



Three Kinds of Rising-Falling Contours in German *wh*-Questions: Evidence From Form and Function

Katharina Zahner-Ritter^{1,2*}, Marieke Einfeldt², Daniela Wochner², Angela James², Nicole Dehé² and Bettina Braun²

¹ Phonetics, Department II, University of Trier, Trier, Germany, ² Department of Linguistics, University of Konstanz, Konstanz, Germany

OPEN ACCESS

Edited by:

Eleanor Chodoff,
University of York, United Kingdom

Reviewed by:

Pärtel Lippus,
University of Tartu, Estonia
Oliver Niebuhr,
University of Southern
Denmark, Denmark

*Correspondence:

Katharina Zahner-Ritter
k.zahner-ritter@uni-trier.de

Specialty section:

This article was submitted to
Language Sciences,
a section of the journal
Frontiers in Communication

Received: 18 December 2021

Accepted: 22 February 2022

Published: 14 April 2022

Citation:

Zahner-Ritter K, Einfeldt M,
Wochner D, James A, Dehé N and
Braun B (2022) Three Kinds of
Rising-Falling Contours in German
wh-Questions: Evidence From Form
and Function.
Front. Commun. 7:838955.
doi: 10.3389/fcomm.2022.838955

The intonational realization of utterances is generally characterized by regional as well as inter- and intra-speaker variability in f_0 . Category boundaries thus remain “fuzzy” and it is non-trivial how the (continuous) acoustic space maps onto (discrete) pitch accent categories. We focus on three types of rising-falling contours, which differ in the alignment of L(ow) and H(igh) tones with respect to the stressed syllable. Most of the intonational systems on German have described two rising accent categories, e.g., L+H* and L*+H in the German ToBI system. L+H* has a high-pitched stressed syllable and a low leading tone aligned in the pre-tonic syllable; L*+H a low-pitched stressed syllable and a high trailing tone in the post-tonic syllable. There are indications for the existence of a third category which lies between these two categories, with both L and H aligned within the stressed syllable, henceforth termed (LH)*. In the present paper, we empirically investigate the distinctiveness of three rising-falling contours [L+H*, (LH)*, and L*+H, all with a subsequent low boundary tone] in German *wh*-questions. We employ an approach that addresses both the **form** and the **function** of the contours, also taking **regional variation** into account. In Experiment 1 (**form**), we used a delayed imitation paradigm to test whether Northern and Southern German speakers can imitate the three rising-falling contours in *wh*-questions as distinct contours. In Experiment 2 (**function**), we used a free association task to investigate whether listeners interpret the pragmatic meaning of the three contours differently. Imitation results showed that German speakers—both from the North and the South—reproduced the three contours. There was a small but significant effect of regional variety such that contours produced by speakers from the North were slightly more distinct than those by speakers from the South. In the association task, listeners from both varieties attributed distinct meanings to the (LH)* accent as opposed to the two ToBI accents L+H* and L*+H. Combined evidence from **form** and **function** suggests that three distinct contours can be found in the acoustic and perceptual space of German rising-falling contours.

Keywords: intonation, pitch accent, category, fuzziness, imitation, meaning, German

INTRODUCTION

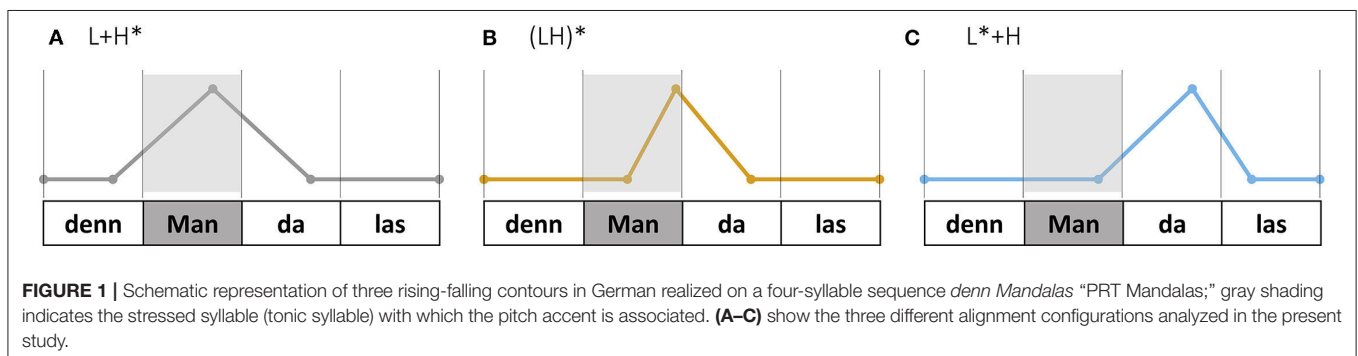
In spoken communication, speakers use intonation, primarily cued by f_0 , to mark sentence type (e.g., question vs. statement), information structure (e.g., focus and topic), information status (e.g., given vs. new information), hierarchical discourse structure, and attitudinal meaning (cf. Lehiste, 1975; Ladd, 2008; Prieto, 2015). The intonational realization of utterances is characterized by a lot of variability in f_0 —both within and across speakers (Atkinson, 1976; Gandour et al., 1991; Niebuhr et al., 2011; Grice et al., 2017), as well as across regional varieties, as shown, for instance, for German (Atterer and Ladd, 2004; Ulbrich, 2005; Braun, 2007; Mücke et al., 2009), or English (Grabe, 2004; Fletcher et al., 2005; Smith and Rathcke, 2020). Such variability includes, among others, the alignment of tonal targets [i.e., the position of low (L) and high (H) turning points with respect to the segmental string], their scaling (i.e., the tonal height), or the shape of intonational events (e.g., the slope or curvature). Category boundaries hence remain “fuzzy” and the question of whether and how the acoustic space can be split into distinct categories is non-trivial (cf. Arvaniti, 2019; Lohfink et al., 2019, for discussion). On the one hand, these categories need to do justice to the variability in the signal; on the other hand, they need to allow for generalizations. In the present paper, we contribute to this debate by addressing the distinctiveness of rising-falling f_0 contours in German in an integrative approach that accounts for both **form** and **function**.

Previous research has demonstrated that nuclear intonation contours in German crucially differ with respect to f_0 -peak alignment (Kohler, 1991b; Grice et al., 2005; Niebuhr, 2022). The f_0 peak may either precede the stressed syllable (H+L*, early-peak accent), or follow it (L*+H, late-peak accent), or be aligned within the stressed syllable (L+H*, medial-peak accent). While in H+L* the accentual movement falls onto the stressed syllable, L+H* and L*+H accents are considered rising accents, with the rising movement being perceptually very prominent (Baumann and Röhr, 2015; Baumann and Winter, 2018). In the present paper, we focus on the two rising accents L+H* and L*+H with a subsequent low boundary tone, along with a third rising-falling contour that lies between the two [henceforth (LH)*, which is our own descriptive label], see **Figure 1**. An earlier study has highlighted the potential existence of a third category

between L+H* and L*+H in German (termed “late-medial peak,” Kohler, 2005; cf. Niebuhr, 2022). The present study is designed to corroborate this preliminary evidence and sharpen the scope of the category, specifically with reference to rhetorical questions where (LH)* was recently observed in production (Braun et al., 2019).

Rising-falling contours in German have received different phonological representations in intonational phonology (Kohler, 1991a; Mayer, 1995; Grice et al., 2005; Peters, 2014). Most of these descriptions distinguish between two kinds of rising-falling contours (**Figures 1A,C**), transcribed as L+H* and L*+H in the German ToBI system (Grice et al., 2005). These accents have been related to differences in meaning: new information vs. self-evident information/information conflicting a speaker's belief (Kohler, 1991b; Grice and Baumann, 2002; Niebuhr, 2007b; Kügler and Gollrad, 2015) or attitudinal information, such as sarcasm (Lommel and Michalsky, 2017). Recent production data on German rhetorical questions (Braun et al., 2019) and verb-first exclamatives (Wochner, 2021) reveal another accent type which falls between the two more established ones (**Figure 1B**). In this contour, both the low and the high tonal target are realized within the stressed syllable (**Figure 1B**), similar to the late-medial peak reported in Kohler (2005, p. 90). This alignment pattern differentiates (LH)* from the more established accents L+H* and L*+H. Here, we take a fresh look at the acoustic and interpretative space of German rising-falling contours to discuss whether there is evidence to model three kinds of rising-falling contours: L+H*, L*+H, and (LH)*. To this end, we employ combined evidence from imitative productions (**form**) and judgments on the connotative meaning (**function**) to determine whether (LH)* is a pitch accent category on its own in German or, alternatively, whether (LH)* might be a variant of one of the two other pitch accents (L+H* or L*+H). If there are three distinct pitch accents in speakers' mental grammars, we expect three distinct contours in production (**form**, Experiment 1) and different connotative meaning attributions in perception (**function**, Experiment 2). The overall aim of our study is hence to probe the fuzziness in German rising-falling contours and discuss ways to model them appropriately.

In section “Background”, we first provide background information on rising-falling contours in German in the different systems of intonational description, before we review approaches



attempting to model variability in intonational contours from a broader perspective. Section “The Present Study: Rationale and Hypotheses” outlines the rationale of our study and our hypotheses. In sections “Experiment 1: Delayed Imitation Study” and “Experiment 2: Paraphrasing of connotative question meaning”, we present the two experiments before discussing the combined experimental results in section “General Discussion”.

BACKGROUND

Rising-Falling Contours in German

According to the autosegmental-metrical (AM) theory of intonation (Arvaniti and Fletcher, 2020 for overview; Pierrehumbert, 1980; Ladd, 2008), pitch accents are represented by sequences of low and high tonal targets. In intonation languages, pitch accents are associated with the metrically stressed syllable, which is a lexical property of the word functioning as an anchor point for pitch accents (highlighted with gray shading in **Figure 1**). The actual alignment of the tonal targets with regard to the position in the stressed syllable varies, which results in different pitch accent types. Different models of German intonation, i.e., *German Tones and Break Indices* (GToBI, Grice et al., 2005), *The Kiel Intonation Model* (KIM, Kohler, 1991a; Niebuhr, 2022), *Intonationsgrammatik des nördlichen Standarddeutschen* “Intonation grammar of the Northern Standard German” (Peters, 2006, 2014) and *Transcription of German Intonation: The Stuttgart System* (STGTsystem, Mayer, 1995) have separated the space of possible pitch accents in rising-falling contours differently: The stylized realizations in **Figures 1A,C** are modeled after the tonal contrast in GToBI (Grice et al., 2005), which is widely used in research on German intonation and which is easily comparable to other ToBI systems in different languages. The realization depicted in **Figure 1B**, i.e., the contour found in rhetorical questions and exclamatives, (LH)*, does not occur in this system. The STGTsystem (Mayer, 1995), an alternative version of GToBI developed in Stuttgart, lists only one accent type for rising-falling contours (L*HL), whose phonetic description resembles the stylization in **Figure 1C** (L*+H in GToBI). Peters (2014, pp. 45–48) describes the contour in **Figure 1A** as a fall (H*L) and the contour in **Figure 1C** as L*H. KIM (Kohler, 1991a; Niebuhr, 2022) is a contour-based account and distinguishes so called “medial peaks” from “late peaks,” whereby the medial peak can be projected onto H*/L+H* and the late peak onto L*+H in an AM framework (cf. Niebuhr and Ambrazaitis, 2006; Niebuhr, 2007b). Importantly, KIM added a contour to its system, based on subsequent research done on the model (Kohler, 2005; Niebuhr, 2022, for overview): In particular, Kohler (2005) used semantic scales to show that the peak alignment continuum contains an additional category that falls between the medial (L+H*) and late peak (L*+H)—a contour called “late-medial peak” in KIM and described as (LH)* in the present paper.

Recent production data further provide evidence for a consistent and meaningful use of (LH)* as a pitch accent signaling rhetorical illocution in *wh*-questions (Braun et al., 2019). This accent was specific to rhetorical questions and did not occur in information-seeking *wh*-questions (Dehé et al., 2022).

Crucially, it did not only occur in contexts of tonal crowding, but also when post-tonic syllables were available. (LH)* has also been observed with verb-first exclamatives, while string-identical information-seeking questions were realized with a high-rising contour (Wochner, 2021). The occurrence of (LH)* in these (non-canonical) utterance types, as opposed to information-seeking questions, suggests that it may be phonemic, rather than a phonetic variant of another accent. At the same time, (LH)* has been described as an allophonic variant of the established accents described above, occurring in suboptimal segmental contexts, in which there are not enough syllables to realize the late-peak contour (cf. Mayer, 1995; Kügler, 2007; Peters, 2014), hence resembling the configuration in **Figure 1B**. In practice, (LH)* realizations caused difficulties in transcription because they share the alignment of the L tone with L*+H and that of the H tone with L+H*. Taken together, the status of (LH)* in German (i.e., whether it is phonetic or phonological) is by far not clear and it raises issues for the mapping between acoustic realization and phonological categories. The specific question of the present paper is how many distinct (meaningful) contours need to be modeled within the broad category of rising-falling contours in German.

A complicating factor for this question is that natural productions are not as clearly distinct as the stylizations in **Figure 1** may suggest, but are subject to variability both within and across speakers (Atkinson, 1976; Gandour et al., 1991; Niebuhr et al., 2011; Grice et al., 2017; Lohfink et al., 2019; Roessig et al., 2019; Roessig, 2021).¹ Clearly, such individual variation blurs the boundaries of intonational categories. Regional variety, which is one of the foci of the present paper, additionally pushes the notion of categories to its limits as distributions between categories might overlap (Atterer and Ladd, 2004; Grabe, 2004; Gilles, 2005; Peters, 2006; Braun, 2007; Mücke et al., 2009): Indeed, a main discriminating aspect of the above-cited models on German intonation are their geographical origins. On a north-south axis, KIM (Kohler, 1991a) is located farthest in the north, followed by the system developed by Jörg Peters in Oldenburg (Peters, 2006, 2014). The STGTsystem (Mayer, 1995), in turn, originates in the South of Germany (Stuttgart). GToBI is a collaborative approach developed at universities in Saarbrücken, Stuttgart, Munich, and Braunschweig (Grice et al., 2005, p. 62). It is possible that the apparent differences in intonation labels and pitch accent contrasts are in part influenced by differences in regional variety (cf. Gilles, 2005; Peters, 2006; Kügler, 2007).

In fact, there is experimental evidence that Southern German speakers produce pitch accents in declarative sentences differently from Northern German speakers, at least in prenuclear position: Atterer and Ladd (2004), for instance, reported that Southern German speakers (from Bavaria) aligned

¹Niebuhr et al. (2011), for instance, showed that German speakers differ in the magnitude of the alignment with which they differentiate H* from H+L*, with a weaker alignment contrast being compensated by adjustments in contour shape. Grice et al. (2017) showed that speakers consistently use the phonetic cues *alignment* and *scaling* when differentiating focus types, however only for some of the speaker did these differences lead to a difference in the intonational event (H* vs. L+H*), see also Braun (2006) on similar findings for contrastive vs. non-contrastively used prenuclear accents.

prenuclear accentual rises significantly later than speakers from the North-West of Germany (cf. Braun, 2007; Mücke et al., 2008); in nuclear position, alignment differences went in the same direction but were not significant (Mücke et al., 2009). In his analysis of the tonal inventory of Swabian, an Alemannic variety in the South of Germany, Kügler (2007) shows that speakers predominantly produced L^*+H accents in declarative sentences (see also Kügler, 2004). Distributional analyses of Northern German speakers (Kiel), in turn, reveal medial peaks to occur more frequently than late peaks (Peters et al., 2005). This suggests that the distribution frequency in tonal inventories might also differ across regions (cf. Fitzpatrick-Cole, 1999; Leemann, 2012, on Swiss German), such that the acoustic space of rising-falling contours in Southern German speakers is shifted toward the right end of the spectrum [recall that the southern STGSystem (Mayer, 1995) only accounts for one rising-falling contour]. In the present paper, we directly compare speakers from two different regions (North vs. South) on the three-way tonal alignment contrast.

Modeling Intonational Categories

The question of how phonological representations—typically thought of as distinct categories—and phonetic modification—typically understood as a gradual change—interrelate has been an issue of on-going debate (e.g., Ohala, 1990; Niebuhr, 2007a; Pierrehumbert, 2016; Arvaniti, 2019; Barnes et al., 2021; Roessig, 2021). It is uncontroversial that some form of generalization is necessary to systematize interfaces with other core areas, such as semantics or pragmatics. At the same time, clear-cut boundaries cannot be maintained given the variability in the speech signal and the fuzziness of the mapping between acoustic **form** and phonological category.

In intonational research, different tasks have been employed to study the relation between the continuous signal and intonational categories (cf. Prieto, 2012 for overview): Focusing on **intonational form**, identification and discrimination tasks have been used in classic categorical perception paradigms (Kohler, 1987, 1991b; Ladd and Morton, 1997; Schneider and Lintfert, 2003; Niebuhr, 2007b). Kohler (1991b), whose work is directly related to our question, showed categorical perception for early vs. medial peaks (i.e., $H+L^*$ vs. $L+H^*$), two accents that differ in the direction of the accentual movement. The difference between the two rising-falling contours (medial vs. late peaks, i.e., $L+H^*$ vs. L^*+H), in turn, was less clear-cut. Categorical perception results were similar for speakers from Northern and Southern Germany (Kiel vs. Munich, Kohler, 1991b, p. 149ff.). Beyond tonal alignment, the shape of the contour also seems to influence the categorical perception of rising-falling contours, leading to a more or less clear-cut perception between $L+H^*$ and L^*+H (Niebuhr, 2007a, for effects of peak shape and intensity transitions); see also Barnes et al. (2021) for a study corroborating the relevance of the shape of the interpolation between L and H for the distinction between $L+H^*$ and L^*+H in English. Another paradigm testing the distinctiveness in intonational form is imitation, which is based on the idea of a perception-production loop (e.g., Pierrehumbert and Steele, 1989; Braun et al., 2006; Dillely and Brown, 2007; Dillely, 2010;

Chodroff and Cole, 2019b; Petrone et al., 2021). In imitation tasks, participants are typically presented with one stimulus at a time and have to imitate it. The productions are analyzed in terms of the overlap (or non-overlap) in the distributions of relevant parameters (such as tonal alignment) or overall shape. In imitation, task difficulty or working memory seem to affect outcome patterns: Braun et al. (2006), for instance, employed an iterative imitation paradigm in which speakers first imitated a set of randomly generated f_0 tracks and then iteratively repeated their previous productions. They showed that speakers retained some detail in immediate imitation, which was lost, however, over successive repetitions. The authors argue for attractors in the perceptual space of intonation that function as a perceptual magnet. Participants in Chodroff and Cole (2019b), American English speakers, had to imitate one of eight nuclear contours and transfer the respective contour to a novel sentence with the same rhythmic structure, hence making generalization necessary. In their study, speakers primarily maintained the distinction between rising and falling contours, similar to the attractor contours in Braun et al. (2006). Petrone et al. (2021) showed that when working memory capacity is smaller, speakers have difficulties in reproducing contours correctly: Specifically, speakers with high working memory capacity were more accurate in the imitation of phonological events, both for obligatory events (pitch accents and boundary tones) and optional events. In sum, the harder the imitation task (due to either task demands or cognitive capacities), the smaller the set of reproduced contours. A challenging imitation task hence seems to us an appropriate method for the question of whether there are three distinct rising-falling contours.

Studies that have addressed the **functional distinction** between intonational contours have employed semantic scales (e.g., Dombrowski, 2003; Kohler, 2005; Dombrowski and Niebuhr, 2010; Kügler and Gollrad, 2015; Wochner, 2021), free association tasks (Kohler, 1991b), acceptability judgment tasks (Baumann and Grice, 2006), or psycholinguistic methods such as eye-tracking (e.g., Braun and Biezma, 2019). Kügler and Gollrad (2015), for instance, showed that German listeners differentiated between a contrastive and a broad focus reading based on differences in the scaling of the H tone (the L tone did not affect perceptual ratings, but see Ritter and Grice, 2015). Based on a free association task, Kohler (1991b) reports that medial peaks were associated with information that was new to the discourse in declaratives and with an information-seeking notion in questions. Late peaks also signaled new information (similar to medial peaks) but also added attitudinal meanings, such as astonishment or self-evidence (see also Grice et al., 2005; Lommel and Michalsky, 2017). Kohler (2005) corroborated these findings using semantic differentials; the contour that falls between the medial and late peak, the late-medial peak, tended to be associated with unexpectedness or surprise. Braun and Biezma (2019) used an eye-tracking paradigm to investigate the contrastive nature of nuclear $L+H^*$, prenuclear $L+H^*$, and prenuclear L^*+H . They showed that listeners interpreted prenuclear L^*+H and nuclear $L+H^*$ contrastively (more fixations to a referent that contrasted with the accented word) as opposed

to prenuclear L+H*. Here, we use a combined approach of **form** and **function** to understand the sources of fuzziness surrounding rising-falling accents in German and to model it successfully.

THE PRESENT STUDY: RATIONALE AND HYPOTHESES

To test the distinctiveness of the three kinds of nuclear rising-falling contours in German [L+H*, L*+H, (LH)*; cf. **Figure 1**], we employ *wh*-questions and make use of two tasks: a delayed imitation task and a free association task. The delayed imitation task requires a kind of storage (beyond access to echoic memory), and hence taps into phonological representations of intonational contours. The working memory model by Baddeley and Hitch (1974) assumes that acoustic information decays after ~2 s (phonological short-term memory, cf. Plomp, 1964; Gathercole et al., 1997), unless it is refreshed by a sub-vocal articulatory rehearsal process (Baddeley and Hitch, 1974; Baddeley, 1986, 2003). Moreover, Crowder (1982) reports that in terms of discrimination accuracy for vowel formants “[t]he auditory memory loss seems to be asymptotic at about 3 s” (Crowder, 1982, p. 197). Hence, there seems to be a threshold of about 2 to maximally 3 s up to which acoustic information is readily available and after which acoustic information decays. Based on this threshold, we designed our delayed imitation task with a 2,000 ms delay and a following sine tone with a duration of 500 ms. The free association task seems to be the best-suited paradigm for our study since the functional scope of (LH)* is not clear yet, which makes it hard to establish pre-defined connotative meanings required in other tasks.

For both experiments, speakers from Southern and Northern Germany were recruited in order to investigate regional variation (Mayer, 1995; Atterer and Ladd, 2004; Ulbrich, 2005; Braun, 2007; Kügler, 2007; Mücke et al., 2009). Speakers were allocated to either the Northern or the Southern German group according to where they were born and grew up. The Northern German group comprised speakers north of the Benrath line, an isogloss separating Low German and High German dialects (based on the High German consonant shift). The Southern German group comprised speakers from Baden-Wuerttemberg and Bavaria (south of the Speyer line, an isogloss that additionally separates Upper German dialects from Central German dialects), cf. Waterman (1991/1966).

In **Experiment 1 (form)** participants imitate three resynthesized nuclear rising-falling contours on *wh*-questions [L+H*, (LH)*, and L*+H] in a delayed imitation paradigm addressing phonological processing (cf. Baddeley and Hitch, 1974; Crowder, 1982; Baddeley, 1986, 2003). Methodologically, we input a three-way alignment contrast, and analyze the productions of Experiment 1 holistically, using general additive mixed models (GAMMs, Wood, 2006, 2017) on time-normalized utterances. This method allows us to capture the f0 contours as a whole and compare when in time two contours differ from each other significantly (cf. Wieling, 2018; van Rij et al.,

2019; Sóskuthy, 2021). Using GAMMs hence not only provides information about tonal alignment (Atterer and Ladd, 2004), but also about tonal onglides (Ritter and Grice, 2015; Roessig et al., 2019), f0 excursions and scaling, and the overall shape of the contour (Niebuhr, 2007b; Niebuhr et al., 2011; Barnes et al., 2012, 2013, 2021). GAMMs furthermore allow us to test for interactions between intonation condition and regional variety over time, hence informing us on whether regional variation affects the distinctions between contours differently. In that sense, GAMMs represents an ideal statistical technique to disambiguate the fuzzy data patterns existent in f0 contours in order to unravel the meaningful underlying structure of intonational phonology.

Participants’ imitative productions will be informative on how many **distinct contours** we need to model in the acoustic space of German rising-falling contours: Three distinct rising-falling contours in the imitative productions of the speakers will provide evidence for (LH)* as a third kind of rising-falling contour in German next to the two more established L+H* and L*+H contours, hence corroborating the three-way-contrast initially laid out in Kohler (2005) and also observed in Braun et al. (2019). Given that our delayed imitation task requires storage of the contours, the evidence would go beyond phonetic details and clearly speak in favor of phonological processing. If, on the other hand, speakers reproduce two contours in their imitative productions, this will provide evidence in favor of collapsing the range of rising-falling contours into two contours (cf. Braun et al., 2006, on English; Chodroff and Cole, 2019b). Reproduction of only one contour would suggest that the task is too hard (since there is plenty of independent evidence in favor of two rising-falling contours in German, see sections “Introduction” and “Background”). With respect to **regional variation**—although a direct comparison of studies reporting occurrence frequency is difficult, medial-peak contours (H*/L+H*) have been shown to be more frequent than late-peak contours (L*+H) for Northern German speakers (Peters et al., 2005). For Southern German speakers, in turn, rising accents with a late L and H alignment have been reported to occur frequently (described as L*+H in Kügler, 2004; see also Truckenbrodt, 2007, for the prenuclear position). Based on these differences in occurrence frequency, it is conceivable that L+H* functions as a perceptual attractor (magnet) for Northern German speakers, while L*+H serves this function for Southern German speakers, along the lines of what is known on magnets on the segmental level (Anderson et al., 2003; cf. Braun et al. (2006) and Roessig et al. (2019) for attractor-based accounts of intonation). Given that (LH)* may be less strongly anchored in the intonational grammar due to its more restricted function and hence less frequent occurrence, it may be yet more prone to merger effects (cf. Braun et al., 2006). Under this assumption, we predict the distinction of contours to differ between regions, with a smaller distinction between L+H* and (LH)* in the North than in the South [L+H* as merger with (LH)*], and conversely, a smaller distinction between L*+H and (LH)* in the South than in the North [L*+H as merger with (LH)*].

Experiment 2 (function) tests whether the three rising-falling contours in *wh*-questions are interpreted differently. To this

end, we conducted a qualitative study in which participants, different from the ones in Experiment 1, paraphrased the connotative meaning of the stimuli in Experiment 1 in their own words. The paraphrases of Experiment 2 were recoded into superordinate categories and analyzed using conditional inference trees (CTrees, Hothorn et al., 2006). This method allows us to test whether and how intonation condition and regional background affect participants' responses. We predict that if there are in fact three distinct contours, they will lead to different interpretations: Drawing on the available literature, we expect that L+H* leads to descriptions relating to an information-seeking nature (Kohler, 1991b; Baumann and Grice, 2006; Braun et al., 2019), while L*+H is expected to trigger descriptions related to contrast and/or attitudinal meanings (Grice et al., 2005; Niebuhr, 2007b; Lommel and Michalsky, 2017); (LH)* is hypothesized to be interpreted as rhetorical (Braun et al., 2019), or to signal surprise or obviousness (Wochner, 2021), or unexpectedness (Kohler, 2005). In terms of regional variation, we cannot make strong predictions regarding meaning—recall that Kohler (1991b, p. 149ff.) showed constant semantic judgments for different contours across Northern and Southern German listeners. If anything, we expect a merger effect in terms of meaning for the most frequent accent type (L+H* in Northern and L*+H in Southern German speakers).

EXPERIMENT 1: DELAYED IMITATION STUDY

Methods

Participants

In total, 28 monolingual native German participants, half from Northern Germany (mean age = 25.7 years, SD = 5.0 years, 10 female, 4 male) and half from Southern Germany (mean age = 25.5 years, SD = 4.4 years, 1 diverse, 8 female, 5 male), who had not learned a second language before the age of six, took part in the imitation study. Speakers from the Southern German group spent most of their lives in Baden-Wuerttemberg ($N = 14$), while speakers in the Northern German group came from Berlin ($N = 1$), Brandenburg ($N = 1$), Hamburg ($N = 1$), Mecklenburg-Vorpommern ($N = 1$), Lower Saxony ($N = 3$), North Rhine-Westphalia ($N = 2$), and Schleswig-Holstein ($N = 5$), all north of the Benrath Line. Due to restrictions imposed by COVID-19, testing started in the lab for eight Southern German speakers and then was continued via the online platform *SosciSurvey* (<https://www.sosciurvey.de>, Leiner, 2018) for all Northern speakers and the six remaining Southern German speakers.

Materials

Four target *wh*-questions were constructed that consisted of the *wh*-word *wer* “who,” a monosyllabic verb, the particle *denn*, and a trisyllabic object noun with initial stress, see (1).

(1)

- a. *Wer heißt denn Melanie?* ['mɛ.la.ni] (“Whose name is Melanie?”)
- b. *Wer spielt denn Libero?* ['li:.bɛ.ʁo] (“Who plays sweeper?”)

- c. *Wer malt denn Mandalas?* ['man.da.las] (“Who draws colors mandalas?”)
- d. *Wer trinkt denn Malibu?* ['ma:.li.bu] (“Who drinks Malibu cocktails?”)

Nouns with two post-tonic syllables were chosen to avoid tonal crowding (Prieto, 2011; Hanssen, 2017; Rathcke, 2017); also, their segments were as sonorous as possible, especially in the first two syllables, to ease f_0 analysis. The propositions of the questions were chosen so that they did not elicit strong (positive or negative) feelings but were mainly perceived as neutral². The four questions were recorded by a female native speaker from Northern Germany (31 years at time of recording), who grew up with a Southern German parent and who is familiar with intonational phonology. She produced the *wh*-questions in two conditions: (i) with a nuclear L+H* accent and (ii) with a nuclear L*+H accent (see **Supplementary Material S1** for acoustic analysis). She was instructed to focus on the alignment of the tonal targets. The recordings were then manipulated in three steps (splicing, duration manipulation, f_0 manipulation) using *Praat* (Boersma and Weenink, 2016). First, splicing ensured that pitch accent realizations were not affected differently by the preceding part of the *wh*-question. To this end, for each item, the auditorily best “precontext” (*wh*-word, verb, particle) was selected. Likewise, the best productions of the object nouns (one for L+H*, one for L*+H for each item) were selected and cut at positive zero-crossings. Both parts were scaled to 63 dB. To reduce variability across items, the precontexts were manipulated in terms of duration using PSOLA resynthesis. This way, the constituents had an equal average duration for each item (in the three intonation conditions). The same was true for the three syllables of the noun. Second, the precontexts were cross-spliced to the nouns. Finally, the alignment of the tonal targets of the noun (L1: start of the f_0 rise, H: f_0 peak, L2: end of the f_0 fall) was manipulated, based on the alignment of the naturally recorded stimuli for L+H* and L*+H and the values reported in Braun et al. (2019) for (LH)*, see **Table 1**. **Table 1** shows the locations of the three tonal targets (L1, H, L2) within the rising-falling contour (in the particle *denn*, the first, second or third syllable in the object noun). Percentages refer to the total duration of the respective unit, e.g., the f_0 peak (H) occurred after 71% of first syllable of the noun in L+H*, and after 94% in (LH)*; for L*+H, it occurred after 71% of the second syllable of the noun. Note that **Figure 1** shows a visual representation of **Table 1**. The f_0 values in the rising-falling contours were set at 166 Hz for L1, at 273 Hz for H, and at 170 Hz for L2 (based on the mean values in natural productions), leading to a pitch range of 8.6 semitones (st) for the rising part and 8.2 st for the falling part of the contour.

²Materials underwent a check in which the propositions of eight questions (four of which were the selected target questions and four of which were filler questions) were presented to 15 listeners (native speakers of German, mostly student assistants) who judged each proposition [e.g., *drinking Malibu* (target item), *going to the cinema* (filler) etc.] as either positive (*I like*), negative (*I don't like*), or neutral (*I don't have an opinion*). The propositions of the selected questions were predominantly judged as neutral (68%); in 27% participants had a positive attitude and in 5% of the cases a negative attitude toward the proposition of the sentence.

TABLE 1 | Alignment of tonal targets (L1, H, L2) in rising-falling contours in experimental stimuli; L+H* and L*+H values based on natural recordings, (LH)* values based on Braun et al. (2019).

	L1	H	L2
L+H*	In [n] from <i>denn</i> (22.0%)	In syllable 1 of noun (71.2%)	In syllable 2 of noun (69.6%)
(LH)*	In syllable 1 of noun (45.4%)	In syllable 1 of noun (94.4%)	In syllable 2 of noun (69.6%)
L*+H	In syllable 1 of noun (76.3%)	In syllable 2 of noun (71.1%)	In syllable 3 of noun (31.6%)

Percentages refer to the total duration of the respective unit.

The f_0 contours of naturally produced L+H* accents and naturally produced L*+H accents (4 items each) were resynthesized into the three intonation conditions, leading to a total of 24 test stimuli (4 items \times 3 target contours \times 2 manipulation origins). We used two manipulation origins to exclude the possibility that spectral effects could have affected the imitations, which was not the case (see below). Natural recordings of (LH)* were avoided because this contour might be realized with breathy voice in *wh*-questions (Braun et al., 2019), which might be a confounding cue. Furthermore, recording the contours at the ends of the continuum (**Figures 1A,C**) will allow us to resynthesize intermediate steps in future studies. In the present study, we start with three contours, the two more established L+H* vs. L*+H, and one intermediate contour (LH)*. We further selected four additional *wh*-questions to be used as practice trials. They had the same syntactic structure but different target words (*Thymian* “thyme”, *Komiker* “comedian”, *Kolibris* “hummingbirds”, *Tombolas* “tombolas”). These questions were resynthesized into the more established accents L+H* and L*+H.

There were two experimental lists with a pseudo-randomized order of trials to avoid priming of contours between trials. Lists did not contain sequences with the same item or the same contour in a row. The second experimental list was a mirror list of the first list such that the first trial in list 1 was the last in list 2. This was done to avoid order effects. Experimental lists were randomly assigned to the participants. Prior to the 24 experimental trials, there were four practice trials to familiarize participants with the procedure and voice of the speaker.

Procedure

Each trial was initiated by a sine tone (at 300 Hz, 500 ms duration) to signal the beginning of the trial. Participants listened to the questions via headphones. Each target question was also orthographically displayed on screen. Participants were instructed to imitate the utterances as closely as possible with a special focus on their speech melody. They were told to choose a pitch register that appeared suitable for them. This was done to avoid a mimicry of pitch and vocal characteristics of the speaker.

Each utterance was played only once, followed by a 2,000 ms period of silence and a sine tone of 500 ms (presented pseudo-randomly at 450 or 150 Hz) before participants started to imitate the question. The sine frequencies meet the floor and ceiling register frequencies of the speaker who produced the stimuli; the sine tones were played to overwrite any acoustic trace that might be kept after the 2,000 ms silence. We used two

TABLE 2 | Overview of imitated productions per group in final dataset of Experiment 1.

	Northern German speakers	Southern German speakers
L+H* contour	108	112
(LH)* contour	111	111
L*+H contour	110	111
Sum	329	334
Total		663

different frequencies for the sine tone, in random order, so that participants could not anticipate and adapt to it. After participants had imitated the respective utterance, they pressed a key to proceed to the next trial. Recordings were done via the microphone of the participants' computers in the remote setting. In the lab setting, recordings were done with a headset microphone (DPA 4088F) onto a MacBookPro in a sound-attenuated booth.

Data Processing and Statistical Analysis

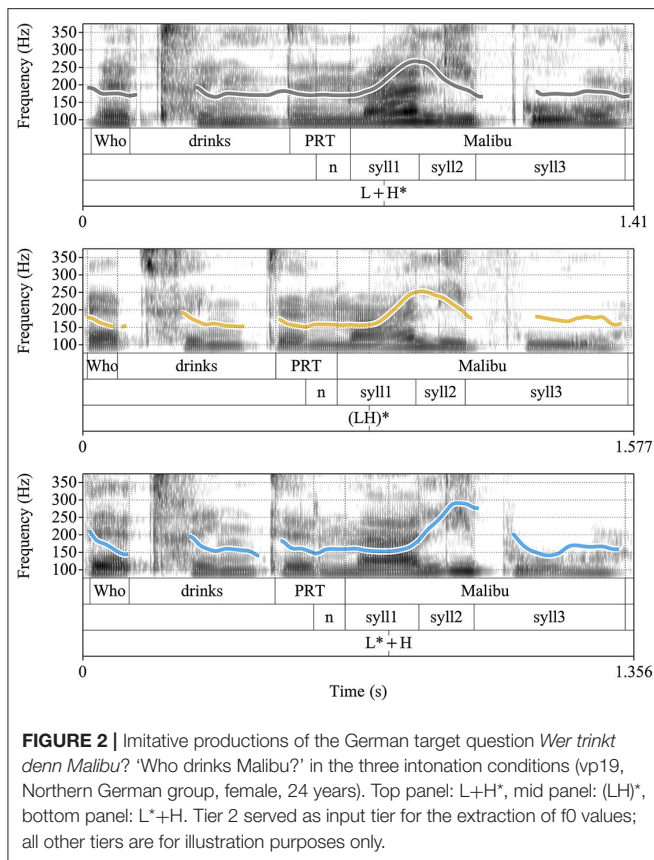
Dataset

In total, we collected 672 sound files (28 participants \times 24 imitated questions). Note that each sound file has one imitation. Nine files were excluded due to mispronunciations, bad sound quality, or a technical error on the online platform that led to data loss. The final data set for the analysis consisted of 663 sound files, see **Table 2** for a breakdown of the distribution of files across different groups.

Data Processing

Sound files were first segmented semi-automatically using the software *Web Maus* (Kisler et al., 2017) with boundaries being manually adjusted according to standard segmentation criteria (Turk et al., 2006). The critical segments were [n] in the particle *denn*, and the first, second, and third syllables of the object nouns. Subsequently, f_0 values were extracted for these four intervals using *Prosody Pro* (Xu, 2013).³ **Figure 2** shows an actual imitation of one target question in the three conditions, along with the annotation. We used 50 measurements per

³To reduce erroneous f_0 values, the Hz range for higher pitched voices (mostly persons that identified as female) was set to 100–500 Hz, while it was changed to 50–300 Hz for lower pitched voices (persons that identified as male).



time interval. To detect and remove f0-tracking errors, which primarily occurred in word-final fricatives and word-medial stops, we used a custom-made algorithm in Python that replaced likely octave jumps by “NA” so that they were excluded from the analysis. In R (R Development Core Team, 2015), raw f0 values were transformed into semitones to ease interpretation of (perceptible) differences across contours and downsampled (10 values per interval for statistical analysis).⁴

Statistical Analysis

We used GAMMs (Wood, 2006, 2017) to test the distinctiveness of the three different intonational contours. GAMMs were chosen as they allow for a direct comparison between f0 contours by modeling non-linear dependencies of a response variable (here f0) and different predictors (here *intonation condition* and *region*) over time via smooth functions. They use a pre-specified number of base functions of different shapes (Baayen et al., 2018; Wieling, 2018; van Rij et al., 2019; Sóskuthy, 2021). GAMMs also allow us to model interactions over time (e.g., *condition* × *region*), which test whether the distinctiveness of contours differs between speakers of Northern and Southern German (cf. van Rij et al., 2019, p. 8ff.; Wieling, 2018, p. 106ff.). For the

⁴For semitone conversion, the following formula was used: $st = 12 \cdot \log_2(f_0/f_{ref})$. Based on visual inspection of the distribution of f0 values per gender, f_{ref} was set to 175 Hz for higher pitched voices (mostly persons that identified as female) and 100 Hz for lower pitched voices (persons that identified as male), resulting in mostly positive semitone values (mean = 2.6 st, sd = 3.0 st).

model fitting of the GAMMs, we used the R package *mgcv* (Wood, 2011, 2017); the package *itsadug* was used to plot the model results (van Rij et al., 2017), which is essential to interpret model outputs.

The response variable was the f0 value (in st) at different time points (10 values per interval), which was roughly normally distributed. All models were corrected for autocorrelation in the f0 data using an autocorrelation parameter rho, determined by the *acf_resid()*-function from the package *itsadug* (van Rij et al., 2017).⁵ Models were initially fitted using the maximum likelihood (ML) estimation method in order to be able to compare models with different complexity (Sóskuthy, 2021, p. 16; Wieling, 2018, p. 89). We first tested whether the modeling of different curves for the three intonation conditions over time is warranted. Since this was the case, we then assessed the interaction between *intonation condition* and *region* (see below for details). Model fits were checked using *gam.check()* and the number of base functions (k) was adjusted if necessary. Also, models were re-run with the scaled *t* distribution (family = “scat”, closely following the suggestion in van Rij et al., 2019, p. 17) due to tailed residuals. All steps of the analyses can be found in the **Supplementary Materials** to this paper (<http://doi.org/10.17632/yhv7nmjmgf.2>).

Results

Figure 3 shows the raw data, i.e., the average f0 contours on time-normalized utterances (in st) of imitated productions in the different intonation conditions for Northern German (left) and Southern German speakers (right). Data from both manipulation origins (i.e., whether the contours were resynthesized from L+H* or L*+H) were collapsed since the resynthesis procedure did not affect the realization of contours (see analysis on Mendeley for details). Note that syllable durations in the imitated questions did not differ across intonation conditions (all $p > 0.12$).

The initial GAMM included *condition* and *region* as parametric effects along with a smooth for the interaction of *intonation condition* over (normalized) time, $s(\text{Normtime}, \text{by} = \text{intonation condition})$, and factor smooths for *participants* and *items*. Model comparisons using the function *compareML()* revealed that this model was superior to a simpler model without the smooth for condition over time [$\chi^2_{(4.00)} = 1323.09, p < 0.0001$], corroborating the existence of different contours. We then assessed the interaction between *intonation condition* and *region* over time to test whether the distinction of contours differed across regions. To do so, we refitted the model including an interaction variable *RegCond* (6 levels, 2 regions × 3 intonation conditions). The interaction model had a better fit [$\chi^2_{(8.00)} = 91.40, p < 0.001$], indicating that the speakers from the North made

⁵Autocorrelation can also be reduced by fitting smooths for an event variable (i.e., a unique time series for each subject on each trial), cf. van Rij et al. (2019). However, this was computationally not possible. Instead, we fitted factor smooths for subject and item and controlled for autocorrelation.

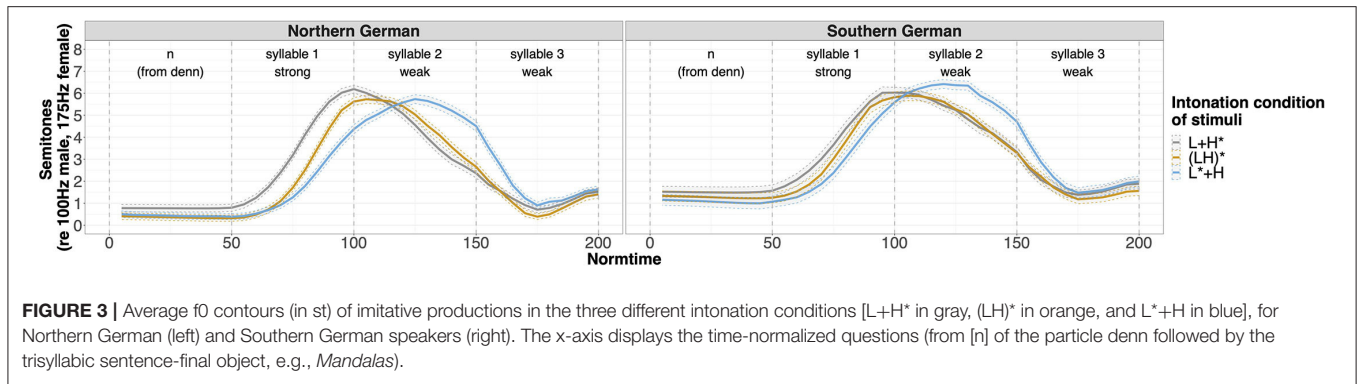


FIGURE 3 | Average f_0 contours (in st) of imitative productions in the three different intonation conditions [L+H* in gray, (LH)* in orange, and L*+H in blue], for Northern German (left) and Southern German speakers (right). The x-axis displays the time-normalized questions (from [n] of the particle *denn* followed by the trisyllabic sentence-final object, e.g., *Mandalas*).

different distinctions than speakers from the South. The best-fitting model was re-run with the scaled t distribution specified (family = “scat”). Based on this new model, we again determined a value for the rho parameter to account for autocorrelation. The outcome of this final model⁶ is visualized in **Figure 4**. It explained 68.5% of the deviance, see **Supplementary Material S2.1** for details on model evaluation in terms of residuals.⁷

Figure 4A shows the f_0 contours in the three intonation conditions as predicted by the final GAMM, split by *region* (Northern German speakers are shown in the left and Southern German speakers in the right panel). The difference between two contours is directly displayed in so called difference curves, where f_0 values in one condition are subtracted from f_0 values in the other condition, see **Figure 4B** for the three pairwise contour comparisons for the two regions:

- **Comparison L+H* vs. (LH)*.** For speakers from Northern Germany, the imitative productions significantly differed already in the pre-tonic interval (the [n] of the particle *denn*), with L+H* contours being slightly higher than (LH)* contours. L+H* furthermore had an earlier low turning point than (LH)* in Northern German speakers (i.e., an earlier start of the rise), which accounted for a large difference in the stressed syllable. For speakers from Southern Germany, by contrast, the difference in alignment of the low turning point was less obvious and contours only differed for a small part of syllable 1 of the noun. The difference was around half a

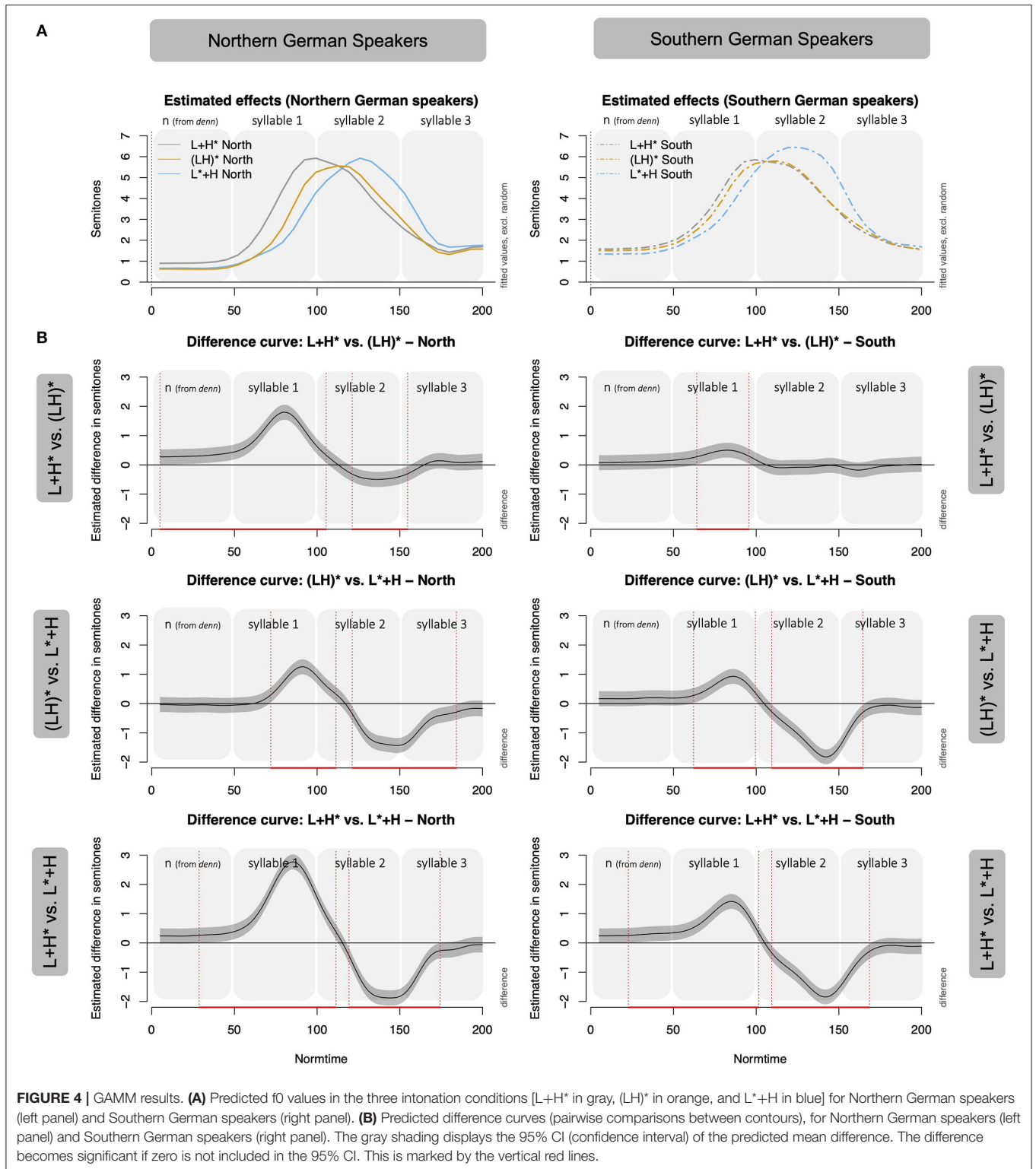
semitone, which is very subtle (Batliner, 1989). Similarly, a later alignment of the peak in (LH)* than in L+H* led to differences in the contour at the end of syllable 1 and the beginning of syllable 2 in the noun in the Northern German group. This difference was again less pronounced in the Southern German group. Overall, the distinction between L+H* and (LH)* seems to be larger for Northern German than for Southern German speakers (see **Supplementary Material S.2.2** for interaction model). Importantly though, both speaker groups distinguished between the two contours.

- **Comparison (LH)* vs. L*+H.** For both speaker groups, the two contours differed significantly in syllables 1 and 2 of the noun. The alignment of the low turning point was comparable but the (LH)* had a steeper rise with an earlier peak than the L*+H accent, which led to considerable differences in the stressed syllable and the post-tonic syllable. **Supplementary Material S.2.2** show that regional differences are minor for this distinction.
- **Comparison L+H* vs. L*+H.** As expected, the imitative productions of the two established accents differed significantly in syllables 1 and 2 of the noun, with a later alignment of the low turning point and the peak for L*+H as compared to L+H*. The f_0 of the L+H* hence rose earlier than for L*+H, leading to strong differences in realization. Interestingly, the contours deviated already in the pre-tonic syllable ([n] of the particle *denn*), similarly in both varieties. The distinction between contours was generally more pronounced for Northern German speakers than for Southern German speakers (see **Supplementary Material S.2.2** for interaction model).

Taken together, the two more established accent types L+H* vs. L*+H are clearly distinct in their **form**, showing significant differences across both syllable 1 and 2 in both regional varieties (the difference was larger for Northern German speakers). Crucially though, the f_0 contour in the (LH)* condition is also significantly different from the f_0 contour in L+H* (syllable 1) and from the f_0 contour in L*+H (syllables 1 and 2)—for both speaker groups, but the distinction between L+H* and (LH)* is smaller for speakers from the South than for speakers from the North; in fact, these two contours are very similar in the productions of Southern German participants. The distinction between L*+H and (LH)* is more pronounced

⁶The final model was specified as follows: `model <- bam(st ~ Regcond + s(Normtime, by = Regcond, k = 20) + s(Normtime, vp_index, bs = “fs”, m = 1) + s(Normtime, manip_item, bs = “fs”, m = 1), data = delayed_imitation_sorted, family = “scat”, discrete = T, method = “fREML”, rho = rhoval, AR.start = delayed_imitation_sorted$start_event)`.

⁷To corroborate the interaction between *condition* and *region* indicated in our final model, we constructed additional models containing binary difference smooths terms that capture the difference of the difference over time between two predictors, and hence their interaction (closely following the procedure described in van Rij et al., 2019, pp. 11–13; Wieling, 2018, p. 109 ff.), see **Supplementary Material S.2.2** for details. Again, these binary difference smooths support the interpretation that the difference between intonation conditions over time is different for speakers from Northern vs. Southern Germany but also, that speakers from both regions produced three distinct contours, **Supplementary Material S.2.2**.



and clearly maintained in both regions, with differences in contours occurring both in syllable 1 and 2 of the object noun. The significant differences in contours across all pairwise

comparisons suggest that speakers maintain a three-way contrast in rising-falling-contours, with regional variety modulating the extent of the distinction.

Interim Discussion

Experiment 1 tested the distinctiveness in realization of three rising-falling contours in Northern and Southern German speakers in a delayed imitation task that addressed phonological processing (Chodroff and Cole, 2019b; Petrone et al., 2021). Our findings show that all three contours are distinguished from each other. However, it is not unambiguously clear at this point whether participants imitated the tonal targets (i.e., pitch accent categories), a communicative function associated with the different contours (e.g., information-seeking vs. rhetorical question), or differences in perceived prominence. In prominence perception tasks, steeper slopes are judged more prominent than shallower ones (Rietveld and Gussenhoven, 1985; Baumann and Röhr, 2015; Baumann and Winter, 2018). Clearly, imitative productions had to be retrieved from stored representations, which may consist of aspects of tonal alignment properties, prominence, and meaning—possibly also a combination of the three.

Regional variety mediated the extent of the distinction between contours such that speakers from Northern Germany had more distinct productions than speakers from Southern Germany, especially regarding the distinction between L+H* vs. (LH)*. In section “The Present Study: Rationale and Hypotheses”, we hypothesized about mergers toward the more frequent accent type, i.e., mergers toward L+H* in Northern German and toward L*+H in Southern German speakers due to a high occurrence frequency of these complementary accents in the respective varieties (Kügler, 2004, 2007; Peters et al., 2005)—an account that had predicted less clear-cut distinctions between L+H* and (LH)* in the North and between L*+H and (LH)* in the South. This prediction was clearly not borne out: Instead, Northern German speakers were more distinct in all pairwise comparisons than Southern German speakers, especially with regard to the distinction L+H* and (LH)*, which almost seemed to converge for large parts of the contours in speakers from the South. We will discuss this finding and its implications in more detail in the General Discussion, including results from the association task in Experiment 2. Summarizing the main findings from Experiment 1, our results indicate that speakers maintain a three-way contrast in their imitative productions. Crucially, (LH)* significantly differs in its **form** from the more established accents L+H* and L*+H in all experimental conditions, in particular for speakers from the North.

EXPERIMENT 2: PARAPHRASING OF CONNOTATIVE QUESTION MEANING

In Experiment 2, we tested whether the three rising-falling contours evoke different connotative meanings. To this end, listeners paraphrased the pragmatic meaning they associated with the stimuli from Experiment 1.

Methods

Participants

Overall, 66 native speakers of German were included in the study. None of the speakers had participated in Experiment 1.

Twenty-eight of them were from Southern Germany (mean age: 24.1 years, SD = 4.3 years, 24 female, 4 male), that is, Baden-Wuerttemberg ($N = 20$) and Bavaria ($N = 8$), and 38 were from Northern Germany (mean age: 25.5 years, SD = 8.2 years, 33 female, 5 male), which includes the states of Saxony-Anhalt (1), Lower-Saxony ($N = 18$), Hamburg ($N = 3$), Bremen ($N = 1$), North-Rhine Westphalia ($N = 5$), and Schleswig-Holstein ($N = 10$). Data from five additional speakers was not considered since these participants could not unambiguously be assigned to the Southern or the Northern group [i.e., participants who were born in the North but grew up in the South, or vice versa ($N = 2$), or came from the Central German dialect area ($N = 3$)].

Materials

We selected 12 *wh*-interrogatives from the material set used in Experiment 1. Since the direction of resynthesis of stimuli (manipulated from L+H* vs. L*+H) did not have an effect on the imitation results in Experiment 1 (see Mendeley), we reduced the number of stimuli by taking only one manipulation direction into account. That is, for L+H* we used those stimuli that were originally recorded as L+H*; similarly, for L*+H we used those stimuli that were originally recorded as L*+H. For (LH)* contours, which were resynthesized half from originally recorded L+H* and half from L*+H in Experiment 1, we chose two of the four questions to be originally recorded as L+H*, and two as L*+H contours. This resulted in 12 items (4 items \times 3 intonation conditions, only one manipulation direction). The 12 *wh*-interrogatives were ordered such that the stimuli with the same intonation condition and stimuli with the same lexicalization were separated by at least one other item to avoid priming.

Procedure

Participants were asked to paraphrase the intention they thought a speaker conveyed in the question. They were told that interrogative sentences may not only be used for inquiring information but can also serve other purposes (which were not further specified). Participants were furthermore explicitly instructed to focus on *how* the respective utterances sounded, that is, which connotative meaning the utterances expressed, disregarding their propositional content. The four different target *wh*-questions [cf. (1) above] were presented in written form in the instructions. This was done to familiarize participants with the syntactic and lexical composition of the target sentences and to focus them on the intonational realization of the utterances.

On each trial in the actual experiment, participants clicked on a “play” button to listen to one *wh*-question at a time. They then had to paraphrase the intention of the speaker in a free response field. The experiment was self-paced and participants were allowed to listen to a question as often as they wanted but were instructed to respond intuitively. In case they associated different intentions with the questions, they were allowed to give multiple responses; conversely, if participants did not identify an intention, they typed “NA” in the description field or left it blank. Participants moved on to the next trial by pressing a “continue” button. The study was conducted as a web-based experiment, which was created via *SoSci Survey*

TABLE 3 | Number of responses split by experimental item and intonation condition.

	Libero	Malibu	Mandalas	Melanie	Total
L+H*	59 (50)	66 (54)	55 (42)	81 (78)	261 (224)
(LH)*	75 (73)	88 (84)	71 (69)	72 (68)	306 (294)
L*+H	58 (49)	52 (47)	71 (57)	60 (54)	241 (207)
Total	192 (172)	206 (185)	197 (168)	213 (200)	808 (725)

Values in brackets indicate the number of instances that were statistically analyzed (after response categories with fewer than 12 instances per category had been removed).

(www.socscisurvey.com, Leiner, 2018), and ran on an in-house server. Participants took between 10 and 15 min to complete the study.

Data Treatment and Analysis

In total, 808 responses were given [261 for L+H*, 306 for (LH)* and 241 for L*+H], see **Table 3** for the distribution of responses across items and intonation condition. Listeners gave on average 12.2 responses for the 12 trials (SD = 3.7), with individuals ranging between 3 (i.e., responses to only a fourth or the trials) and 21 responses (i.e., almost two responses to every trial). There were 135 questions to which participants did not provide a response at all [50 times for L+H*, 29 times for (LH)*, and 56 times for L*+H].

We extracted superordinate categories from participants' responses. To this end, we took a sample of about a third of the data ($N = 262$) and grouped the responses into a set of superordinate categories (e.g., "information-seeking," "surprise," "p is odd"). These categories were generated bottom-up (i.e., data-driven) and were often explicitly mentioned by a number of participants (e.g., scepticism). The grouping was done together by two coders (a consensual coding between first and third author). **Table 4** lists the nine superordinate categories that contained $N \geq 12$ instances each. It also includes example responses for each category. To verify the objectivity of the superordinate categories and the reliability with which they can be coded, a third coder (fourth author) independently coded a subset of 220 responses based on the keywords in **Table 4**. Agreement between the third coder and the consensual coding of the first two coders was assessed by calculating Cohen's kappa (Cohen, 1960) with the *irr* package in R. The interrater agreement was 88.2% ($\kappa = 0.83$), i.e., "almost perfect" (Landis and Koch, 1977, p. 165). The set of keywords was then used to code the remaining items by one of the three coders (first, third, and fourth author).

In total, responses fell into 27 superordinate categories, with instances in individual categories ranging between 1 and 370 responses. To keep the number of categories feasible for statistical analysis, we excluded categories with fewer than 12 instances per category, in total excluding 83 responses in 18 different categories (10.3% of the data). The statistical analysis was based on the nine different response categories of **Table 4** ($N = 725$ responses), see values in brackets in **Table 3** for distribution across condition and items.

TABLE 4 | Nine most frequent superordinate categories inferred from participants' responses.

Superordinate category	Exemplar responses
Aversion ($N = 77$)	<i>Abschätzigte Meinung zu Mandalas;</i> ("pejorative opinion on mandalas") <i>Auf Mandalas malen als Beschäftigung wird herabgesehen;</i> ("disdaining coloring mandalas as an activity")
Information-seeking ($N = 370$)	<i>Wer malt gerade ein Mandala?;</i> ("Who is coloring a mandala at the moment?") <i>Herausfinden, wer Melanie heißt;</i> ("find out which of the persons is called Melanie") <i>Tatsächliches Interesse;</i> ("actual interest")
Irony ($N = 17$)	<i>Ironische Frage;</i> ("ironic question") <i>Ablehnung/ Spott ausdrücken;</i> ("to express rejection and mockery")
Negative attitude ($N = 54$)	<i>Melanie ist kein schöner Name;</i> ("Melanie is not a nice name.") <i>Kritische Äußerung zu Mandalas;</i> ("critical statement toward Mandalas")
P is odd ($N = 53$)	<i>Melanie ist ein ungewöhnlicher Name;</i> ("Melanie is an unusual name.") <i>Mandalas malen ist ungewöhnlich;</i> ("Drawing mandalas is unusual.") <i>Dass Leute, die Mandalas malen, komisch sind;</i> ("That people who draw mandalas are weird.")
Positive attitude ($N = 13$)	<i>Bewusst geduldig und freundlich auftreten;</i> ("to intentionally appear patient and friendly"), <i>Malibu wird positiv bewertet;</i> ("Malibu is rated positively.")
Rhetorical meaning ($N = 47$)	<i>Niemand heißt Melanie;</i> ("Nobody is called Melanie.") <i>Melanie als Name wird infrage gestellt. Wer heißt denn schon so?;</i> ("The name Melanie itself is questioned. Who is called Melanie?") <i>Rhetorische Frage;</i> ("rhetorical question")
Scepticism ($N = 12$)	<i>Zweifel;</i> ("doubt") <i>Skepsis;</i> ("scepticism")
Surprise ($N = 82$)	<i>Libero spielen ist etwas, das man nicht erwarten würde;</i> ("To play in the sweeper position is not something one would expect.") <i>Verwunderung;</i> ("astonishment/surprise")

Original responses for the superordinate category are shown on the right. Categories are presented in alphabetical order. The number in brackets gives the total number of responses in this category.

We used Conditional Inference Trees (CTrees) to test whether there was a significant clustering of response categories based on our two predictors *intonation condition* and *region*. CTrees are a non-parametric class of regression trees, applicable to all kinds of response variables (Hothorn et al., 2006). Different

from other regression tree algorithms such as CART-based trees, CTree employ a significance test procedure that grows only statistically significant splits. Hence, tree pruning is not needed in this approach [ctree() function description, Hothorn and Zeileis, 2015]. To fit the trees, we used the *party* package in R (Hothorn et al., 2006; Hothorn and Zeileis, 2015). To evaluate the generalization of the CTree, we used a 10-fold cross validation procedure: We split the data in 10 randomly sampled sets, training the tree on 85% of the data and testing it on the 15% of unseen data. For evaluation of the tree, the R package *caret* (Kuhn, 2020) was used.

Results

Figure 5 shows the distribution of the nine response categories across the three different intonation conditions, split by region ($N = 725$ responses).

Most of the questions were paraphrased as “information-seeking” ($N = 370$ of all 808 responses, 45.8%), which is not unexpected given their interrogative syntax. However, the specification “information-seeking” was more often ascribed with L+H* and L*+H accents, as compared to (LH)* accents: 59.4% of the L+H* accents, 59.8% of the L*+H accents, compared to 23.2% of the (LH)* accents. A similar distribution, but with much lower numbers, was found for the category “positive attitude.” Conversely, all other response categories occurred more frequently in the (LH)* condition than in the L+H* and L*+H accents. That is, (LH)* was often paraphrased as “aversion,” “surprise,” “negative attitude,” or “p is odd”. A “rhetorical meaning” was also attributed to (LH)*, more often than for the two other accent types. However, this connotation was rare overall. The CTree shows one significant split only (**Figure 6**), which is caused by intonation condition, separating the (LH)* accent on the one hand from the L+H* and L*+H on the other. The L+H* and L*+H accents were not further subdivided. The factor *region* was not considered by the CTree, which mirrors the similar meaning attributions across regions shown in **Figure 5**. The evaluation of the unseen test set (15% of the data) revealed a mean accuracy of 51.8%, 95% CI [42.1%; 61.5%].⁸

It was surprising that the two established GToBI accents L+H* and L*+H were not distinguished in the meaning task. After all, these contours have often been claimed to differ in their communicative **function**: L+H* has been associated with new or contrastive information (Kohler, 1991b; Grice et al., 2005; Baumann and Grice, 2006), while L*+H has been associated with contrast and/or certain attitudinal meanings (Grice et al., 2005; Niebuhr, 2007b; Lommel and Michalsky, 2017). These meaning attributions refer to information structure, the information status of referents and to attitudes that are typically conveyed in utterances with a declarative syntax. It is conceivable that these intonational meanings are less obvious in the *wh*-question structure employed in our experimental

sentences. An alternative explanation could be that the meaning contrast was not captured well by the superordinate categories. To follow up on these possibilities, we conducted a *post-hoc* study ($N = 15$ participants, 5 from Northern and 10 from Southern Germany) in which we used the same recordings of the target words (*Mandala*, *Malibu*, *Melanie*, and *Liberio*), but spliced onto a declarative-sentence structure (*Das ist der/die* “That is the”).⁹ The instructions and the experimental procedure were the same as in Experiment 2. Participants’ responses in this follow-up study were coded into keywords (which partly differed from the ones for *wh*-questions) and analyzed by the third and fourth author in a consensus coding. In this follow-up study, results revealed differences in interpretation between L+H* and L*+H: L+H* was more often paraphrased as “correction”, “enforcement”, “statement”, “p is new”, and “information-giving” than L*+H. Conversely, L*+H was more often paraphrased as “surprise” and “aversion” than L+H*. In line with the results of Experiment 2, (LH)* was interpreted more often as “correction”, “surprise”, and “aversion” than the other two accents (see **Supplementary Material S.3** for more details).

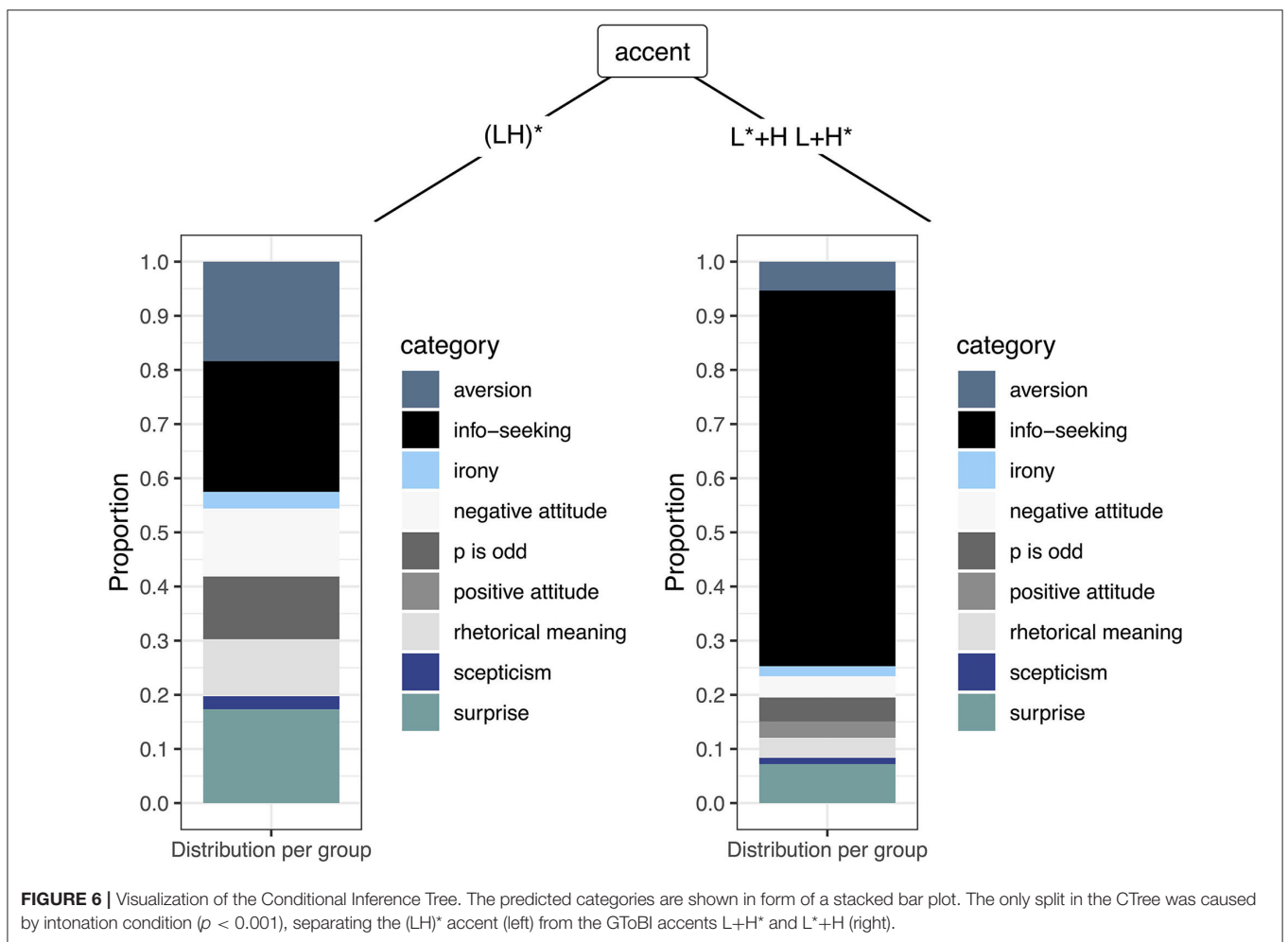
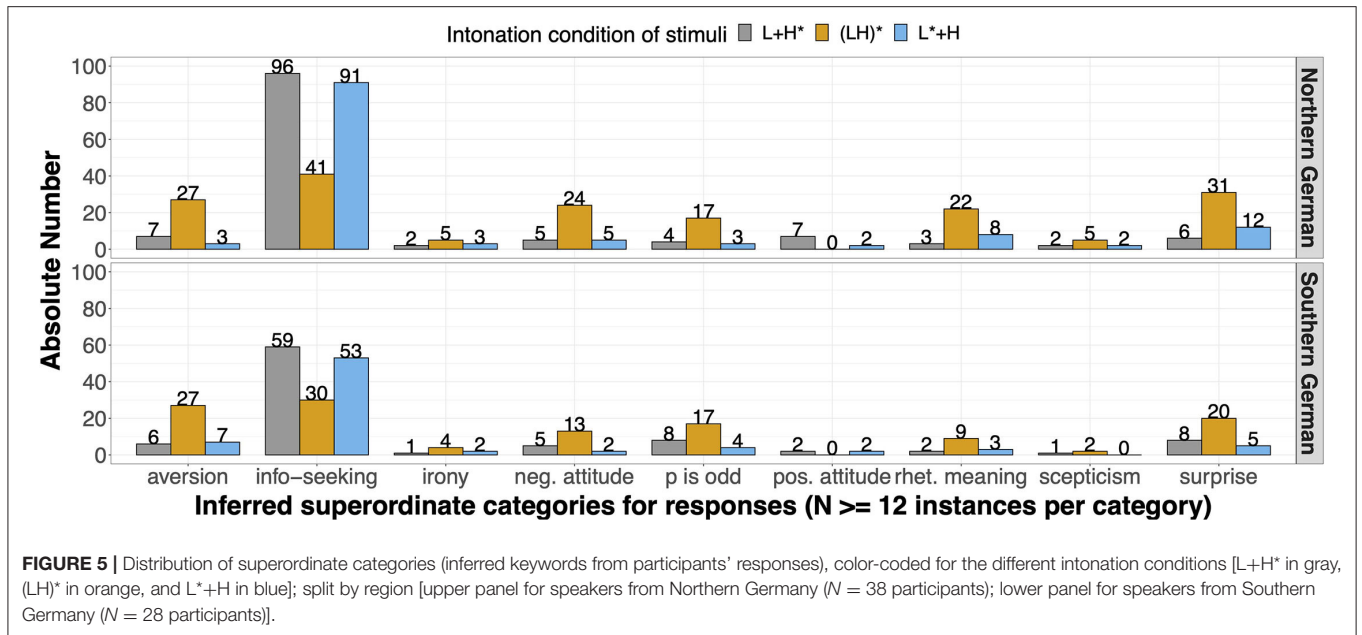
Interim Discussion

In Experiment 2, we assessed the connotative meanings listeners associate with the three different kinds of rising-falling contours, L+H*, (LH)*, and L*+H, schematized in **Figure 1**. This was done in a qualitative study in which participants freely paraphrased the perceived intention of the speaker. Our results showed that *intonation condition* clearly affects the connotative meanings associated with the questions, causing the only split in the CTree (see **Figure 6**). Importantly for the question of whether (LH)* forms its own category, the connotative meanings evoked by (LH)* were distinct from the two other accent types: While L+H* and L*+H contours were equally paraphrased as “information-seeking” in most of the *wh*-questions, (LH)* received more diverse meaning attributions, which were often paraphrased as “aversion”, “surprise”, “negative attitude”, or “p is odd”. An explicit “rhetorical” meaning was also ascribed to (LH)*, but this association was comparatively rare. Importantly, the pattern of results was the same across regions, suggesting that speakers from Northern and Southern Germany share the same set of connotative meanings for the three rising-falling contours.

We first discuss the connotations ascribed to the (LH)* accent before we turn to the finding that the meaning attributed to L+H* and L*+H did not differ in *wh*-questions. From a phonetic point of view, (LH)* differs from the other two accents in our study in that it exhibits a steeper slope of the rising movement since both the low and the high tonal target occur within the stressed syllable. This might have increased the perceptual salience of this accent type. As discussed briefly in section “Interim Discussion” of Experiment 1, previous prominence rating tasks have shown that rising nuclear accents are perceived

⁸Note that reducing the number of categories (excluding those with only few instances: irony, positive attitude and scepticism) led to the same results in the CTree, i.e., one single split grouping L+H*/L*+H on the one hand, and (LH)* on the other.

⁹For the item “Mandalas” we had to remove the final [s] to change it from a plural to a singular word form to match it to the referential expression *das* and the verb *ist*.



as more prominent than falling ones, H* accents as more prominent than L* accents, steeper slopes as more prominent than shallower ones, and larger f0 excursions as more prominent than smaller f0 excursions (Rietveld and Gussenhoven, 1985; Baumann and Röhr, 2015; Baumann and Winter, 2018). This implies the following decreasing prominence order among the accents of this study: (LH)* > L+H* > L*+H. This ranking is reproduced in only three of the nine categories (“p is odd,” “negative attitude,” “aversion”). Hence, while the perceptual prominence may have affected listeners’ interpretations to some degree, differences in prominence alone cannot explain the findings. Clearly though, the steeper slope perceptually stands out. Higher peaks and concomitant steeper slopes have been shown to affect meaning interpretation: In particular, a steeper slope in rising movements has been associated with surprise in the literature (Ladd and Morton, 1997; Chen, 2009).

To further interpret the findings concerning the meaning attributions to (LH)*, it helps to access the pragmatics literature: As mentioned earlier, the accent (LH)* has been observed in a study on rhetorical questions (Braun et al., 2019). Several connotative meanings mentioned by the listeners are in fact compatible with a rhetorical question interpretation: Rhetorical questions are often described to have the illocutionary force of assertions (Han, 2002) or to be assertion-like (Caponigro and Sprouse, 2007; Biezma and Rawlins, 2017). The speaker of a rhetorical question commits her interlocutors to the proposition presupposed by the rhetorical question (Biezma and Rawlins, 2017). For positive *wh*-questions used in this paper, the presupposition denotes the empty set (e.g., *niemand* “nobody” for *Wer mag Mandalas?* “Who likes mandalas?”). At the same time, the speaker of a rhetorical question signals that the answer to the rhetorical question is obvious and she expects all interlocutors to know that it is obvious. A rhetorical question is not, a priori, connected to any specific kind of speaker emotion. However, it can convey a large range of emotional or attitudinal load. It may be used positively (e.g., *Mach dir keine Sorgen. Wer ist denn nicht nervös vor einer Prüfung?* “Don’t worry. Who isn’t nervous right before an exam?”) or negatively (e.g., *Was weiß der schon?* “What does he know, after all?”). Hence, it is conceivable to assume that a rhetorical question may also trigger emotional stances such as aversion or negativity or a certain sense of surprise or oddness. The indeterminacy in attitudes also explains the larger variability in paraphrases in the (LH)* accent compared to the L+H* and L*+H accents. A frequent category for the (LH)* condition was “surprise” (e.g., that the speaker is surprised that there are people who like drawing mandalas). The surprise aspect ties in with observations that rhetorical questions may be marked by “mirativity markers” in some languages (for Basque: Alcázar, 2017). Also, rhetorical questions in English have been associated with surprise (Celle, 2018). Taken together, both the phonetic composition of (LH)* as well as pragmatic approaches explain why (LH)* evokes a different meaning than the two other accents. We now turn to the lack of distinction between L+H* and L*+H.

Contrary to our hypothesis, L+H* and L*+H contours in *wh*-interrogative structures did not lead to overall different judgments of meaning but were both predominantly interpreted

as conveying the intent of requesting information from the addressee (“information-seeking”). The syntactic question form, i.e., the *wh*-verb-second-interrogative form, may have inflicted a strong bias, in particular for the two established accents L+H* and L*+H. Given ample evidence on different meaning contributions for L+H* and L*+H (Grice et al., 2005; Kohler, 2005; Niebuhr, 2007b; Lommel and Michalsky, 2017; Braun and Biezma, 2019), the lack of a distinction is indeed surprising, and may theoretically cast doubt on the validity of the study. This is not the case, however: A follow-up study conducted with the same design but with declarative sentences instead of *wh*-interrogatives showed meaning differences between L+H* and L*+H. Our findings corroborate current reports in the literature that challenge the view of a one-to-one mapping between intonational form and pragmatic meaning (Chodroff and Cole, 2019a; Roettger et al., 2019; Orrico and D’Imperio, 2020), but clearly show that differences in tonal alignment interact with propositional content and sentence structure to evoke different meaning interpretations.

Taken together, (LH)* differed in its interpretation from the GToBI accents L+H* and L*+H. The data may be partly explained in terms of a link between phonetic emphasis (steepness of the slope of the rise) and surprise. From a functional perspective, there is evidence that the (LH)* accent is interpreted differently from the L+H* and L*+H accents and that the latter two, regarding their meaning, do not differ in *wh*-questions—but, corroborating previous research, in declaratives.

GENERAL DISCUSSION

In the present study, we examined the distinctiveness in **form** and **function** of German nuclear rising-falling intonation contours. We focused on three rising-falling contours, which have been described in several different intonational frameworks and empirical studies on German intonation. Only one system previously discussed a three-way contrast (Kohler, 2005; Niebuhr, 2022); the other systems mostly provide a two-way distinction (Kohler, 1991a; Grice et al., 2005; Peters, 2006, 2014). The contours investigated in the present study are termed L+H* (H aligned in stressed syllable, L in preceding syllable), L*+H (L aligned in stressed syllable, H in following syllable), and (LH)* (both L and H aligned in the stressed syllable). In Experiment 1, we tested whether German speakers are able to imitate these three distinct rising-falling contours using a delayed imitation task, which taps into phonological processing (Baddeley and Hitch, 1974). Imitative productions were complemented by a qualitative study on the perceived intention of the speaker (Experiment 2). In both paradigms (**form** and **function**), we further investigated whether the regional background of the participants (Northern vs. Southern Germany) affects the ability to distinguish between the three contours in production and perception. The factor regional variety suggests itself (a) because Southern and Northern German speakers were found to align tonal targets differently in prenuclear rising accents—with later tonal targets in the South than the North (e.g., Atterer and Ladd, 2004) and (b) because

Northern German speakers use L+H* as most frequent accent type while Southern German speakers predominantly use L*+H (Peters et al., 2005; Kügler, 2007).

With respect to **form**, our results show that speakers of both varieties produced three distinct rising-falling contours. While imitated contours differed significantly in all experimental conditions, they were more distinct in Northern German speakers than in Southern German speakers, especially with regard to the contrast between L+H* and (LH)*. In none of the varieties did the contours totally converge onto a single or two contours. Based on the previous literature, one may have expected that frequent categories act as attractors (Anderson et al., 2003; Braun et al., 2006; Chodroff and Cole, 2019b). In this study, L+H* is the most frequent accent in Northern German (Peters et al., 2005), while L*+H is frequently used in Southern German (Kügler, 2007; Truckenbrodt, 2007). An attractor account would have predicted more mergers toward L+H* in Northern German and more mergers toward L*+H in Southern German (cf. Braun et al., 2006). However, we do not see evidence that one of these frequent accents acts as a perceptual magnet that is able to warp the perceptual space. It rather seems that even less frequent accent types can easily be held in memory and retrieved for production. It is likely that adult speakers of both regions have accumulated enough experience with different pitch accents, which allowed them to form the respective three categories (cf. Zahner et al., 2016, showing that already children who grow up in Southern Germany are exposed to the full German accent inventory from early on). Our data hence call for a three-way distinction between rising-falling contours for both regional varieties.

Let us nevertheless briefly speculate about the smaller distinction between L+H* and (LH)* contours for Southern German as compared to Northern German speakers. Since these two regional varieties have been shown to differ in the alignment of tonal targets, one may assume that this tendency is the cause for the small difference between L+H* and (LH)* in Southern German speakers. However, if tonal alignment differences uniformly applied to all accents in this speaker group, we would not have observed differences in the distinction of the contours across regions (but a main effect of region, later alignment in Southern than in Northern German speakers throughout). We argue that the contextually more restricted (LH)* accent is also more restricted in terms of its realization and needs to have both tonal targets realized within the stressed syllable. This requirement does not allow for variety-specific alignment differences. If (LH)* has fixed alignment (L toward the middle and H at the end of the stressed syllable), then a later aligned L+H* for Southern German speakers may lead to an overlap in production with (LH)*. The most important finding we take from Experiment 1 is that all contours significantly differed from each other, suggesting that the German acoustic space of rising-falling contours may be best described as a three-fold partition.

The presence of three distinct contours is further corroborated by the data from Experiment 2 (**function**). Listeners associated different meanings with (LH)* on the one hand, and with L+H* and L*+H on the other. We argued that the steep slope of

the rise in (LH)* (both L and H in the stressed syllable with the same f₀ excursion as L+H* and L*+H) may have evoked the perception of surprise (cf. Kohler, 2005; Chen, 2009). It is also conceivable that the increased prominence of this accent (Baumann and Röhr, 2015; Baumann and Winter, 2018) triggers implicatures through the effort code, such that an increased effort signals pragmatic relevance (cf. Hirschberg, 2002 on modeling intonational meaning in terms of implicatures). The attitudinally loaded meanings (aversion, negative attitude, unexpectedness, etc.) furthermore have been argued to be compatible with the pragmatics of rhetorical questions. In contrast, L+H* and L*+H were frequently paraphrased as information-seeking, but with no further differences in meaning, which was unexpected but seems to be due to the use of *wh*-questions (as differences were found when using declarative sentences). This, in turn, suggests that intonational meaning cannot entirely be dissociated from sentence type. In sum, we take the combined results from imitation and perception experiments as evidence for three types of distinct rising-falling contours in German.

This brings us to the question of whether and how we need to model three distinct contours in the acoustic space of German rising-falling contours. The present paper supports models that contain a three-way contrast (Kohler, 2005; Niebuhr, 2022) by providing combined evidence from production (**intonational form**) and perception (**function**). Our results show that (LH)* is distinct in **form** and **function**. For answering the question on the number of contours we need to model, it may be helpful to keep an open eye for the (LH)* pitch accent in future transcriptions of German intonation to learn more about its distribution and the (phonetic, phonological, syntactic, or pragmatic) conditions in which it occurs. As the meaning data suggest, its use is likely not restricted to German rhetorical questions (cf. Kohler, 2005; Braun et al., 2019; Wochner, 2021). Pilot data of a study on the realization of sarcastic irony suggest that this type of accent, (LH)*, also frequently occurs in sarcastic utterances of the sort *Das klappt ja super* “That works PRT great,” accent on *super* (Fünfgeld et al., 2021). As those utterances are clearly attitudinally loaded, it is not surprising that speakers also employ (LH)* in ironic situations. Also, recent data suggest that (LH)* may also be found in exclamative sentences, e.g., *Kann die Lene malen* “Can Lene paint!” (Wochner and Dehé, 2018; Wochner, 2021), which have been described to express an attitude of surprise in the sense that a speaker conveys that the proposition of an utterance is unexpected (e.g., Fries, 1988). As the current experiment investigated solely *wh*-questions with a very homogeneous structure, more research needs to be done to answer the question of whether or not the (LH)* constitutes its own phonological category or not. With the current knowledge, it seems justifiable to posit three kinds of rising-falling accents in German.

An alternative view is that (LH)* is a meaningful modification of the more established accents L+H* or L*+H. The small differences between L+H* and (LH)* for Southern German speakers, in particular, might in fact allow such an interpretation. The idea of phonetic modifications of pitch accent types is not new (Ladd and Morton, 1997; Gussenhoven, 2004; Ladd, 2008, p. 155f). Other studies have also called for a gradual mapping

between intonation and meaning, cf. Chen and Gussenhoven (2008) on aspects of gradience in the encoding of emphasis in a tone language, Orrico and D'Imperio (2020) on gradience in intonation-meaning mapping in biased questions or Dorokhova and D'Imperio (2019) on gradience in the interpretation of final rises in French. In terms of modeling, Ladd and Morton (1997) labeled rising-falling accents that differ regarding their pitch range with a binary feature [\pm emphatic]. The difference between “normal” and “emphatic” accents may be interpreted categorically (e.g., to signal uncertainty or incredulity) but not perceived as categorically different kinds of pitch accents (Ladd and Morton, 1997, p. 339). For our purposes, the feature [\pm emphatic] would serve the purpose to single out the (LH)* accent from the other two accents. Alternatively, acoustic features, such as [+ steep slope] could be used for achieving the (LH)* from an L*+H or L+H* accent. This proposal would gain support if we found this kind of modification, ideally with similar contributions to meaning, also for other accent types or contexts in German. Future research needs to test this proposal.

Taken together, our study addressed the question of how many rising-falling contours are needed to best describe the German perceptual space, thus serving as an attempt to clear up the “fuzzy” space for these contours. We hence addressed the distinctiveness in **form** and **function** of three different contours, two widely established accents L+H* vs. L*+H, and an accent which lies in-between, (LH)*. Our data show that speakers can differentiate between these three contours, both in perception and production, suggesting the existence of three kinds of rises. The most prototypical connotative meaning of the (LH)* accent is “surprise,” which was frequently mentioned in *wh*-questions and declaratives for this accentual realization. Generally, the intonational contrasts are employed and interpreted in a consistent and meaningful way. Advocating (LH)* as a third category (beyond the two more established ones) might appear premature, as the same effect may be achieved by employing modifications of the other two accents. However, as it stands, it is difficult to decide whether (LH)* would be a variant of L*+H or of L+H*. If anything, the data from Southern German speakers, in particular, suggest that (LH)* is more likely to be a variant of L+H* than of L*+H. In future research, we plan to present contours that are more variable in the phonetic space (with more phonetic variability in the alignment and shape of accents and the steepness of the slope) to better map the “white spots” between contours. We further plan to test a wider variety of accentual realizations and sentence types to gain more insights into a broader range of contexts.

DATA AVAILABILITY STATEMENT

The datasets analyzed for this study, along with analyses scripts, can be found in the Mendeley repository: <http://dx.doi.org/10.17632/yhvn7nmjmgf.2>.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by IRB Konstanz, 05/2021. The participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

KZ-R and ME processed and annotated the data for Experiment 1. KZ-R, DW, and AJ for Experiment 2. KZ-R led stimulus preparation (resynthesis), data analysis (annotation and processing), and statistical analyses (supported by BB). KZ-R drafted the manuscript. All authors contributed to the idea of using a combined approach (form-function) to study the distinctiveness of rising-falling contours and collectively developed the study designs. All authors contributed to the article, edited and wrote parts of the manuscript, and approved the submitted version and are responsible for it.

FUNDING

This research was funded by the DFG as part of research unit Questions at the Interfaces (FOR 2111, P6), Grant Numbers BR 3428/4-1,2 and DE 876/3-1,2.

ACKNOWLEDGMENTS

We are very grateful to the speaker of our experimental stimuli, Clara Huttenlauch. We also thank Johannes Heim, Jan Michalsky, Andrea Pešková, and Meg Zellers for distributing the link to the online experiment in Northern Germany, and Massimiliano Canzi and Kajsa Djärv for advice on the Conditional Inference Trees. We thank Moritz Jakob for testing (in the lab) and the implementation of the imitation study on SoSci Survey, Friederike Hohl for data processing with MAUS and the generation of the algorithm for f0 correction, and Justin Hofenbitzer, Katharina Hölzl, Elena Schweitzer, Friederike Hohl, Meike Rommel, Johanna Schnell, Meiqiao Li, and Svenja Willenborg for help with segmental annotation. We are very grateful to Cesko Voeten for practical advice on GAMM modeling, and to Melanie Loes and Vivien Meyer for help with checking the bibliography. We are grateful to the audiences at LabPhon2020 (Satellite Workshop “Situating phonological contrast within the production-perception loop”) in Vancouver (virtual conference) for discussion and feedback on preliminary results of the imitation study.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fcomm.2022.838955/full#supplementary-material>

REFERENCES

- Alcázar, A. (2017). "A syntactic analysis of rhetorical questions," in *Proceedings of the 34th West Coast Conference on Formal Linguistics* (Somerville, MA), 32–41.
- Anderson, J. L., Morgan, J. L., and White, K. S. (2003). A statistical basis for speech sound discrimination. *Lang. Speech* 46, 155–182. doi: 10.1177/00238309030460020601
- Arvaniti, A. (2019). "Crosslinguistic variation, phonetic variability, and the formation of categories in intonation," in *Proceedings of the 19th International Congress of Phonetic Sciences (ICPhS)* (Melbourne, VIC), 1–6.
- Arvaniti, A., and Fletcher, J. (2020). "The autosegmental-metrical theory of intonational phonology," in *The Oxford Handbook of Language Prosody*, eds C. Gussenhoven and A. Chen (Oxford: Oxford University Press), 78–95.
- Atkinson, J. E. (1976). Inter- and intraspeaker variability in fundamental voice frequency. *J. Acoust. Soc. Am.* 60, 440–445. doi: 10.1121/1.381101
- Atterer, M., and Ladd, D. R. (2004). On the phonetics and phonology of "segmental anchoring" of F0: evidence from German. *J. Phon.* 32, 177–197. doi: 10.1016/S0095-4470(03)00039-1
- Baayen, R. H., van Rij, J., de Cat, C., and Wood, S. N. (2018). "Autocorrelated errors in experimental data in the language sciences: some solutions offered by Generalized Additive Mixed Models," in *Mixed Effects Regression Models in Linguistics*, eds D. Spelman, K. Heylen, and D. Geeraerts (Berlin: Springer), 49–69.
- Baddeley, A. D. (1986). *Working Memory*. Oxford: Oxford University Press.
- Baddeley, A. D. (2003). Working memory and language: an overview. *J. Commun. Disord.* 36, 189–208. doi: 10.1016/S0021-9924(03)0019-4
- Baddeley, A. D., and Hitch, G. J. (1974). "Working memory," in *The Psychology of Learning and Motivation*, Vol. 8, ed G. H. Bower (London: Academic Press), 47–90.
- Barnes, J., Brugos, A., Shattuck-Hufnagel, S., and Veilleux, N. (2013). "On the nature of perceptual differences between accentual peaks and plateaux," in *Understanding Prosody: The Role of Context, Function and Communication*, ed O. Niebuhr (Berlin; New York, NY: de Gruyter), 93–118.
- Barnes, J., Brugos, A., Veilleux, N., and Shattuck-Hufnagel, S. (2021). On (and off) ramps in intonational phonology: rises, falls, and the Tonal Center of Gravity. *J. Phon.* 85. doi: 10.1016/j.wocn.2020.101020
- 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. *Lab. Phonol.* 3, 337–383. doi: 10.1515/lp-2012-0017
- Batliner, A. (1989). "Wieviel Halbtöne braucht die Frage? Merkmale, Dimensionen, Kategorie [How many semitones does a question need? Characteristics, dimensions, category]," in *Zur Intonation von Modus und Fokus im Deutschen [On the Intonation of Mode and Focus in German]*, Vol. 234, eds H. Altmann, A. Batliner, and W. Oppenrieder (Tübingen: Niemeyer), 111–153.
- Baumann, S., and Grice, M. (2006). The intonation of accessibility. *J. Pragmat.* 38, 1636–1657. doi: 10.1016/j.pragma.2005.03.017
- Baumann, S., and Röhr, C. (2015). "The perceptual prominence of pitch accent types in German," in *Proceedings of the 18th International Congress of Phonetic Sciences (ICPhS)* (Glasgow).
- Baumann, S., and Winter, B. (2018). What makes a word prominent? Predicting untrained German listeners' perceptual judgments. *J. Phonetics* 70, 20–38. doi: 10.1016/j.wocn.2018.05.004
- Biezma, M., and Rawlins, K. (2017). Rhetorical questions: severing asking from questioning. *Proc. SALT* 27, 302–322. doi: 10.3765/salt.v27i0.4155
- Boersma, P., and Weenink, D. (2016). *Praat: Doing Phonetics by Computer*. Version 6.1.42 (version depended on labeller) [Computer Program].
- Braun, B. (2006). Phonetics and phonology of thematic contrast in German. *Lang. Speech* 49, 451–493. doi: 10.1177/00238309060490040201
- Braun, B. (2007). "Effects of dialect and context on the realisation of German prenuclear accents," in *Proceedings of the 16th International Congress of Phonetic Sciences (ICPhS)* (Saarbrücken), 961–964.
- Braun, B., and Biezma, M. (2019). Prenuclear L*+H activates alternatives for the accented word. *Front. Psychol.* 20, 1993. doi: 10.3389/fpsyg.2019.01993
- Braun, B., Dehé, N., Neitsch, J., Wochner, D., and Zahner, K. (2019). The prosody of rhetorical and information-seeking questions in German. *Lang. Speech* 62, 779–807. doi: 10.1177/0023830918816351
- Braun, B., Kochanski, G., Grabe, E., and Rosner, B. (2006). Evidence for attractors in English intonation. *J. Acoustical Soc. Am.* 119, 4006–4015. doi: 10.1121/1.2195267
- Caponigro, I., and Sprouse, J. (2007). "Rhetorical questions as questions," in *Proceedings of the Sinn und Bedeutung* (Barcelona: Universitat Pompeu Fabra), 121–133.
- Celle, A. (2018). "Questions as indirect speech acts in surprise contexts," in *Tense, Aspect, Modality and Evidentiality. Crosslinguistic Perspectives*, eds D. Ayoun, A. Celle, and L. Laure (Amsterdam; Philadelphia, PA: John Benjamins), 213–238.
- Chen, A. (2009). Perception of paralinguistic intonational meaning in a second language. *Lang. Learn.* 59, 367–409. doi: 10.1111/j.1467-9922.2009.00510.x
- Chen, Y., and Gussenhoven, C. (2008). Emphasis and tonal implementation in Standard Chinese. *J. Phon.* 36, 724–746. doi: 10.1016/j.wocn.2008.06.003
- Chodroff, E., and Cole, J. (2019a). "The phonological and phonetic encoding of information structure in American English nuclear accents," in *Proceedings of the International Congress of Phonetic Sciences (ICPhS)* (Melbourne, VIC), 1570–1574.
- Chodroff, E., and Cole, J. (2019b). "Testing the distinctiveness of intonational tunes: Evidence from imitative productions in American English," in *Proceedings of the Proceedings of the 20th Annual Conference of the International Speech Communication Association (Graz: Interspeech 2019)*, 1966–1970.
- Cohen, J. (1960). A coefficient of agreement for nominal scales. *Educ. Psychol. Meas.* 20, 37–46. doi: 10.1177/001316446002000104
- Crowder, R. G. (1982). Decay of auditory memory in vowel discrimination. *J. Exp. Psychol. Learn. Memory Cogn.* 8, 153–162. doi: 10.1037/0278-7393.8.2.153
- Dehé, N., Braun, B., Einfeldt, M., Wochner, D., and Zahner-Ritter, K. (2022). The prosody of rhetorical questions: a cross-linguistic view. *Linguistische Berichte*, 3–42.
- Dilley, L. C. (2010). Pitch range variation in English tonal contrasts: continuous or categorical? *Phonetica* 67, 63–81. doi: 10.1159/000319379
- Dilley, L. C., and Brown, M. (2007). Effects of pitch range variation on f0 extrema in an imitation task. *J. Phon.* 35, 523–551. doi: 10.1016/j.wocn.2007.01.003
- Dombrowski, E. (2003). "Semantic features of accent contours: Effects of F0 peak position and F0 time shape," in *Proceedings of the 15th International Congress of Phonetic Sciences* (Barcelona), 1217–1220.
- Dombrowski, E., and Niebuhr, O. (2010). "Shaping phrase-final rising intonation in German," in *Proceedings of the Fifth International Conference on Speech Prosody* (Chicago, IL), 1–4.
- Dorokhova, L., and D'Imperio, M. (2019). "Rise dynamics determines tune perception in French: The case of questions and continuations," in *Proceedings of the 19th International Congress of Phonetic Sciences (ICPhS)* (Melbourne, VIC), 691–695.
- Fitzpatrick-Cole, J. (1999). "The alpine intonation of Bern Swiss German," in *Proceedings of the 14th International Congress of the Phonetic Sciences (ICPhS)* (San Francisco, CA), 941–944.
- Fletcher, J., Grabe, E., and Warren, P. (2005). "Intonational variation in four dialects of English: the high rising tune," in *Prosodic typology. The Phonology of Intonation and Phrasing*, ed S.-A. Jun (Oxford: Oxford University Press), 390–409.
- Fries, N. (1988). "Ist Pragmatik schwer! - Über sogenannte Exklamativsätze im Deutschen [Pragmatics is difficult - On so called exclamatives in German]," in *Sprache und Pragmatik. Veröffentlichung des Lunder Projektes "Sprache und Pragmatik"*, ed I. Rosengren (Lund: Germanistisches Institut der Universität Lund), 1–18.
- Fünfgeld, S., Braun, A., and Zahner-Ritter, K. (2021). *The Phonetics of Verbal Irony. A Contrastive Study of Two German Regional Accents*. Talk at the Common Phonetics Colloquium Trier University and Saarland University, Saarbrücken; Trier.
- Gandour, J., Potisuk, S., Ponglorpisit, S., and Dechongkit, S. (1991). Inter- and intraspeaker variability in fundamental frequency of Thai tones. *Speech Commun.* 10, 355–372. doi: 10.1016/0167-6393(91)90003-C
- Gathercole, S. E., Hitch, G. J., Service, E., and Martin, A. J. (1997). Phonological short-term memory and new word learning in children. *Dev. Psychol.* 33, 966–979. doi: 10.1037/0012-1649.33.6.966

- Gilles, P. (2005). *Regionale Prosodie im Deutschen: Variabilität in der Intonation von Abschluss und Weiterweisung [Regional Prosody in German: Variability in the Intonation of Terminality and Continuation]*, Vol. 6. Berlin: de Gruyter. doi: 10.1515/9783110201611
- Grabe, E. (2004). "Intonational variation in urban dialects of English spoken in the British Isles," in *Regional Variation in Intonation*, eds P. Gilles and J. Peters (Tübingen: Niemeyer), 9–31.
- Grice, M., and Baumann, S. (2002). Deutsche Intonation und GToBI [German intonation and GToBI]. *Linguistische Berichte* 191, 267–298.
- Grice, M., Baumann, S., and Benz Müller, R. (2005). "German intonation in autosegmental-metrical phonology," in *Prosodic Typology. The Phonology of Intonation and Phrasing*, ed J. Sun-Ah (Oxford: Oxford University Press), 55–83.
- Grice, M., Ritter, S., Niemann, H., and Roettger, T. (2017). Integrating the discreteness and continuity of intonational categories. *J. Phon.* 64, 90–107. doi: 10.1016/j.wocn.2017.03.003
- Gussenhoven, C. (2004). *The Phonology of Tone and Intonation*. Cambridge: Cambridge University Press.
- Han, C.-H. (2002). Interpreting interrogatives as rhetorical questions. *Lingua* 112, 201–229. doi: 10.1016/S0024-3841(01)00044-4
- Hanssen, J. (2017). *Regional Variation in the Realization of Intonation Contours in the Netherlands*. Utrecht: LOT.
- Hirschberg, J. (2002). "The pragmatics of intonational meaning," in *Proceedings of the 1st International Conference on Speech Prosody, April 11-13 2002 (Aix-en-Provence)*, 65–68.
- Hothorn, T., Hornik, K., and Zeileis, A. (2006). Unbiased recursive partitioning: a conditional inference framework. *J. Comput. Graph. Stat.* 15, 651–674. doi: 10.1198/106186006X133933
- Hothorn, T., and Zeileis, A. (2015). partykit: a modular toolkit for recursive partitioning in R. *J. Mach. Learn. Res.* 16, 3905–3909. Available online at: <http://jmlr.org/papers/v16/hothorn15a.html>
- Kisler, T., Reichel, U. D., and Schiel, F. (2017). Multilingual processing of speech via web services. *Comput. Speech Lang.* 45, 326–347. doi: 10.1016/j.csl.2017.01.005
- Kohler, K. (1987). "Categorical pitch perception," in *Proceedings of the 11th International Congress of Phonetic Sciences (ICPhS)* (Tallinn), 331–333.
- Kohler, K. (1991a). A model of German intonation. *Arbeitsberichte des Instituts für Phonetik und Digitale Sprachverarbeitung der Universität Kiel (AIPUK)* 25, 295–360.
- Kohler, K. (1991b). 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, 115–185.
- Kohler, K. (2005). Timing and communicative functions of pitch contours. *Phonetica* 62, 88–105. doi: 10.1159/000090091
- Kügler, F. (2004). "The phonology and phonetics of rising pitch accents in Swabian," in *Regional Variation in Intonation*, eds P. Gilles and J. Peters (Tübingen: Niemeyer), 75–98.
- Kügler, F. (2007). *The Intonational Phonology of Swabian and Upper Saxon*. Tübingen: Niemeyer.
- Kügler, F., and Gollrad, A. (2015). Production and perception of contrast: the case of the rise-fall contour in German. *Front. Psychol.* 6, 1254. doi: 10.3389/fpsyg.2015.01254
- Kuhn, M. (2020). *Classification and Regression Training (caret). R package, Version 6.0-86*. Available online at: <http://cran.r-project.org/web/packages/caret/index.html>
- Ladd, D. R. (2008). *Intonational Phonology, 2nd Edn*. Cambridge: Cambridge University Press.
- Ladd, D. R., and Morton, R. (1997). The perception of intonational emphasis: continuous or categorical? *J. Phon.* 25, 313–342. doi: 10.1006/jpho.1997.0046
- Landis, J. R., and Koch, G. G. (1977). An application of hierarchical kappa-type statistics in the assessment of majority agreement among multiple observers. *Biometrics* 363–374. doi: 10.2307/2529786
- Leemann, A. (2012). *Swiss German Intonation Patterns*. Amsterdam; Philadelphia, PA: John Benjamins Publishing Company.
- Lehiste, I. (1975). "The phonetic structure of paragraphs," in *Structure and Process in Speech Perception*, eds A. Cohen and S. G. Nooteboom (Berlin: Springer), 195–203.
- Leiner, D. J. (2018). *SoSci Survey (Current Version: Version 3.2.05-i in 2020)*. Available online at: <http://www.socisurvey.com>
- Lohfink, G., Katsika, A., and Arvaniti, A. (2019). "Variability and category overlap in the realization of intonation," in *Proceedings of the 19th International Congress of Phonetic Sciences (ICPhS)* (Melbourne, VIC), 701–705.
- Lommel, N., and Michalsky, J. (2017). "Der Gipfel des Spotts. Die Ausrichtung von Tonhöhengipfeln als intonatorisches Indiz für Sarkasmus [Peak alignment as intonational cue to sarcasm]," in *Diversitas Linguarum*, Vol. 42, eds N. Levkovich and A. Urdze (Bremen: Universitätsverlag Dr. N. Brockmeyer), 33.
- Mayer, J. (1995). *Transcription of German Intonation - The Stuttgart System*. University of Stuttgart. Available online at: <https://www.ims.uni-stuttgart.de/institut/arbeitsgruppen/ehemalig/ep-dogil/joerg/labman/STGSystem.html>
- Mücke, D., Grice, M., Becker, J., and Hermes, A. (2009). Sources of variation in tonal alignment: Evidence from acoustic and kinematic data. *J. Phon.* 37, 321–338. doi: 10.1016/j.wocn.2009.03.005
- Mücke, D., Grice, M., Hermes, A., and Becker, J. (2008). "Prenuclear rises in northern and southern German," in *Proceedings of the 4th International Conference on Speech Prosody 2008* (Campinas), 245–248.
- Niebuhr, O. (2007a). "Categorical perception in intonation: a matter of signal dynamics?" in *Proceedings of the 8th Annual Conference of the International Speech Communication Association (Interspeech)* (Antwerp), 109–112.
- Niebuhr, O. (2007b). The signalling of German rising-falling intonation categories - the interplay of synchronization, shape, and height. *Phonetica* 64, 174–193. doi: 10.1159/000107915
- Niebuhr, O. (2022). "The Kiel intonation model," in *Prosodic Theory and Practice*, eds J. Barnes and S. Shattuck-Hufnagel (Cambridge: MIT Press), 287–318.
- Niebuhr, O., and Ambrazaitis, G. I. (2006). "Alignment of medial and late peaks in German spontaneous speech," in *Proceedings of the 3rd International Conference of Speech Prosody* (Dresden), 161–164.
- Niebuhr, O., D'Imperio, M., Gili Fivela, B., and Cangemi, F. (2011). "Are there 'shapers' and 'aligners'? Individual differences in signalling pitch accent category," in *Proceedings of the 17th International Congress of Phonetic Sciences (ICPhS)* (Hong Kong), 120–123.
- Ohala, J. (1990). There is no interface between phonology and phonetics: a personal view. *J. Phon.* 18, 153–171. doi: 10.1016/S0095-4470(19)30399-7
- Orrico, R., and D'Imperio, M. (2020). Individual empathy levels affect gradual intonation-meaning mapping: the case of biased questions in Salerno Italian. *Lab. Phonol.* 11, 1–39. doi: 10.5334/labphon.238
- Peters, B., Kohler, K., and Wesener, T. (2005). "Melodische Satzakkentmuster in prosodischen Phrasen deutscher Spontansprache - Statistische Verteilung und sprachliche Funktion [Melodic sentence accent patterns in prosodic phrases of German spontaneous speech - Statistical distribution and linguistic function]," in *Prosodic Structures in German Spontaneous Speech (AIPUK 35a)*, eds K. Kohler, F. Kleber, and B. Peters (Kiel: IPDS), 185–201.
- Peters, J. (2006). *Intonation deutscher Regionalsprachen*. Berlin: de Gruyter.
- Peters, J. (2014). *Intonation*. Heidelberg: Winter.
- Petronne, C., D'Alessandro, D., and Falk, S. (2021). Working memory differences in prosodic imitation. *J. Phon.* 89, 101100. doi: 10.1016/j.wocn.2021.101100
- Pierrehumbert, J. B. (1980). *The phonology and phonetics of English intonation* (PhD thesis). Massachusetts Institute of Technology; Department of Linguistics and Philosophy, Boston, MA, United States.
- Pierrehumbert, J. B. (2016). Phonological representation: beyond abstract versus episodic. *Annu. Rev. Linguist.* 2, 33–52. doi: 10.1146/annurev-linguistics-030514-125050
- Pierrehumbert, J. B., and Steele, S. A. (1989). Categories of tonal alignment in English. *Phonetica* 46, 181–196. doi: 10.1159/000261842
- Plomp, R. (1964). Rate of decay of auditory sensation. *J. Acoust. Soc. Am.* 36, 277–282. doi: 10.1121/1.1918946
- Prieto, P. (2011). "Tonal alignment," in *The Blackwell Companion to Phonology*, eds M. v. Oostendorp, C. Ewen, B. Hume, and K. Rice (Malden, MA; Oxford: Blackwell), 1185–1203.
- Prieto, P. (2012). "Experimental methods and paradigms for prosodic analysis," in *The Oxford Handbook of Laboratory Phonology*, eds A. Cohn, C. Fougeron, and M. Huffman (Oxford: Oxford University Press), 528–537.

- Prieto, P. (2015). Intonational meaning. *Wiley Interdiscipl. Rev. Cogn. Sci.* 6, 371–381. doi: 10.1002/wcs.1352
- R Development Core Team (2015). *R: A Language and Environment for Statistical Computing*. Vienna: R Foundation for Statistical Computing.
- Rathcke, T. (2017). How truncating are 'truncating languages'? Evidence from Russian and German. *Phonetica* 73, 194–228. doi: 10.1159/000444190
- Rietveld, A. C. M., and Gussenhoven, C. (1985). On the relation between pitch excursion size and prominence. *J. Phon.* 13, 299–308. doi: 10.1016/S0095-4470(19)30761-2
- Ritter, S., and Grice, M. (2015). The role of tonal onglides in German nuclear pitch accents. *Lang. Speech* 58(Pt 1), 114–128. doi: 10.1177/0023830914565688
- Roessig, S. (2021). "Categoriality and continuity in prosodic prominence," in *Studies in Laboratory Phonology*, Vol. 10 (Berlin: Language Science Press).
- Roessig, S., Mücke, D., and Grice, M. (2019). The dynamics of intonation: Categorical and continuous variation in an attractor-based model. *PLoS ONE* 14, e0216859. doi: 10.1371/journal.pone.0216859
- Roettger, T. B., Mahrt, T., and Cole, J. (2019). Mapping prosody onto meaning - the case of information structure in American English. *Lang. Cogn. Neurosci.* 34, 841–860. doi: 10.1080/23273798.2019.1587482
- Schneider, K., and Lintfert, B. (2003). "Categorical perception of boundary tones in German," in *Proceedings of the 15th International Congress of Phonetic Sciences (ICPhS)*, 631–634.
- Smith, R., and Rathcke, T. (2020). Dialectal phonology constrains the phonetics of prominence. *J. Phon.* 78, 100934. doi: 10.1016/j.wocn.2019.100934
- Sóskuthy, M. (2021). Evaluating generalised additive mixed modelling strategies for dynamic speech analysis. *J. Phon.* 84, 101017. doi: 10.1016/j.wocn.2020.101017
- Truckenbrodt, H. (2007). "Upstep on edge tones and on nuclear accents," in *Tones and Tunes. Volume 2: Experimental Studies in Word and Sentence Prosody*, eds C. Gussenhoven and T. Riad (Berlin: Mouton de Gruyter), 165–172.
- Turk, A. E., Nakai, S., and Sugahara, M. (2006). "Acoustic segment durations in prosodic research: a practical guide," in *Methods in Empirical Prosody Research*, eds S. Sudhoff, D. Lenertová, R. Meyer, S. Pappert, P. Augurzky, I. Mleinek, N. Richter, and J. Schließer (Berlin; New York, NY: Walter de Gruyter), 1–27.
- Ulbrich, C. (2005). *Phonetische Untersuchungen zur Prosodie der Standardvarietäten des Deutschen in der Bundesrepublik Deutschland, in der Schweiz und in Österreich [Phonetic studies on prosody in the Standard varieties in Germany, Switzerland, and Austria]*. Frankfurt am Main: Peter Lang.
- van Rij, J., Hendriks, P., an Rijn, I., H., Baayen, R. H., and Wood, Simon, N. (2019). Analyzing the time course of pupillometric data. *Trends Hearing* 23, 1–22. doi: 10.1177/2331216519832483
- van Rij, J., Wieling, M., Baayen, R. H., and van Rijn, H. (2017). *itsadug: Interpreting Time Series and Autocorrelated Data Using GAMMs. R Package Version, 2*.
- Waterman, J. C. (1991/1966). *A History of the German Language*. Long Grove, IL: Waveland Press Inc. (by arrangement with University of Washington Press).
- Wieling, M. (2018). Analyzing dynamic phonetic data using generalized additive mixed modeling: a tutorial focusing on articulatory differences between L1 and L2 speakers of English. *J. Phon.* 70, 86–116. doi: 10.1016/j.wocn.2018.03.002
- Wochner, D. (2021). *Prosody meets Pragmatics: rhetorical questions, exclamatives and assertions* (PhD thesis). University of Konstanz, submitted to the Department of Linguistics, Konstanz.
- Wochner, D., and Dehé, N. (2018). "Prosody meets pragmatics: a production study on German verb-first sentences, in *Proceedings of the 9th International Conference on Speech Prosody* (Poznan). 418–422.
- Wood, S. N. (2006). *Generalized Additive Models: An Introduction With R*. Boca Raton, MA: CRC Press.
- Wood, S. N. (2011). Fast stable restricted maximum likelihood and marginal likelihood estimation of semiparametric generalized linear models. *J. R. Stat. Soc. Ser. B* 73, 3–36. doi: 10.1111/j.1467-9868.2010.00749.x
- Wood, S. N. (2017). *Generalized Additive Models: An Introduction With R, 2nd Edn*. Boca Raton, MA: CRC Press.
- Xu, Y. (2013). "ProsodyPro - a tool for large-scale systematic prosody analysis," in *Proceedings of the Tools and Resources for the Analysis of Speech Prosody (TRASP 2013)* (Aix-en-Provence), 7–10.
- Zahner, K., Schönhuber, M., Grijzenhout, J., and Braun, B. (2016). "Konstanz prosodically annotated infant-directed speech corpus (KIDS corpus)," in *Proceedings of the 8th International Conference on Speech Prosody* (Boston, MA), 562–566.

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2022 Zahner-Ritter, Einfeldt, Wochner, James, Dehé and Braun. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.