

## A METHOD FOR THE ANALYSIS OF PROSODIC REGISTERS

*Gregor Möhler & Jörg Mayer*

Institute of Natural Language Processing, University of Stuttgart  
Azenbergstr.12, 70174 Stuttgart, Germany  
{moehler,joemayer}@ims.uni-stuttgart.de

### ABSTRACT

The position of a speaker's pitch range conveys important information about the structure of the discourse. In this study we extract the pitch range of a speech database and compare the results with a register description of global prominence. The automatic extraction of the pitch range uses a parametric description of pitch accents and boundaries. We focused on the analysis of the pitch range properties during a sub-topic change, and found that the phonetic properties extracted from a database confirmed the predictions of the phonological model of global prominence: In phrases introducing a sub-topic, the lower margin is found to be at an average level, whereas the upper margin is clearly raised (extra-high register). In messages concluding a sub-topic, the pitch range is compressed and at a low level (low-compressed register).

### 1. INTRODUCTION

Long discourses are structured, i.e. they consist of discourse segments, each of which contributes in a specific way to the dynamic semantic representation of the discourse. Individual segments are glued together by a variety of semantic relations to form a coherent whole. Such relations are for example continuation, elaboration, contrast, etc. It has been shown that the prosodic realization of spoken discourse reflects to some degree the inherent semantic structure of the discourse. Grosz & Hirschberg show that pause duration correlates with the rough topic structure of a discourse [1]. Another parameter under consideration is pitch range [2]. But most of these studies are based on relatively short stretches of speech from few speakers. In view of recent efforts in corpus-based linguistics, it appears that research on prosody, and particularly on discourse prosody, yields more reliable results when large databases are used. In addition, theories of discourse structure and phonological models of prosody have to be as ex-

PLICIT as possible to obtain insightful generalizations on the prosody-semantics interface.

In the next section we introduce a phonological model of register. Register is assumed to be a phonological entity, which is phonetically interpreted as pitch range, and which reflects semantic relations among discourse segments as described in [3]. The formal interface between register and dynamic semantics was developed in [4] and will not be discussed in detail here. In the subsequent sections we describe a method for fundamental frequency ( $F_0$ ) parametrization and for testing the predictions of the register model. We present first results on a crucial issue of discourse interpretation, the transition between an old and a new sub-topic.

### 2. GLOBAL PROMINENCE AND PITCH RANGE

In the model described in [4], the pitch range of a speaker is phonologically analyzed as being divided into two categorical register levels {l} and {h}. These underlying or primary register features are associated with intonation phrases, and they reflect the position of the phrase within a discourse segment. Segment-final phrases are associated with the established finality marker {l}, whereas non-final phrases are associated with the feature {h}. Since the phonetic interpretation of register features is pitch range, every intonation phrase is obligatorily associated with at least one primary register feature. Primary {h} is a relatively high portion of the speaker's overall pitch range, while primary {l} is relatively low. Phonological registers may and do overlap in their phonetic realization.

Primary registers can be further subdivided into subregisters, e.g. extra-high (which reaches the upper edge of the speaker's range) or low-compressed (which is a narrow interval at the bottom of the speaker's range). These subregisters are represented by modifying or secondary register features. Extra-high is analyzed as {h}-modified primary phrase-ini-

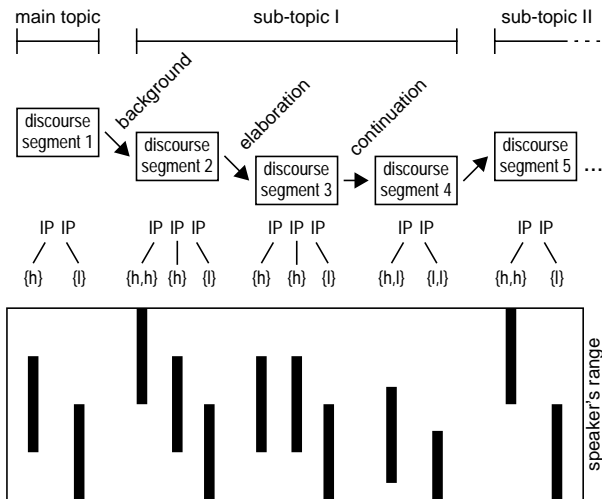


Figure 1. Summary of Mayer's Global Prominence Model: Discourse structure (*top*), intonation phrases (IP) with register features assigned (*middle*), and a diagrammatic illustration of the resulting pitch range expansion and position in the speaker's overall pitch range (*bottom*).

tial {h}-register, represented as {h,h} (first symbol: primary, second symbol: secondary register feature). Low-compressed is represented as {l,l}, i.e. {l}-modified primary phrase-final {l}-register. Mayer states that register modifications reflect semantic relations among discourse segments, such as background, elaboration, contrast, etc. For example, given the main topic of a discourse, an intonation phrase that introduces a background information sequence (i.e. a sub-topic) becomes {h}-modified. On the other hand, phrases constituting a discourse segment that terminates a sub-topic sequence become {l}-modified. Figure 1 gives an overview of the various aspects of the global Prominence Model.

### 3. F<sub>0</sub> PARAMETRIZATION

The automatic extraction of the pitch range is based on the parametric description of intonation events as proposed in [5]. Five parameters determine the principal movement of F<sub>0</sub> within a two-syllable window around pitch accents and intonation boundaries (Figure 2). The parameters have been chosen to allow a phonetic interpretation of the underlying F<sub>0</sub> movement. The shape parameter  $p$  signifies whether F<sub>0</sub> is mainly rising ( $p \approx 1$ ), falling ( $p \approx -1$ ) or whether F<sub>0</sub> has a peak contour ( $p \approx 0$ ). The two parameters  $d$  and  $s$  correlate with the position and steepness of the main slope within the two-syllable analysis window. The parameters  $l$  and  $h$  describe the base and height, respectively, of the particular movement. We calculate the pitch range based on these two parameters.

## 4. DEFINITION AND EXTRACTION OF PITCH RANGE

The definition of the pitch range used in this study is based on the tone-sequence model of intonation [6]. In this approach the intonation is seen as a sequence of high (H) and low (L) targets. Following this model we define the pitch range of an intonation phrase as the distance between the lowest L-target and the highest H-target within a the intonation phrase (following the definition in [4]).

To automatically calculate the pitch range of a specific intonation phrase we use the base ( $l$ ) and height ( $h$ ) parameters of all pitch movements corresponding to accented and phrase-final syllables. We excluded outliers, based on the inter-quartile range of the parameters' distributions, to achieve a more robust analysis. Such outliers can be caused by statistical errors as well as by emphatic accents ('expressive raising' of an accent's H-target). We define the lower margin of the pitch range to be equal to the lowest base parameter  $l$  found in the corresponding intonation phrase, which is not a statistical outlier. The same method is carried out for the upper margin, which uses the sum of the base and the height parameters.

## 5. DATABASE DESCRIPTION

The analysis is based on a German radio news speech corpus recorded from satellite broadcast. The database contains 67 minutes of news messages read by a professional male news speaker. The recordings were segmented into chunks, each containing one single news message. One message consists of 11 in-

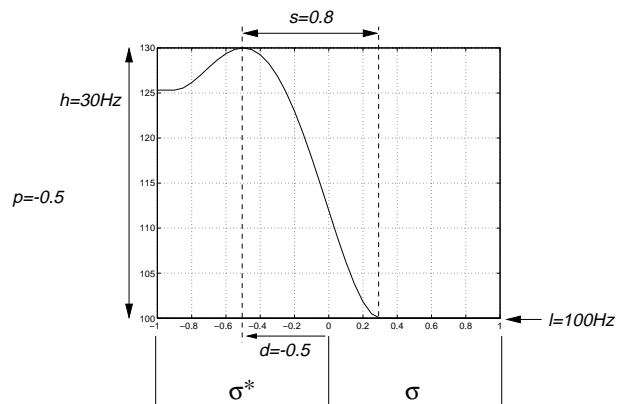


Figure 2. Example of a mainly falling F<sub>0</sub> movement ( $p \approx -0.5$ ) with a high target in the middle of the accented syllable ( $d \approx -0.5$ ). Its slope extends over  $s \approx 0.8$  of the syllable unity. The movement has its baseline at  $l = 100$  Hz and an amplitude of  $h = 30$  Hz.

tonation phrases on average. The speech data was augmented with phonetic and syllabic transcriptions using techniques of forced alignment. The prosody was manually annotated based on the German ToBI transcription system. For more details about this database and its transcriptions see [7].

## 6. RESULTS

We are interested in the phonetic analysis of discourse segments that mark the change of a sub-topic. Our goal is to calculate the pitch range of intonation phrases that introduce or conclude a sub-topic, and to compare it against the speaker’s global pitch range.

In the case of news messages we can easily extract the intonation phrases that are required for this analysis. Phrases that introduce sub-topics can be found in the initial phrase of every news story. The last phrase of a particular sub-topic is equivalent to the last intonation phrase in a story. In Figure 3 the pitch range values for several news stories are displayed.

### 6.1. Phrases introducing sub-topics

We extracted the pitch range of all intonation phrases that start a new message (N=71) and compared them to the overall distribution of pitch ranges for the whole corpus (N=805). In message-initial phrases, the lower margin of the pitch range is found to be at an average level (Figure 4a), whereas the upper margin is clearly raised (Figure 4b).

These findings are compared to the predictions of the phonological model of global prominence proposed by [4]. For discourse segments that introduce a sub-topic the model predicts an {h}-modification of the underlying {h}-register. This register is called *extra-high* register. The expected phonetic properties of the extra-high register are in agreement with the findings of this study.

### 6.2. Phrases concluding sub-topics

For phrases that precede a topic change, a *low-compressed* register is predicted by the phonological model of global prominence ({l} modification of an underlying {l}-register). As the name suggests, we can expect a reduced pitch range situated low in the speaker’s range.

To evaluate this prediction on our database we compared the pitch range of all phrases that conclude a news message (N=71) with the general distribution of the pitch range. We found that the lower margin of the pitch range is slightly below the average (Figure 4c), whereas the upper margin is clearly reduced (Figure 4d), indicating the compressed characteristics of the register.

## 7. CONCLUSIONS

Our study shows a convenient way to evaluate the predictions of the proposed model of global prominence proposed in [4]. We examined the relationship

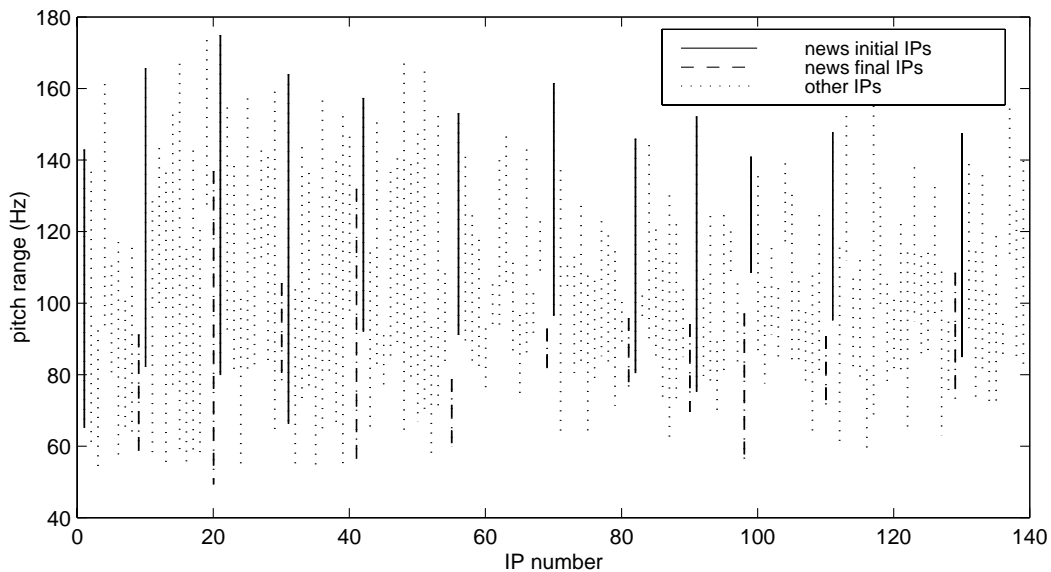


Figure 3. Pitch range values for several news stories extracted from the German news corpus.

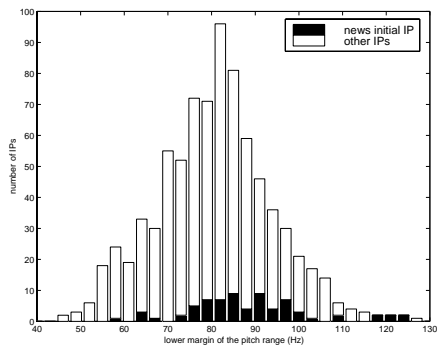


Figure 4a. Lower margin of the pitch range of *news-initial* phrases compared to other phrases.

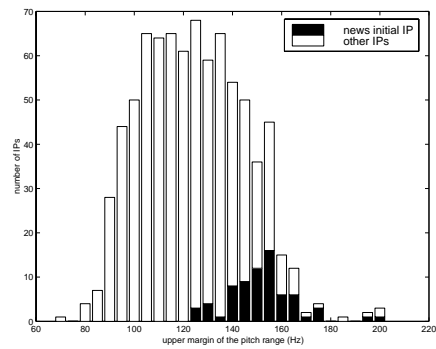


Figure 4b. Upper margin of the pitch range of *news-initial* phrases compared to other phrases.

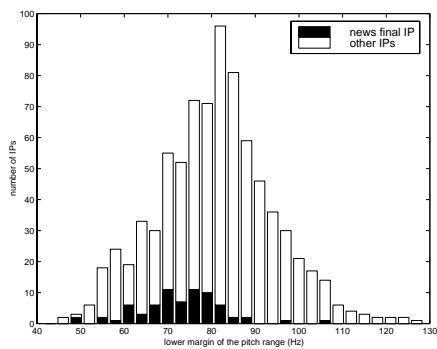


Figure 4c. Lower margin of the pitch range of *news-final* phrases compared to other phrases.

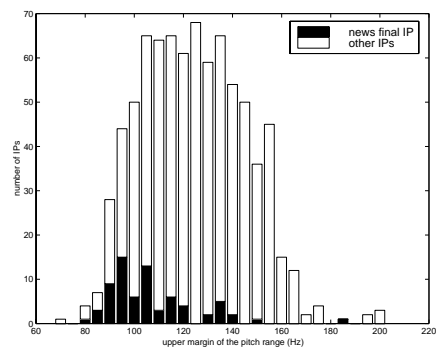


Figure 4d. Upper margin of the pitch range of *news-final* phrases compared to other phrases.

between two specific discourse relations and their phonetic realizations in terms of pitch range. Discourse elements that introduce a sub-topic are realized in the upper part of the speaker's pitch range, indicating an extra-high register. Those discourse elements preceding a topic change exhibit a low-compressed register realized in the lower frequency range. These results confirm the prediction of the global prominence model of prosody described in [4]. The method described here can, furthermore, be used to build an efficient topic change detector based on the properties of the pitch range within the discourse. This detector can be applied to speech recognition in dialog situations.

By using the same analysis method the effects of other discourse functions could be explored and quantified. Since such discourse functions are harder to detect within a database, this task would require semantically annotated corpora. Hence, the authors suggest to extend their work to areas such as dialog speech for which annotated corpora are already available.

## REFERENCES

- [1] Grosz, B.J. & Hirschberg, J. (1992). Some intonational characteristics of discourse structure. Proceedings of the 2<sup>nd</sup> International Conference on Spoken Language Processing, Banff, Canada: 429-432.
- [2] Ayers, G.M. (1994). Discourse functions of pitch range in spontaneous and read speech. Ohio State University Working Papers in Linguistics 44: 1-49.
- [3] Asher, N. (1993). Reference to abstract objects in discourse. Dordrecht: Kluwer Academic Publishers.
- [4] Mayer, J. (1999). Prosodische Merkmale von Diskursrelationen. Linguistische Berichte, 177: 65-86.
- [5] Möhler, G. (1998). Theoriebasierte Modellierung der deutschen Intonation für die Sprachsynthese. Dissertation. University of Stuttgart. Institute of Natural Language Processing.
- [6] Pierrehumbert, J. (1980). *The Phonology and Phonetics of English Intonation*. Ph.D. Thesis. MIT, Cambridge, MA.
- [7] Rapp, S. (1998). Automatische Erstellung von Korpora für die Prosodieforschung. Dissertation. University of Stuttgart. Institute of Natural Language Processing.