TEST ENVIRONMENT FOR THE TWO LEVEL MODEL OF GERMANIC PROMINENCE

Gregor Möhler & Grzegorz Dogil
email: moehler,dogil@ims.uni-stuttgart.de
University of Stuttgart, Institute of Natural Language Processing
Azenbergstr. 12, 70174 Stuttgart
GERMANY

ABSTRACT

In this work we present a test bed designed to verify the two level model of Germanic prominence. We give an introduction to the linguistic background of the model and derive the features that the test environment should possess. Finally we describe the details of the implementation in the *ESPS/xwaves*[1] environment. The implementation is based on resynthesis using PSOLA algorithm. As a linguistic application the Tone Sequence Model (TSM) is implemented and tested.

Keywords: Prosody, Modelling of F0, PSOLA, Tone-Sequence-Model (TSM).

1. LINGUISTIC-PHONETIC BACKGROUND

Recent developments in the theory of prominence [1][3] rigorously restrict the number of categories upon which prominence relations may be expressed. Of the number of prosodic constituents proposed in the early 80´s (see [4] for an overview) practically only two could be validated in experimental studies: the *stress foot* (the basic linguistic category at the level of *word* prosody [word stress]), and the *intonational phrase* (the basic linguistic category at the level of *phrasal* prosody). In Germanic languages the evidence that has been recently accumulated suggests that the primary acoustic correlate for prominence at the stress foot level is *duration* of the designated syllable [5], whereas at the intonational phrase level it is the fundamental frequency categorized as *pitch-accent*.
It has also been observed that these two levels of prominence are fairly independent [6]. If a syllable is accented, it will *necessarily* be *much longer* than the syllable in a non-accentable position. However, if a syllable is associated to the *pitch-accent* within the phrase, other aspects of its production (like duration) *may* be, but do not have to be affected. This state of linguistic phonetic research calls for a new computational model, in which the prosody implementation is parallel and modular rather than hierarchical and incremental.[2]

2. THE PSOLA TEST ENVIRONMENT

The two level model for Germanic prominence predicts *pitch* movements for stress on the phrase level, and modifications of the *phone durations* on the word level. A resynthesis tool that could be used to investigate the theory should give us free hand in modifying the pitch and the duration of pre-recorded speech data.

2.1. The PSOLA Algorithm

The pitch synchronous overlap & add (PSOLA) algorithm as described in [10] meets our claim of independent variation of pitch and duration. Moreover, the intensity of the speech signal can be modified by a simple amplitude scaling of the output.

2.1.1. The Algorithm

We describe the algorithm only in few words here and suggest [10] for further reference.
First the speech signal to be resynthesized is cut into segments of two pitch periods overlapping in one pitch period. Pitch synchronous labels are required for this purpose, which we create using the algorithm implemented in the speech analysis program *xwaves* [11].
To achieve *pitch variation* the segments are now shifted against one another in time and then added again. The *duration* on the other hand is changed by

[1] ESPS™ / waves+™ (Version 3.1), Entropic Research Laboratory, Inc., AT&T Bell Laboratories, 1993

[2] See [8] and [9] for two leading implementations of different prosodic models ("peak prominence KIM model" and Fujisaki's accent model, respectively) for German.

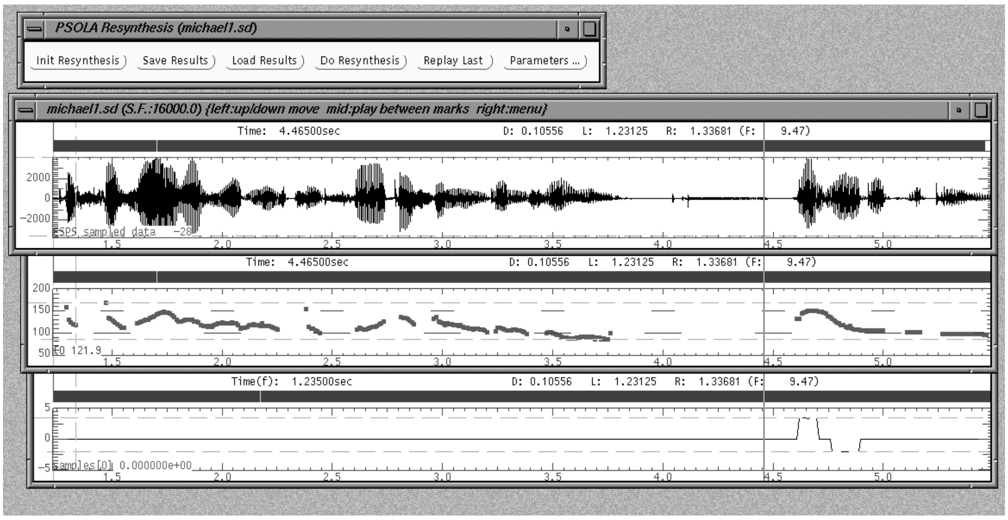

Figure1. Screen dump of the testing tool (*xwaves* environment) with F0 and duration modification window

doubling or leaving out certain segments before the final process of addition.

2.1.2. Characteristics of PSOLA

The PSOLA algorithm preserves the naturalness of the speech utterance like no other resynthesis means and it does it at extremely low computational cost. We implemented the algorithm in *C* and integrated it into the *xwaves* environment.

2.2. Implementation within *xwaves*

Our aim was to create a tool that could be easily used by linguists in an environment that provides a variety of speech analysis methods. That is why our test bed became a part of the *ESPS/xwaves* environment, a worldwide used speech analysis program available for various types of workstations.

2.2.1. Tool Layout

Easy usage could be achieved by displaying one window each for pitch, for duration and, if required, for intensity modification (Fig. 1. without intensity modification window). All windows show an initial curve the shape of which can be changed by mouse movements.

2.2.2. Pitch, Duration and Intensity Modification

After initialization of the tool the F0 curve of the original signal is shown in the pitch variation window. It can then be changed to the desired shape. The PSOLA pitch modification factor is calculated as the ratio between the two curves. An F0 algorithm based on [11] which is implemented in the *xwaves* program is used for the F0 analysis. This algorithm returns acceptable F0 estimations. For speakers with creaky voice, however, differences could occur between the calculated F0 and the auditive perception. Here the PSOLA algorithm would lead to undesirable results. For this reason speakers with little laryngalisation are chosen for our experiments.
For duration and intensity the modification factor itself is displayed in the window in logarithmic scale. This dB-representation proofed helpful for both factors.

2.2.3. Extras

Overall modification factors valid for the whole file could be entered by a parameter window. All functions provided by the tool are called by pressing the appropriate button in the resynthesis panel (Fig. 1).

3. APPLICATION: THE TONE-SEQUENCE-MODEL

The tool has been developed to test the two level model of Germanic Prominence. While on the word stress level no specific model has been so far tested, the *Tone-Sequence-Model* (TSM) has been chosen for testing on the phrase level. The TSM will be taken as an example of how to use our tool for testing of a specific linguistic model. However, the tool is not restricted to any specific prominence model. It can be used to test any prosodic hypothesis which requires independent control of pitch, duration and intensity.

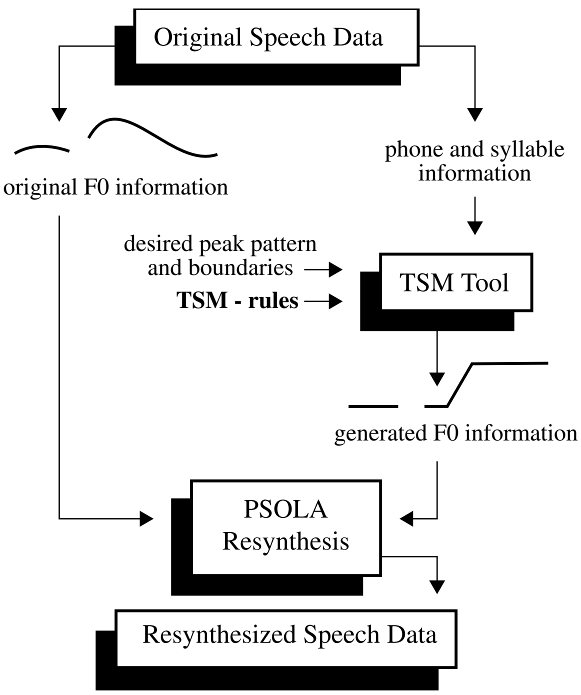

Figure 2.
Overview of the test tool with automatic F0 generation according to the Tone-Sequence-Model

3.1. Manual Usage of the Tool

The first step towards the investigation of the Two Level Model is manual modification of the prosodic parameters.
In case of the *phrase accents* TSM peak patterns are applied by drawing them with the mouse in the F0 modification window. Listening to the resynthesized result will give an impression of whether the peak's shape [12] (and height [13]) properly realizes the appropriateness criteria described by the theory.
Manual testing the *word stress level* is achieved by modifying the duration (and the integral of the intensity over the syllable duration) in order to change stress (and rhythm) pattern.

3.2. Implementation of the Tone-Sequence-Model

For the TSM we built a tool that generates the desired F0 curve following the rules of the TSM. This engine is plugged upon the PSOLA resynthesis tool as shown in figure 2.

3.2.1. The TSM for German

The TSM was originally introduced for English by Pierrehumbert [12] and was adapted for German by Féry [14]. In this implementation we apply the extension of Féry's model (cf. Mayer [15]), which is used at our institute for prosodic labelling.

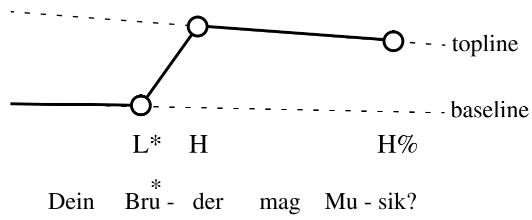

Figure 3. Simple example the tone sequence model (notation and appropriate F0 curve)

Within the TSM intonation curves are considered as a sequence of high (H) and low (L) tones. In German each of these tones is assigned to a single syllable. All high tones lie on a descending line called topline, low tones lie on the baseline. Topline and baseline embrace the pitch range. German has 4 different peak patterns and 3 boundary types [15]. Figure 3 shows a simple example of a sentence with one peak (L*H) and a boundary tone H%. The asterisk designates the tone of the accented syllable, the %-sign indicates the tone of a phrase boundary. Hence, we have a low tone on an accented syllable followed by a high tone on the following syllable (so-called *trail tone*). The phrase is terminated by a high tone boundary. All tones are connected by linear interpolation.

3.2.2. TSM-Implementation within xwaves

Our program converts the TSM notation into fundamental frequency values. This conversion is formulated in term of rules. The refinement of these rules is subject of linguistic research.
Since the peak contour depends on the syllable structure of the accented elements, phone and syllable information has to be at our disposal at the input to the rule engine (cf. Fig. 2). Additionally we require the information about the peak and boundary types in TSM notation. As the program is implemented within the *xwaves* environment we use the *xwaves* label facilities for data input.
The program itself has been developed in *C++*. This leads to clear data structures and rules that can be easily interpreted and modified by linguists.

4. RESULTS AND PERSPECTIVES

We have developed a tool for independent pitch and duration modifications of speech data. So far, the pitch accent side of our tool has been used to evaluate the Tone-Sequence-Model for German. First results are encouraging both from the tool operation point of view and from the point of view of the accuracy of the TSM model. Phrases and long stretches of read speech ("Buttergeschichte" on [16]) resynthesized with various peak patterns sounded naturally and the prominence impression came close to what has been predicted by the theory. To evaluate the degree of appropriateness of the TSM, experiments will be carried out in near future. We plan to build interfaces between our tool and higher linguistic structures to investigate theories of discourse prosody. Pitch range has been considered as an promising cue to discourse structure [17]. Slight extensions of the TSM part of the tool would give us control over the pitch range. A further extension within this research framework would be creating an interface to Discourse Representation Theory (DRT, [18]) in order to develop a restricted and testable model of discourse prosody.

REFERENCES

[1] M. Halle & W. Idsardi. "General Properties of Stress and Metrical Structures". 1995, in [2], pp 403-444.

[2] J. Goldsmith (ed.). *The Handbook of Phonological Theory*. Blackwell. Oxford, 1995.

[3] E. Selkirk. "Sentence Prosody: Intonation, Stress, and Phrasing". 1995 in [2], pp 550-570.

[4] M. Nespor & I. Vogel. *Prosodic Phonology*. Foris. Dordrecht, 1986.

[5] Phonetik AIMS volume 2. *Word Stress*. Working Papers of the Institute of Natural Language Processing, University of Stuttgart, 1995.

[6] M. Beckmann & J. Edwards. "Articulatory evidence for differentiating stress categories". 1994, in [7], pp 7-34.

[7] P. Keating (ed.). *Phonological structure and phonetic form: Papers in Laboratory Phonology III*. Cambridge University Press. Cambridge, 1994.

[8] K. Kohler. "Prosody in Speech Synthesis", *Journal of Phonetics.* No.19, 1991, pp 121-138.

[9] B. Möbius. *Ein quantitatives Modell der deutschen Intonation*. 1993. Max Niemayer Verlag. Tübingen.

[10] E. Moulines & F. Charpentier. "Pitch-synchronous waveform processing techniques for text-to-speech synthesis using diphones". *Speech Communication*, Vol 9, 1990, pp. 435-467.

[11] B. G. Secrest & G. R. Doddington. *An Integrated Pitch Tracking Algorithm for Speech Systems*. Proceedings ICASSP, 1993, pp 1352-1355.

[12] J. Pierrehumbert. *The Phonology and Phonetics of English Intonation.* MIT PhD Dissertation. 1980.

[13] D. R. Ladd. „Constraints on the Gradient Variability of Pitch Range, or, Pitch Level 4 lives!“. 1994, in [7], pp. 43-63.

[14] C. Féry, *German Intonational Patterns.* Niemayer. Tübingen. 1993.

[15] J. Mayer. *Transcription of German Intonation: The Stuttgart System.* 1995. accessible through WorldWideWeb: http://www.ims.uni-stuttgart.de/phonetik/joerg/labman/STGTsystem.html.

[16] *The Kiel Corpus of Read Speech.* CD-ROM. Institut für Phonetik und Digitale Sprachverarbeitung Kiel, 1994.

[17] E. Buchberger. *Prosodic Cues for Inferring and Realizing Discourse Relations*. Proceedings ICPhS Stockholm, 1995.

[18] H. Kamp & U. Reyle. *From Discourse to Logic*. Kluwer, Dordrecht, 1993