IDENTIFYING INTONATIONAL FOREIGN ACCENT WITH THE HELP OF DIFFERENT METHODS OF F_O GENERATION

Matthias Jilka
Institute of Natural Language Processing, University of Stuttgart, Germany

ABSTRACT

Fo generation is an effective instrument for identifying intonational foreign accent (IFA), as it can be applied to determine the relevance of observed deviations in the intonation patterns produced by non-native speakers of a language, This study aims to show that the method of intonation description that is used and that also serves as the basis for F_0 generation determines the form of the observed IFA which is examined both in the German speech of native speakers of American English and the American English productions by Germans. Two approaches to intonation description and generation are introduced: a category-based approach which identifies IFA in terms of tonal categories and a data-based approach which represents the F_0 contour in terms of an approximation function determined by six parameters and thus focuses on the phonetic aspects below the category level.

1. INTRODUCTION

The identification of the phonetic causes of intonational foreign accent (IFA) is strongly dependent on the used framework of intonation description. Aspects that are not explicitly covered by a particular type of description will not be perceived.

This study attempts to show that the use of different methods of intonation description and generation considerably improves the ability to identify the relevant causes of IFA. For this purpose German utterances produced by native speakers of American English as well as American English utterances produced by native speakers of German were recorded (see section 2.).

For the category-based view of intonation shown in section 3 intonational errors can be expected to concern the type and placement of tonal categories, as well as the phonetic realization of these categories, although the representation of the latter aspects is somewhat arbitrary in part.

An alternative approach to intonation description and generation introcduced in section 4. may facilitate the identification of other aspects of IFA. The data-based model introduced in this study is based on an F_0 parametrization (6 parameters) with an especially designed approximation function. Different features of the phonetic realizations of tonal categories such as steepness of fall and/or rise or alignment of the peak are measured automatically and are easily accessible for analysis. The data-based approach allows a more accurate and reliable identification of intonational deviations in the phonetic realization of tonal categories and also offers the possibility of examining questions concerning the acceptable degree and frequency of variation in these phonetic realizations as they manifest themselves in the parameter values.

2. DATA

Native speakers of American English with good command of German (all with academic backgrounds and a sojourn in Germany of at least six months) were asked to read two short stories, "Die Buttergeschichte" (240 words) and "Das dicke Kind" (100 words). as well as repeat (not imitate) short phrases played to them from news reports and the Verbmobil corpus.

The German speakers met the corresponding criteria for American English and read news reports (ca. 350 words) from the Boston Radio News Corpus [1] as well as repeated short phrases played to them from the BRNC and the ToBI training corpus [2]. In conversations with the instructor all speakers also produced spontaneous speech. A recording session lasted aproximately 30 minutes. Recordings were made in the anechoic chamber of the Institute of Experimental Phonetics Stuttgart on DAT (48 KHz sampling rate).

3. CATEGORY-BASED APPROACH

This approach to the identification of IFA is based on the categorical representation of intonation events. Intonation patterns are interpreted as tonal categories that are phonetically realized by intonation rules. Thus, this approach to intonation description and generation is compatible with the established segment-based theories of second language acquisition (e.g., [3], [4]).

3.1. Intonation description

Intonation patterns are described using the ToBI (Tones and Break Indices) system of prosodic transcription . Intonation events are interpreted as tonal categories that are in turn designated by ToBI tone lables.

ToBI was originally developed for the description of American English intonation [5], but there is also an inventory of tonal categories for the description of German intonation [6].

	American English	German
Pitch accents	H*, L*, L+H*, L*+H, H+!H*	H*L, L*H, (L*HL, HH*L, H*M)
Phrase accents	L-, H-	- (default)
Boundary Tones	L%, H%	% (default)

Table 1. Tone inventories of American English and German

The major differences lie in the number of pitch accents (German is usually limited to a falling and a rising pitch accent) and the phrasing structure which is also less complex in German (spreading of the nuclear pitch accent to a default boundary tone). Thus

the compositions of the inventories already imply to a certain extent which types of intonational errors concerning tonal categories are to be expected.

3.2. F_0 generation

Both intonation description and generation are based on the Tone Sequence Model [7] which represents an F_0 contour as a sequence of high and low target points that are connected by interpolation.

Thus, for the concrete generation process each tone label describing a category is expressed by phonetic realization rules in terms of H and L target points that are defined along the dimensions of position in the voiced part of the marked syllable and pitch range (baseline = 0%; topline = 100%). The phonetic realizations of the tonal categories differ depending on their position in the intonation phrase, as well as the tonal and segmental environment (for a more detailed description see [8]). The generation process is summarized in Figure 1 (see also [9]).

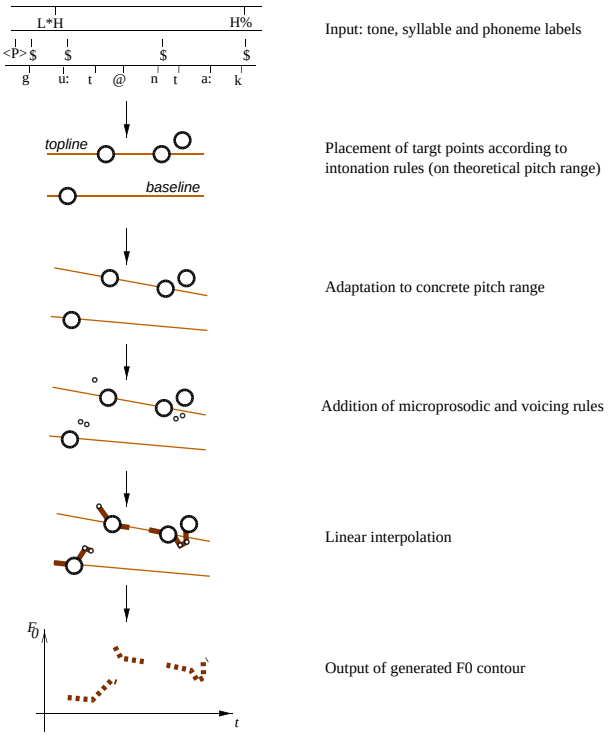

Figure 1. Process of category-based F_0 generation

The generated contours are then resynthesized with PSOLA, allowing a direct perceptual evaluation of different intonation patterns as they are expressed by ToBI tone labels. This analysis-by-synthesis approach is instrumental in determining the relevance of observed deviations in the F_0 contour for the perception of IFA.

3.3. Results

The analysis of utterances by non-native speakers shows that there is a number of aspects of IFA that manifest themselves in the form of categorical mistakes.

The most obvious forms of such an intonational error are the use of a wrong category and/or the wrong placement of a tonal category, both of which need not be motivated.

Other intonational errors, however, can be explained as a simple transfer of a tonal category from a speaker's native language to the respective second language. IFA occurs if the tune appropriate for a certain discourse situation is different in the two languages.

These factors acting in combination can also account for the observation of higher-level characteristics of foreign-sounding intonation such as the fact that there is more up-and-down movement in the German spoken by the Americans. It is caused mainly by the use of additional pitch accents and the transfer of more complex categories. Some of these factors are demonstrated in Figure 2 which depicts the F_0 contour of the read sentence *Ich hatte in diesem Winter angefangen* ... (This winter I had started ...). A comparison of the top contour produced by a native speaker of German and the middle contour produced by an American clearly shows the additional movement, which is due to an additional rising (L*H) pitch accent on ***Winter*** and especially the transfer of the tonal category appropriate for the discourse phenomenon of a continuation rise in American English. A continuation rise at the end of an intonation phrase signals that more information on the addressed subject is to follow. In German a continuation rise is realized simply by a rising nuclear pitch accent which spreads to the default boundary tone. In American English, on the other hand, the continuation rise is realized by a low phrase accent and a high boundary tine (L-H%) indepedently of the preceding pitch accent, so that if the rising pitch accent on ***angefangen*** is retained there is an additional fall and rise movement compared to the German version. The bottom contour shows an improved version of the American speaker's utterance generated on the basis of the tone labels describing the German speaker's intonation.

Similarly to the tone inventories the intonation rules for both languages also presuppose differences in the phonetic realization of equivalent categories. Nevertheless, despite being based on a thorough analysis, these rules are somewhat arbitrary and only reflect average values. While some tendencies concerning the timing of rising and falling pitch accents are detected, there are no definitive insights as to whether there are any consistently identifiable deviations in the phonetic realizations of corresponding categories.

4. DATA-BASED APPROACH

The data-based approach to intonation description and generation presented in this section allows a greater emphasis on the aspect of phonetic realization, but also relies on an intonation description by means of ToBI category labels.

4.1. Intonation analysis and generation

In the data-based approach the F_0 contour is analyzed in such a way that every labelled intonational event is represented by an approximation function (consisting of the sum of two sigmoids) within a 3-syllable window around the accented syllable. The starting values of the approximation process are based on detected turning points in the normalized F_0 contour. The concrete shape of the function as depicted in Figure 3 is determined by 6 parameter values. A more detailed description of the data-based approach and the parametric intonation events (PaIntE) can be found in [10]:

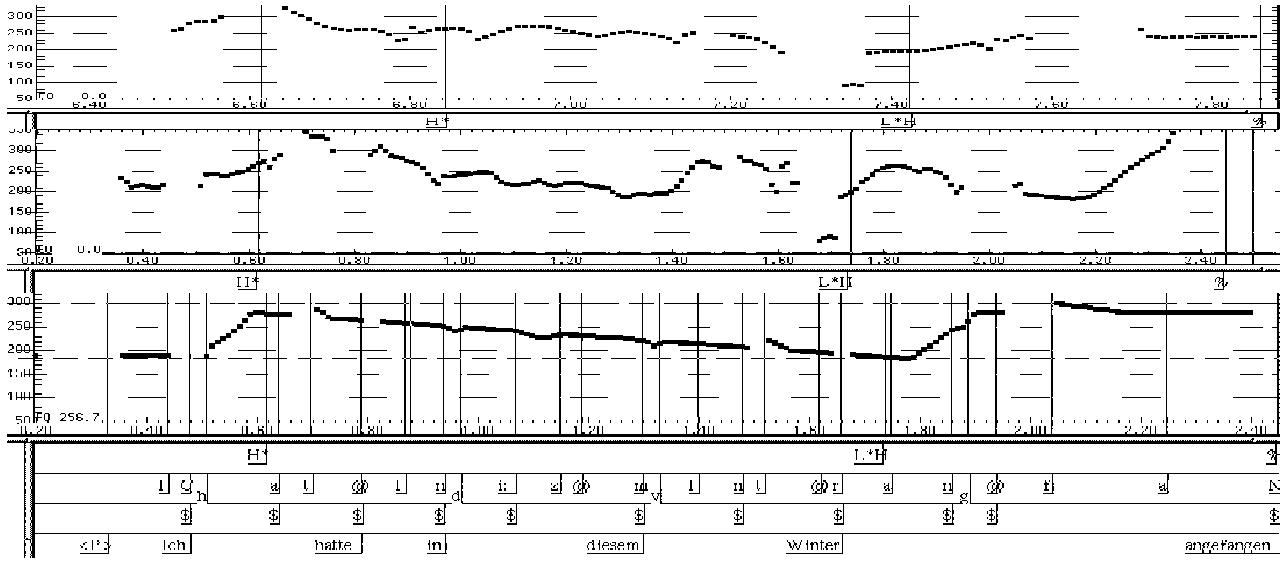

Figure 2. F_0 contour of the read utterance *Ich hatte in diesem Winter angefangen ...* (This winter I had started ...); top: reading by German speaker, middle: reading by American speaker, bottom: generated version of American speaker's reading.

a1, a2 : steepness of rise and fall respectively
b : time alignment of peak
c1, c2 : amplitudes of rise and fall in Hz
d : peak value in Hz

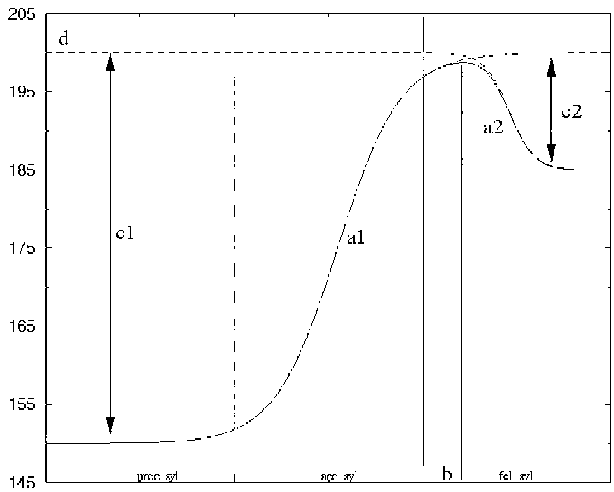

Figure 3. The approximation function (with a rising and a falling sigmoid) as determined by the parameter values within a 3-syllable window around the accented syllable.

The parameter values offer an alternative method of describing aspects of the phonetic realization of tonal events. In the original application of the data-based approach in intonation synthesis (e.g., within a text-to-speech system) the parameter values for the respective accents are trained on a large database. However, for the identification of IFA it is sufficient to examine the parameter values for certain tonal categories produced by individual speakers. There is also the possibility of adding more parameters, if they are of help in the analysis of specific categories (see section 4.2.)

The actual analyis, including the approximation of the function and the extraction of the parameters relies on an F_0 contour that is obtained from the raw F_0 by means of interpolation of the voiceless regions and subsequent smoothing of the whole contour.

Figure 4 shows an example utterance with the raw F_0 (top contour) and the interpolated and smoothed F_0 (middle contour). The bottom contour depicts the approximation of the tonal categories that provides the parameter values.

4.2. Application to the identification of IFA

In order to give a first impression of how the data-based approach may be used to examine non-native intonation categories and what types of results are to be expected, a small-scale comparison was made between the phonetic realizations of the four L*H pitch accents in the utterance *Natürlich kannte ich die meisten dieser Kinder* (Of course I knew most of these children) produced by two American and five German speakers (the original version by an American speaker is depicted in the top contour of Figure 4).

The examination of an additional parameter (not used in the regular set of parameters) describing the time alignment of the starred low tone proves to be interesting, as it seems to be consistently realized earlier for the American speakers than for the Germans. The additional parameter represents the relative position of the pitch accent's F_0 minimum in the voiced part of the accented syllable, the average value for the Americans being -0.074 and 0.375 for the Germans.

It can also be observed that the parameter a1 (steepness of rise) is lower for the Americans' versions of the L*H pitch accent (38.6) compared to the native speakers' mean value of 47.6, also implying that the rise is longer in the Americans' productions.

The result of this very limited analysis does of course not allow any general conclusions about the timing of the low element in every L*H pitch accent and the steepness of the corresponding

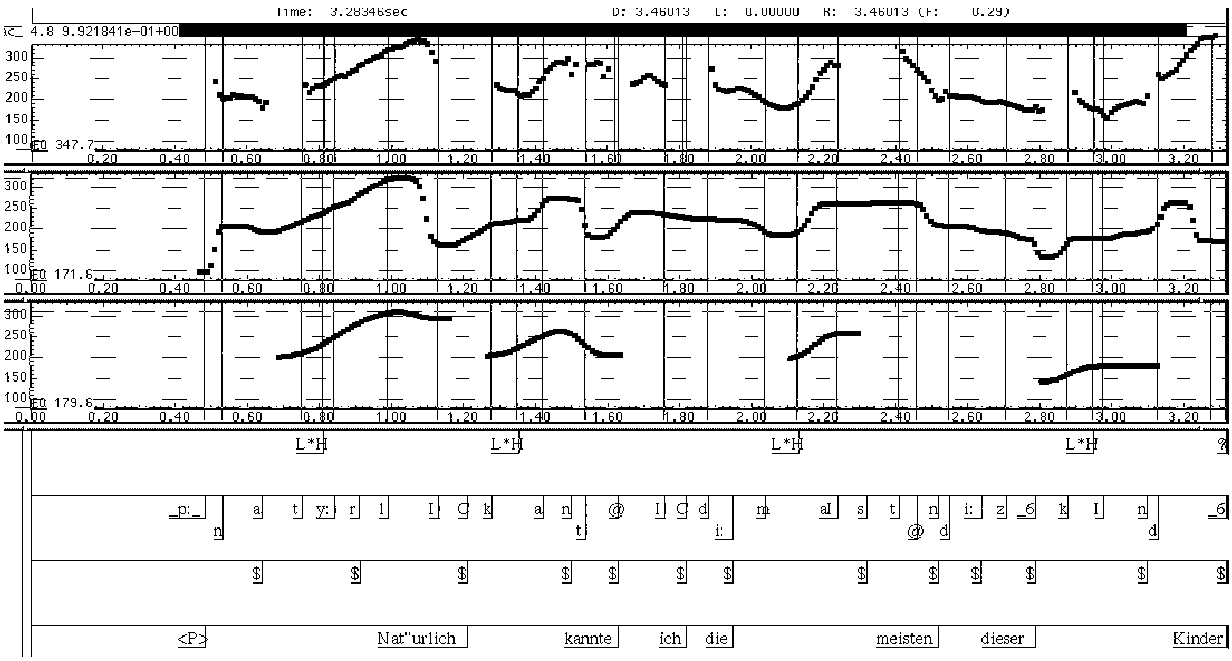

Figure 4. F_0 contour of the utterance *Natürlich kannte ich die meisten dieser Kinder* (Of course I knew most of these children); top: raw F_0 contour; middle: interpolated and smoothed F0 contour; bottom: F_0 of individual tonal events reconstructed on the basis of the parameters.

rise, but is rather intended to show how this approach to intonation description and generation can be helpful in the identification of IFA once more data is analyzed.

5. CONCLUSION

The identification of the relevant causes of intonational foreign accent is quite complex, since languages allow more variation in intonation than in segmental articulation. It is necessary to take context, discourse situation and other utterances by the same speaker into account when determining whether an observed deviation is a sign of foreign accent or still an acceptable variation . The impression of IFA may be caused mainly by an accumulation of small deviations that, if perceived individually, would still be within the range of permissible variation.

Nevertheless several important aspects of IFA are covered by a representation of intonation patterns based on the principles of the Tone Sequence Model. The presented category- and data-based methods of intonation description and generation are capable of detecting both higher-level categorical aspects and lower-level phonetic aspects of IFA.

The underlying representation of intonation as a linear sequence of tonal events has to (if at all possible) express characteristics of IFA that concern the shape of the overall contour in terms of categories (see the accumulation of up-and-down movements described in section 3.3.)

Consequently, a non-linear, but rather superpositional view of intonation might determine completely different aspects to be responsible for the perception of IFA.

While for a thorough identification of of IFA further analysis of more data is necessary, it has been demonstrated that the use of F_0 generation and resynthesis is very effective in the analysis of intonational foreign accent.

REFERENCES

[1] Ostendorf, M.., Price, P.J., and Shattuck-Hufnagel, S. 1995. *The Boston University Radio News Corpus*. Technical Report ECS-95-001, Electrical, Computer and Systems Engineering Department, Boston University, Boston, MA.

[2] Beckman, M. and G. Ayers 1994. *Guidelines to ToBI Labelling.* Version 2.0.

[3] Flege, J. 1995. Second language speech learning: Theory, findings and problems. In Strange, W. (ed.), *Speech perception and Linguistic Experience: Theoretical and Methodological Issues*Timonium, MD: York Press, 233 - 277.

[4] Kuhl, P. and P. Iverson 1995. Linguistic Experience and the "Perceptual Magnet Effect". In Strange, W. (ed.), *Speech perception and Linguistic Experience: Theoretical and Methodological Issues*Timonium, MD: York Press, 121 - 154.

[5] Beckman, M. and J. Hirschberg 1994. *The ToBI Annotation Conventions.*

[6] Féry, C. 1993. *German Intonational Patterns*. Niemeyer, Tübingen.

[7] Pierrehumbert, J. 1980. *The phonology and phonetics of English intonation*. Ph.D. Dissertation, M.I.T., Cambridge.

[8] Jilka, M., Möhler, G. and Dogil G. 1999. Rules for the generation of ToBI-based American English intonation*Speech Communication* (forthcoming)

[9] Möhler, G. 1998. *Eine theoriebasierte Modellierung der deutschen Intonation für die Sprachsynthese*. Doctoral Dissertation, University of Stuttgart.

[10] Möhler, G. and A. Conkie 1998. Parametric modeling of intonation using vector quantization. *Proceedings of 3rd ESCA Workshop on Speech Synthesis*, Jenolan Caves, Australia, 311-316.