Pitch Scaling as a Perceptual Cue for Questions in German

Jan Michalsky
Department of German Studies, University of Oldenburg
jan.michalsky@uni-oldenburg.de

Abstract
Recent studies on the intonation of German suggest that the phonetic realization may contribute to the signaling of questions. In a previous production study polar questions, alternative questions and continuous statements were found to differ by a gradual increase in pitch scaling of a phonologically identical final rising contour [1]. Based on similar findings for Dutch Haan [2] concludes that the meaning signaled by the phonetic realization indicates an attitude rather than a categorical function. This is supported by Chen’s [3] perception studies on question intonation in Dutch, Hungarian and Mandarin Chinese as well as early findings for German by Batliner [4]. This paper investigates whether the phonetic realization of intonation in questions signals the categorical pragmatic function of ‘interrogativity’ or rather a ‘questioning’ attitude. Additionally, we investigate, which phonetic parameter is the decisive cue to this meaning. Three perception studies are reported: a combination of an identification and discrimination task, an imitation task, and a semantic rating task. Results suggest that the phonetic implementation of intonation in German questions signals an attitude rather than a linguistic category and that this function is primarily signaled by the offset of the final rising pitch movement.

Index Terms: questions, pitch perception, phonetic implementation, paralinguistics

1. Introduction

Early studies of intonation [5, 6, 7], including German [8, 9, 10, 11], investigated the association of certain intonational patterns with grammatical functions like the commonly assumed rising question intonation. It can be shown that such a grammatical intonation does not exist. In German prototypical rising intonation is limited to certain question types, like polar questions, and excludes frequent question types, like wh-questions (cf. [12]). Additionally, rising intonation is a common means of signaling continuation in statements (cf. [13]). Lastly, corpus studies of spontaneous speech have shown that even in polar questions rising intonation can be described as a preference at best [14, 15, 16, 17]. It has been suggested that intonation operates on a more abstract semantic level, which allows every sentence type to be realized with every contour of a languages intonation system [13, 14].

Since there are many ways to ensure that an utterance is perceived as a question, like syntax, pragmatic context, gesture, or facial expression, cues on the intonational level are not necessary [18, 19, 20]. Nevertheless, there is evidence that the choice of contour might be supplemented by features of the phonetic realization cueing ‘interrogativity’ as a pragmatic function (e.g. for German [4, 8, 9], for Dutch [2]). This has been supported by a recent production study [1]. It was found that continuous statements, alternative questions, and polar questions with the same syntactical, lexical and tonal structure showed significant differences in the phonetic implementation of the final rising movement. Questions were regularly realized with a larger excursion of the final rise than statements.

Furthermore, it was shown, that the phonetic scaling not only distinguished questions from statements but also final questions from non-final questions through a larger excursion. On the one hand, this might suggest that the phonetic scaling signals two sets of linguistic categories: ‘interrogativity’ and ‘finality’ (cf. biological codes [21]). On the other hand, this might indicate that the phonetic realization does not signal categorical pragmatic functions at all but rather a continuous meaning dimension [2, 3]. One such attitude can be described as ‘questioning’ in the sense of the speaker’s gradual demand for information [4]. Accordingly, alternative questions are distinguished from statements by having a demand for information, and polar questions are distinguished from alternative question by having a more immediate demand for information. In this paper, we investigate if pitch scaling of the final rise excursion is perceived as a continuum or as distinct phonological categories. A perceptual continuum would support the theory of a continuous meaning dimension while distinct categories would support the theory of distinct linguistic functions. For this purpose we carried out, an experiment in the style of a categorical perception paradigm [22] and an imitation experiment [23].

Additionally, the production study [1] shows that the phonetic effects of interrogativity show no variation for prenuclear accents. Accordingly, by raising the final rise and keeping the prenuclear region constant four phonetic parameters change at once: the final rise height, the final rise offset, the difference between nuclear and prenuclear pitch excursions and the slope of the tonpline between f0-maxima. To answer the question, which of the four parameters contributes to the degree of perceived ‘questioning’ a third experiment was conducted with resynthesized stimuli to isolate the four features and their respective influence. The two goals of this paper are therefore to investigate if the phonetic scaling of the final rising movement is perceived continuously or categorically and which dimension of the phonetic implementation contributes most to the perception of ‘interrogativity’ or ‘questioning’.

2. Experiment 1

2.1. Material

The stimuli for all three experiments were resynthesized from the recordings of a female speaker from the federal state of Lower Saxony in northern Germany. The stimulus for experiment 1 included a single sentence with declarative syntax and exclusively sonorous material in the nuclear region.
were no prenuclear accents. The nuclear contour was realized as a low-rise L*H H%, which corresponds to L*H-H% in classical ToBI [24]. A realization with a natural final rising movement close to the mean value of the desired target stimuli was used for manipulation. Resynthesis was carried out with PSOLA using the Praat package for audio editing [25]. The f0-Onset, the onset of the final rise, and the high leading tone were kept constant. The final offset was varied in steps of 1 semitone ranging from a final rise of 4 semitones to 10 semitones resulting in 7 equidistant levels (see figure 1).

Figure 1: Stylized test sentence for experiment 1 with tonal annotation and the 7 levels for the final rise offset.

For the identification task 10 repetitions of each stimulus were brought into randomized order resulting in 70 stimuli. All stimuli were concatenated with an inter-stimulus interval of 4 seconds. For the discrimination task every stimulus was paired with itself, the two stimuli that were one semitone apart in rise excursion and the two stimuli that were two semitones apart (e.g. 6+6, 6+5, 6+7, 6+4, 6+8). The ‘same’ stimuli pairs were multiplied by 15, the ‘different’ stimuli pairs by 5 resulting in 105 ‘same’ and 110 ‘different’ stimuli pairs for a total of 215. The stimuli of each pair were concatenated with an inter-stimulus interval of 750 ms. All stimuli pairs were again brought into a randomized order and concatenated with an inter-stimulus interval of 4 seconds.

2.2. Speakers

The first experiment was conducted with 20 students (10 female, 10 male) from the University of Oldenburg. All subjects were between 18 and 30 years old, born and raised in the north-western part of Lower Saxony and monolingual speakers of German.

2.3. Procedure

According to the classical categorical perception paradigm [4, 22, 26] two experiments were conducted. In an identification task subjects received a questionnaire with a written presentation of the test sentence and two possible interpretations, one with a question mark indicating a question, the other with a possible continuation of the sentence indicating an incomplete statement. The subjects listened to the 70 stimuli via headphones and were instructed to decide for every stimulus whether they perceived it more as a declarative question or a continuous statement. In the subsequent discrimination task the subjects listened to the 215 stimulus pairs with the instruction to decide whether they perceived the presented pair as ‘same’ or ‘different’ via forced choice.

2.4. Results

Figure 2 illustrates the results for both the identification and the discrimination task. To account for individual differences the data were centered at the crossing point of the identification curve for every subject. The identification curve shows ambiguous results regarding the research question. On the one hand, there is a rather abrupt jump from 30 % to 75 % in question judgement. On the other hand, the transition from a definite (90 %) statement to a definite question decision takes from ca. 45 % to 80 % on the centralized scale which corresponds to 3 to 4 semitones. Additionally, the discrimination curve shows a relatively smooth horizontal trajectory without any peak thus indicating a lack of a categorical boundary. The identification task furthermore shows that subjects were capable to attribute the differences in phonetic implementation to the two pragmatic functions when faced with a forced choice decision. The discrimination curve additionally shows that stimuli pairs with a difference of two semitones in the final rise excursion were correctly identified as different with a chance of 80 % to 90 %.

Figure 2: Question response for the identification task and correct discrimination for the 1 semitone and 2 semitones stimulus pairs.

3. Experiment 2

3.1. Material

For the imitation experiment the same 70 stimuli were used as in the identification task in experiment 1 with the same randomized order and the same inter-stimulus interval.

3.2. Speakers

The second experiment was conducted with 17 students (10 female, 7 male) from the University of Oldenburg who did not participate in the first experiment. The criteria were the same as for the first experiment.

3.3. Procedure

Subjects listened to the list of 70 stimuli via headphones. The subjects were instructed to pay close attention to the speech melody. After each stimulus the subjects imitated the heard stimulus, again with the instruction to try to imitate the speech melody as precisely as possible. Recordings were made in a sound-proof booth in the University of Oldenburg. A portable digital recorder (Tascam HD P2) was used at a sampling rate of 48 kHz and 16 bit resolution with a head-mounted microphone (DPA 4065 FR).
3.4. Acoustical analysis

For the acoustical analysis f0 was measured at the lowest point in the accented syllable as well as at the last traceable f0 value at the final phrase boundary. To account for inter-individual and sex differences all hertz values were logarithmically transformed into a semitone scale.

3.5. Statistical analysis

For statistical analysis we used linear mixed effects models. STIMULUS was used as a fixed factor, SPEAKER as a random factor, and final rise excursion as a dependent variable. In a post-hoc analysis the parameter estimates for every neighboring pair of steps were compared applying the bonferroni correction.

3.6. Results

Figure 3 shows a gradual increase of the produced final rise excursion depending on the excursion of the reference stimulus. Table 1 shows that every level of the imitated continuum is significantly distinguished from the previous and the following level thus showing no signs of establishing categories in the phonetic continuum. The only exception is between stimulus 5 and 6 which constitutes the transition from a plateau to a slight rise.

![Figure 3: Final rise excursion in semitones of the reference stimulus and the subject responses.](image)

Table 1: Mean values and statistical effects for STIMULUS of the pairwise comparisons using the bonferroni correction.

<table>
<thead>
<tr>
<th>Stimulus (ST)</th>
<th>4</th>
<th>5</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Response (ST)</td>
<td>4.30</td>
<td>4.58</td>
<td>n.s.</td>
</tr>
<tr>
<td>Stimulus (ST)</td>
<td>5</td>
<td>6</td>
<td>p</td>
</tr>
<tr>
<td>Response (ST)</td>
<td>4.58</td>
<td>5.58</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Stimulus (ST)</td>
<td>6</td>
<td>7</td>
<td>p</td>
</tr>
<tr>
<td>Response (ST)</td>
<td>5.58</td>
<td>6.47</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Stimulus (ST)</td>
<td>7</td>
<td>8</td>
<td>p</td>
</tr>
<tr>
<td>Response (ST)</td>
<td>6.47</td>
<td>7.83</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Stimulus (ST)</td>
<td>8</td>
<td>9</td>
<td>p</td>
</tr>
<tr>
<td>Response (ST)</td>
<td>7.83</td>
<td>9.18</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Stimulus (ST)</td>
<td>9</td>
<td>10</td>
<td>p</td>
</tr>
<tr>
<td>Response (ST)</td>
<td>9.18</td>
<td>10.27</td>
<td>&lt;.001</td>
</tr>
</tbody>
</table>

3.7. Experiment 3

4.1. Material

The test sentence for experiment 3 differed from the previous two by addition of an equally segmentally controlled prenuclear stretch. The prenuclear region received a H*L pitch accent.

Resynthesis was again carried out with PSOLA using the Praat package for audio editing [25]. The f0 onset and the low prenuclear leading tone were fixed. The prenuclear peak, the rise onset and the rise offset were synthetically varied in two steps of two semitones in relationship to the reference value (see figure 4).

![Figure 4: Stylized f0 contour for the stimulus of experiment 3 with three levels for prenuclear peak, rise onset and rise offset.](image)

The three values for each of the three variables resulted in 27 possible combinations, excluding the 2-2-2 stimulus as a reference stimulus. Every stimulus followed the reference stimulus after 750 ms. The 26 stimuli pairs were randomized for five different lists and concatenated with an inter-stimulus interval of 4.5 seconds.

4.2. Speakers

Again, the third experiment was conducted with 10 students (5 female, 5 male) from the University of Oldenburg. The criteria were identical with the first two experiments and again a different group of subjects was used.

4.3. Procedure

All subjects listened to the stimuli via headphones in a sound-proof booth in the University of Oldenburg. The subjects were instructed to rate on a 7-point scale, which stimulus of each pair they perceived as more ‘questioning’. The first three points indicated that the reference stimulus was more ‘questioning’, the last three points indicated that the test stimulus was more ‘questioning’ and the midpoint indicated that both stimuli were perceived as equally ‘questioning’.

4.4. Statistical Analysis

For the statistical analysis linear mixed effect models were used. We included PRENUCLEAR PEAK, RISE ONSET, RISE OFFSET, and SPEAKER SEX as fixed factors and SPEAKER as a random factor. The dependent variable was the degree of perceived ‘questioning’.

4.5. Results

Table 2 shows the statistical results for the three variables. Figure 5 illustrates the effects for each of the three factors independently in a simplified way. FINAL RISE OFFSET shows significant influences on perceived ‘questioning’ in a way that a higher offset indicates a higher degree of perceived ‘questioning’. The PRENUCLEAR PEAK shows significant effects on the degree of perceived ‘questioning’ but, as can be seen in figure 5 and table 2, the differences in the mean values are small.
and opposite to the expected direction. Raising of the PRENUCLEAR PEAK does not decrease but increase the degree of perceived ‘questioning’. The RISE ONSET shows only marginal effects and does not reach statistical significance. Additionally, both raising and lowering increases perceived ‘questioning’. No effects for SPEAKER SEX were found.

<table>
<thead>
<tr>
<th>Female</th>
<th>Male</th>
<th>P</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>4.83</td>
<td>4.74</td>
<td>0.22</td>
<td>n.s.</td>
</tr>
<tr>
<td>Prenuclear peak 1</td>
<td>Prenuclear peak 2</td>
<td>Prenuclear peak 3</td>
<td>4.63</td>
</tr>
<tr>
<td>Rise onset 1</td>
<td>Rise onset 2</td>
<td>Rise onset 3</td>
<td>4.83</td>
</tr>
<tr>
<td>Rise offset 1</td>
<td>Rise offset 2</td>
<td>Rise offset 3</td>
<td>3.49</td>
</tr>
</tbody>
</table>

Table 2: Mean values and statistical effects for Prenuclear peak, rise onset, rise offset and speaker sex on the degree of perceived ‘questioning’.

![Figure 5: Effects of rise offset, prenuclear peak and rise onset on the degree of perceived ‘questioning’](image)

The horizontal line marks an evaluation of the stimuli as equal.

## 5. Discussion

Experiment 1 shows that the scaling effects in the final rising movement can be perceived as a cue to questions at least when presented with a binary forced choice decision thus supporting previous findings of Batliner [4] for German and Chen [3] for Dutch, Hungarian and Mandarin Chinese among others. While a previous production study [1] shows that the mean difference between questions and continuous statements ranges from 2 to 6 semitones, the discrimination task suggests that the magnitude of the production effects is perceptively sufficient.

Regarding the first research question we reach the following conclusion: The identification curve allows no unambiguous interpretation for a continuous or a categorical distinction. The discrimination task on the other hand shows no peak for either stimuli pair condition indicating that there is no category boundary. Thus, the sharp rise in the identification curve suggests pragmatic prototypes associated with a phonetic and potentially semantic continuum. This interpretation is supported by the results of experiment 2. The imitation data shows that subjects are able to reproduce a phonetic continuum with levels as small as one semitone apart kept distinct. This supports the assumption of a paralinguistic continuum. This point is lastly reinforced by the results of experiment 3, where subjects uniformly chose a continuous scale to evaluate the perceived feature of questions as continuous though having the opportunity to make a binary choice.

Regarding the second research question the third experiment answers which of the four possible phonetic features, final rise excursion, final rise offset, excursion difference, or topline slope, contributes the most to the interpretation of ‘interrogativity’ or ‘questioning’. The fact that the lowering of the final offset significantly decreases and the raising significantly increases the perceived degree of ‘questioning’ shows that at least one of the four parameters correlates with the perception of questions and thus indicates the perceptual usability of the phonetic effects found in the production studies. Furthermore, the results show that raising of the prenuclear peak increases the perceived degree of ‘questioning’. This contradicts the expectations since a previous production study showed no significant effects for scaling in the prenuclear region [1]. This suggests that both the topline and the excursion differences co-occur with raising of the final rise but do not constitute cues to questions on their own. Both lowering and raising of the rise onset also slightly increases the degree of perceived ‘questioning’. This shows that although increasing the final rise excursion does contribute to the perceived questioning, a decrease of the final rise excursion does not diminish it as long as the offset stays the same. These results suggest that the absolute value of the final rise offset constitutes the primary cue to questions in German.

The finding that raising of the rise onset and the prenuclear peak can contribute to the perception of ‘questioning’ was unexpected since both effects did not occur in the previously conducted production study [1]. On the other hand, a perception study by Petrone and Niebuhr [27] showed significant effects in the prenuclear region that could not be confirmed in a production as well [1, 28]. This shows the potential dangers of using perception experiments without additional production experiments since parameters can be relevant for perception without necessarily being used in production. This is consistent with Chen’s [3] findings that native speakers of Chinese and Hungarian were able to identify Dutch cues to interrogativity that are not used in either one of the two languages. Furthermore, these results are compatible with the concept of the frequency code as a typological constant [29] suggesting that high pitch is an innate cue to questions and can be interpreted as such even in forms that are not actually used in a language like the prenuclear rise excursion in German.

## 6. References


