

Integrating the discreteness and continuity of intonational categories

Martine Grice, Simon Ritter, Henrik Niemann, Timo B. Roettger

Abstract

It has already been observed that there is no one-to-one mapping between intonational categories and the pragmatic functions they are used to express. For instance, in German a particular pitch accent (L+H*) is often used to express contrastive (corrective) focus, but neither is the use of this pitch accent confined to this function nor is this the only pitch accent used to express it. In particular, there are considerable *differences* across speakers in the use of pitch accents and the functions they express. In this paper we look at the phonetic parameters that are characteristic of each of these pitch accents (f0 peak alignment, tonal onglide and target height) and observe a striking *similarity* across speakers: All speakers modulate each parameter in the same direction, e.g. the f0 peak is aligned later for contrastive focus than for narrow focus. Whereas for some speakers this is transcribed as two different pitch accents (L+H* vs. H*), for others it is not, and the peak alignment is treated as phonetic variation within one accent type (H*). To capture both the differences and similarities in intonation, we therefore argue for an integrated analysis of the discrete phonological pitch accents and the modulation of continuous phonetic parameters that characterise them.

Key words

Pitch accent; intonation; prosody; tonal alignment; category.

1 Introduction

In Autosegmental-Metrical phonology (Ladd 2008, Pierrehumbert 1980), intonation contours are decomposed into sequences of postlexical tones associated with prosodic and metrical structures. These tones can be culminative or delimitative in nature, corresponding to descriptive classes such as pitch accents and boundary tones, respectively (e.g. Beckman and Venditti 2011). Here we are concerned with pitch accents, tones that co-occur – either singly or in combination – with metrically strong positions in the phrase, and that commonly lend prominence to the words they are on. In languages such as German and English, the inventory of pitch accents ranges from four to six, depending on the model (e.g. compare for German: Grice et al. 2005, Peters 2014, and Kügler et al. 2015; and for English: Grabe et al. 2001, Beckman et al. 2005, and Dillely and Heffner 2013). It is true for all of these models, however, that whilst establishing the *presence or absence* of a

pitch accent is relatively straightforward, distinguishing between different types of pitch accent is not.

In the above intonation models, what is taken to be a pitch accent category is characterised by a bundle of phonetic properties (Ladd 2008) that serve a specific communicative function in a given context. To establish an inventory of pitch accents thus involves the analysis of a number of continuous phenomena into discrete categories, which are taken to be phonological representations. Transcribers are trained to pay attention to those phonetic properties that are deemed to be the main characteristics of each accent. This is also the case for the German GToBI transcription system used in this study (Grice et al. 2005, see also Grice and Baumann 2002, 2016). Thus, transcribers attend to gradual dimensions of the speech signal, and are trained to draw a line between pitch accent categories based on multiple gradual acoustic and auditory dimensions.

These dimensions typically involve *f0 peak alignment*, which is taken to be a major cue to pitch accent type in a number of languages (see Arvaniti 2011, D'Imperio 2011, and Prieto 2011, for extensive overviews). Early work on English by Pierrehumbert and Steele (1989) showed that a distinction could be made between L+H* and L*+H, simply by adjusting the timing of the *f0* peak (in a rising-falling-rising contour from medial to late) in relation to the segments. Dilley and Heffner (2013) showed that a distinction between H+L* and H* could also be made by adjusting the position of the *f0* peak (from early to medial). Such a distinction between medial and early peaks was also found in German by Kohler (1987). Subsequent work has shown that the alignment of an *f0* peak can interact with the shape of the *f0* curve (Knight 2008, Barnes et al. 2012), and can be subject to considerable speaker specific variation (Niebuhr et al. 2011). Moreover, it can be strongly affected by the segmental makeup of the words (Barnes et al. 2012).

Target height, i.e. the height of the *f0* target corresponding to the starred tone, is also a dimension that plays a role in pitch accent categorisation, although to a lesser extent than alignment (Ladd and Morton 1997 for English, Kügler and Gollrad 2015 for German). Target height is of particular importance when comparing H* to L+H*. The perceived height of an *f0* target can also be affected by alignment, such that peak delay has been treated as a substitute parameter for a raised peak (Gussenhoven 2004).

A further dimension that has been shown to play a role in the categorisation of pitch accents is the *tonal onglide* (Ritter and Grice 2015). This is the *f0* movement leading towards the *f0* target on the accented syllable and thus takes into account not only the *f0* on the accented syllable but also on the syllable before the accented syllable. This part of the *f0* contour is generally accounted for phonologically in terms of a leading tone, such as the L in L+H*, or the first H in H+!H*.

In German, as in English, different pitch accents typically express information structure (the division of sentences into focus and background) and information status (the degree of givenness of a referent in the current

discourse): A pitch accent with a late (and high) f_0 peak and a rising onglide (L+H*) generally signals contrastive focus and one with a medial peak and shallow rising onglide (H*) new information, while a pitch accent with an early f_0 peak and a falling onglide (H+!H*) tends to signal broad focus, or given information (cf. Kohler 1991, 2005, Grice et al. 2005, Baumann et al. 2007, Féry and Kügler 2008, Ritter and Grice 2015). A similar form-function mapping has been argued for English too (Pierrehumbert and Hirschberg 1990). However, studies looking at speaker specific behaviour in detail have found that these statements do not hold for all speakers of a language, or even for one speaker all of the time (Peppé et al. 2000, Grice et al. 2009, Cangemi et al. 2015). As Cruttenden (1986) pointed out, there are usually alternatives to the more common, preferred intonation patterns, which at the time of writing his book were usually neglected. More recent work takes an approach that provides information as to the distribution of alternative realisations of a given function (Grabe 2004 and Yoon 2010 for English, Baumann 2006, Grice et al. 2009, and Baumann et al. 2015 for German, see also Savino and Grice 2011 and Cangemi and Grice 2016 for Italian).

This paper is concerned with a corpus of read speech produced by five speakers in German. The corpus was designed to investigate the distinction between different types of focus, in terms of both supralaryngeal articulation and intonation. We shall first discuss the corpus with regard to categorical annotations of pitch accent types. It has already been shown for this corpus that there is no consistent relation between pitch accent categories and pragmatic function at this level (Mücke and Grice 2014). In this paper we examine the relation between the phonetic parameters characterising pitch accent categories assigned by the transcribers, and how they are distributed across the different functions and speakers.

The main reason for exploring this particular corpus is that the productions of all five speakers have been subjected to perceptual testing in which it was shown that all five speakers are able to convey their pragmatic intentions to listeners in a comparable way (Krüger 2009; Cangemi et al. 2015). Taking all listeners together, the intended focus structure was correctly identified on average between 64% and 67% of cases. This indicates that the different speakers were comparable in their adequacy to carry out the task, thus motivating further exploration of this dataset. We discuss the perception results in detail in section 5.

In this exploration we do not *a priori* attempt to reject theoretically driven null-hypotheses, but rather aim to analyse the data post-hoc and to generate hypotheses for further study. In line with standards of reproducible research (Peng 2011), the obtained data table and the R scripts produced for data processing and plotting are made available on GitHub: <https://github.com/troettge/grice-et-al-2017-german-focus>.

The goal of this paper is to undergo a detailed analysis of the f_0 contour and how aspects of this contour (peak alignment, target height and tonal onglide) relate to the focus structures expressed and perceived. More specifically, we show that there is no one-to-one-mapping between the pitch accent type and

the expressed focus structure. Even though speakers differ in their choice of a pitch accent type for an intended focus type, the intended focus type is almost equally well perceived. In other words, listeners perceive the focus type not exclusively on the basis of the given accent type but are sensitive to more fine-grained details such as the peak alignment, target height and tonal onglide. These continuous measures show the same pattern across all speakers and can be linked to the perception of a focus type.

The paper is structured as follows: Section 2 presents the methods including the speech material and the analyses focusing on the quantitative evaluation of the pitch accent types produced by the speakers. While Section 3 examines the distribution of discrete pitch accent types as transcribed by the labellers, Section 4 examines the produced intonational patterns in terms of three continuous parameters (F0 peak alignment, tonal onglide, target height) and relates these to articulatory parameters. Section 5 takes a closer look at individual differences in the mapping between intended focus structure and perceptual recovery of these functions, showing that pitch accent type cannot solely account for the perceptual results. Section 6 explores the implications for intonation theory and calls for an integrated analysis of the discrete phonological pitch accents and the modulation of continuous phonetic parameters that characterise them.

2 Method

2.1 Speakers and recordings

In this paper we closely examine a dataset produced by five native speakers of Standard German (3 female, 2 male, mean age = 28, age range = 22-37). Speakers F1, F3, and M2 were from the Franconian area, and F2 and M1 from the Western Low German area. Acoustic and articulatory recordings were made at the phonetics laboratory at the University of Cologne, employing a Digital Audio Tape recorder (TASCAMDA-P1) and a condenser microphone (AKG C420 head set) at a sampling rate of 44.1 kHz (16 bit), along with an Articulograph (AG100). Mücke and Grice (2014) reported on the articulatory data, touching only briefly on the pitch accent analysis, whereas this paper focuses on the intonation, both in terms of pitch accent types as well as three continuous parameters contributing towards the categorical analysis (peak alignment, target height and tonal onglide) and how they relate to the focus structures expressed.

2.2 Speech material

The speech material was designed such that the three target words (nonce names) Bieber [bi:bə], Bahber [ba:bə] and Bohber [bo:bə] appeared in three different focus structures: either in broad, narrow or contrastive focus. All target words were disyllabic trochaic words. The different focus structures were elicited using question-answer pairs in which the answer was the target sentence. Examples for the context and elicited focus types are shown in Table 1.

Questions	Answers	test word /ba:bə/ in:
1. Was gibt's Neues? <i>What's new?</i>	Melanie will Doktor Bahber treffen. <small>focus</small>	broad focus
2. Wen will Melanie treffen? <i>Whom does Melanie want to meet?</i>	Melanie will Doktor Bahber treffen. <small>focus</small>	narrow focus
3. Will Melanie Dr. Werner treffen? <i>Does Melanie want to meet Dr. Werner?</i>	Melanie will Doktor Bahber treffen. <small>focus</small>	contrastive focus

Table 1: Speech material example, target word <Bahber> /'ba:bə/. Focus constituents are shaded.

Our dataset consisted of 310 tokens (3 target words x 3 focus structures x 7 repetitions x 5 speakers = 315 tokens; 5 tokens were excluded due to recording or processing issues). In contrast to the data in Mücke and Grice (2014) we excluded target words produced as part of the background (where the nuclear accent was at the beginning of the phrase, on Melanie), as none of them were produced with a pitch accent, there thus being no peak or target to measure.

2.3 Procedure

Subjects were presented with a question both auditorily and orthographically on a computer screen. Questions were presented via the computer's built-in loudspeakers and were spoken by a professional male speaker¹. Subsequently, subjects were presented with the answer orthographically and instructed to read it aloud in a contextually appropriate manner. Each recording block was preceded by a training block of five question-answer pairs. The question-answer pairs in the test materials were pseudo-randomised, making sure that no two consecutive target sentences contained the same test word.

2.4 Analysis

Two trained transcribers independently placed GToBI labels using their auditory impressions and the f0 contour in PRAAT (Boersma and Weenink 2016), using the GToBI training materials for orientation (Grice and Baumann 2002, Grice et al. 2005, website for GToBI: <http://www.gtobi.uni-koeln.de>). These training materials make transcribers aware of the specific characteristics of each pitch accent type, as shown in Table 2.

¹ The questions were produced with a consistent intonation contour for each condition. For broad focus there was a high accent (H*) on <Neues> followed by low boundary tones (L-%). For narrow and contrastive focus there was a low accent (L*) on both <Melanie> and <Werner>, followed by high boundary tones (H-^H%).

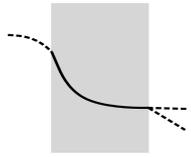
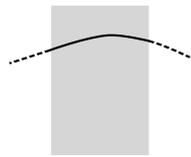
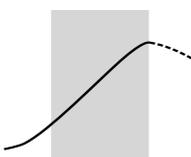
Accent	Schematised f0	Characteristics
H+!H*		<ul style="list-style-type: none"> ▪ Fall from previous syllable onto accented syllable ▪ Mid target on accented syllable ▪ Accented syllable sounds mid
H*		<ul style="list-style-type: none"> ▪ High accent with a shallow rise (sagging interpolation) ▪ Medial f0 peak ▪ Accented syllable sounds high (not as high as L+H*)
L+H*		<ul style="list-style-type: none"> ▪ Low tone before accented syllable and rise to accented syllable ▪ The extent of the rise is great ▪ Late f0 peak ▪ Accented syllable sounds high

Table 2: Main characteristics of the GToBI accents H+!H*, H* and L+H*, adapted from the GToBI training materials, <http://www.gtobi.uni-koeln.de>. Accented syllables are highlighted in grey.

In addition to the GToBI standard labels, individual targets for bitonal pitch accents were labelled separately (e.g. H+!H* has a separate label for each H). Transcribers agreed in 80% of the cases (or 84% of the cases if the focus type “background” was included). We tested the interrater reliability by calculating Cohen’s kappa (Cohen, 1960) for pitch accent categorization (H+!H*, H*, L+H*) using the R package irr (Gamer, Lemon, and Singh 2012), revealing substantial agreement between transcribers (0.712). Apart from cases of uncertainty, where one labeller did not place a label, disagreement involved H* being confused with other accents: either with L+H* (49% of disagreements) or with H+!H* (42% of disagreements). The cases where they did not agree were subject to further scrutiny, leading to a consensus transcription, which was used for further analysis.

For the quantitative evaluation of the contours, we calculated three phonetic parameters explicitly used for the categorisation of pitch accents (see Table 2): the alignment of the f0 peak, the tonal onglide (cf. Ritter and Grice 2015), and the height of the f0 target on the accented syllable. These are explained in detail below and schematically depicted in Figure 1.

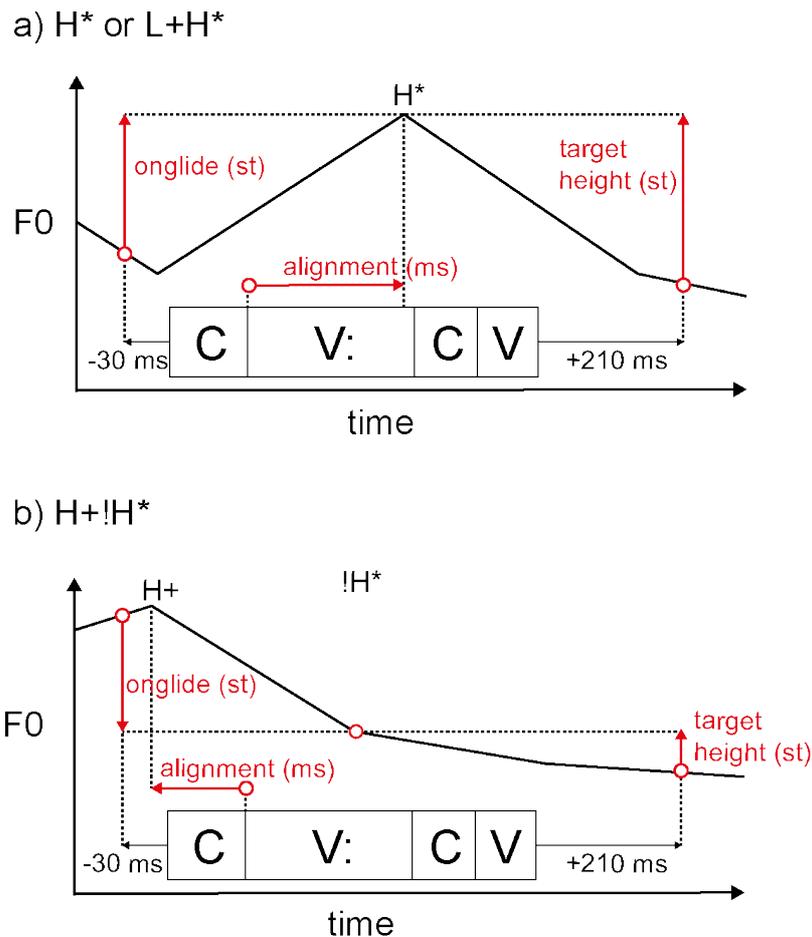


Figure 1: Schematic depiction of the three measurements involved for (a) rising pitch accents (H^* and $L+H^*$) and (b) falling pitch accents ($H+!H^*$).

F0 Peak Alignment

The peak alignment measure indicates the temporal alignment of the f_0 peak in relation to the onset of the nuclear accented vowel (in milliseconds). The f_0 peak can occur either within or after the accented syllable, as in the case of H^* and $L+H^*$ accents (Figure 1a), or before the accented syllable, as in the case of $H+!H^*$ (Figure 1b). Positive alignment values correspond to f_0 peaks occurring within or after the accented syllable and negative values correspond to f_0 peaks occurring before the accented syllable.

Tonal Onglide

The tonal onglide measure indicates whether the f_0 is rising or falling towards the accented syllable. It is calculated by relating the f_0 value in the nuclear accented syllable (at the location where the GToBI label for the starred tone was placed) to a reference point 30 ms before the start of the syllable (in semitones, cf. Ritter and Grice 2015). The starred tone label was placed on the f_0 peak in the case of rising pitch accents (Figure 1a). In the case of falling pitch accents, it was placed on a change of gradient in the contour or in the middle of the syllable if the f_0 trajectory had no inflection point (Figure 1b). A

positive tonal onglide value indicates a rise up to the peak while a negative value indicates a fall to the tonal target on the accented syllable.

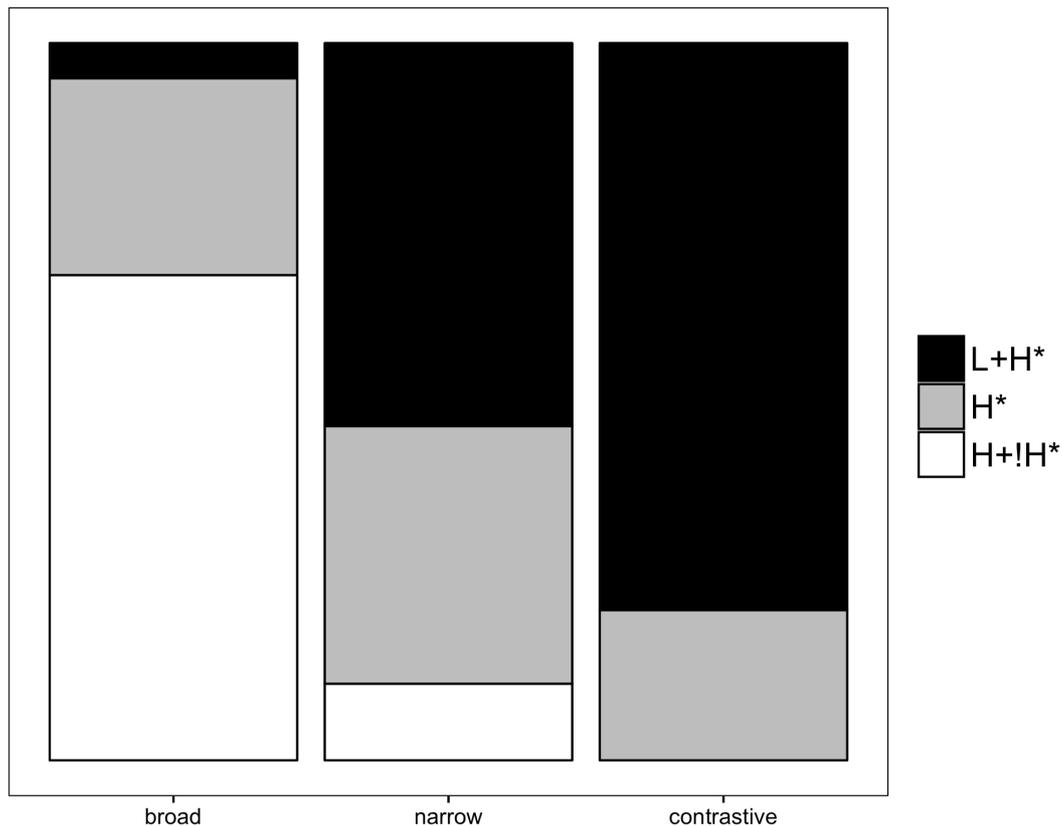
The reference point 30 ms before the accented syllable was chosen in order to obtain a point in the low central vowel of the last syllable of “Doktor” [ˈdɔktɔ], i.e. the syllable preceding the accented syllable of the target word. This point within the centralised vowel [ɐ] provided a consistent f₀ measurement and a good approximation to the position of the acoustic realisation of a potential leading tone, if there was one.

Target height

The target height measure is operationalised as the difference in semitones between the tonal target on the accented syllable, corresponding to the starred tone, and a following low reference point towards the end of the phrase, corresponding to a low boundary tone. In cases of a low boundary tone, the sentence-final low point is generally assumed to be stable for a speaker across different contexts (cf. Pierrehumbert 1980, and Liberman and Pierrehumbert 1984 for English, and Féry and Kügler 2008, Kügler and Féry 2016, and Grabe 1998 for German). Since transcribers, however, faced problems when determining the position at which to place the low boundary tone label, due to microprosody and creaky voice, we automatically labelled a constant reference point 210 milliseconds after the end of the target word. See Appendix 1 for information as to the constancy in F₀ values of this reference point for each speaker.

3 Results in terms of pitch accent types

Although the results reported by Mücke and Grice (2014) concentrated on supralaryngeal kinematics, they also looked at transcribed pitch accent placement and type. In all three focus conditions (broad, narrow, contrastive) there was always a pitch accent but no one-to-one mapping between the focus type and the transcribed accent type. There were, however, probabilistic preferences.



Focus Type

Figure 2: Distribution of pitch accent types according to focus structure, pooled over speakers.

Figure 2 shows the distributions for each focus condition pooled over speakers. Here the most common pitch accent type in the broad focus condition was the early peak pitch accent, H+!H* (66% of cases), exemplified in Figure 3a. This accent type has an “early peak” (H) before the target word, and a downstepped target (!H*) on the accented syllable. The most common pitch accent in the contrastive focus condition was L+H* (79%), exemplified in Figure 3c, with a peak late in the accented syllable.

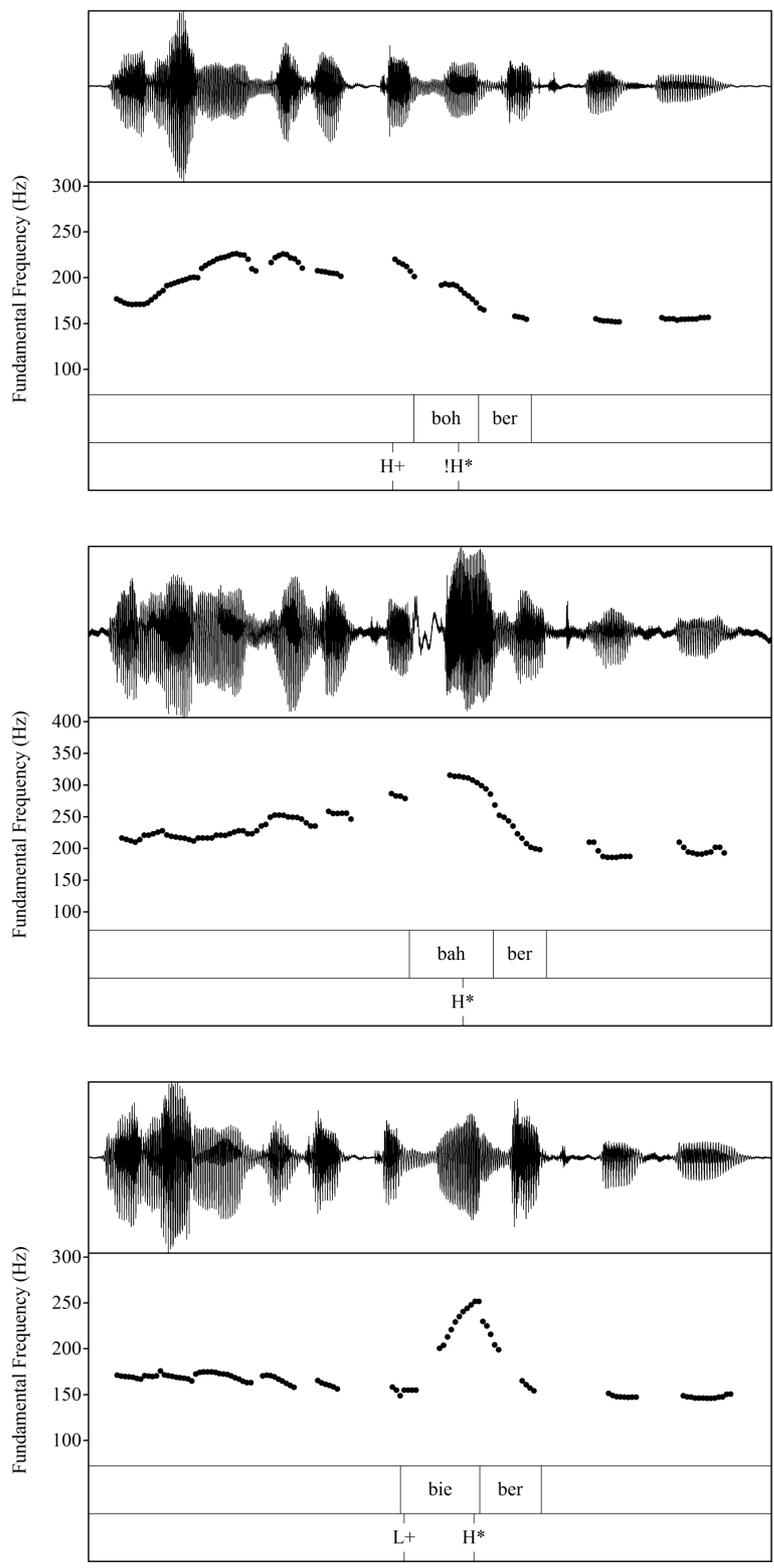


Figure 3: Representative waveforms and f0 trajectories for three pitch accent types (a) H+!H*, top panel; (b) H*, middle panel; (c) L+H* , bottom panel, all produced by speaker F2.

Although there is a predominance of H+!H* for broad focus and L+H* for contrastive focus, the percentages displayed in Figure 2 also reveal that both

of these focus conditions had alternative realisations. These were often H^* , see the example in Figure 3b (middle panel), although there were a few cases of $L+H^*$ in the broad focus condition. The narrow focus condition had a slightly more even distribution of accent types: 53% for $L+H^*$, 35% for H^* and 12% for $H+!H^*$. These results show that there was no condition with an exclusive pitch accent realisation, nor one particular pitch accent used exclusively in one of the conditions. This supports the observation made in Grice et al. (2009) that there is considerable variation within and across speakers with regard to accent choice when expressing pragmatic functions of this type. The only form-function generalisation we can make for this dataset is that $H+!H^*$ never occurs in the contrastive condition (see also Kügler and Gollrad 2015 for similar results).

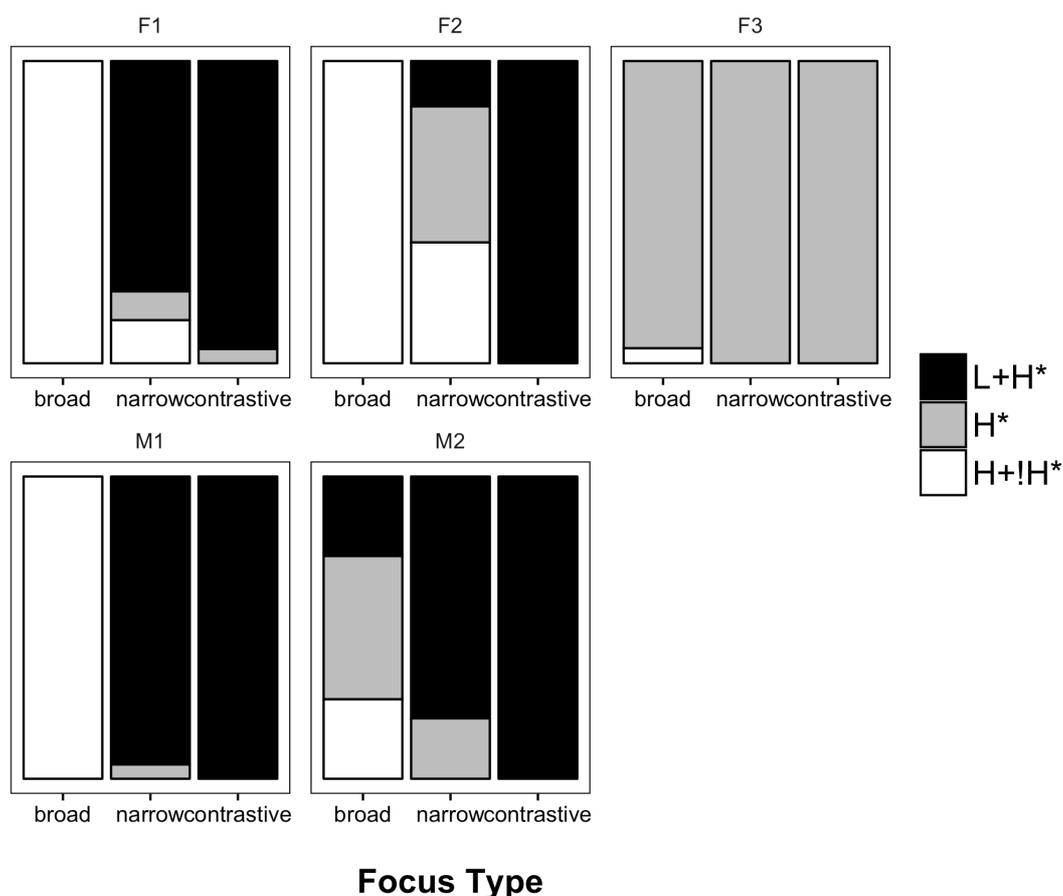


Figure 4: Distribution of pitch accent types across focus types for each speaker separately.

Looking at individual speakers' productions (see Figure 4), it becomes clear that not all speakers make use of each of the three pitch accent types, accounting for some of the variation in the distributions. One speaker (F3) almost exclusively uses only one category of pitch accent (H^*) across all three focus types while two speakers (F1, M1) hardly use this pitch accent at all.

All of the distributions relating pitch accent type to focus condition rely on a discrete decision made by the transcribers to categorise what they observe as one pitch accent type over another. In what follows, we take a step back from the discrete analysis into pitch accent categories and examine the global

contours and the continuous parameters that are taken to correspond to the characteristics of each pitch accent.

4 Results in terms of continuous parameters

4.1 Global contours

Looking at the superimposed f_0 contours in Figure 5, the three focus conditions show a great deal of speaker-specific variation. Some speakers make a clear distinction between at least two conditions. Take for example speaker M1: The global contours he produces for signalling broad focus (a hat pattern) are clearly distinct from those signalling narrow or contrastive focus. This difference corresponds to the distribution in nuclear pitch accents as analysed by the transcribers (shown in Figure 4), where for this speaker, broad focus was produced consistently with H+!H* and narrow and contrastive focus with L+H*.

Speakers F1 and F2 appear to distinguish between broad and contrastive focus (a hat pattern for the former, a low prenuclear stretch in the latter) but produce more variation in the contours for narrow focus than in the other conditions. On the other hand, speakers F3 and M2 use what appears to be the same global contour to mark all three focus conditions (broad, narrow and contrastive).

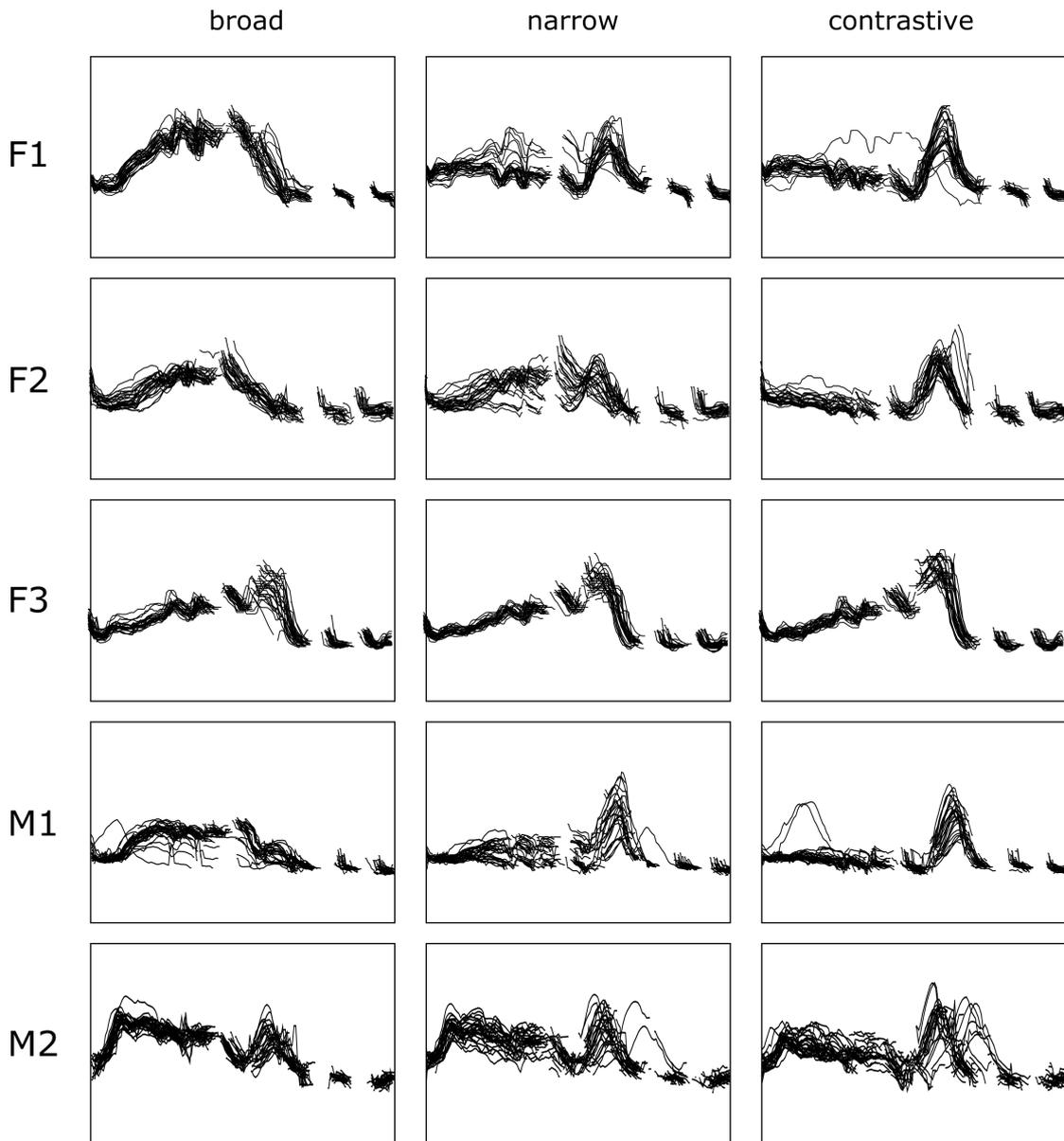


Figure 5: Time-normalised, superimposed f0 contours for each intended focus type (from left to right: broad focus, narrow focus and contrastive focus, displayed for each speaker separately).

In the next section we explore whether measuring selected parameters can reveal quantifiable differences that might be missed by the categorical pitch accent analysis and which can only be impressionistically observed in global contours.

4.2 Continuous f0 parameters

A closer look at the three parameters, f0 peak alignment, tonal onglide and the f0 target height on the accented word explores how individual speakers modify each parameter, and the relation of this modification to the transcribed categories. In the following we will descriptively explore the three continuous parameters using violin plots, which, based on kernel density estimations, show the probability density of data points at different values in a similar way

to a histogram. The estimated likelihood of a value is proportional to its “bulginess”. In addition, black dots and dashed lines indicate mean values of distributions for better comparison across focus types within and across speakers. Due to the multimodal nature of distributions, the mean values provided should, however, be treated with caution. Here, they merely serve the purpose of visually guiding the reader to notice differences in central tendencies.

F0 Peak alignment

Figure 6 displays the results for peak alignment, i.e. the alignment of the f0 peak relative to the vowel onset².

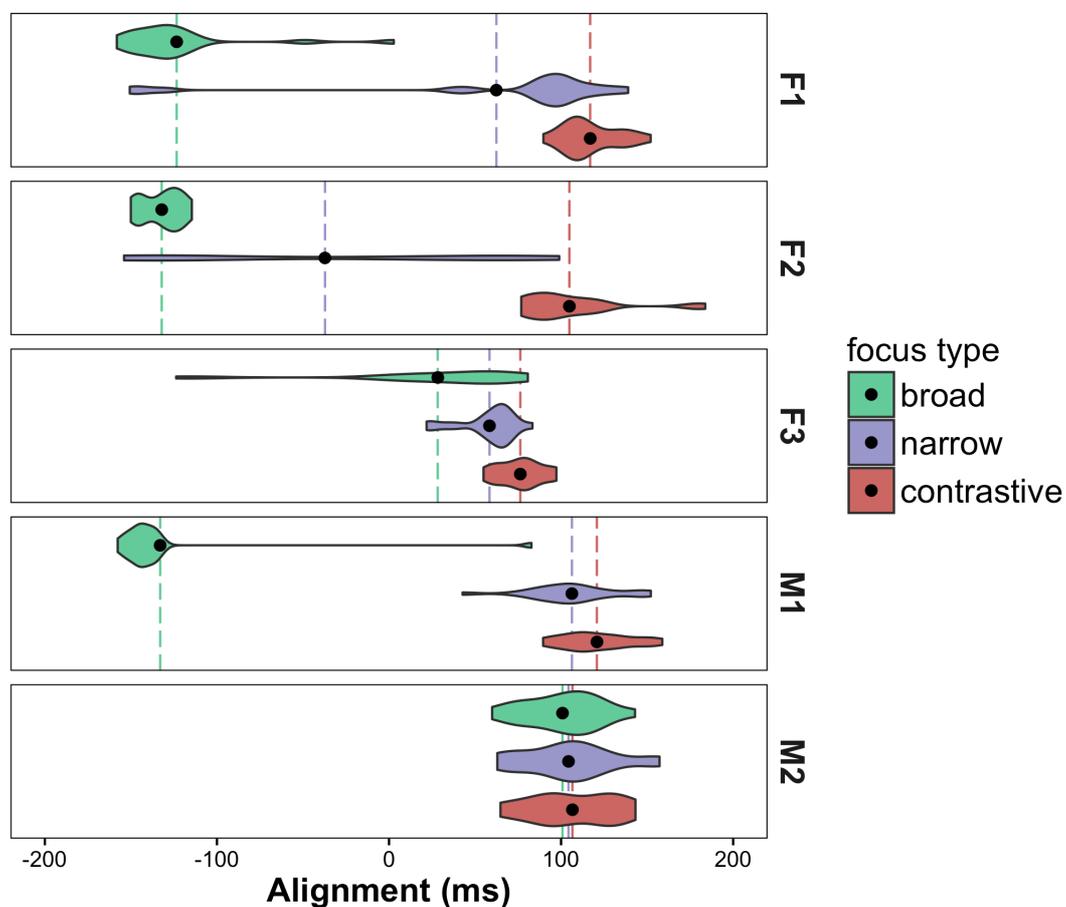


Figure 6: Violin plots of alignment data as a function of focus type (colour coded) and speaker. Mean values are indicated by black dots and coloured dashed lines for comparison across focus types.

We can observe that speakers F1, F2, and M1 make a clear distinction in the alignment of the peak, with broad focus almost exclusively having peaks well before the beginning of the accented syllable (negative values, corresponding to H+!H*), and narrow and contrastive focus almost exclusively having peaks late in this syllable (positive values corresponding to H* and L+H*). For these

² Alignment is presented here in absolute terms. Relative alignment is presented in Appendix 2.

speakers, values for narrow focus fall between these two but are highly variable, both within and across speakers, indicated by the large spread of the purple distribution. F1 and M1 do not appear to differentiate clearly between narrow and contrastive focus, but each speaker's distributions, albeit overlapping heavily, are different across the two conditions. Both speakers exhibit later peak alignment in contrastive focus compared to narrow focus. F1 and F2 clearly show multimodal distributions of alignment values in the narrow focus condition, with some values negative and others positive, reflecting the variation in the assigned pitch accent transcriptions shown in Figure 4. For both speakers pitch accents in the narrow focus condition consisted of both early peaks (H+!H*) and medial and/or late peaks (H*/L+H*).

For speakers F3 and M2, the difference in alignment is subtler (barely detectable for M2). Speaker F3's contrastive focus peaks are somewhat earlier than the average peak alignment values for the other speakers in this condition, contributing towards the transcription of H* rather than L+H*. Across the three conditions, there are only subtle differences in alignment, with broad focus showing the largest spread in distribution, with peaks aligned with either the accented vowel or before it. Speaker M2 shows similar distributions across focus types, however, the weight of data points slightly shifts, with narrow focus further to the right than broad focus, and contrastive focus slightly more to the right than narrow focus. Comparing the sequence of dashed lines indicating the means of the distributions, all speakers show the same relative mapping of peak alignment and focus types (green – purple – red).

Tonal onglide

Similar to peak alignment, the onglide results in Figure 7 show a clear differentiation between broad focus and the other two conditions for the same three speakers (F1, F2 and M1). The broad focus condition exhibits negative values reflecting the falling nature of the early peak accent. While F1 shows a large spread of distributions, F2 and M1 show a clearer separation at least for broad and contrastive focus. F2, again, shows a clearly separated distribution for narrow focus as compared to contrastive focus, while the distributions for F1 and M1 in these two conditions overlap heavily. Regardless of this overlap, the weighting of distributions nevertheless consistently tends towards larger onglide values for contrastive focus, resulting in overall higher mean values.

The accents of speakers F3 and M2 almost exclusively have rising onglides, but they exhibit a subtler differentiation within the rising accents that mirror the global trends: greater rising onglides in narrow focus than in broad focus and even greater rises for contrastive focus. M2 shows substantial overlap across the three focus conditions. However, distribution weights clearly follow the same global trend. Again the subtle differences in the productions of these two speakers mirror neatly the rather categorical patterns found for the other three speakers.

Comparing the sequence of dashed lines indicating the means of the distributions, it is clear that all speakers show the same relative mapping of onglide values and focus types. As with the alignment distributions, the green

line is more to the left than the purple line, which, in turn, is more to the left than the red line.

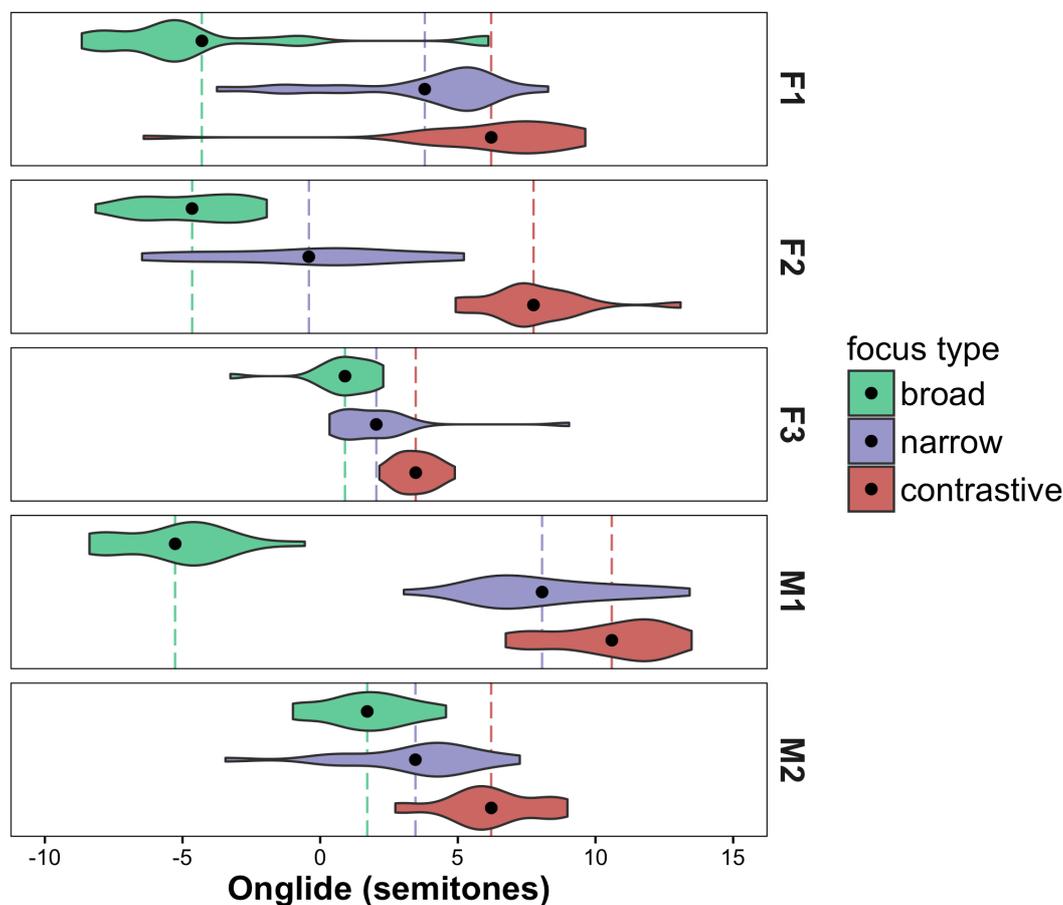


Figure 7: Violin plots of tonal onglide values (in semitones) as a function of focus type (colour coded) and speaker. Mean values are indicated by black dots and coloured dashed lines for comparison across focus types.

Target height

Finally, looking at the f_0 height of the target on the accented syllable (see Figure 8), again the same pattern emerges. Distributions for this measurement overlap heavily as the spread is large. Regardless of whether a falling or a rising accent is used, f_0 is higher in narrow focus than in broad focus and even higher in contrastive focus. This is true for speakers using contours that were transcribed as categorically different pitch accents, such as (F1, F2 and M1) as well as for speakers using contours categorised as having the same pitch accent types (F3 and M2). Again, speakers F1, F2 and M1 show distributions with no or minimal overlap for broad and contrastive focus. For two of the speakers (F1 and F2), narrow focus is in between, overlapping with both of the other conditions, whereas for M1 narrow focus and contrastive focus are highly overlapping. For the subtler speakers, F3 and M2, all three focus types overlap substantially. However, again, the weight of distributions shifts systematically from broad through narrow to contrastive focus. Except for narrow vs. contrastive focus of speaker M1, the mean differences of target height values go in the same direction for all five

speakers. The green line is more to the left than the purple line, which, again, is more to the left than the red line.

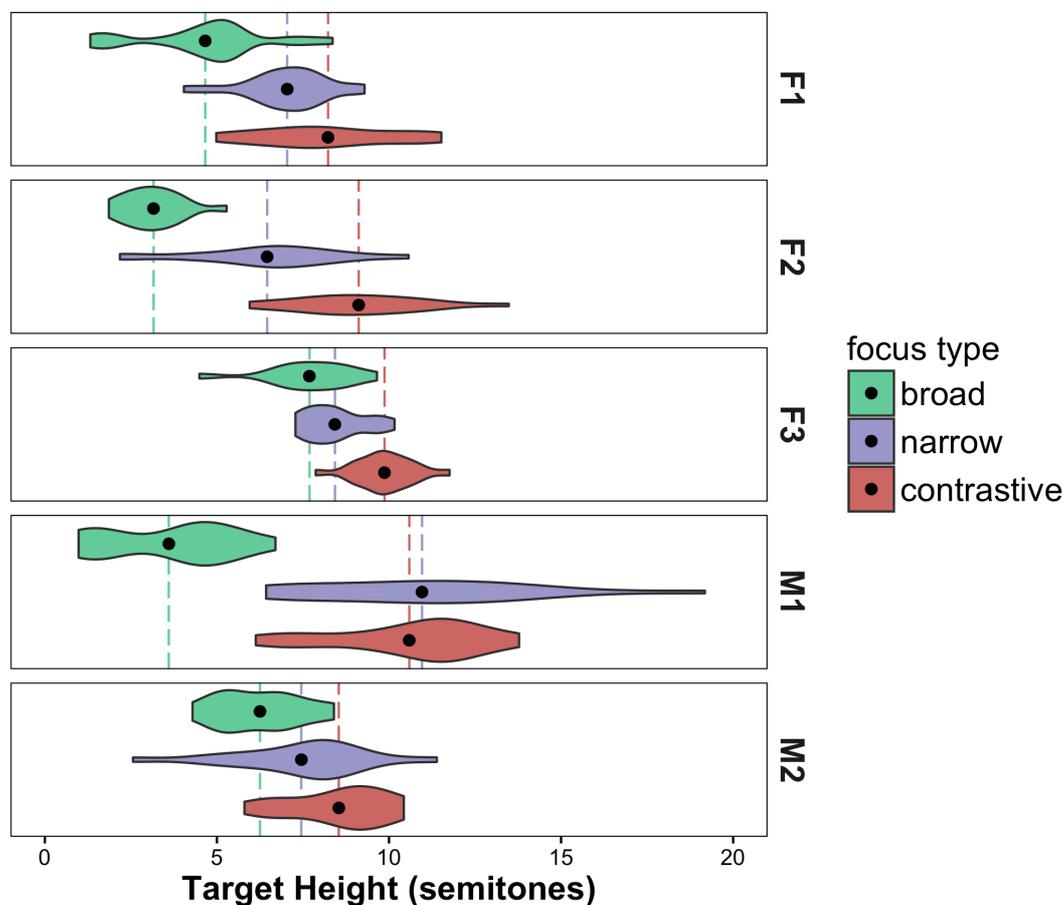


Figure 8: Violin plots of target height values (in semitones) as a function of focus type (colour coded) and speaker. Mean values are indicated by black dots and coloured dashed lines for comparison across focus types. Target height is operationalized in relation to a low reference point for each speaker and condition separately.

It is interesting to note that transcribers disagreed most when labelling speaker M2, especially the broad focus and narrow focus conditions, but not in the contrastive focus condition in which the L+H* label was transcribed by both labellers almost all of the time. Recall that in the contrastive condition this speaker produced slightly higher targets with a greater onglide (despite there being barely any difference in alignment), possibly making the realisations easier to categorise.

4.3 Discussion of f0 parameters and their relation to articulation

We have shown that speakers vary greatly in the extent to which they use intonational parameters to distinguish between the different focus conditions. For instance, some speakers make a clear difference between broad and narrow focus or broad and contrastive focus, whereas others make rather subtle differences. The question arises as to whether the intonationally more subtle speakers make more extensive use of other parameters instead. To address this question we examined individual productions in terms of the two

most informative kinematic parameters found in Mücke and Grice (2014), i.e. the duration and displacement of the opening gesture. Specifically, the duration of the opening gesture refers to the time between the maximum lip closure in the word-initial [b] and the maximum lip opening during the following vowel in the target word ([ba:be], [bo:be], [bi:be]). The displacement refers to the interlip distance, i.e. the spatial distance between lower- and upper lips. More specifically, the displacement of the opening gesture refers to the lip distance difference between the maximum closure and the maximum opening corresponding to the production of the word-initial [b] and the following vowel, respectively.

For comparison with the intonation results, violin plots are presented in Figure 9 (duration of opening gesture) and Figure 10 (displacement of opening gesture). Looking at these plots, it is evident that there is considerable overlap across the three focus types for all five speakers for both articulatory parameters. Mücke and Grice (2014) paid particular attention to the differences in these parameters when comparing broad and contrastive focus. They found that articulatory expansion (longer and larger lip opening gestures) not only depended on whether there was an accent or not, but was also affected by other factors, such as degree of prominence (as a function of the focus structure). These differences can be observed when comparing the means (dots and dashed lines) and distributions for broad and contrastive focus in the violin plots (green vs. red), the latter always tending towards being longer (greater duration of the opening gesture) and larger (greater displacement of the opening gesture)

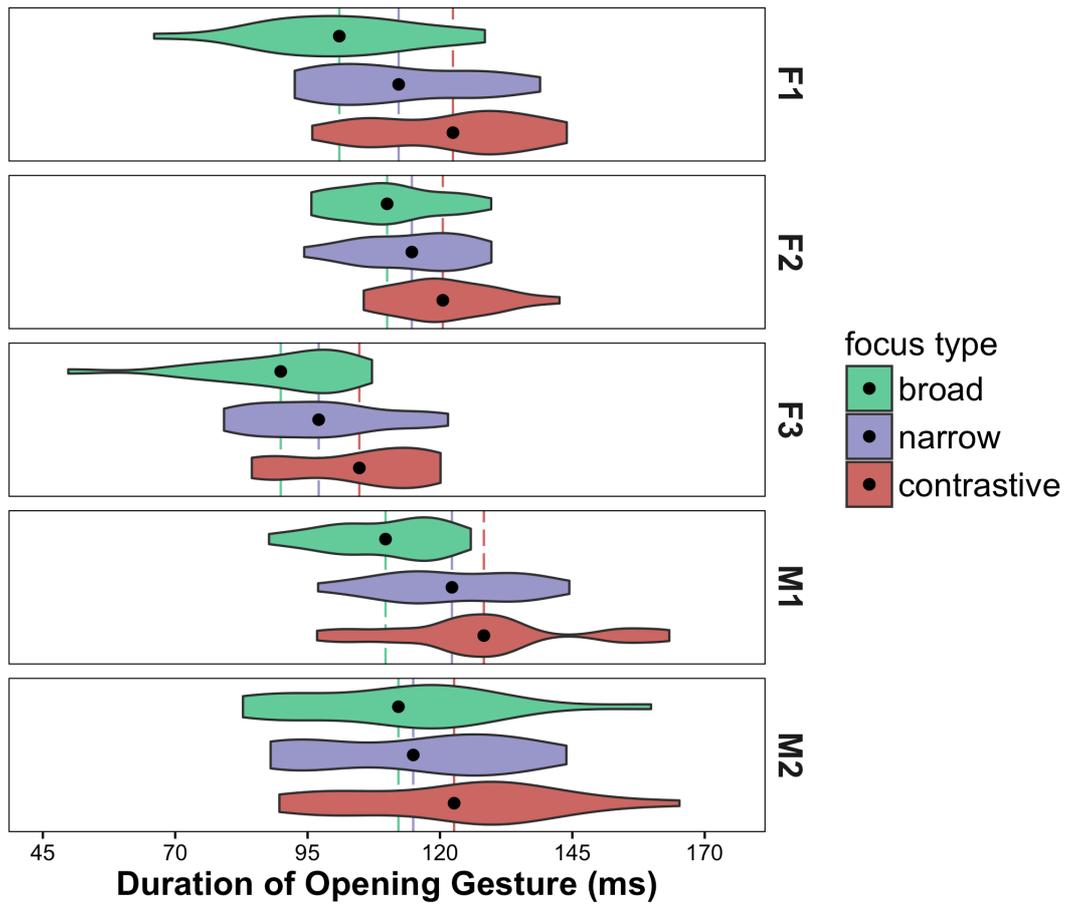


Figure 9: Violin plots of duration of opening gesture (in ms) as a function of focus type (colour coded) and speaker. Mean values are indicated by black dots and coloured dashed lines for comparison across focus types.

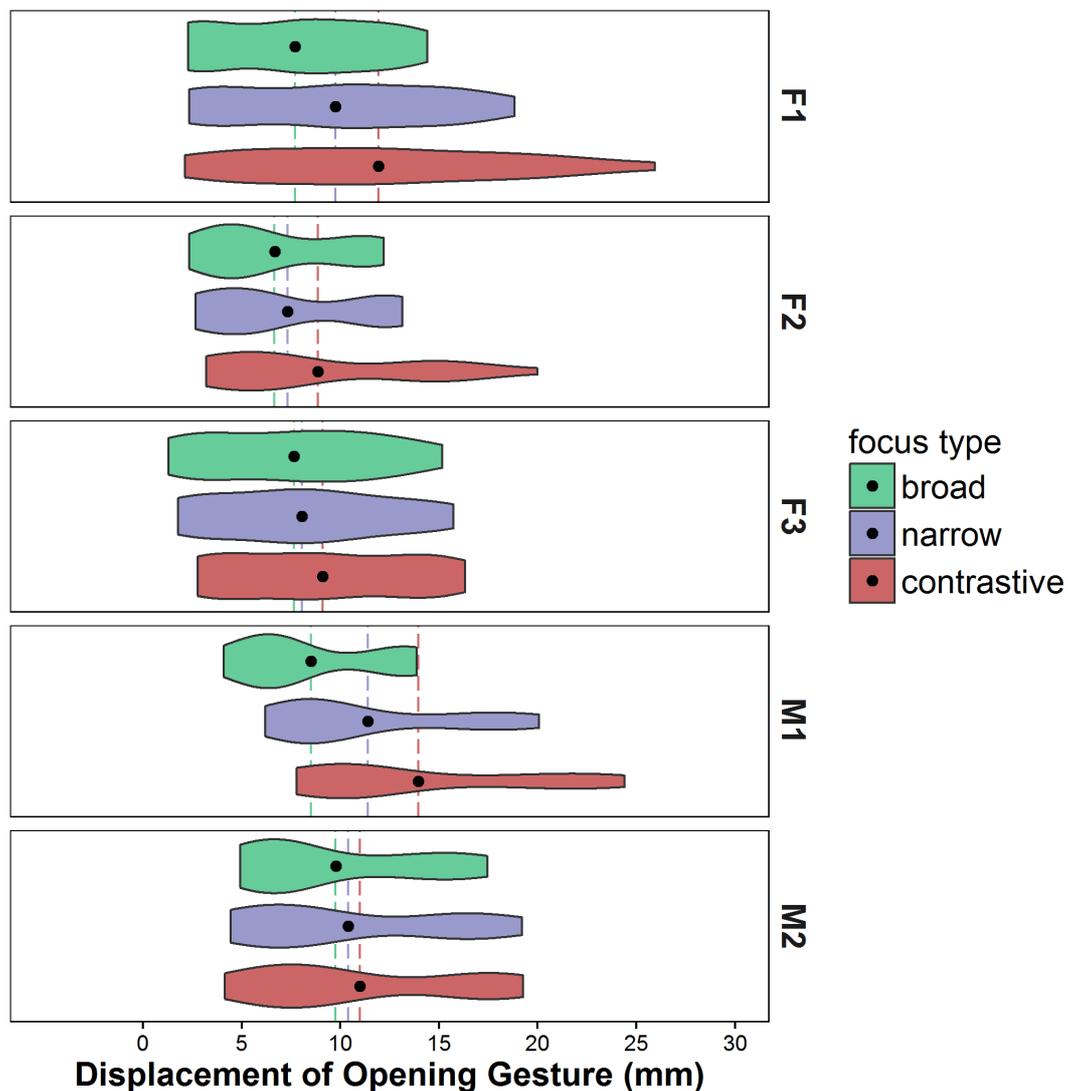


Figure 10: Violin plots of displacement of opening gesture (in mm) as a function of focus type (colour coded) and speaker. Mean values are indicated by black dots and coloured dashed lines for comparison across focus types.

Even when comparing broad and narrow focus, we can observe the same trend across all speakers. That is, all speakers tend to produce longer and larger opening gestures in narrow focus than in broad focus. Likewise, all speakers tend to produce longer and larger opening gestures in contrastive focus as compared to narrow focus. Overall, speakers use articulatory parameters to express focus types in a more homogeneous manner than they use intonational parameters.

Looking at individual speakers, F2 makes a clear distinction between the three focus types in the *intonation*, whereas speaker M2 makes very subtle adjustments (see Figures 6-8). However, when looking at the *articulation*, both speakers have a comparable separation of the three focus types. Moreover, F3, who also made fairly subtle adjustments in the *intonational* parameters, tends to distinguish between the focus types in the duration of the opening gesture, although to a smaller extent in displacement. M1, who did not clearly

separate narrow and contrastive focus intonationally (especially in terms of alignment and target height, Figures 6 and 8 respectively), tended to differentiate these in terms of displacement. Thus, the distributions presented in this section suggest that the high variability in the realisation of *intonational* parameters in this dataset is neither mirrored nor compensated for by *articulatory* parameters.

In the following section we return to the pitch accent distributions for each speaker and relate them to how listeners perceived intended focus conditions. It will become apparent that pitch accent type alone cannot account for the perceptual results, and that thus some of the continuous parameters examined above play a role in the signalling of these pragmatic functions.

5 Perception of focus type and pitch accent categories

Recall that it has been reported for this dataset that all five speakers are able to convey their pragmatic intentions to listeners in a comparable way.³ Averaging over all listeners, the intended focus structure was correctly identified to a similar extent for each speaker. However, beyond overall identification performances, the following paragraphs will provide more detailed information as to the pragmatic meanings confused.

Figure 11 shows matrices depicting intended focus type by speaker on the y-axis plotted against either the transcribed pitch accent category (left column) or the perceived focus type by all listeners from the perception experiment (right column). The darker the shade, the greater the match between intended focus type and pitch accent, or intended focus type and perceived focus type respectively. Pitting pitch accent distributions against the perceptual confusion of categories allows us to evaluate how informative the transcribed pitch accent type really is for listeners' assessment of communicative intentions.

In an ideal situation in which pitch accent and focus type would map on to each other in a one-to-one fashion, we would expect a black diagonal from top left to bottom right. As is described in Section 3, however, pitch accent distribution is highly speaker specific. This is particularly apparent for speaker F3, who almost exclusively produces H* accents for all focus types, or speaker M1, who produced H+!H* accents for broad focus and L+H* accents for both narrow and contrastive focus.

Despite the high degree of variability in the mapping between intended focus type and pitch accent type, the perceptual confusions look strikingly similar. As might be expected, there is frequent confusion between broad and narrow focus and between narrow and contrastive focus, whereas broad and contrastive focus are seldom confused with each other. The mismatch between the pitch accent types used and the perceptual results is particularly

³ The perception experiment involved 20 naïve listeners, who were tasked with matching the test sentences heard to one of four questions, reflecting the broad focus, narrow focus and contrastive focus contexts, plus an additional context, where the target word was out of focus (referred to as the background condition). For more details on the perception experiment see Krüger (2009) and Cangemi et al. (2015).

evident when looking at speaker F3 where confusion patterns are comparable to those for other speakers, despite the fact that this speaker only produces one pitch accent pattern for all intended focus types. We conclude that, in our sample, discrete pitch accent categories are only a poor approximation for how listeners comprehend intended pragmatic meanings. The confusion matrices suggest that irrespective of the assigned pitch accent type, listeners show similar patterns of confusion, an observation that is very much in line with the idea that listeners are sensitive to continuous acoustic parameters in recognising speakers' intentions.

However, recall that speakers made similar use of the continuous parameters investigated, although they differed substantially with regard to the extent to which they modulated these parameters: Some speakers had distinct distributions while others had highly overlapping distributions. In light of the perceptual confusion matrices, it appears that listeners are able to tune into the speaker-specific mapping of continuous phonetic parameters and intended focus types.

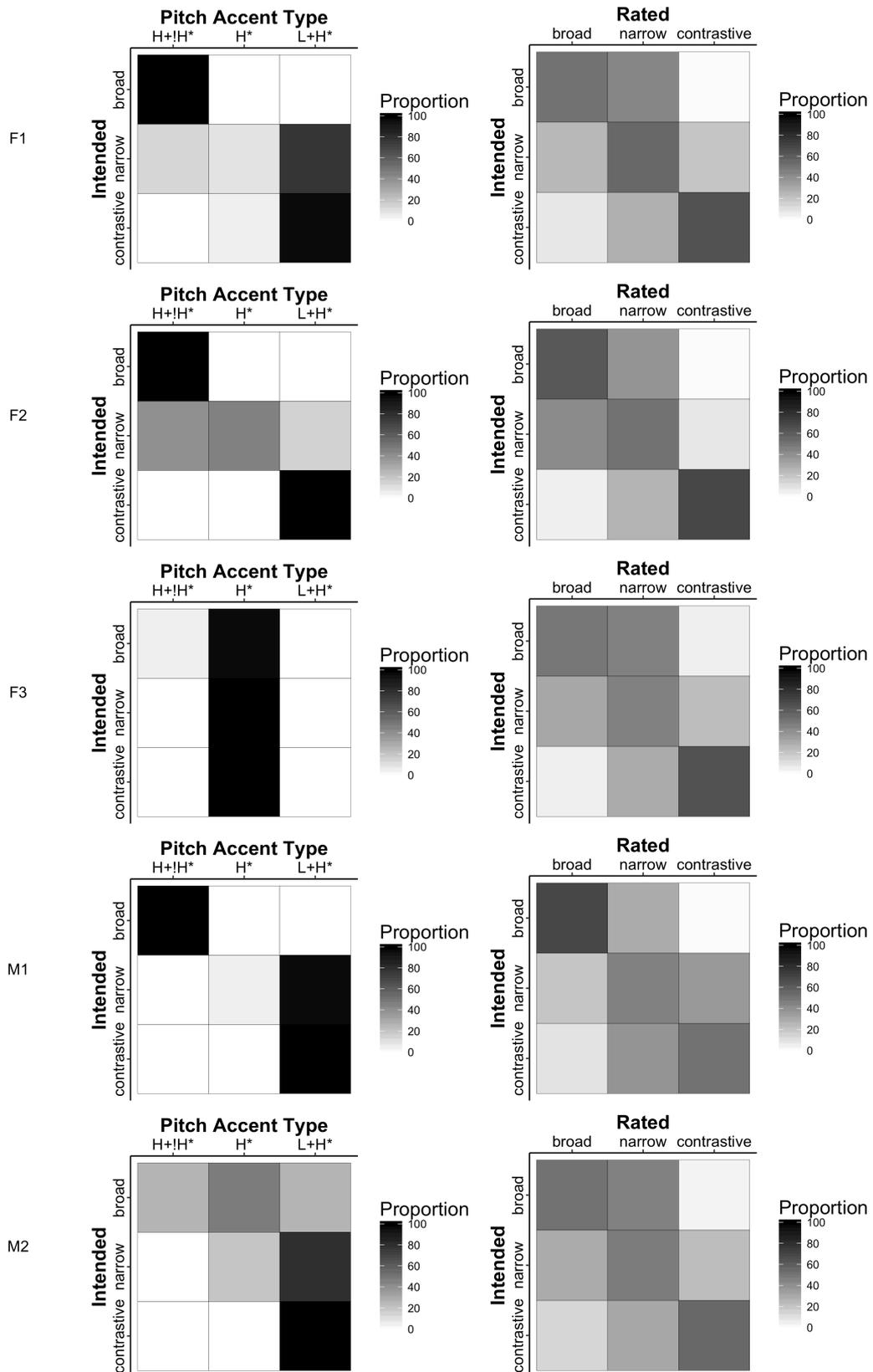


Figure 11: Matrices for production (left) and perception (right). Production matrices show proportions of coded pitch accent by intended focus; perception matrices show rated focus type by intended focus type.

6. General Discussion

To summarise the results, looking at the distributions of pitch accent types as categorised by the transcribers, it appears that certain focus types are typically expressed by certain pitch accents. However, there is variation, both within and across speakers, with no one-to-one mapping between focus structure and pitch accent type.

Looking at the means of the phonetic parameters analysed, all five speakers have strikingly similar tendencies, although for some speakers the differences between realisations for each focus type are greater than for others. For some speakers, these parameters are large enough (either singly or in combination) to correspond to a discrete shift in the choice by the transcribers of one category over another; in other cases, the change appears to stay within the limits of one category. In these cases, different focus types are expressed by instances of the same pitch accent type. Most importantly, regardless of the mapping onto pitch accent categories, all speakers show the same relative pattern: All speakers exhibit a later peak alignment, a greater tonal onglide and a higher target for contrastive focus than for narrow focus, and likewise for narrow focus in relation to broad focus, regardless of whether they use different pitch accent types or not.

A purely categorical (pitch accent-based) account would miss the continuous differences across and within speakers. Crucially, such an approach would also miss the *similarities* in the expression of the different focus conditions. By looking at the parameters behind the phonological categories and by keeping in mind which categories they are involved in distinguishing, we can see that all speakers are modulating the continuous parameters in the same way, even if they do so to different extents. We can thus conclude that all five speakers essentially use the same relative system.

Our data indicate that it is not possible to provide a clear-cut categorisation of the intonation contour as relating to a particular function. Nevertheless, aggregating continuous phenomena into descriptive bins, such as the GToBI pitch accents, allows us to examine distributions indicating relative frequencies of such categories. Here we see that not only are there alternative realisations but also general preferences for certain contours to express a given function. By looking at continuous parameters, we can supplement this information, showing that some speakers manipulate parameters in such a subtle way that they are not reflected in the categorisation of the analyst (the transcriber). The preferred categories for each function (the ones that are most frequent across speakers) involve parameters that are at the extreme end of the scale for that category.

The results discussed above, are difficult to reconcile with traditional linguistic descriptions that focus entirely on abstract phonological categories, not least because these categories are by definition devoid of gradient phonetic information (Uhmann 1991, Féry 1993, Grabe 1998, Grice et al. 2005). This has both practical and conceptual implications for phonological accounts of intonation.

In cases of well-described languages such as German, an intonational system, i.e. an inventory of contrastive intonational events represented as symbols (H+!H* vs. L+H*), has already been established. These symbols are used by transcribers to annotate corpora such as the present one. Even though transcribers are guided by a number of different parameters, they usually only transcribe the result of their categorisation. If only these rigid symbolic representations are considered when analysing and comparing intonational phenomena, we falsely assume that certain speakers do not express certain pragmatic functions intonationally. As described above, some speakers used the same pitch accent type to express all focus conditions. This conclusion, however, misses important generalisations that can be observed through inspection of continuous parameter values.

In cases of language documentation, in which a system has to be established bottom-up, these generalisations can easily be missed and might even lead to an ill-defined inventory of descriptive categories. As the present study has shown, looking at both a categorical assessment in terms of distributions of symbolic representations, as well as continuous modulations of certain acoustic parameters within these categories, allows a more general assessment of an intonation system. Thus, the transcription of pitch accents would benefit from a combined approach, complementing the abstract symbolic representations with the quantification of continuous parameters that define these symbolic representations. The extraction of a number of parameters can be carried out automatically if each target point is transcribed as a separate label (e.g. one for L and one for H* in L+H*, a practice that has been suggested for the new consensus German transcription system (Kügler et al. 2015)). This is in line with recent arguments made by Cole and Shattuck-Hufnagel (2016) in favour of a combined approach to the transcription of prosody in general.

Apart from the practical implications for language documentation, these findings have important consequences for our conceptualisation of the knowledge a speaker has to have to use intonation. Generally, this issue lies at the threshold of the long-debated relation between phonology and phonetics. Traditionally, it has been assumed that grammatical computation is based on discrete categories (e.g. H* vs. L+H*). These categories are manifested in the continuous substance of the speaking event, the language's phonetics (e.g. the actual alignment of the f₀ peak). Because of this disparity, phonology and phonetics are assumed to be fundamentally separate but related through a process of translation – or mapping – from discrete symbols to continuous properties in the phonetic signal. This is the view behind most prominent work on language in particular, and, in fact, on cognitive science in general (Fodor 1975, Newell and Simon 1976, Fodor and Pylyshyn 1981, Haugeland 1985, Harnad 1990).

Our results provide evidence from intonation in favour of a view that maintains that what we refer to as “phonology” and “phonetics” are two sides of the same coin, best understood as a single system. Such a system could be described using a formal language able to express both discrete and

continuous phenomena, in which the key constructs are not symbol strings (representations) and algorithms for their manipulation (discrete computation), but rather laws stated in the form of differential equations. This approach has been successfully used to account for various aspects of cognition (Haken 1977, Kelso, 1995, Port and van Gelder 1985). In phonology this approach was first used within the framework of Articulatory Phonology to represent coordination patterns within the speech apparatus (among others: Fowler et al. 1980, Saltzman and Munhall 1989, Browman and Goldstein 1986). Gafos and Benus (2006) have applied a similar approach to capture quantitative and qualitative aspects of final devoicing in German and vowel harmony in Hungarian.

Further research is needed before we can model quantitative and qualitative aspects of intonation and its interaction with grammatical and non-grammatical aspects of language. Although this is beyond the scope of the present paper, the data presented provide an indication that such a dynamical approach might adequately account for the patterns of intonation we have described so far.

7. Conclusion

This paper has explored the mapping of pragmatic functions onto intonation in terms of the discrete choice of pitch accent type made by the transcriber as well as continuous phonetic parameters that contribute towards the pitch accent categorisation. Although distributions indicating relative frequencies of intonational categories (accent types) expressing each function can be shown to provide a first approximation of the variation found, an important generalisation is missed, namely that there is a parallel between quantitative and qualitative effects: Some speakers' productions lead to a categorical distinction (reflected in the analysis as a different pitch accent category), others being more subtle (resulting in no difference in the assigned category). What is particularly striking is that regardless of the mapping onto proposed categories, all speakers show the same relative pattern.

Focusing on the similarities (rather than the differences in intonational categories used) allows us to provide a plausible account for the fact that all five speakers were perceived at a similar level of accuracy: It appears that listeners are able to "tune in" to the productions of each speaker. This would involve simply calibrating their perception in terms of the degree to which each parameter is modulated, rather than learning new parameters for each speaker.

Moreover, these results indicate the need for a model of intonation that treats continuous modulation of individual parameters as contributing towards the discrete interpretation in terms of intonational categories.

8. References

- Arvaniti, A. (2011). Segment-to-Tone Association In Cohn, A., Fougeron, C., and Huffman, M. (Eds), *Handbook of Laboratory Phonology* (pp. 265-274). Oxford University Press.
- Barnes, J., Veilleux, N., Brugos, A., and Shattuck-Hufnagel, S. (2012). Tonal Center of Gravity: A global approach to tonal implementation in a level-based intonational phonology. *Laboratory Phonology*, 3(2). doi:10.1515/lp-2012-0017
- Baumann, S. (2006). *The intonation of givenness: Evidence from German*. Tübingen: Niemeyer.
- Baumann, S., Becker, J., Grice, M., and Mücke, D. (2007) Tonal and Articulatory Marking of Focus in German. *Proceedings of the 16th ICPHS*, Saarbrücken, Germany, 1029-1032.
- Baumann, S., Röhr, C.T., and Grice M. (2015). Prosodische (De-)Kodierung des Informationsstatus im Deutschen. *Zeitschrift für Sprachwissenschaft* 34(1), 1-42.
- Beckman, M. E., Hirschberg, J., and Shattuck-Hufnagel, S. (2005). The original ToBI system and the evolution of the ToBI framework. In Jun, S.-A. (Ed.), *Prosodic typology: The phonology of intonation and phrasing* (pp. 9-54). Oxford University Press.
- Beckman, M. E., and Venditti, J. J. (2011). Intonation. In Goldsmith, J., Riggle, J., and Yu, A. (Eds), *The Handbook of Phonological Theory, Second Edition* (485-532). Wiley-Blackwell.
- Boersma, P., and Weenink, D. (2016). Praat: doing phonetics by computer [Computer program]. Version 6.0.13, retrieved 31 January 2016 from, <http://www.praat.org/>
- Browman, C. P., and Goldstein, L. M. (1986). Towards an articulatory phonology. *Phonology*, 3(01), 219-252.
- Cangemi, F., and Grice, M. (2016). The importance of a distributional approach to categoriality in autosegmental-metrical accounts of intonation. *Journal of the Association for Laboratory Phonology*.
- Cangemi, F., Krüger, M. and Grice, M. (2015). Listener-specific perception of speaker-specific production in intonation. In S. Fuchs, D. Pape, C. Petrone, and P. Perrier (Eds.), *Individual Differences in Speech Production and Perception* (pp. 123-145). Frankfurt am Main: Peter Lang.
- Cohen, J. (1960). A coefficient of agreement for nominal scales. *Educational and Psychological Measurement*, 20, 37–46.
- Cole, J., and Shattuck-Hufnagel, S. (2016). New methods for prosodic transcription; capturing variability as a source of information. *Journal of the Association for Laboratory Phonology*.
- Cruttenden, A. (1986). *Intonation*. Cambridge: Cambridge University Press.
- Dilley, L. and Heffner, C. (2013): The role of f0 alignment in distinguishing intonation categories. evidence from American English. In: *Journal of Speech Sciences* 3 (1), S. 3-67.
- D'Imperio, M. (2011). Tonal Alignment. In Cohn, A., Fougeron, C., and Huffman, M. (Eds), *Handbook of Laboratory Phonology* (pp. 275-287). Oxford: Oxford University Press.
- Féry, C. (1993). *German intonational patterns*. Tübingen: Niemeyer.
- Féry, C., and Kügler, F. (2008). Pitch accent scaling on given, new and focused constituents in German. *Journal of Phonetics*, 36(4), 680-703. doi:10.1016/j.wocn.2008.05.001
- Fodor, J. A. (1975). *The language of thought* (Vol. 5). Harvard University Press.
- Fodor, J. A., and Pylyshyn, Z. W. (1981). How direct is visual perception?: Some reflections on Gibson's "ecological approach". *Cognition*, 9(2), 139-196.
- Fowler, C., Rubin, P.E., Remez, R.E., and Turvey, M.T. (1980). Implications for speech production of a general theory of action. In Butterworth, B. (Ed.), *Language Production, Vol. 1: Speech and Talk* (pp. 373-420). New York: Academic Press.
- Gafos, A. I., and Benus, S. (2006). Dynamics of Phonological Cognition. *Cognitive Science*, 30(5), 905-943. doi:10.1207/s15516709cog0000_80
- Gamer, M., Lemon, J., Fellows, I., Singh, P. (2012). irr: Various Coefficients of Interrater Reliability and Agreement. R package version 0.84. <https://CRAN.R-project.org/package=irr>
- Grabe, E. (1998). Pitch accent realization in English and German. *Journal of Phonetics*, 26(2), 129-143. doi:10.1006/jpho.1997.0072

- Grabe, E. (2004). Intonational variation in urban dialects of English spoken in the British Isles. In Gilles, P., and Peters, J. (Eds), *Regional Variation in Intonation*. Linguistische Arbeiten (9-31). Tübingen: Niemeyer.
- Grabe, E., Post, B., and Nolan, F. (2001). Modelling intonational variation in English. The IViE system. In Puppel, S., and Demenko, G. (Eds), *Proceedings of Prosody 2000*, 51-57.
- Grice, M., and Baumann, S. (2002). Deutsche Intonation und GToBI. *Linguistische Berichte* 191, 267-298.
- Grice, M., Baumann, S., and Benz Müller, R. (2005). German intonation in Autosegmental-Metrical Phonology. In Jun, S.-A. (Ed.), *Prosodic typology: The phonology of intonation and phrasing* (55-83). Oxford University Press.
- Grice, M., and Baumann, S. (2016). Intonation in der Lautsprache: Tonale Analyse. In Primus, B. and Domahs, U. (Eds.), *Handbuch Laut, Gebärde, Buchstabe*. (pp. 84-105). De Gruyter, Reihe Handbücher Sprachwissen, Band 2.
- Grice, M., Baumann, S. and Jagdfeld, N. (2009): Tonal association and derived nuclear accents-The case of downstepping contours in German. In: *Lingua* 119, S. 881-905.
- Gussenhoven, C. (2004). *The phonology of tone and intonation*. Cambridge: Cambridge University Press.
- Haken, H. (1977). *Synergetics. An Introduction. Nonequilibrium Phase Transitions and Self-organization in Physics, Chemistry, and Biology*. Berlin.
- Harnad, S. (1990). The symbol grounding problem. *Physica D: Nonlinear Phenomena*, 42(1-3), 335-346. doi:10.1016/0167-2789(90)90087-6
- Haugeland, J. (1985). *Artificial intelligence: The very idea*. Cambridge, MA: MIT Press.
- Kelso, J. A. (1995). *Dynamic patterns: The self-organization of brain and behavior*. Cambridge, MA: MIT Press.
- Knight, R. (2008). The Shape of Nuclear Falls and their Effect on the Perception of Pitch and Prominence: Peaks vs. Plateaux. *Language and Speech*, 51(3), 223-244. doi:10.1177/0023830908098541
- Kohler, K.J. (1987). Categorical pitch perception, *Proceedings of the 11th ICPHS*, Tallinn, Estonia, 331-333.
- Kohler, K. J. (1991). Terminal intonation patterns in single-accent utterances of German: phonetics, phonology and semantics. *Arbeitsberichte des Instituts für Phonetik und digitale Sprachverarbeitung der Universität Kiel (AIPUK)*, 25(1), 15-185.
- Kohler, K. J. (2005). Timing and Communicative Functions of Pitch Contours. *Phonetica*, 62(2-4), 88-105. doi:10.1159/000090091
- Krüger, M. (2009). *Produktion und Perzeption von Fokus im Deutschen. Unpublished Magisterarbeit*, IfL Phonetik, University of Cologne.
- Kügler, F., Féry, C. and van de Vijver, R. (2003): Pitch accent realization in German. In: M. Solé, Recasens D. and Romero, J. (Hg.): *Proceedings of the 15th International Congress of Phonetic Sciences*. Barcelona: Univ. Autònoma de Barcelona, S. 1261-1264.
- Kügler F. and Féry, C. (2016): Postfocal downstep in German. In: *Language and Speech*, S. 1-29. DOI: 10.1177/0023830916647204.
- Kügler, Frank., Gollrad, Anja (2015): Production and Perception of Contrast. The case of the rise-fall contour in German. In: *Frontiers in psychology* 6 (1254), S. 1-18. DOI: 10.3389/fpsyg.2015.01254.
- Kügler, F., Smolibočki, B., Arnold, D., Baumann, S., Braun, B., Grice, M., Jannedy, S., Michalsky, J., Niebuhr, O., Peters, J., Ritter, S., Röhr, C.T., Schweitzer, A., Schweitzer, K., and Wagner, P. (2015). DIMA-Annotation guidelines for German intonation. In *Proceedings of the 18th International Congress of Phonetic Sciences*.
- Ladd, D.R. (2008). *Intonational Phonology (Cambridge Studies in Linguistics)*. Cambridge University Press.
- Ladd, D.R., and Morton, R. (1997). *The perception of intonational emphasis: continuous or categorical?* *Journal of Phonetics*, 25, 313-342.
- Liberman, Mark; Pierrehumbert, Janet B. (1984): Intonational invariance under changes in pitch range and length. In: Mark Aronoff und Richard T. Oehrle (Hg.): *Language Sound Structure. Studies in Phonology presented to Morris Halle*. Cambridge: MIT Press, S. 157-233.

- Mücke, D., and Grice, M. (2014). The effect of focus marking on supralaryngeal articulation – Is it mediated by accentuation? *Journal of Phonetics*, 44, 47-61. doi:10.1016/j.wocn.2014.02.003
- Newell, A., and Simon, H. A. (1976). Computer science as empirical inquiry: Symbols and search. *Communications of the ACM*, 19(3), 113-126.
- Niebuhr, O., D'Imperio, M., Gili Fivela, B., and Cangemi, F. (2011). Are there „shapers“ and „aligners“? Individual differences in signalling pitch accent category. *Proceedings of the 17th ICPHS, Hong Kong*, 120-123.
- Pierrehumbert, Janet B. (1980): The phonology and phonetics of English intonation. PhD Thesis. MIT, Cambridge, Massachusetts.
- Peppé, S., Maxim, J., and Wells, B. (2000). Prosodic variation in southern British English. *Language and Speech*, 43(3), 309-334.
- Peters, J. (2014). *Intonation*. Heidelberg: Winter.
- Peng, Roger D. "Reproducible research in computational science." *Science* 334.6060 (2011): 1226-1227.
- Pierrehumbert, J. B. (1980). *The phonology and phonetics of English intonation* (Doctoral dissertation, Massachusetts Institute of Technology).
- Pierrehumbert, J., and Hirschberg, J. (1990). The meaning of intonational contours in the interpretation of discourse. *Intentions in communication*, 271, 311.
- Pierrehumbert, J. B., and Steele, S. A. (1989). Categories of tonal alignment in English. *Phonetica*, 46(4), 181-196.
- Port, R. F., and Van Gelder, T. (1995). *Mind as motion: Explorations in the dynamics of cognition*. MIT Press.
- Prieto, P. (2011). Tonal alignment. In Van Oostendorp, M., Ewen, C., Hume, B., and Rice, K. (Eds), *Companion to Phonology* (pp. 1185-1203). Wiley-Blackwell.
- Ritter, S., and Grice, M. (2015). The Role of Tonal Onglides in German Nuclear Pitch Accents. *Language and Speech*, 58(1), 114-128. doi:10.1177/0023830914565688
- Saltzman, E. L., and Munhall, K. G. (1989). A Dynamical Approach to Gestural Patterning in Speech Production. *Ecological Psychology*, 1(4), 333-382. doi:10.1207/s15326969eco0104_2
- Savino, M. and Grice, M. (2011) The perception of negative bias in Bari Italian questions. In S. Frota, E. Gorka & P. Prieto (Eds.). *Prosodic categories: production perception and comprehension*, Studies in Natural Language and Linguistic Theory 82. 187-206.
- Uhmann, S. (1991). *Fokusphonologie: Eine Analyse deutscher Intonationskonturen im Rahmen der nicht-linearen Phonologie*. Tübingen: Niemeyer.
- Yoon, T. J. (2010). Speaker consistency in the realization of prosodic prominence in the Boston University Radio Speech Corpus. *Proceedings of Speech Prosody*, Chicago, Illinois.

Acknowledgements

This project was funded by the projects DASS (Degrees of Activation and Focus-Background Structure in Spontaneous Speech) and IndiBe (Individual Behaviour in the Encoding and Decoding of Prominence) awarded by the German Research Foundation (GR 1610/5-2 and SFB 1252/1) to Martine Grice.

Appendix 1: Low reference point for calculation of target height

The reference point used for calculating the height of a f₀ target on the accented syllable was a low point 210 ms after the end of the target word. This roughly corresponds to the midpoint of the vowel [ɛ] in [tʁɛf(ə)ŋ]

("treffen"), the last word in the carrier phrase (avoiding the voiceless fricative and final nasal, which are both susceptible to microprosodic effects).

Figure A1 displays the F0 values at this reference point across focus types for each speaker separately. As expected, the female speakers show a higher F0 than the male speakers. Crucially, this reference point is not affected by the intended focus type, i.e. at this reference point, speakers return to a similar F0 value, regardless of whether they produce the target word in a broad, narrow or contrastive focus context.

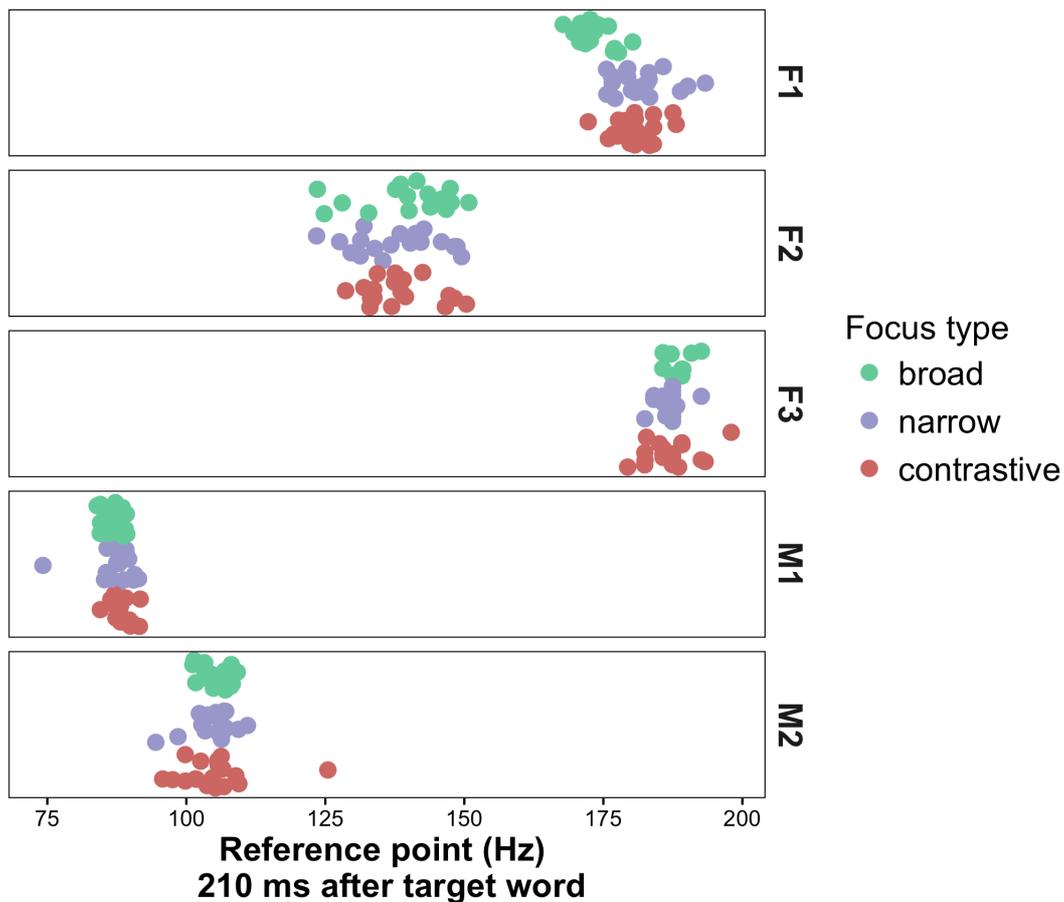


Figure A1: F0 values at the reference point across focus types for each speaker separately.

Appendix 2: Relative Alignment

Alignment was calculated in relation to the duration of the stressed syllable and is presented in violin plots in Figure A2. The results do not reveal a substantially different picture compared to the alignment measure in terms of absolute values. For instance, speaker M2 makes little or no distinction in alignment between the three focus conditions, both in absolute and relative terms. This speaker modulates the remaining parameters to a greater extent in these contexts.

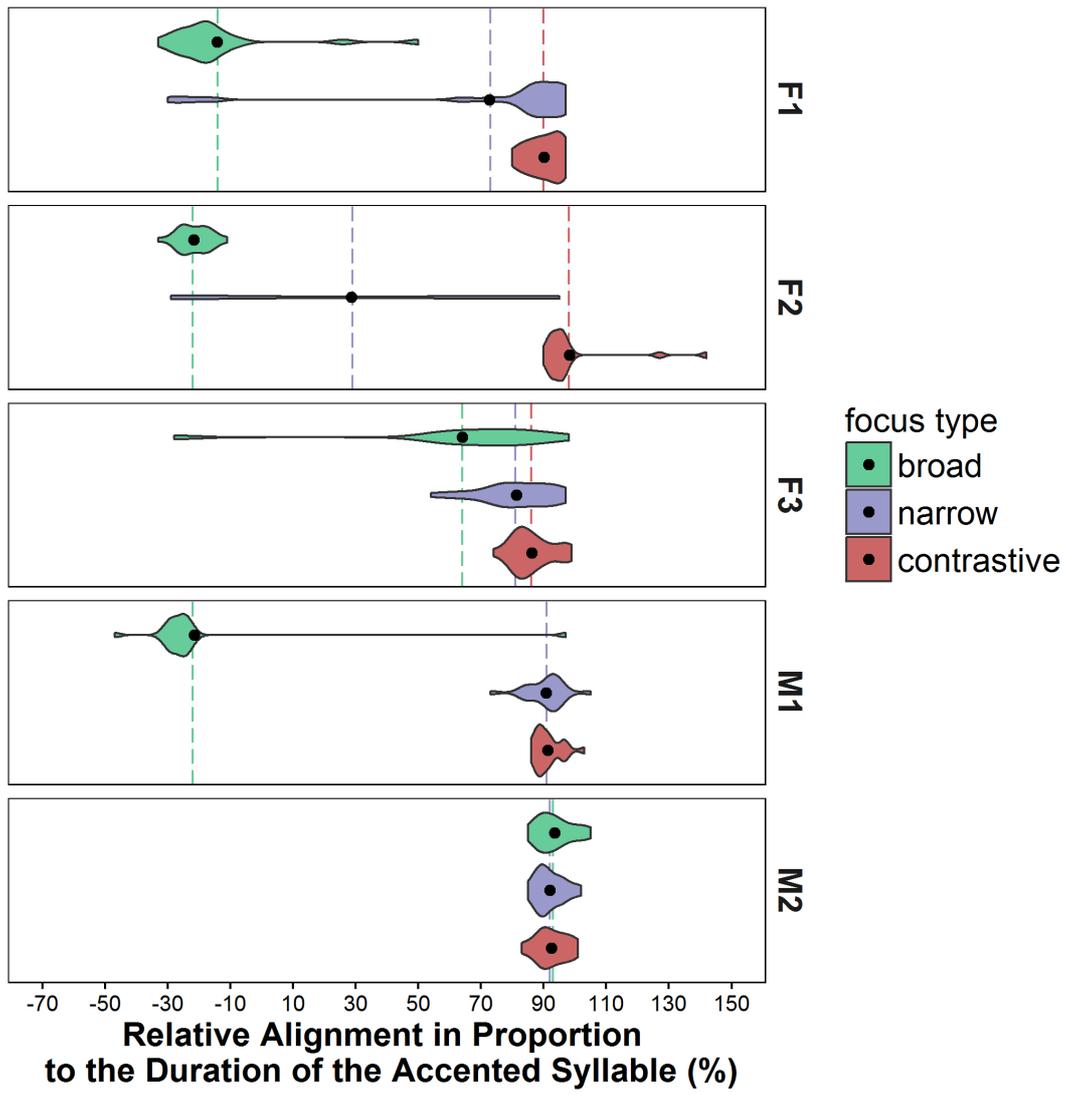


Figure A2: Alignment in relation to the duration of the syllable.