Form and Function of Falling Pitch Contours in English

Felicitas Kleber

Institute of Phonetics and Digital Speech Processing (IPDS)
Christian-Albrechts-University of Kiel, Germany
fk@ipds.uni-kiel.de

Abstract

This paper presents the results of a set of perception experiments concerning the phonological status of early, medial and late F_0 peak synchronization in English and the nature of the contrast between these categories. By means of one *identification* and two *discrimination tasks*, it has been shown that subjects perceive a categorical-like change when the F_0 maximum of a peak is shifted into the stressed vowel and a gradual change when the F_0 maximum is moved into the following unstressed vowel. Therefore, we conclude that the early peak constitutes a phonological category as opposed to medial peaks; late peaks form a phonetic continuum.

1. Introduction

In a study by Kohler ([5, 6]), it has been shown that the paradigm of categorical perception (for an overview see [12]) also applies to intonation contours in German. By means of perception experiments, two types of *discrimination* and of *identification tests*, [5, 6] found that listeners perceive a change when a F_0 peak is moved into the stressed vowel. This change is clearly categorical. Any other changes listeners perceived are gradual. [5, 6] has shown that an early peak constitutes a phonological category in opposition to a medial peak, whereas late peaks form a phonetic continuum. Furthermore, it has been verified that falling intonation contours which are synchronised with different positions in the stressed syllable of the word carrying sentence stress can be associated with different semantics and pragmatics: early peaks with *established facts* or *final summing up*, medial peaks with *new facts*, and late peaks with *unexpectedness*.

These findings have, among others, resulted in the setting up of the categories of early, medial, and late peaks in the *Kiel Intonation Model* (KIM) [7] with early and medial peaks being minimal pairs only differing in their phonetic alignment. Early peaks are synchronized with the segmental string preceding the accented syllable nucleus, medial peaks are synchronized with the accented syllable rhyme, and late peaks are synchronized with the following unaccented syllables. We follow Ladd's [8] definition of alignment being "a *phonetic* property of the relative timing of events in the F_0 contour and events in the segmental string" ([8]: 55). As for the phonological status of differences in alignment, we follow Kohler's terminology of synchronisation [7], which accounts for the phonological property of a connection between an event in the F_0 contour and a phonological element such as the syllable rhyme but also the syllable nucleus.

For English, Pierrehumbert and Steele [10] showed by means of an imitation experiment the presence of two F_0 peak categories. Although speakers produced a continuum of varying F_0 maximum alignments when confronted with a continuum of rising-falling-rising intonation contours ranging from early peak alignment to late peak alignment (represented as L+H* vs. L*+H in the Pierrehumbert system [9]), they favoured the production of two distinct F_0 maximum synchronizations, being either aligned with the accented or the following unaccented syllable. Note that in that study early alignment refers to early alignment with the syllable nucleus, i.e. medial as used here. The distribution of these two intonation patterns, however, showed that the productions of medial peaks outweighed those of late peaks by far. The authors concluded that the F_0 maximum alignment difference proposed in [9] must be phonological. Furthermore, they related the differences in synchronization with differences in meaning, medial peaks being associated with *assertion* and late peaks being associated with *incredulity*. The form-function relation, however, was not tested in this study.

Chen [3] took this experiment as a starting point to shed light on the form-function relation of the phonological contrasts found by [10] from a perceptual point of view. In this study again the focus was only on differences between F_0 peaks that are synchronized with the accented syllable or with syllables following the accented syllable. The discreteness of the two intonation patterns, however, was not confirmed. Rather, results suggest that the alignment difference between medial and late aligned F_0 peaks is gradual. Thus, the study differed from that of [10] with respect to the experimental set-up. Whereas in [10] the F_0 maximum as well as the F_0 minima preceding and following the F_0 maximum were shifted, in [3] only the F_0 maximum was shifted. Thus, peak shapes of all stimuli differed in [3].

Redi [11] has investigated both the discrete contrast between early and medial peaks (represented as H+L* and H*, respectively, in the ToBI system [1]) and the non-discrete contrast between medial and late peaks (also represented as H* in the ToBI system). The investigation was done within the production paradigm set out in [10]. As to the early/medial contrast, subjects favour two intonation patterns which are in accordance with the proposed two phonological categories. Her findings for this distinction are the first reported for English. As for the medial/late contrast, subjects tended to produce varying intonation patterns, ranging from medial to late peak patterns. The results for the early/medial and medial/late contrasts cannot be interpreted as both showing category boundaries. Rather, they suggest that only the contrast between early and medial peaks is discrete, while the contrast between medial and late peaks is gradual.

Thus previous studies lead to new research questions for the following reasons: (1) the aspect of the form-function relation has been discarded in [10, 11], (2) findings on the medial/late contrast diverge in [10, 3], (3) findings concerning the medial/late contrast in [11] can be interpreted differently and therefore need clarification.

The investigation presented here is done within the KIM framework [7]. Phonological categories are not postulated *a priori*. The studies discussed above suggest the following intonational synchronization contrasts in English: (1) there is

a categorical intonational contrast between F_0 maxima which are aligned with the accented syllable onset and F_0 maxima which are aligned with the accented syllable rhyme and (2) there is a gradient intonational contrast between F_0 maxima which are aligned with the accented syllable rhyme and F_0 maxima which are aligned with the following unaccented syllable(s). On the basis of the above-mentioned research questions the following hypotheses are formulated:

(1) Shifting an F_0 peak with a fall through a single-accented utterance from the start of the accented syllable onset to a medial position yields a categorical change in perception when the F_0 maximum enters the accentual vowel. The categorical change correlates with a semantic change along the dimension *old information*/*new information.*

(2) The corresponding shift from a medial position of the nucleus of the accented syllable to the offset of the following unaccented syllable yields a gradual change in the perception of medial and late peaks. The gradual change correlates with a semantic change along the dimension *new information*/*incredulity.*

(3) Physically identical stimuli are judged the same more often than physically non-identical stimuli.

With regard to the findings of [5, 6] as well as [4], it is necessary to divide the syllable further into its constituents instead of just speaking of accented and unaccented syllables. Thus, the onset and the nucleus of the accented syllable may trigger significantly different results for different alignment positions. Therefore, for the rest of this paper, we differentiate between syllable onset and syllable rhyme. A further distinction between nucleus and coda is rejected for this investigation, since we believe that the VC transition does not present the listener with further acoustic cues. However, the distinction made between onset and rhyme, we claim, does trigger different reactions no matter whether the listener is confronted with an accented or an unaccented syllable. But the rejection of the latter distinction is only preliminary, i.e. further investigation will be needed.

2. Method

The hypotheses were tested within the peak-shift and semantic contextualization paradigm set out in [5, 6].

2.1. Stimuli and experimental set up

Both, the test sentence *She's gone to Malaga* and the context sentence *I almost forgot* were spoken by a male native speaker of Southern British English. The informant produced a number of tokens along a phonetic continuum of the F_0 maximum from early to late. One token was chosen which had the F_0 maxima central on the syllables <*Mal*> and <*got*>. The context sentence provides the required communicative frame for judgments in the *identification test.* We believe that the stimuli containing early peaks will not fit the context sentence, stimuli containing medial peaks will fit the context and stimuli containing late peaks will fit the context sentence to a lesser extent.

One token of the test sentence with an auditorily classified medial peak served as the basis for stimuli generation. All stimuli were generated using *praat* 4.0.50 [2] and resynthesized using the PSOLA-technique. The F_0 contour of the original utterance was removed completely and replaced by a new stylized contour with the following five F_0 values (in temporal order): (1) 111 Hz, (2) 114 Hz, (3) 140 Hz, (4) 96 Hz, and (5) 76 Hz. Whereas (1) and (5) constitute temporarily fixed time points, (2), (3), and (4), i.e. the values modelling the peak contour were shifted in parallel. (1) and (5) were located at the beginning and end of voicing of the entire utterance. The time span between (2) and (3) as well as (3) and (4) amounted to 128 ms giving pitch differences of 27,8 st/s for the rising slope and 51,03 st/s for the falling slope. As a point of departure, (3) was placed at 42 ms after the CV boundary /mæ/ and then shifted together with (2) and (4) in 4 and 6 equal steps of 29 ms to the left and to the right, so that the first and the last stimulus of the continuum were aligned with the onset of /m/ and with the offset of the second /æ/ in *Malaga*, respectively. The utterances were then resynthesized creating a peak alignment continuum. Stimulus 11 differs in that the falling slope of the peak contour is compressed. However, after auditory evaluation, this stimulus was included in the alignment continuum. Figure 1 illustrates the F_0 peak shape and the alignment shift schematically.

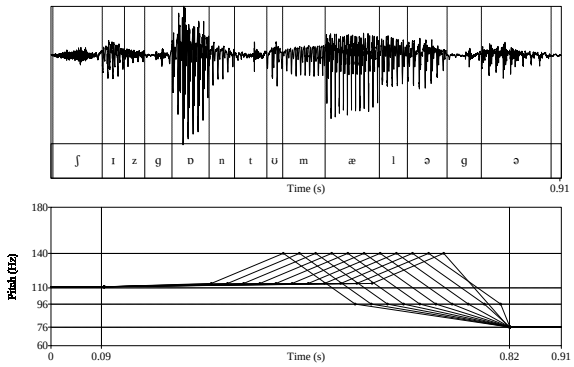

Figure 1: *Schematic illustration of the peak contour shift.*

Table 1 gives the temporal values of the F_0 maximum alignment in relation to one of the two possible CV transitions either into the stressed or the following unstressed vowel.

Table 1: *Temporal positions of F_0 maxima in relation to the CV transitions* /mæ/ *and* /lə/.

Stimulus	Segment	Peak position in relation to nearest CV transition [ms]
1	/m/	-74
2	/m/	-45
3	/m/	-16
4	/æ/	+13
5	/æ/	+42
6	/æ/	+71
7	/æ/	+100
8	/l/	-22
9	/ə/	+7
10	/ə/	+36
11	/ə/	+64

Following the peak shift paradigm set out in [5, 6] a *serial* (A) and a *randomized AX or XA discrimination test* (B) as well as an *identification test* (C) were prepared. In (A), stimuli were presented in ascending order; in (B), stimuli with varying, i.e. stimuli with a step size of 2 through the series in both directions, as well as identical (only odd numbers as control pairs) F_0 peak positions were paired; in (C), stimuli

were paired with the context sentence that was kept the same throughout all pairs.

Both, the 24 pairs of the *randomized discrimination test* as well as the 11 pairs of the *identification test* were each reproduced 10 times and randomized.

2.2. Subjects and Task

A group of 25 phonetically naïve speakers (20 female and 5 male) who speak a variety of Southern Standard English participated in the perception experiments. All subjects grew up and were educated in the south of the UK. They were students at the University of Cambridge. None of the subjects reported any speech or hearing pathology. Subjects were paid at standard compensation rate.

The perception experiments took place in a lecture room at the University of Cambridge. The stimuli were presented to the group of listeners via loudspeakers.

In the *serial discrimination test*, listeners were presented with the series once for each of the following questions: the subjects were asked to indicate on prepared answer sheets (1) whether they perceived any change(s) in the melody of the utterance from one stimulus to the next, (2) in the case of perceivable changes, at which stimulus in the sequence they perceived the first melodic change compared to the preceding stimuli, at which stimulus they perceived any further change(s), and (3) which meanings they associated with the original utterance and the perceived change(s). In order to answer the third task, subjects were first presented stimulus 1 separately and afterwards with the series ranging only from stimulus 2 to 11.

In the *randomized discrimination test*, subjects were asked to mark on prepared answer sheets whether they perceive a difference between the members of a pair.

In the *identification test*, the subjects were asked to indicate on prepared answer sheets whether test and context sentence match semantically.

Two practice runs were done before each of the last two experiments proper. The experiments took one hour. The *serial discrimination test* is to be taken as more informal since elicitation matters are not uncontroversial. Nevertheless, results were taken into consideration.

3. Results

Because of too many missing responses in both the *randomized discrimination test* and in the *identification test*, data from two subjects were excluded from the analysis.

Results from the first task are as follows: 12% of the subjects perceived no change, 48% perceived one change and 40% perceived several changes.

Table 2: *Frequency distribution of change has occurred judgments. N = 25. (c. p. = change perceived).*

	2	3	4	5	6	7	8	9	10	11
1st c. p.			13	6	2		1	1		
Further c. p.				3	2		5	6	3	1
Total	0	0	13	9	4	0	6	7	3	1

In Table 2 results from the *serial discrimination task* are presented. Discrimination function shows a major peak (52%) at stimulus 4 and a minor peak (24%) at stimulus 9. Both peaks coincide with the first stimuli after a CV transition. The change in the perception of early and medial peaks is categorical, whereas the change in the perception of medial and late peaks is gradual.

The results for the third task suggest that subjects associated, in most cases, an *established fact* and *dejection* with early peaks, a *new fact* with medial peaks, and *incredulity* and *disbelief* with late peaks.

Figure 2 gives the discrimination functions from the *randomized discrimination task*. Discrimination curves for the non-identical pairs in ascending order show a major peak around stimuli 4 and 5 and a minor peak at stimulus 10. Listeners perceived a difference between members of a pair when the stimuli span the CV transition, but also when stimuli do not span the CV transition but differ as to the amount of the fall or rise of the slope within the accented vowel, as in the case of the stimulus pair 4/6. The change in the perception of stimulus pairs 1/3 versus 2/4 is categorical; and this coincides with the fact that there is a CV boundary in 2/4 as against 1/3. As for the discrimination curve of non-identical pairs in descending order, a major peak around stimulus pairs 6/4 and 7/5 and a minor peak around the stimulus pair 9/7 occur. Here, however, changes in the judgments of stimulus pairs are gradual. Thus, a strong order effect has to be taken into account, i.e. perception depends on the presentation modus. Subjects judged identical pairs throughout the acoustic continuum as more often perceived as the same than physically non-identical stimuli.

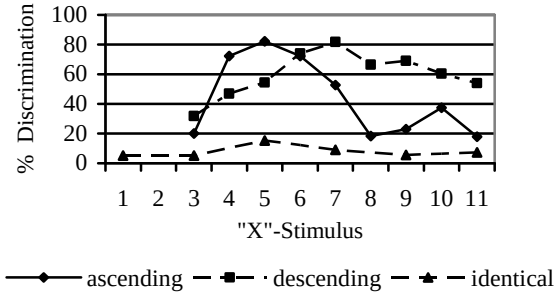

Figure 2: *Discrimination functions in the randomized AX or XA discrimination test. N = 23.*

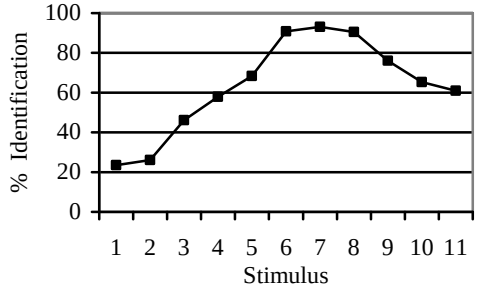

Figure 3: *Identification functions in the identification test. N = 23.*

Figure 3 presents the identification function for the *identification experiment*. It shows an identification maximum around stimuli 6, 7, and 8 and an identification minimum around stimuli 1 and 2. The change between the judgments is gradual. The identification function decreases with respect to late stimuli but does not change to non-matching judgments. Less than 50% of the subjects tended to opt for *do not match*

when it came to the judgment of stimulus 3 and more than 50% of the subjects tended to opt for *do match* when it came to the judgment of stimulus 4. Nevertheless, the difference between listeners' reactions with regard to stimuli 3 and 4 is not significant.

4. Discussion and Conclusions

With regard to the paradigm of categorical speech perception and the findings of [5, 6], we can draw the following conclusion. As for the *serial discrimination test*, results appear to be similar to those found for German (cf. [5, 6]). The different meanings associated with different peak categories correspond with the semantics found for early, medial and late peaks in German and late peaks in English (cf. [5, 6, 10]).

Hypothesis (1) can only be partially verified: The shift of an F_0 peak with a fall through a single-accented utterance from the start of the accented syllable onset to a medial position yields a categorical change in the perception when the F_0 maximum enters the accentual vowel. The categorical change correlates with a semantic-pragmatic change along the dimension *old information*/*new information*'. This categorical change is reflected in the discrimination functions showing significant local maxima for stimuli in the *serial discrimination task* and for stimulus pairs in ascending order in the *randomized AX or XA discrimination test*, whose F_0 maxima span the CV transition of the stressed syllable. However, no such discrimination maximum can be found for stimulus pairs in descending order as would have been expected as a function of categorical speech perception. Moreover, the identification function shows no abrupt changes between the judgements of the stimuli spanning the CV transition of the stressed vowel. That is, categorical speech perception in its traditional sense cannot be confirmed. Nevertheless, we cannot deny the form-function relation of falling contours in association with diverging synchronization patterns.

Table 3: *Summary of subjects' perception behaviour in the identification experiment. Cat = categorical, grad = gradual.*

group	synchronization category	contrast	subjects
1	early vs. medial	cat	5
2	early vs. medial	grad	4
3	medial vs. late	cat	0
4	medial vs. late	grad	0
5	early vs. medial vs. late	cat/cat	2
6	early vs. medial vs. late	grad/grad	1
7	early vs. medial vs. late	cat/grad	2
8	early vs. medial vs. late	grad/cat	2
9	no grouping possible		7
Total			23

As regards hypothesis (2), the change has proven gradual (cf. [3]) and it correlates with a semantic change along the dimension *new information/incredulity* (cf. [10]). A closer look at the identification curves for individual speakers showed that subjects vary in their perception of categories and category boundaries (see Table 3). Against the background of hypotheses (1) and (2) it is in particular interesting to note that two subjects perceive a clear category boundary between medial and late aligned peaks but none between early and medial peaks. 7 subjects cannot be assigned to any group. The numerically second strongest group consists of subjects who perceive early and medial peaks as clearly different.

Hypothesis 3 can be verified, too.

The nature of the contrast between early and medial as well as medial and late peaks cannot be treated as being the same as was done in [11]. As to the former, we are presented with a relatively clear category boundary, whereas for the latter no such clear category boundary occurs. Results support the assumption of three categories of early, medial and late peaks with early and medial peaks constituting phonological categories opposed to each other (cf. [3]) and late peaks being phonetically but not phonologically different from medial peaks (cf. [3]).

It has been proven that listeners react to very subtle differences in the alignment once the peak is shifted into the vowel (13ms). This finding is in accordance with the interpretation of [5, 6] that the perception of only a small proportion of a rising F_0 contour is perceived when it co-occurs with a vowel - both stressed and unstressed (cf. [4]). Results suggest that the perception of intonational contrasts is a matter of the synchronization with the stressed nucleus and not only with the accented syllable. Changes in perception occur when the F_0 maximum of the peak is moved into a vowel. This change is particularly prominent when moved into the vowel of the accented syllable. This supports theories in favour of the CV-transition (cf. [5, 6]) as being important for perceiving semantic differences of intonation.

The differences in the reaction to the presentation mode, i.e. descending and ascending, illustrate clearly the need of further investigation of the phenomenon of order effects in the perception of intonational contrasts.

5. References

[1] Beckman, M. E.; Ayers-Elam, G., 1997. Guidelines for ToBI labelling. URL: http://www.ling.ohio-state.edu/research/phonetics/E_ToBI.

[2] Boersma, P.; Weenik, D. Praat: doing phonetics by computer. URL: http://www.fon.hum.uva.nl/praat/.

[3] Chen, A., 2003. Reaction time as an indicator of discrete intonational contrasts in English. *Proc. Eurospeech 2003*, 97-100.

[4] House, D. 1996. Differential perception of tonal contours through the syllable. *Proc. ICSLP* 2048–2051.

[5] Kohler, K.J., 1987. Categorical pitch perception. *Proc. 11th ICPhS, vol. 5, Tallinn*. 331-333.

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

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

[8] Ladd, D.R., 1996. *Intonational Phonology*. Cambridge: Cambridge University Press.

[9] Pierrehumbert, J. B., 1987. *The Phonology and Phonetics of English Intonation.* PhD. Diss., MIT.

[10] Pierrehumbert, J.B.; Steele, S.A., 1989. Categories of tonal alignment in English. *Phonetica* 46, 181-196.

[11] Redi, L.C., 2003. Categorical effects in production of pitch contours in English. *15th ICPhS, Barcelona, Vol. 3*, 2921-2924.

[12] Repp, B. H., 1984. Categorical perception: Issues, methods, findings. In *Speech and Language: Advances in Basic Research and Practice*, vol. 10, N.J. Lass (ed.). Orlando FL: Academic Press, 243–335.