Perceptual study of timing variables in F0 peaks

Oliver Niebuhr
Institute for Phonetics and Digital Speech Processing (IPDS)
Christian-Albrecht-University, Kiel, Germany

ABSTRACT

This paper is concerned with the role of alignment and peak shape of F0 peaks in the perception of intonational categories. The term intonation is restricted to speech melody. Several experiments on early and medial peaks in German show that both dimensions are relevant in this intonational contrast, mirroring the findings for comparable intonational contrasts in other languages. Hence, my findings support the assumption that all these intonational contrasts originate from a common psychophonetic mechanism which is linked to holistic contours rather than to its local features.

1. INTRODUCTION

F0 peak contours can be described in terms of three phonetic dimensions. First, in the alignment dimension laryngeal activity forming F0 peaks is coordinated with articulatory activity producing syllables. For describing their alignment, both units are represented by structurally significant points: peak maximum and vowel onset. With regard to these points, three alignment categories have to be distinguished. In left alignment the peak maximum is reached before vowel onset. Thus F0 is falling into the vowel. The opposite is true for right alignment. In central alignment peak maximum and vowel onset are synchronized. Alignment refers to the external timing of F0 peaks.

Second, differences in the slope of the rising and falling movements are related to the durational structure of F0 peaks and are thus a matter of internal timing. This is covered by the peak shape dimension.

While alignment and peak shape are both timing variables in F0 peaks, a third dimension, peak height, is a frequency variable. It describes the distance between top and bottom of F0 peaks.

Studies in different languages indicate that these phonetic dimensions of F0 peaks can be importmant acoustic cues for phonological contrasts in intonation. But so far, only the the relevance of alignment has been widely recognized across different languages and intonation categories. This is also true for the *Kiel Intonation Model* (KIM), developed by Kohler [1]. My research is based on this contour defined model of German intonation, which postulates different categories of *peaks* and *valleys*.

These categories emerged from comprehensive perception experiments, summarized by Kohler in [2]. Using a series of equal-sized temporal shifts of F0 peaks across the constant accented syllable /lo/ in the utterance "*Sie hat ja ge**lo**gen"* (*"She's been lying"*), Kohler finds a categorical change in perception. These two intonation categories are integrated in KIM as 'early' and 'medial' peaks, being associated with parallel changes along the semantic dimension from 'established' to 'new'. Kohler did not investigate systematic influences of peak shape. Therefore, the change from early to medial peak is ascribed exclusively to the alignment dimension, corresponding to the difference between rising and falling F0 after vowel onset. For medial peaks a rising F0 of about 60 ms after vowel onset is needed. Thus central or left aligned F0 peaks can only signal early peaks.

On the other hand, D'Imperio and House [3] as well as Gósy and Terken [4] demonstrate the interaction of alignment and peak shape in the perception of intonational contrasts akin to that of early and medial peaks. In both studies, shifting the F0 peaks from left or central to right alignment results in a categorical change from the perception of statements to the perception of questions. Additionally, [4] show that F0 peaks with slow rise and fast fall enhance the perception of statements while F0 peaks with fast rise and slow fall increase the perception of questions. The results of [3] indicate that in comparison with a blunt peak shape a pointed peak shape (at least one steeper slope) increases the statement judgements for left aligned F0 peaks and leads to an later and more abrupt change from statement to question judgements. Furthermore, larger standard deviations indicate less uniform judgements for blunt F0 peaks.

Apart from influences of peak shape, the findings of [2,3,4] are comparable. The F0 peak shift from left to right, for expample, induces the categorical change when the F0 rise into the vowel exceeds about 60 ms. In comparision with differently aligned plateau contours, Kohler [2] points out that F0 movements and not F0 levels are decisive for perceiving the contrastive intonation categories. This agrees with findings of D'Imperio and House [3]. These comparable findings support the view already proposed by Kohler [2] that intonational contrasts like these share a common psychophonetic basis which is functionally exploited in different ways. Therefore, it is expected that the interactive effects of alignment and peak shape on the perception of the contrastive peak contours noticed by [3] and [4] are also valid for early and medial peaks in German.

On the basis of this argumentation, my study introduces the peak shape dimension and explores the interaction of alignment and peak shape in the perception of early and medial peaks in German intonation.

 ISBN 1-876346-48-5

2. METHOD

To examine this hypothesis, the continuously voiced utterance *“Sie war mal **Ma**lerin”* (*“She was once a painter”*) was produced with an accent on /ma:/ of “***Ma**lerin”* (*“painter”*). Using *praat* [8] as the experimental tool, the plain F0 contour of this utterance was completely replaced by another contour, stylized at five contour points. Two of them determined the beginning (112 Hz) and the end (90 Hz) of the F0 contour. This slight declination is interrupted by three contour points, defining the frequency values of an F0 peak. It is composed of a rise from 105 Hz to 150 Hz (6,2 semi tones, st) and a fall from 150 Hz to 110 Hz (5,4 st).

Without changing the frequency values of this F0 peak, four different peak shapes were constructed. For this purpose, two different speeds were cross-combined with rise and fall of the F0 peak, resulting in two nearly symmetrical and two clearly asymmetrical peak shapes. The fast speed was to represent the physiological limits for changing F0. With reference to Xu and Sun [5] two slightly different gradients were chosen for the rise of 6,2 st and for the fall of 5,4 st (Fig. 1). The slow speed was to be discriminated from the fast one. Following the outcomes of several perception experiments, e.g. [6,7], the gradients of the fast speed were divided by two to get the gradients of the slow speed. This again yielded two slightly different gradients (Fig. 1).

The four different peak shapes are referred to as slow rise and fast fall (s/f), fast rise and slow fall (f/s), slow rise and slow fall (s/s), fast rise and fast fall (f/f). Figure 1 shows the gradients of the peak shapes and its durations.

A frame for an alignment continuum was created that spreads over the accented syllable /ma:/ in *“**Ma**lerin”*. The frame extended in 11 equal-sized steps of 20 ms from the onset of /m/ to 100 ms after the onset of /a/. Thus it covered all three alignment categories (left, central, right). Applying this frame to all 4 peak shapes resulted in four stimuli series (s/f, f/s, s/s and f/f), each with 11 resynthesized stimuli. In every series stimulus 6 was aligned centrally

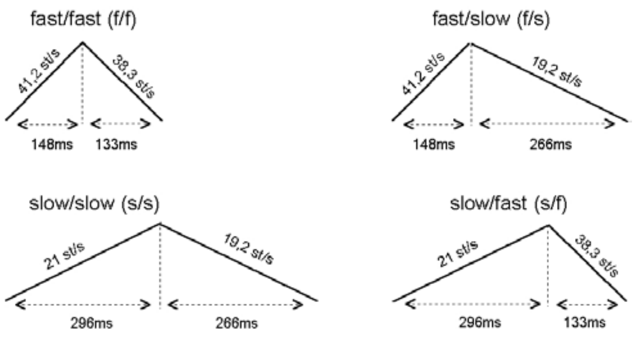

Figure 1: Gradients and transition durations of rises and falls of the four peak shapes used in perception experiments

For each stimuli series one AX-test and one identification test were run. The AX-test consisted of 9 nonidentical pairs and of corresponding 5 identical pairs. The stimulus numbers of the 14 pairs are listed in Table 1.

Table 1: The 9 nonidentical pairs and the corresponding 5 identical pairs in the AX-test of each stimuli series.

Nonidentical pairs	1/3	2/4	3/5	4/6	5/7	6/8	7/9	8/10	9/11
Identical pairs	1/1		3/3		5/5		7/7		9/9

The 14 pairs occurred ten times in a randomized order. This order was different for every stimuli series. The 140 pairs were judged as ‘same’ or ‘different’. Because the non-identical pairs contained only ascending stimulus numbers, effects of presentation order known from other perception experiments, e.g. [2,9], cannot be met in this study. But since it is expected that the interaction of alignment and peak shape itself is comparable for both presentation orders, such effects may be neglected here.

In the identification test, the 11 stimuli of each series were presented in the constant preceeding context *“Jetzt ver**steh**’ ich das erst”* (*“Now, I understand”*), which was naturally produced with a medial peak on *“-**steh**”*. Its lexical and intonational composition make the listener expect a new information. The context was produced by the same speaker as *“Sie war mal **Ma**lerin”*. F0 level, voice quality, speech rate and pause duration were suitable for perceiving both utterances as connected. The 11 pairs of each identification test were repeated 10 times in an incidental order. Subjects had to judge whether context and stimulus matched or not. Supposing that early peaks convey known information and medial peaks express new information, listeners will judge the pairs as ‘matching’ if they identify the F0 peak as a medial peak. Otherwise, they will judge the pairs as ‘non matching’. For this reason, I call this an identification test.

28 native speakers of German (24 females and 4 males), average age 21, participated in the perception experiments. They were divided into 5 groups. Each group took part in the AX- and identification test of all four stimuli series. While both tests of a stimuli series were never seperated, the presentation order of the four series was permuted across the five groups. The subjects listened to the stimuli from a loudspeaker in a sound treated room. They were instructed to judge as soon as possible (at most 3 seconds after each stimulus) by pressing buttons.

3. RESULTS

The results of both tests for each stimuli series are shown in Figures 2 and 3. Figure 2 displays the percentage of ‘different’ judgements for the ten repetitions of each stimulus as a function of F0 peak alignment. Figure 3 presents the same for the ‘matching’ judgements. From both Figures it is apparent that peak shape as well as alignment had an evident influence on the ‘different’ and ‘matching’ judgements of the 28 subjects.

Tables 2 and 3 exhibit the results of statistical analyses of the discrimination and identification data. Because the data do not fulfil the distributional requirements essential for parametric testing, nonparametric tests were performed.

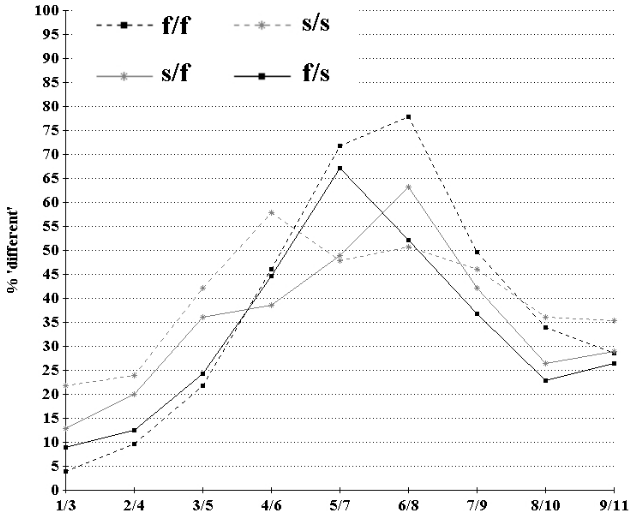

Figure 2: Discrimination functions of the four stimuli series s/f, f/s, s/s and f/f of the nonidentical pairs in the AX-test, 28 sbs., n=280.

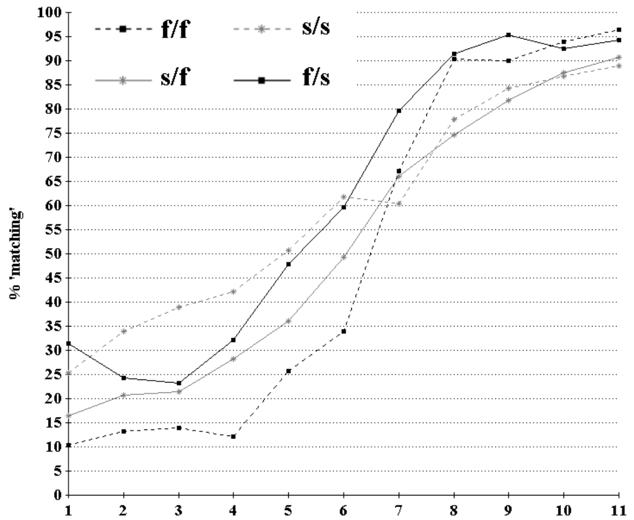

Figure 3: Identifikation functions of the four stimuli series s/f, f/s, s/s and f/f, 28 sbs., n=280.

Table 2: Results of an analysis of interaction structure [11] (two-tailed) on the basis of the mode distribution of 'different' judgements over the 9 nonidentical pairs of each stimuli series. For each subject, the first mode of 'different' judgements in the ascending order of the nonidentical pairs was taken. For the mode distribution, the first and the last three pairs of the ascending order were grouped together.

		χ^2	P
rise slow vs. fast	(f/f, f/s) vs. (s/f, s/s)	25,01	< 0,001
fall slow vs. fast	(f/f, s/f) vs. (f/s, s/s)	12,56	< 0,05
Rise and fall slow vs. fast	s/f vs. f/s vs. s/s vs. f/f	40,54	< 0,001

Table 3: Critical differences (D) and probabilities of alpha errors (P) for the Wilcoxon-Wilcox multiple comparisons test (two-tailed) [10] on the basis of the sums of 'matching' judgements from each subject before and after vowel onset.

		Before vowel onset (stimulus 1-6)		After vowel onset (stimulus 7-11)	
		D	P	D	P
Fast rise	f/s vs.f/f	41	< 0,01	6	n.s.
Slow rise	s/s vs. s/f	20	n.s.	7	n.s.
Fast fall	s/f vs. f/f	24	n.s.	32,5	< 0,01
Slow fall	f/s vs. s/s	3	n.s.	31,5	< 0,01
Rise and fall different	s/f vs. f/s	17	n.s.	38,5	< 0,01
	f/s vs. s/s	44	< 0,01	25,5	< 0,05

The four discrimination functions (Fig. 2) show that pairs from the centre of the alignment continuum were discriminated more frequently than pairs from both edges. Comparing the sums of 'different' judgements of all 5 identical pairs with those of the corresponding 5 nonidentical pairs of each series (Tab. 1), prooves that the nonidentical pairs were judged significantly more frequently as 'different' than the identical pairs (not shown here). Figure 3 illustrates a majority of 'non matching' judgements for the left aligned stimuli 1-4, whereas the right aligned stimuli 8-11 were predominantly judged as 'matching'. These common features of the functions in both Figures can be attributed to the alignment dimension.

Apart from these influences of F0 peak alignment, there are obvious differences between the discrimination and identification functions, pointing to influences of the peak shape dimension. In opposition to stimuli pairs with fast rises and/or falls, stimuli pairs with slowly rising and/or falling F0 peaks enhanced the discrimination in the first half of the alignment continuum (cp. s/s and s/f in Fig. 2). On the basis of mode distributions, these differences are significant (Tab. 2). Consequently, it can be seen from Figure 2 that only f/f clearly supported the discrimination of pairs in the second half of the alignment continuum. In comparison with fast rising F0 peaks in right alignment, the identification functions of slowly rising F0 peaks reveal a significant reduction of 'machting' judgements after vowel onset (cp. Fig. 3, Tab. 3). On the other hand, slowly falling F0 peaks increased the number of 'matching' judgements before vowel onset. Comparing f/s with f/f and f/f with s/s, these differences also yielded significance (Tab. 3).

Additionally, the data indicate that the peak shape with slow rise and fall (s/s) was judged with greater disagreement between the 28 subjects than the other peak shapes. In the discrimination function of s/s, this causes a broader distribution and a less pronounced maximum than in the remaining functions (Fig.2). For the same reason, the identification function shows less variation and a more gradual rise from 'non matching' to 'matching' judgements (Fig 3).

ISBN 1-876346-48-5 © 2003 UAB

4. CONCLUSIONS

In categorical perception, discrimination functions show a local peak with a clearly marked maximum, and identification functions are characterized by abrupt switches from one judgement to another at the point of maximum differentation. Although the results of none of my stimuli series fit strictly into these patterns, they are quite distinct from what is expected for a gradual change in perception. For this reason, I conclude that in all four stimuli series the alignment dimension induces a categorical change from early to medial peak. In the case of f/f, this categorical change is a feature of individual subjects rather than the group. Table 4 lists the stimuli which represent the approximate boundaries between early and medial peak categories for each stimuli series. The stimuli are defined as boundaries, because they are spanned by the pairs with the discrimination maximum and because they are the first in series to receive a majority of ‘matching’ judgements.

Table 4: Boundaries for the perception of early and medial peaks in each stimuli series

Peak shape	Boundary between early and medial	F0 peak maximum relative to vowel onset	Alignment category
f/f	Stimulus 7	+20 ms	Right
s/f	Stimulus 7	+20 ms	Right
f/s	Stimulus 6	0 ms	Central
s/s	Stimulus 5	-20 ms	Left

Besides the alignment dimension, two influences of peak shape dimension on the perception of early and medial peaks must be differentiated. First, peak shape influences the position of the boundary between early and medial peaks. F0 peaks with fast rise and slow fall (f/s) support the perception of medial peaks (cp. Fig. 3), shifting the boundary between both categories towards the syllable onset. The inverse peak shape (s/f) supports the perception of early peaks (cp. Fig. 3), shifting the boundary in the opposite direction, as shown in Table 4. Second, peak shape influences the sharpness of the boundary between early and medial peaks. A blunt peak shape (s/s) produces a less sharp boundary than a pointed peak shape (f/f), primarily caused by greater disagreement between subjects. In comparison with pointed peaks, blunt peaks also shift the boundary towards the syllable onset (Tab. 4).

The results show an interaction of alignment and peak shape dimension of F0 peaks in the perception of early and medial peaks in German intonation, which is consistent with the findings of [3] and [4] for comparable contrasts in other languages. This supports the assumption that the three intonational contrasts are derived from a common psychophonetic mechanism [2]. The results further show that all four peak shapes induced a change from early to medial peak. But peak shape alone could not bring about a change in perception. Different peak shapes only enhanced the perception of early and medial peaks for ambigiously aligned F0 peaks. I will therefore regard the alignment dimension as a more important acoustic cue than the peak shape dimension. Nevertheless, the results summarized in Table 4 refute the view suggested by Kohler [2] that rising F0 after vowel onset is essential for the perception of medial peaks. Finally, the influences of differently rising *and* falling slopes show that the perception of intonational contrasts depends on holistic contours rather than on local features.

5. OUTLOOK

The results of [4] also show an influence of peak height on the perception of the intonational contrast explored. With regard to a common psychophonetic mechanism, it is worth searching for a similar influence in German intonation. The examination of interactions of F0 peak dimensions in the production and perception of intonation categories should be expanded further to other intonational contrasts like that of medial and late peak [2].

REFERENCES

[1] Kohler, K. J. 1991. A model of German intonation. *AIPUK* ***25***, pp. 295-360

[2] Kohler, K. J. 1991. Terminal intonation patterns in single-accent utterances in German: phonetics, phonology and semantics. *AIPUK* ***25***, pp. 117-185.

[3] D’Imperio, M. and D. House. 1997. Perception of questions and statements in Neapolitan Italian. *Proceedings of Eurospeech ‘97*, *Rhodes, Greece*, pp. 251-254.

[4] Gósy, M. and J. Terken. 1994. Question marking in Hungarian: timing and height of pitch peaks. *Journal of Phonetics* ***22***, pp. 269-281

[5] Xu, Y. and X. Sun. 2002. Maximum speed of pitch change and how it may relate to speech. *JASA* ***111/3***, pp. 1399-1413

[6] ‘t Hart, J., R. Collier and A. Cohen. 1990. *A perceptual study of intonation. An experimental-phonetic approach to speech melody*. Cambridge/New York: Cambridge University Press.

[7] Nábelek, I. und I. J. Hirsh. 1969. On the discrimination of frequency transitions. *JASA* ***45/6***, pp. 1510-1519

[8] Boersma, P. and D. Weenink. Praat: doing phonetics by computer. http://www.fon.hum.uva.nl/praat/

[9] Hellström, Ake, Olli Aaltonen, Ilkka Raimo und Erkki Vilkman. 1994. The role of vowel quality in pitch comparison. In: JASA 96/4. S.2133-2139

[10] Sachs, L. 1972. Statistische Auswertungsmethoden. Berlin/New York: Springer.

[11] Fillbrandt, H. 1986. *Verteilungsfreie Methoden in der Biostatistik*. Meisenheim: Hain.