Speech Prosody 2022
23-26 May 2022, Lisbon, Portugal

# Boundary tones in German *wh*-questions and *wh*-exclamatives – a cluster-based approach

*Heiko Seeliger*[1], *Constantijn Kaland*[2]

[1]Department of German Language and Literature I
[2]Institute of Linguistics
University of Cologne, Germany

heiko.seeliger@uni-koeln.de, ckaland@uni-koeln.de

## Abstract

We present data from a production study investigating verb-final German *wh*-questions and *wh*-exclamatives. We focus here on the variety of boundary tones found in the data. While most exclamatives were produced with falling contours and most questions with rising ones, there was substantial overlap between the two speech acts. Many utterance-final pitch offsets and pitch movements were also ambiguous in height and direction, providing a challenge for manual annotation. We apply a contour-based cluster analysis to the data in order to semi-automatically group intonation contours. We show that a cluster analysis based on the entire contour can shed light on the variety of boundary tones, although some clusters are 'mixed' and contain both high and low boundary tones. We then supplement this analysis with another cluster analysis based on the final two syllables of the utterance, which succeeds in splitting the mixed clusters. Finally, we use the combined cluster analysis to identify and describe the ambiguous boundary tones and argue that their form-to-function mapping falls outside the scope of the canonical inventory of GToBI boundary tones. Specifically, we found late falls after high plateaus and level, medium-high plateaus that are distinct both from continuation rises and calling contours.

**Index Terms**: prosody, F0, cluster analysis, boundary tones, questions, exclamatives

## 1. Introduction

The right edges of intonation phrases have attracted a considerable amount of attention and research. The encoding of intonational meaning appears to be especially rich in this domain of boundary tones and phrase accents. Regarding the inventory of boundary tones found in German, there is some disagreement in the literature. GToBI [1] assumes five boundary tones for intonation phrase boundaries: L-% for low boundaries, H-% for (medium-)high boundaries, H-^H% for high plateaus followed by rises to the upper end of the speaker range (or a direct rise to the upper end, in case there is not enough segmental material to realize a plateau), L-H% for low rises to the middle of the speaker range, and !H-% for downstepped high boundaries. H-L% as the label for a fall to low after a high plateau is explicitly excluded for Standard German by [1, p. 69], but is said to occur in some German dialects.

Turning to the functions and domains of occurrence of these boundary tones, L-% represents the unmarked low boundary tone of assertions and neutral *wh*-questions. H-% is officially only described for continuation rises, i.e. turn-internal high boundaries in e.g. lists. The default boundary tone of polar questions in GToBI is H-^H%. L-H% is glossed as 'polite offer' or 'indignation' in [1], depending on the preceding nuclear accent. Finally, !H-% is described only as the second tone of the calling contour.

ToDI / ToGI [2] accounts for all of the boundary tones that GToBI accounts for, but additionally allows for the lack of a boundary tone, transcribed as 0%. Together with three different nuclear accents (H*+L, H*, and L*+H), this results in a total of three so-called stylized contours: stylized falls, stylized high rises, and stylized low rises. GToBI only allows stylized high rises, which correspond to continuation rises in enumerations. The calling contour is also analyzed in terms of a downstepped !H accent, which, unlike in GToBI, is not part of the boundary tone. Concerning the functions of the boundary tones, there is substantial agreement between the two models.

Recently, some of the canonical GToBI boundary tones have been connected to functions other than the ones described in [1]: [3] use H-% to describe comparatively low final rises of rhetorical questions, while [4, p. 26] suggest that H-% is associated with biased questions (of which rhetorical questions are one sub-type). [5] describes a plateau contour that seems to be intermediate in terms of height between the continuation rise (H-%) and the calling contour (!H-%). It is used to signal reluctance on the part of the speaker. [5] points out that this contour cannot be modeled in GToBI, as !H-% is already used for the calling contour, and suggests a holistic modeling of plateau contours and/or stylized contours as a separate intonational mode.

These recent studies point towards an important issue: it is very likely that the description of the inventory of German boundary tones and their functions is incomplete. The foundation of the description in both models of German intonation was based on the most common speech acts: assertions, polar questions and *wh*-questions. It is possible that other speech acts are associated with other boundary tones, and it is also possible that minor functions go unnoticed even in the common speech acts. In this article, we focus on two speech acts that are arguably under-researched: *wh*-exclamatives and embedded *wh*-questions. We use a contour-based cluster analysis, using the procedure described in [6], to semi-automatically group the contours that were produced, and find further support for additional functions of H-%, as well as evidence for the existence of H-L% in German.

## 2. Data

### 2.1. Study design

The data set that we applied clustering to comes from a production study investigating German *wh*-questions and *wh*-exclamatives – specifically the prosodic marking of givenness, new information and contrastive focus. The study serves as a

 10.21437/SpeechProsody.2022-6

follow-up study to [7], which only investigated the given vs. new contrast in these two speech acts. The study was split into two sub-experiments. We focus on the second sub-experiment exclusively here, which manipulated the information-structural status of the object of transitive *wh*-clauses.[1] For reasons of space, we focus only on the difference between the two speech acts in this article.

The two *wh*-structures under investigation were transitive verb-final clauses, see (1) for an example. There were 8 such items. While verb-final word order is possible in German *wh*-exclamatives even in root clauses (1a), it is only possible in (non-echo) *wh*-questions if they are embedded (1b). The embedding clauses in the experimental material were polar questions. Crucially, German polar questions tend to predominantly end in high boundary tones [9]. It is thus possible that the *wh*-questions elicited in this experiment exhibit a greater than normal variety in terms of boundary tones, since the matrix question and the embedded question may enable both high and low boundary tones. Note that the contour of the matrix question was *not* part of the analysis.

(1) a. Wo die schon überall
*where she*-DEM *already everywhere*
Germanen erforscht hat!
*Germanic.tribes researched has*
'The places where she has already researched Germanic tribes!'

b. Weißt du zufällig, [= matrix polar question]
wo die schon überall Germanen erforscht hat?
'Do you happen to know where she has researched Germanic tribes already?'

### 2.2. Procedure and participants

The target sentences were embedded in a dialog between two speakers. The part of the first speaker was presented visually and via a prepared recording. The part of the second speaker was presented only visually, which participants then read out loud. All participants saw all items in all conditions.

The experiment had 18 participants (9 male, 9 female). They were native speakers of German who were recruited from the student population of the Humboldt University of Berlin and were paid for their participation. A total of 43 of 1152 recordings (3.7%) were excluded from the analysis because of disfluencies, technical issues or deviations from the lexical material, leaving 1109 contours for analysis.

### 2.3. Data normalization

Since speakers differ in terms of their F0 range, it is necessary to perform a per-speaker normalization of F0 values before a cluster analysis can be applied. We chose the Octave Median scale (OMe), proposed by [10]. OMe is defined as $log_2(\frac{\text{Hz}}{\text{median}_{\text{speaker}}})$. After normalization, pitch values mostly lie in the interval spanning one octave around the speaker's median pitch, i.e. most values range between $-0.5$ and $0.5$. The median was calculated from all available pitch values per speaker. Crucial for our purposes is the fact that OMe allows direct estimation of a speaker's range: there are very few data points below the $-0.5$ threshold and comparatively few above the 1 threshold, suggesting that these two points correspond to the extremes of a speaker's range.

[1] The other sub-experiment manipulated the information-structural status of the subject. See [8] for an analysis of both sub-experiments.

### 2.4. Data annotation

For the statistical analysis reported in [8], the data were annotated manually by phonetically-trained annotators using Praat [11] according to the DIMA guidelines [12]. There was one deviation from DIMA: The annotation of boundary tones was binary, such that utterance-final pitch movements were labeled as either falling or rising. During annotation, it was noticed that the data contain many ambiguous contours, in particular contours that end in mid-high plateaus in which the perceptual difference between a rise and a fall was very small. This data set thus represents a potential challenge both for manual annotation and for cluster analysis.

The annotators manually corrected pitch tracking errors, allowing the pitch objects generated by Praat to serve as the direct input for the cluster analysis. Overall, the annotators labeled 498 exclamatives as falling and 54 as rising (90.2% vs. 9.8%). For questions, they labeled 481 contours as rising and 76 as falling (86.3% vs. 13.7%). The distinction between rising and falling contours thus seems to coincide with the two speech acts to a large extent, but not categorically so.

### 2.5. Cluster analysis

For the cluster analysis, we used hierarchical clustering, since this variety of cluster analysis does not require an *a priori* specification of the number of clusters and is thus ideal for exploratory, visual data analysis. The distance matrix that served as the input for the cluster analysis was calculated using Euclidean distances. The cluster analysis itself used the complete linkage criterion. The analysis was done in R [13], using the app presented in [6] (available from "https://constantijnkaland.github.io/contourclustering/"). The app uses `stats::hclust`. The pitch object produced by Praat had been manually cleaned of tracking errors during annotation, so no further tracking errors were removed using the app's subsetting feature. Prior to clustering, we smoothed the contours using Praat. Since the raw contours are of different lengths, we linearly interpolated the requisite number of points (see below) using the R function `stats::approx`.

Two different time domains were used in two separate cluster analyses: the whole utterance, with 20 evenly spaced sampling points per contour, and the final two syllables of the utterance, with 10 evenly spaced sampling points per interval. We will refer to the resulting clusters as *total* clusters and *late* clusters, respectively. We chose the final two syllables of the utterance as the late domain because, by GToBI assumptions, the boundary tone of an intonational phrase is realized on the final syllable of the phrase. Visual inspection indicated, however, that plateau contours in particular showed earlier differences. Alternatively, we could have chosen the post-nuclear stretch of each contour as the sampling domain for the late clusters, but since the data contains many early nuclear accents as well as late nuclear accents, this would lead to wildly varying lengths of the sampling domain, which would worsen the distortion of the time domain that is unavoidable with cluster analysis. We instead chose to keep the length of the sampling domain more comparable between contours.

### 2.6. Aims of the analysis

The number of clusters was determined in an exploratory, bottom-up fashion: using the app presented in [6], we increased the number of clusters step-by-step until a new cluster was deemed visually too similar to the cluster that it split from. We then settled on an analysis using the previous number of clusters. We eventually settled on 6 clusters for the total clusters and 5 clusters for the late clusters. This analysis is exploratory – we investigate on the methodological side which domain of analysis works best for semi-automatically grouping boundary tones: the entire utterance, a late portion of the utterance, i.e. in the case of the present study the final two syllables of the utterance, or a combination of the two. Once a satisfactory clustering is achieved, we then see to which extent the contour types found in the data can be mapped onto the canonical GToBI inventory of boundary tones and nuclear contours.

# 3. Analysis

## 3.1. Results

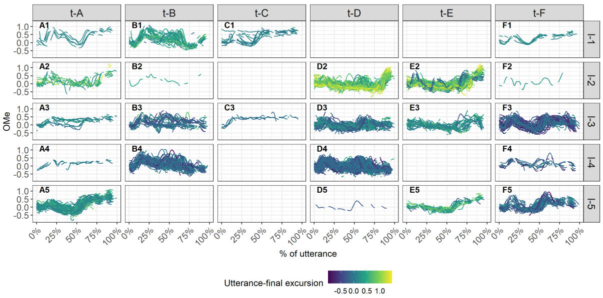


Figure 1: *Comparison of clusters based on the total contour (columns; labeled 't' for 'total' and letters A–F) to clusters based on the last two syllables of the utterance (rows; labeled 'l' for 'last' and numbers 1–5). Utterance-final excursion is calculated as the difference between pitch offset of the final syllable and pitch onset of the penultimate syllable.*

Fig. 1 gives an overview of contours split up by the two clustering domains – the entire utterance (ordered along the top left-to-right) vs. only the final two syllables (ordered along the right top-to-bottom). The total clusters have been labeled using letters, while the late clusters have been labeled using numbers. Sub-clusters are labeled with a combination of these labels in Fig. 1. As a starting point, note that Clusters B, D and F are mixed in the sense that they all contain substantial amounts of both falling and rising contours, while Clusters A, C and E consist mostly of rising contours. Table 1 shows how the speech acts and annotated falls and rises are distributed across clusters. Also note that the late clusters manage to separate the mixed clusters reasonably well into falling and rising contours, which can be seen most clearly for Clusters B1, D2, and E2. Table 2 shows the distribution of the particularly mixed clusters in the combined cluster analysis.

Even after the combined cluster analysis, two clusters remain mixed in terms of annotated falls and rises, but also visually in terms of utterance-final pitch height and movement: D3 (23% annotated rises) and F5 (46% annotated rises). The contours in F5 contain clear, medium-high plateaus after prominent accents on the object, while the contours in D3 do not display clear plateaus or clear rises, but pitch is overall slightly higher than expected in the post-object stretch of most utterances.

Turning to the question of how the clusters map to the inventory of boundary tones, there are some clear correspondences between the clusters and the standard GToBI inventory of nuclear contours: Cluster B1 corresponds to L+H* L-H%;

Table 1: *Distribution of the two speech acts and of annotated boundary heights across clusters. Percentages in brackets give the proportion of annotated rises within each combination of speech act and cluster.*

| Cluster | Exclamatives | Questions | Falls | Rises |
|---|---|---|---|---|
| A | 5 (100%) | 132 (98%) | 3 | 134 |
| B | 131 (17%) | 21 (100%) | 109 | 43 |
| C | 1 (0%) | 15 (100%) | 1 | 15 |
| D | 302 (4%) | 185 (83%) | 321 | 166 |
| E | 11 (27%) | 152 (89%) | 24 | 139 |
| F | 102 (11%) | 52 (52%) | 116 | 38 |

Table 2: *Distribution of the two speech acts and of annotated boundary heights in the mixed clusters*

| Cluster | Exclamatives | Questions | Falls | Rises |
|---|---|---|---|---|
| B1 | 14 | 16 | 1 | 29 |
| B2 | 0 | 1 | 0 | 1 |
| B3 | 31 | 3 | 26 | 8 |
| B4 | 86 | 1 | 82 | 5 |
| D2 | 5 | 132 | 4 | 133 |
| D3 | 75 | 34 | 84 | 25 |
| D4 | 221 | 19 | 232 | 8 |
| D5 | 1 | 0 | 1 | 0 |
| F1 | 0 | 6 | 0 | 6 |
| F2 | 1 | 0 | 1 | 0 |
| F3 | 67 | 25 | 81 | 11 |
| F4 | 6 | 3 | 9 | 0 |
| F5 | 28 | 18 | 25 | 21 |

E2 mostly consists of L* H-^H% or L+H* H-^H%, D2 of L* H-^H% with compressed rises (i.e. the rise does not go through a plateau because of a lack of segmental material). There are, however, also some clusters that contain contours that do not seem to have clear correspondences to the standard GToBI inventory of nuclear contours. In the following, we focus on two of these: F3 and F5.

Fig. 2 gives a closeup view of these two clusters, split up by speech act. The commonality between the two clusters is that all contours feature an upward F0 jump in the middle of the utterance. The difference between the two is that many instances of F3 feature late and steep falls, while the contours in F5 are level after the jump. Within each cluster, there are also slight differences between the speech acts. In both clusters, the exclamatives contain many more early prominences (on the d-pronoun). More relevant for the issue of boundary tones, the questions in F3 end with steeper falls than the exclamatives. In F5, the rise in the middle of the utterance reaches a higher point in exclamatives than in questions, and is followed by a slight dip to roughly the middle of the speaker's range.

## 3.2. Auditory description of the contours

The auditory impression of these contours is as follows. The steep falls in the questions in F3 are entirely located within the final syllable and perceptually seem to end at a medium-low level, even if pitch lowers all the way to the level of L*(+H) accents. Presumably this comparatively high perceived pitch is a result of the short duration of the falling movements. The boundary tones sound turn-ending, but their intonational mean-

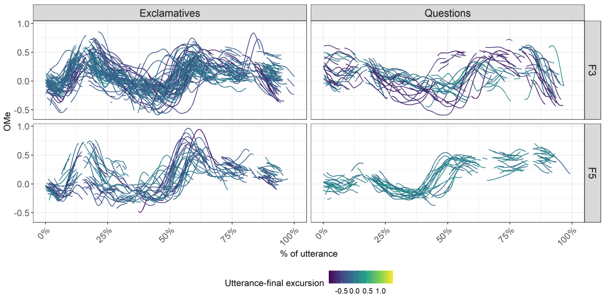


Figure 2: *Clusters F3 and F5 split up by speech act*

ing is otherwise unclear. It is notable that the contours all start with pitch above the speaker's overall median values, whereas most other contours start around the median. It is possible that this indicates the absence of a phrase boundary between the matrix question and the embedded question. As for the exclamatives in F3, they generally sound less steeply falling than the questions, although most of them were indeed perceived as falls (cf. Table 2). The final pitch level sounds medium-high. The speakers usually sound admiring or amazed.

In F5, the questions sound high-level. The absence of an upstepped ˆH% tone sounds slightly hesitant, as if speakers are unsure if they are questioning the right proposition. In the exclamatives, the speakers sound like they are 'trailing off', which intuitively sends a turn-yielding signal (i.e. the contours in F5 do not sound like continuation rises). Crucially, this contour does not sound like the calling contour, either.

### 3.3. Phonological analysis

The final fall in the contours of F3 is too far removed from the upward pitch movement to be analyzed as L+H* L-% or L*+H L-%, the only two possibilities that GToBI offers for utterance-final rise-falls (compare cluster E3, in which it *is* possible to analyze the nuclear contour as L+H* L-%). Instead, there appears to be tonal spreading between two high targets, one immediately after the nuclear accent and one in the final syllable. Alternatively, the late falls in F3 might be analyzed as an H+L* accent on the final syllable, i.e. the auxiliary verb. However, no accents were annotated on any auxiliary verb in the entire experiment, suggesting that it was not perceived as a nuclear accent, and we are indeed dealing with a boundary tone. [14] analyze plateaus in the Eastern European Question Tune (which is generally L* H-L%, with varying alignment of the H- phrase accent) by way of copying the H-% phrase accent to a position immediately following the nuclear accent. We propose that the boundary tone in cluster F3 is indeed best analyzed as H-L%. Whether this is a dialectal feature, e.g. of Berlin German, is left for future research.

Turning to the plateau contours in F5, the slight fall in exclamatives sounds to small for !H-%, while in the questions the pitch is too level to be analyzed as H-ˆH%. The contours also do not correspond, in either speech act, to the already described functions of similar contours (calling contours and continuation rises). For the questions, H-% appears to be the most appropriate description, with the caveat that the nuclear contours do not sound or function like the continuation rise, i.e. we seem to be dealing with yet another function of H-%. For the exclamatives, whose contours feature a slight drop after the utterance-medial pitch rise, H-% is harder to motivate because there are overall three tonal targets in the relevant part of the contour. [2] derives level plateaus from the absence of boundary tones, transcribed as 0%. 0% occurs in lists, but also when an utterance is presented as a routine that involves repetitive or uniform activities (cf. [15] on stylized contours). Crucially, [2] says that the contrast between low boundary tones and unspecified boundary tones is often marked by raising the final low and level stretch of a 0% boundary tone to a higher level – a description that fits the exclamative contours.

A slightly different analysis could be based on [16], who describes a contour found in Berlin German that is visually identical to those in cluster F5, but whose domain of occurrence and usage is different from the *wh*-structures under consideration here. That contour occurs in monological, narrative sequences of routine events. [16] proposes to analyze this contour as L+H* . . . 0%, where the dots indicate that no declination takes place after the nuclear accent. A variant in which there *is* declination, but only to a medium-high level, is analyzed as L+H* 0%. Both transcriptions could be applied to the contours in F5, in which case both the nuclear accent and the boundary tone would be identical between the speech acts, and they would only differ in the presence of post-nuclear declination.

Both [2] and [16] connect 0% boundary tones to routine in some fashion. At first sight, this is unlikely to play a role for the exclamatives, since exclamatives express surprise at noteworthy events. However, since most of the exclamatives in this experiment expressed surprise at the large amount of places in which the subject of the exclamatives has done things, they might connect to routine, after all: someone who e.g. researches Germanic tribes in many places crucially has a routine habit of doing research. We can therefore tentatively give this semantic motivation for the occurrence of this boundary tone in the data.

## 4. Conclusion

Our study shows that cluster analysis can be used to group intonational contours, and semi-automatically identify e.g. various shapes of plateau contours. Note that boundary tones may be particularly well-suited for cluster analysis, since boundary tone differences usually span the largest F0 range between contours, and bottom-up complete linkage manages to separate large differences with comparatively few clusters. On the other hand, differences unrelated to the boundary tones might not impact the clustering as much when there are large boundary tone differences in the data.

For the specific contours discussed here, we suggest that the medium-high plateaus can be analyzed as H-%, but that they represent yet another function of this boundary tone. It is plausible that [3, 4] are on the right track and that this boundary tone is associated with bias. The slightly falling plateaus appear to be connected to routine, and thus to 0%. They constitute evidence for functions of stylized contours in German beyond the calling contour. For the late falls after high plateaus, we suggest that they are best modeled as H-L%, i.e. they require an expansion of the inventory of German boundary tones. However, note that H-L% was overall quite rare in the data, and that further research – i.e. a perception study – is needed to show that it is indeed perceptually distinct from L+H* L-%.

## 5. Acknowledgements

Funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) – Project-ID 281511265 – SFB 1252.

# 6. References


[1] M. Grice, S. Baumann, and R. Benzmüller, "German intonation in autosegmental-metrical phonology," in *Prosodic Typology: The Phonology of Intonation and Phrasing*, S.-A. Jun, Ed. Oxford: Oxford University Press, 2005, pp. 55–83.

[2] J. Peters, "Phonological and semantic aspects of German intonation," *Linguistik Online*, vol. 88, no. 1, pp. 85–107, 2018.

[3] B. Braun, N. Dehé, J. Neitsch, D. Wochner, and K. Zahner, "The prosody of rhetorical and information-seeking questions in German," *Language and Speech*, vol. 62, no. 4, pp. 779–807, 2019.

[4] A. Arnhold, B. Braun, and M. Romero, "Aren't prosody and syntax marking bias in questions?" *Language and Speech*, vol. 64, no. 1, pp. 141–180, 2021.

[5] O. Niebuhr, "Resistance is futile – the intonation between continuation rise and calling contour in German," in *Proceedings of INTERSPEECH 2013*, 2013, pp. 225–229.

[6] C. Kaland, "Contour clustering: A field-data-driven approach for documenting and analysing prototypical f0 contours," *Journal of the International Phonetic Association*, pp. 1–30, 2021.

[7] S. Repp, "The prosody of wh-exclamatives and wh-questions in German: Speech act differences, information structure, and sex of speaker," *Language and Speech*, vol. 63, no. 2, pp. 306–361, 2020, pMID: 31096841. [Online]. Available: https://doi.org/10.1177/0023830919846147

[8] H. Seeliger and S. Repp, "Contrast and givenness in German wh-exclamatives and wh-questions," in preparation.

[9] J. Michalsky, *Frageintonation im Deutschen. Zur intonatorischen Markierung von Interrogativität und Fragehaltigkeit*. Berlin, New York: de Gruyter, 2017.

[10] C. De Looze and D. Hirst, "The OMe (Octave-Median) scale: a natural scale for speech melody," in *Proceedings of Speech Prosody 7*, N. Campbell, D. Gibbon, and D. Hirst, Eds., Dublin, 2014, pp. 910–914.

[11] P. Boersma and D. Weenink. (2021) Praat: doing phonetics by computer. [Online]. Available: http://www.praat.org

[12] F. Kügler, B. Smolibocki, D. Arnold, S. Baumann, B. Braun, M. Grice, S. Jannedy, J. Michalsky, O. Niebuhr, J. Peters, S. Ritter, C. Röhr, A. Schweitzer, K. Schweitzer, and P. Wagner, "DIMA – annotation guidelines for German intonation," in *Proceedings of 18th ICPhS*. Glasgow: University of Glasgow, 2015, pp. Paper number 317, pp. 1–5.

[13] R Core Team, *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna, Austria, 2021. [Online]. Available: https://www.R-project.org/

[14] M. Grice, D. R. Ladd, and A. Arvaniti, "On the place of phrase accents in intonational phonology," *Phonology*, vol. 17, no. 2, pp. 143–185, 2000.

[15] D. R. Ladd, "Stylized intonation," *Language*, vol. 54, pp. 517–539, 1978.

[16] M. Selting, "Berlinische Intonationskonturen: 'Die Treppe aufwärts'," *Zeitschrift für Sprachwissenschaft*, vol. 20, no. 1, pp. 66–116, 2001.