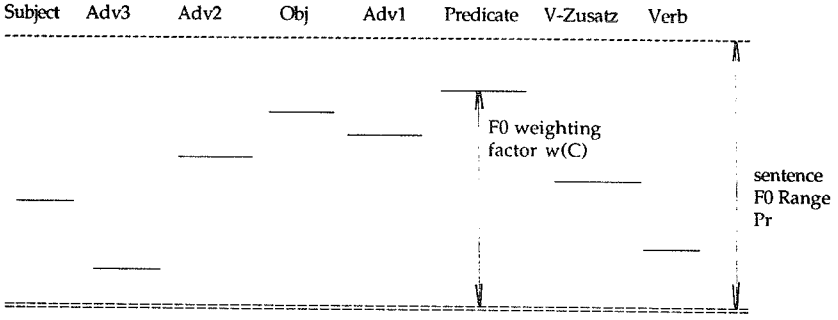
Adv3	<	Verb	<	V-Zusatz	<	Subj	<	Adv2	<	Adv1	<	Obj	<	Predicate
temp								instrument		direct.				
case								locative						

Formula (1)

where the symbol "<" means "weaker than".

The intonation of a sentence depends mainly on the F0 contour, the amplitude contour, and the syllable duration, of which the F0 contour is predominant. The F0 contour of a sentence is closely related and similar to the sentence focus structure. For speech synthesis, we must obtain a quantitative description of the F0 contour structure.

We define Pb to represent the sentence F0 baseline, Pr the sentence F0 range, and Fl(C) the local F0 baseline of the constituent C, C={Adv3, Verb, V-Zusatz, Sub, Adv2, Adv1, Obj, Predicate}. According to Formula (1) and the German sentence structure shown in Figure 1. , we can represent the sentence F0 contour structure as shown in Fig. 2:



"____" indicates local word F0 baseline Fl(C), "====" sentence F0 baseline Pb

Figure 2. The hierarchical structure of the F0 contour of a German sentence

where w(C), the F0 Weighting Factor of the constituent C, is a positive real not larger than 1, i.e., $0.0 \leq w(C) < 1.0$. With sentence F0 baseline Pb, sentence F0 range Pr, and F0 weighting factor w(C), the local F0 baseline of the constituent C can be represented by Formula (2):

$$Fl(C) = Pb + Pr \times w(C), \qquad \text{Formula (2).}$$

where $0 \leq w(C) < 1$ and C = {Adv3, Verb, V-Zusatz, Sub, Adv2, Adv1, Obj, Predicate}.

In selecting the F0 weighting factor w(C), formula (1) must be observed. The weighting factor w(predicate) of the constituent predicate has the highest value in the general case. Considering that stressed syllables have higher F0 than unstressed syllables in a word, the highest weighting factor must be restricted to be less than one. When the focus of a sentence is a constituent other than the predicate, the F0 contour structure shown in Fig. 1 should be changed by modifying the F0 weighting factor w(C). For instance, as an answer to the question 'Wer ist nach Wien gefahren?' ("Who went to Vienna?"), the sentence 'Johann ist nach Wien gefahren.' (" Johann went to Vienna.") has the focus on the subject JOHANN. Consequently, the subject JOHANN rather than the predicate GEFAHREN will have the highest F0 weighting factor. Obviously, all F0 weighting factors w(C) should be changed as the sentence focus shifts. The method of determining the F0 weighting factor w(C) when the focus of a German sentence does not observe Formula (1) will not be discussed in this paper.

III. THE PATTERNS OF SYLLABLE F0 CONTOURS

The F0 contour of a syllable is quite different from that of a sentence. Only one fixed syllable can be stressed in a word, but any words can be focused in a sentence. A word has a specific lexical structure, but a sentence can have many different focus structures depending on the context and the speaker's intention. For speech synthesis, the sentence F0 contour can be obtained by concatenation of word F0 contours which consist of syllable F0 contours. So the reconstruction of a syllable F0 contour is the basic and most important task in sentence intonation synthesis. To reconstruct syllable F0 contours we make the following two assumptions:

1. The sentence focus structure, i.e. the sentence F0 structure, is already known. That means that we must know the sentence F0 baseline, the sentence F0 range, and the local F0 baseline of each word in the sentence. Obviously, the changes of fundamental frequencies of all syllables are based on the local F0 baselines.

2. According to articulatory phonetics, it is assumed that the voiced syllables attempt to maintain the F0 movement in a sentence and that the F0 movement will be interrupted by non-voiced elements.

In order to describe the syllable F0 contour patterns we divide syllables into three groups: the initial syllable, the final syllable, and the intermediate syllables. For example, the word "fundamental" is divided into the groups:

fun	da men	tal
initial syllable	intermediate syllable	final syllable

All monosyllabic words are stressed in German. The syllable of a monosyllabic word is treated as a final syllable in this paper.

The F0 contours of intermediate syllables are affected only by their adjacent syllables within the word, and not by the previous and successive words. However, the initial and final syllables are affected not only by the adjacent syllables in the word, but also by the previous and successive words, respectively. Below we will discuss the three syllable groups in detail and propose two F0 Reconstruction Rules for Stressed syllables (SRR) and nine F0 Reconstruction Rules (FRR), which are used for a German TTS system.

There are three basic F0 movements, i.e. tones: rising("/"), falling("\"), and flat("-"). For example, the stressed syllable in an initial or intermediate syllable group usually has the rising tone("/"). But the final syllable of the last word in a declarative sentence has falling tone("\").

The three basic F0 tones can combine to form several complex tones: falling-flat("_"), rising-flat("/-"), flat-falling("-\"), and flat-rising("_/"). Unstressed initial and intermediate syllables can have any one of these complex tones.

A. F0 RECONSTRUCTION RULES FOR STRESSED SYLLABLES

From the perspective of acoustics, a stressed syllable has higher fundamental frequency. As shown in Figure 3, the peak of the F0 contour of stressed syllable, such as the syllable /AO/ in the word AUTO ("car"), is usually 20 to 50 Hz higher than the local F0 baseline.

[SRR1]:

The peak of the F0 contour of a stressed syllable is usually 20 to 50 Hz. A stressed syllable within a focused word can have a peak 40 to 50 Hz higher than the local F0 baseline. Otherwise, it is about 20 to 30 Hz higher than the local F0 baseline

A stressed syllable usually has rising tone("/"). But in some cases stressed syllables occur with falling tone ("\"), such as the stressed syllable /PUT/ in last word KAPUTT ("broken") of the sentence 'Mein Auto ist kaputt' (" My car is broken") shown in Figure 3.

[SRR2]:

If a stressed syllable is not word-final, it has a rising tone ("/"). Otherwise, the reconstruction follows the rules for final syllables.

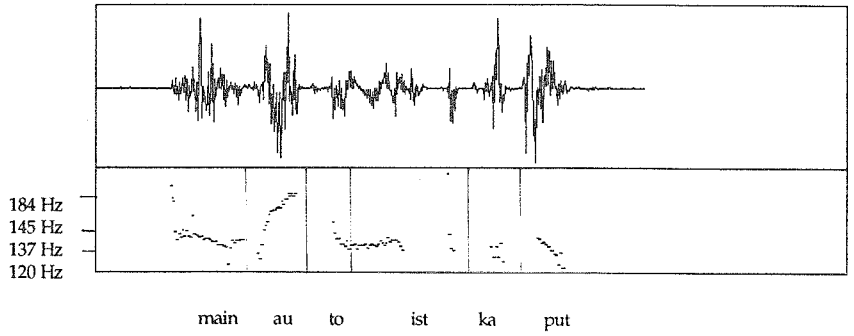


Figure 3. A F0 contour of the sentence 'Mein Auto ist kaputt' ("My car is broken")

B. F0 RECONSTRUCTION RULES FOR FINAL SYLLABLES

We will use complex tone to represent the models of syllable F0 contours because, if necessary, it is easy to change a complex tone to a basic one.

According to the location of a word in a sentence, the final syllable can have one of three basic F0 tones: falling ("\"), rising ("/"), and flat ("-") tone.

[FRR1]:

The final syllable has falling tone ("\") if the word lies at the end of a phrase or a non-question sentence, or at the front of a subordinate clause.

[FRR2]:

The final syllable has rising tone("/") if the word lies at the end of a question sentence or at the beginning of a phrase.

[FRR3]:

If neither of rules FRR1 and FRR2 apply, the final syllable has flat tone ("-").

C. F0 RECONSTRUCTION RULES FOR UNSTRESSED INTERMEDIATE SYLLABLES

Of the three syllable groups, intermediate syllables are the easiest to deal with, because they are affected only by adjacent syllables in the same word.

[FRR4]:

An unstressed intermediate syllable has a falling-flat tone ("_") if the previous one is stressed.

[FRR5]:

An unstressed intermediate syllable has a flat-rising tone ("_/") if the successive one is stressed.

[FRR6]:

If neither of rules FRR4 and FRR5 apply, an unstressed intermediate syllable has flat tone ("-").

D. F0 RECONSTRUCTION RULES FOR UNSTRESSED INITIAL SYLLABLES

An initial syllable is quite different from an intermediate one, since it is affected not only by the successive syllable within the same word but also by the final syllable of the previous word unless the current word is the first one in a sentence.

[FRR7]:

An unstressed initial syllable has a flat-rising tone ("_/") if the following syllable is stressed.

[FRR8]:

If the following syllable is unstressed and the previous syllable has a higher baseline than the syllable itself, the unstressed initial syllable has a falling-flat tone("_").

[FRR9]:

If neither of rules FRR7 and FRR8 apply, an unstressed initial syllable has flat tone ("-").

IV. CONCLUSIONS

We have presented a general intonation model for German sentences based on the hierarchical focus structure of a German sentence and we have given a quantitative description of the intonation structure of a German sentence. We must point out here that the German intonation model showed in Figure 2. is simplified and does not apply if the focus of a sentence is set to constituents other than the predicate. A method of modifying F0 weighting factors so as to take into account the shift of the sentence focus is currently being investigated.

In this paper our main concern were models of syllable F0 contours. When the effects of adjacent syllables are most serious a complex tone can become a basic tone ("\" or "/"). Otherwise, a complex tone can take a flat tone ("-").Within a single word these syllable F0 contour models can be directly applied for speech synthesis. If sentence synthesis is the aim, we will be confronted with problems such as in which cases the successive word should carry on the F0 movement of the previous one, how to transfer the F0 movement trend between adjacent words or, in other words, how to maintain the trend of F0 movement of the previous word. To resolve these problems we need additional F0 reconstruction rules which will be the topic of future work.

For single German words our model obtains a very natural tone. Through our experiments we have found that semivowels and nasals play an important role in the synthesis of sentence intonation. Because at the moment our speech parameter database lacks sufficient information about semivowels and nasals we cannot deal specially with them, which affects the quality of the synthesized sentences.

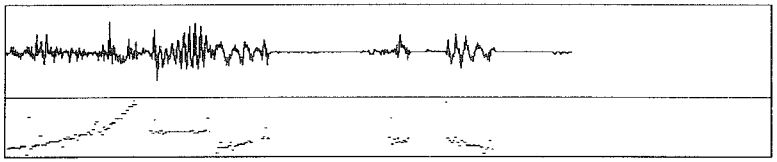


Figure 4. A synthetic intonation of the sentence 'Mein Auto ist kaputt' ("My car is broken")

Finally, in Figure 4, we show the result of a synthetic intonation of the German sentence 'Mein Auto ist kaputt' ("My car is broken"). Here the sentence F0 baseline Pb is selected to be 120 Hz, the sentence F0 range Pr to be 70 Hz. The peak of a focused stressed syllable Sp is set 40 Hz higher than its local F0 baseline, then the F0 weighting factor w(Auto)=(Pr-Sp)/Pr=(70-40)/70=0.43, w(Mein) = 0.2, w(ist) = 0.2, and w(kaputt) = 0.3.

REFERENCES

Féry, Caroline (1993) *German Intonational Patterns* , (Tübingen: Niemeyer , Linguistische Arbeiten).

Fant, G. (1970) *Acoustic Theory of Speech Production*, (Mount, The Hague)

Pheby. J. (1980) *Phonologie Intonation*, Grundzüge einer deutschen Grammatik, Berlin: Akademie-Verlag, 839-897.

Pierrehumbert, J. (1980) *The phonology and phonetics of English Intonation*, MIT Ph.D Dissertation. Distributed by Indiana University Linguistic Club, Bloomington.

Rabiner, L.R. and Schafer, R.W. (1978) *Digital Processing of Speech Signals*, (Prentice-Hall, Inc.)