Prosodic Theory and Practice

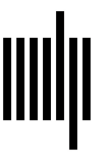
Edited by: Jonathan Barnes, Stefanie Shattuck-Hufnagel

Citation:

Prosodic Theory and Practice
Edited by: Jonathan Barnes, Stefanie Shattuck-Hufnagel
DOI: 10.7551/mitpress/10413.001.0001
ISBN (electronic): 9780262543194
Publisher: The MIT Press
Published: 2022

OA Funding Provided By:

OA Funding from MIT Press Direct to Open



The MIT Press

8 The Kiel Intonation Model—KIM

Oliver Niebuhr

8.1 Introduction

The Kiel intonation model (KIM) was developed as a contour model for Northern Standard German in the late 1980s by Klaus Kohler and his colleagues at the former Kiel Institute of Phonetics and Digital Speech Processing. It has since been refined and extended in a number of ways. These refinements and extensions rely on a growing body of empirical evidence from perception experiments and supplementary acoustic analysis of dialogue recordings, mainly conducted since 2003 by Benno Peters, Tamara Rathcke, Ernst Dombrowski, Klaus Kohler, Rabea Landgraf, and the author of the present chapter.

The ultimate goal of the KIM is to determine all communicative functions of Standard German prosody and relate them to distinctive prosodic features, derived from detailed descriptions, and their contextual as well as individual variation. The wording "all communicative functions" implies that the KIM does not differentiate between so-called linguistic and paralinguistic meanings in order to give the former priority over the latter. Rather, all types of meanings or functions from information structure to attitudinal meanings and argumentation structure to emphatic intensification are taken into account on an equal footing (Kohler 2007).

At the heart of the KIM is a phonological system. It is probably the only system to date that explicitly includes phonological levels of perceptual prominence, next to the phonological categories of intonation and phrasing that are offered by many other models as well. But the KIM as a whole is even larger. It also takes phenomena such as speech rate, pitch register, and types of disfluencies into account that often extend across several prosodic phrases. However, this chapter focuses on phonological categories that are associated with the domain of the prosodic phrase; it does so to comply with the maximum number of words and because almost all research and labeling within the KIM framework so far has been concerned with these phrase-level categories.

As this chapter provides the first extensive summary of the KIM after almost twenty years, the revised phonological system differs quite a bit from Kohler's (1991a, 1991b, 1997) early system presentations. Any changes and suggestions made in this context represent the author's own reflection about what the KIM currently looks like and how it should be developed in the future. The current phrase-level phonology of the KIM distinguishes among three prominence levels, nineteen intonation categories, nine types of phrase boundaries, and fourteen strategies for emphatic intensification. Not least in view of these relatively many categories, the target audience of the KIM comprises those who approach prosody from a general communicative perspective,

and who would like their annotation to be as rich as possible in terms of both forms and functions.

Attached to the KIM is an inventory for prosodic labeling (PROLAB) that translates the phonological categories of intonation, phrasing, prominence, and emphasis into sequences of simple ASCII letters. These symbols are annotated on a single tier with a conventionalized syntactic structure. However, there are tools that can automatically convert these annotations into multiple-tier representations. These tools ensure that PROLAB annotations are largely compatible and interchangeable across different software and transcription systems such as Praat and Emu (John 2012), which were not as popular when PROLAB was developed.

Finally, the KIM is obviously most suitable for German data; however, a number of phonological distinctions have recently also been found in a range of other languages, for example, English (Kleber 2006) Swedish (Ambrazaitis 2005, 2009), and Estonian (Asu 2006), making the KIM potentially also beneficial for a wider international audience.

8.2 A Brief Retrospect

The origins of the KIM lie in a practical rather than theoretical challenge: to develop a set of categories and rules that are able to generate proper prosodic patterns for German text-to-speech (TTS) synthesis within the Rulesys/Infovox framework (Carlson, Granström, and Hunnicut 1990; Kohler 1997). Accordingly, the first key publication of the KIM by Kohler (1991a, 1991b), which is still the major reference to the model, is structured similar to a generative grammar and develops decision trees with binary features of prominence and intonation (see figure 8.1). These feature trees are complemented by twenty-six pages specifying not fewer than 130 rules on the concatenation and context-sensitive implementation of these binary features.

The focus of the KIM changed with the start of the large international Verbmobil project in the mid-1990s (Wahlster and Engelkamp 1992). Verbmobil was supposed to create an application that both detected anger and disfluencies and also provided its users with instant translations of German, English, and Japanese spontaneous speech, for example, during telephone conversations. Implementing the idea of Verbmobil required recording, annotating, and analyzing what at the time were huge corpora of spontaneous speech. The Kiel Phonetics Institute was part of the Verbmobil network, and so the Kiel Corpus of Spontaneous Speech was recorded. Based on an appointment-making scenario, it is still one of the largest German corpora with full segmental and prosodic annotations at multiple phonetic-phonological levels (Peters 2005).

The KIM was used in the Verbmobil project as a resource for phonology-based prosodic labeling (Kohler 1996). The labeling tool that was developed, PROLAB, took the prominence and intonation categories of the KIM and translated them into machine-friendly seven-bit ASCII labels (Kohler 1997). The work with PROLAB in Verbmobil had a crucial side effect: applying the KIM to hours of spontaneous speech revealed quite a few discrepancies between the categories of the model (i.e., their phonetic definitions or assumed communicative functions) and actual speech data. These inadequacies and insufficiencies became the driving force for refining and extending the model's prosodic phonology (Kohler 2007).

This does not mean, however, that KIM categories were simply changed or added with reference to single-case examples or distributional differences between prosodic patterns in the Kiel corpus. Production data always played second fiddle in the KIM's

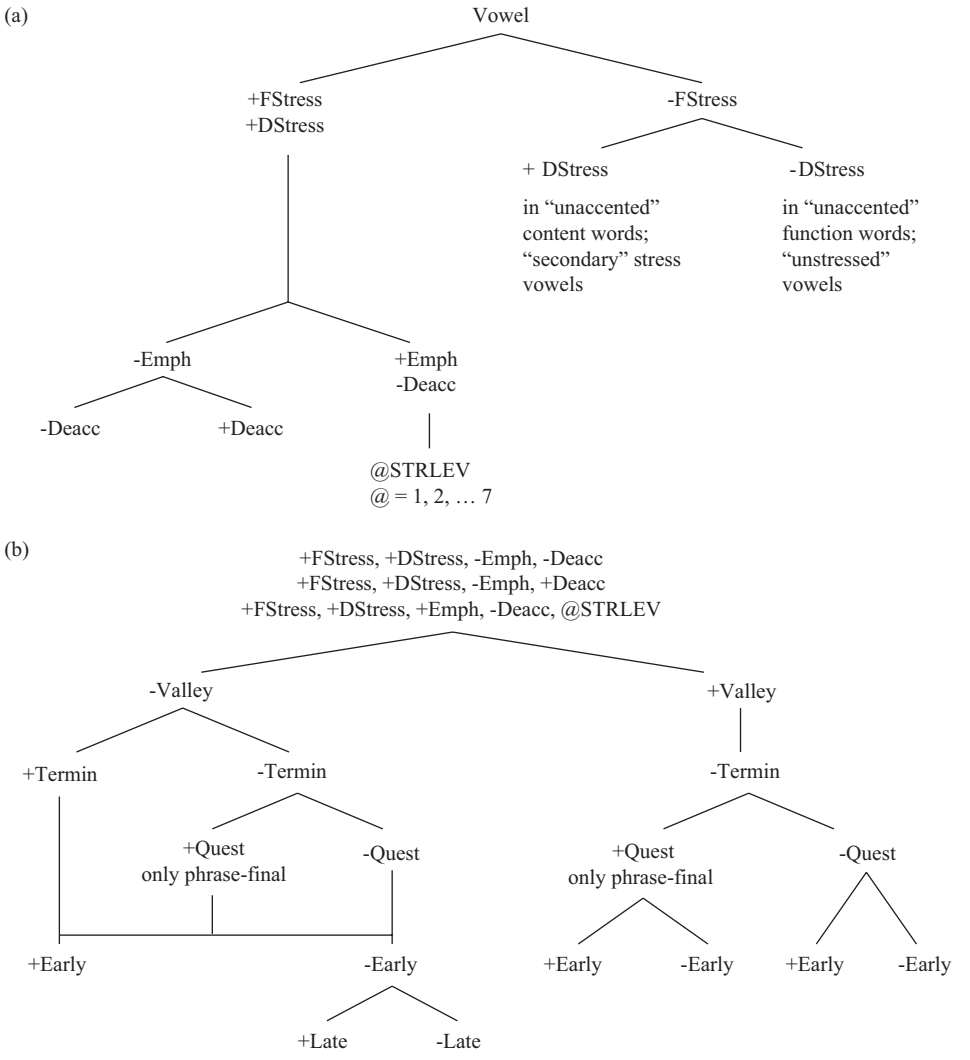


Figure 8.1
Binary distinctive features of (a) prominence and (b) intonation and their corresponding hierarchical organization in the original version of the KIM. Deacc, deaccented; DStress, duration-dominated stress; Emph, emphasis; FStress, f0-dominated stress; Quest, question; STRLEV, stress level; Termin, terminal F0 fall/rise. *Source*: After Kohler (1991a, 1991b, 1997).

prosodic phonology. The methodological fundamentals of the KIM are rooted in the Prague school (Trubetzkoy 1939), which, in turn, is based on the semiotic theory of de Saussure (1916). Unlike in the distribution-oriented, behavioristic American structuralism (Bloch 1948), a basic principle of de Saussure's theory is that phonological categories have to be set up in contrast to one another *and* with direct evidence that this contrast is associated with a change in meaning. As a result, production data were primarily used in the KIM for generating hypotheses about prosodic form-function relationships. Then it was the critical task of perception experiments to test these hypotheses and, on this basis, to revise or add phonological categories of the prosodic model. The listeners'

task in these perception experiments was always to judge a (resynthesized) prosodic parameter space by making direct or indirect reference to meaning (see section 8.3.1).

In line with having its roots in the European tradition of linguistics, the KIM's intonation categories are inspired by the intonation analyses of major British school scholars, for example, Halliday (1967), Jones (1969), and O'Connor and Arnold (1970). Kohler projected their tune structures and some of their paradigmatic oppositions onto German, but reconceptualized them as arising from two basic F0 contour types, each of which can be in different synchronization with the vowel of the accented syllable. The idea about the relevance of synchronization was obviously influenced by the seminal work of Bruce (1977). This becomes even clearer when the following is taken into account. While the KIM is, on the one hand, based on holistic contours, the synchronization-dependent phonological distinctions within each contour type rely, on the other hand, solely on local landmarks within these contours—most importantly F0 minima and maxima (see section 8.4.3).

In summary, having an idea about the KIM's history and development helps in understanding why the KIM is as it is today. Many characteristics that set the KIM apart from other prosodic models were not grounded in a new theoretical approach but rather arose from the fact that the model "grew up" in technical speech application environments, first Rulesys/Infovox and later Verbmobil. These environments motivated the two phonological subsystems, one for prominence and the other for intonation, and determined the PROLAB representation of KIM categories in simple ASCII letters. Furthermore, the intensive engagement with analyzing and processing spontaneous speech over more than twenty years, in combination with adopting the phonological concept of the Prague school, led to a rich, meaning-oriented phonological structure, particularly in terms of intonation.

It can also be considered a consequence of the KIM's technical background that phonological categories have an intermediate degree of abstraction. That is, categories are explicitly defined independently of micro-prosodic and individual variation (Kohler 1991a, 1991b). However, at the same time, specific characteristics of time and frequency serve as distinctive features and are an essential part of the categories' definitions. These features were obviously inspired by both traditional contour models and the tonal-target approach (in the sense of the early work of Bruce). Accordingly, the KIM distinguishes basic contour types, but sets up categories within each contour type on the basis of local F0 landmarks. By this hybrid approach to intonation, the KIM clearly differs from pure contour models such as the Dutch IPO model, named after the Eindhoven Instituut voor Perceptie Onderzoek (Eindhoven Institute for Perception Research). In this author's opinion, the KIM could be called a "pattern model."

8.3 Establishing Phonological Structures

8.3.1 The KIM's Methodological Guideline

Insights and progress in the field of segmental phonology essentially rely on the fact that we know where the boundaries between meaning-related units are and what their communicative (semantic or grammatical) function is. We can make use of this knowledge as a point of reference when we set up a system of paradigmatic contrasts and study variation within and between the categories of this system, for example, in the form of coarticulation and reduction processes.

The starting point in prosodic phonology, and of intonation in particular, is a completely different one (see the discussions in Hirschberg 2002, 2004). The general (and

still widely unresolved) uncertainty about the boundaries between intonational elements is most prominently reflected in the controversy about contour and tonal-target (i.e., level) models (Ladd 1996). Moreover, the difference in meaning conveyed by an audible difference in intonation is often rather difficult to pinpoint, for two reasons. First, intonational meanings are typically quite abstract, and second, these meanings can take different shades in different conversational contexts. It may also be that the second fact is a consequence of the first.

The reason that intonational meanings are "essentially pragmatic in nature" (Hirschberg 2004, 515) could also be related to the kind of meaning conveyed. Intonational meanings are often attributive in the sense that they ascribe a certain evaluative quality to a piece of information, the dialogue partner, or the situation in general. A similar kind of attributive meaning is conveyed by lexical adjectives, and just like intonational meanings, the meanings of lexical adjectives such as *funny, crazy, odd* are also far from straightforward and independent of contextual interpretation. For instance, the synonyms for *funny* given in the *Oxford English Dictionary* (Winchester 2003) range from *humorous* through *unpleasant* and *wrong* to *strange* and *curious*.

Many intonational meanings that were claimed to be nonattributive at first sight turned out to be attributive on closer inspection. A prime example is the difference between sentence-final rising and falling intonation, for instance, in German and Swedish (Petrone and Niebuhr 2013; Strömbergsson, Edlund, and House 2012). This difference was once firmly associated with signaling sentence mode. Today, we know that the rise-versus-fall difference actually conveys an attitudinal meaning difference, indicating whether the speaker super- or subordinates himself or herself to the interlocutor (Kohler 2004). This new meaning difference was suddenly able to account for a lot of mismatches between intonation and sentence mode that had puzzled researchers for quite a long time.

So, intonational meaning variation may appear a bit less exotic if we compare it to the range of meanings expressed by lexical adjectives. Moreover, approaching intonational meanings from an attributive perspective, which goes beyond the "usual suspects" of sentence mode, information, and discourse structure, can make the definition of intonational meanings a more worthwhile endeavor, even if not all intonational meanings turn out to be attributive in the end. Against this background, it is a key assumption of the KIM that phonological categories have meanings, that these meanings can be identified—or at least roughly outlined in contrast to one another—and that on this basis, meanings can be points of reference in establishing a prosodic phonology, analogous to the segmental phonology of the Prague school.

Accordingly, intonational meanings were always determined in the KIM framework in parallel to establishing intonational contrasts. The meaning difference of an intonational contrast had already been the subject of multiple studies, before the specific prosodic characteristics of the categories involved were investigated. In other words, the phonological structures of the KIM were developed according to the principle that "form follows function, not function form" (Xu 2004, 91).

The form-follows-function principle, in the chronological sense of understanding a function before understanding how it is encoded in the speech signal,was implemented on the basis of perception experiments that had the same underlying concept in all cases. First, a prosodic parameter space (i.e., a stimulus continuum) was generated on the basis of a short utterance, typically by means of pitch synchronous overlap and add (PSOLA) resynthesis in Praat (Boersma 2001) or, until the late 1990s, based on the Rulesys TTS system for which the KIM was originally developed (Kohler 1987, 2007).

Varied parameters were, for example, F0-peak synchronization or scaling, the duration of the accented syllable, or the direction and steepness of the phrase-final F0 movement. Then these stimuli were presented in multiple randomized repetitions to groups of naive German listeners. The listeners made direct or indirect judgments about the intonational/prosodic meanings in the stimuli.

8.3.2 Perception Experiments Based on Direct Judgments about Intonational Meanings

Direct judgments are obtained by means of the semantic-differential paradigm (Osgood, Suci, and Tannenbaum 1957). That is, during its occurrences in the experiment, each stimulus is judged on five to nine bipolar scales whose opposite adjectives at the scale ends are selected to cover three semantic dimensions: arousal, potency, and valence. Most studies, such as those of Dombrowski (2003), Dombrowski and Niebuhr (2010), Kohler (2005, 2011), and Niebuhr et al. (2012), used a seven-point Likert scale. However, a number of recent studies (e.g., Niebuhr 2008) also used continuous scales (visual analogue scales), Rietveld and Chen (2006) having shown that these scales trigger more distinct judgments and are more sensitive in detecting judgment differences than the Likert scale and other "equal-appearing interval scales" (288).

8.3.3 Perception Experiments Based On Indirect Judgments about Intonational Meanings

The most important type of perception experiment for the KIM is probably the "indirect identification task," also known as the "semantic task" in terms of Gussenhoven (1999). It was used to find and substantiate intonational categories since Kohler's (1987) seminal paper. The indirect identification test is a further development of the task that Nash and Mulac (1980) introduced. The basic idea behind this test is that a change of the intonation category causes a change in the meaning of the stimulus utterance. This meaning change and, hence, the existence of two separate intonation categories can be indirectly revealed by combining the stimuli with a constant context utterance, which is designed to match with the assumed meaning of only of the intonation categories. Then the listeners simply judge in the perception experiment whether the two presented utterances (stimulus and context) do or do not match. The principle of the indirect identification task is displayed in figure 8.2.

Recently the basic operation principle of the indirect identification test was refined by Landgraf (2014) and Niebuhr (2015) in the following way. Whenever an intonational function clashes with a coinciding lexical meaning, the corresponding utterance is interpreted as ironic (Bryant and Fox Tree 2005; Loevenbruck et al. 2013). So if listeners perceive a stimulus utterance as ironic in combination with intonation contour A, but not in combination with intonation contour B, then we can conclude that the listeners distinguished A from B and, moreover, associated contour A with a meaning that is diametrically opposed to the semantic content of the accented target word or phrase. This operation principle, which was called *irony task*, requires no constant context stimulus, which considerably lengthens the experiment and, due to its invariance, introduces a certain monotony into the task. The irony task has the further advantage that intonational meanings can be defined or tested in more detail, as they are the "negative images" of a specific word or phrase. In addition, it is reasonable to assume, and is supported by informal interviews with listeners, that "ironic" and "not ironic" are more common and intuitively applicable judgment categories than "matching" and "not matching" are.

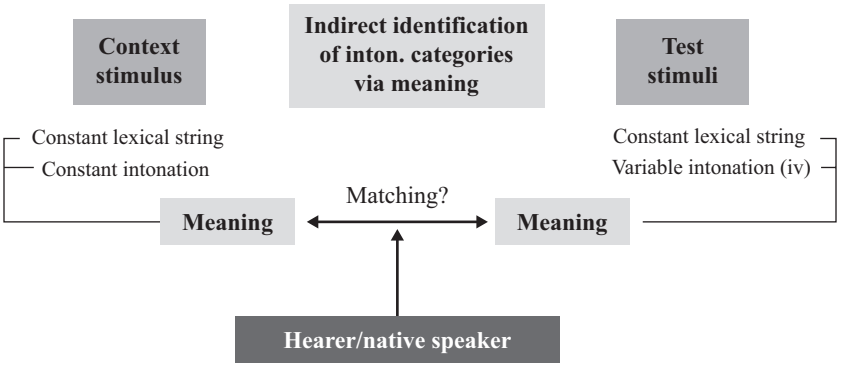


Figure 8.2
Schematic illustration of the makeup of indirect identification tests used to establish phonological categories in the KIM. inton, intonation.

8.3.4 Résumé

These descriptions of the KIM's methodological guideline and the corresponding perception experiments were to flesh out the fact that all phonological elements of the KIM (summarized in the section 8.4 below) are backed up by empirical evidence from at least one of these formal perception experiments. The phrase boundary categories are a small exception: the experiments did not rely on meaning-related judgments, but on similarly qualitative judgments about the degree of perceptual coherence of two adjacent stimulus utterances. As in semantic differential tasks, these judgments were made on a seven-step Likert scale.

Almost all phonological elements of the KIM have in the meantime been supported by consistent evidence from both direct and indirect judgment tasks. Perception experiments also served to identify and scrutinize trade-offs within and across distinctive features. For many categories or phonological contrasts, the evidence from perception was additionally cross-validated by acoustic analysis of production data obtained from specifically tailored read-speech tasks or the Kiel Corpus of Spontaneous Speech, or both. This applies in particular to pitch-accent peaks and valleys, concatenation contours, phrase boundary categories, and prominence levels, including classes and categories of emphasis.

8.4 Phonological Elements

8.4.1 Prominence Levels

Pitch accents are the elements of an utterance tune that stand out in terms of perceptual salience against the surrounding unaccented elements. This is a consensus of probably all intonation models, including the KIM. The KIM, however, goes one step further in that it assumes distinctive prominence levels and, hence, a separate phonological structure of accentual prominence that is orthogonal to the pitch-accent paradigms addressed in section 8.4.3.

The initial TTS-oriented version of the KIM distinguished nine prominence levels. This is surely a lot. But studies showed that listeners are able to project even more finely grained prominence scales onto perceived utterances (Fant and Kruckenberg 1989; Arnold, Wagner, and Möbius 2011). While prominence level 0 meant "unaccented,"

levels 1 to 7 represented different degrees of prominence. This basic framework still exists in the current version of the KIM. However, from the perspective of meaningful prominence differences, it turned out in perception experiments that a quadripartite accent phonology suffices (Kohler 1991a, 1991b). The corresponding PROLAB numbering is <0>, <1>, <2>, <3>.

Prominence level <0> still means "unaccented" and can thus not be associated with a pitch accent. In contrast, prominence levels <1> to <3> have to be associated with a pitch-accent category. Prominence level <1> refers to a secondary or "reduced" accent, level 2 represents a fully-fledged or "default" accent, and level <3> denotes an emphatically enhanced accent production. In the syntactic structure of PROLAB, prominence levels <1>, <2>, and <3> precede the corresponding pitch-accent label. Level <0> is not followed by any other PROLAB label. The communicative differences between the nonemphatic prominence levels <0> to <2> are illustrated in figure 8.3, based on an example of Kohler (1991a).

Figure 8.3 shows three realizations of *Er ist ins Kino gegangen* (He went to the cinema). The first utterance is realized with a fully fletched pitch accent on the initial syllable KI- of *Kino* (cinema), and a secondary/reduced pitch accent on the middle syllable -GAN- of *gegangen* (went). This sequence of prominence levels <2> and <1> on noun and verb is informationally neutral. That is, it puts the entire utterance content in focus. In contrast, reducing the secondary prominence on -GAN- further to <0> shifts the focus to the first prominence and in this way creates a (narrow-focus) contrastive interpretation of the utterance, for example, in the sense of, "He finally went to the cinema, although he actually wanted to go to the theater." Similarly, by increasing the prominence on -GAN- to the same level as in KI-, level <2>, the focus is shifted to the verb, so that the (narrow-focus) contrastive interpretation of the utterance changes, for example, to, "He finally walked to the cinema, although he actually wanted to go there by car."

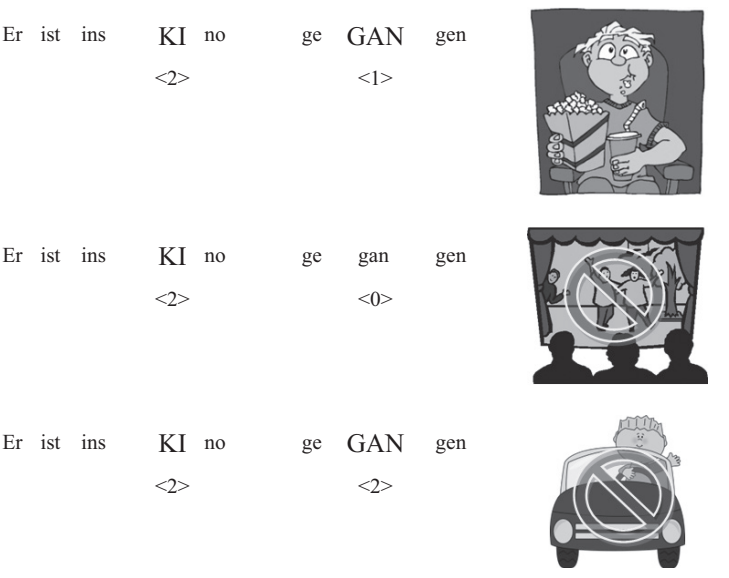


Figure 8.3
Differences in utterance interpretation caused by prominence levels <0> to <2>.

A further increase from level <2> to level <3> on KI- or -GAN- maintains the respective narrow focus interpretation and additionally intensifies a certain aspect of the utterance (see section 8.4.6). For example, changing the level <2> accent on KI-no into an emphatic level <3> accent could signal, "Oh, how I envy him! I also wanted to go there today!" (positive intensification; see section 8.4.6).

In addition to these obvious meaningful effects of changes in relative prominence level, Kügler et al. (2015) recently showed that after a short introductory training, nonexperts are able to reliably apply the KIM's phonological prominence-level distinctions to both spontaneous and read speech data. Interannotator agreement was even slightly higher in spontaneous speech, probably due to the fact that four prominence levels better reflect the complex prominence patterns of this speaking style than just two prominence levels (nonprominent versus prominent). According to feedback from annotators in Kiel, it is especially the availability of the secondary/reduced accent level <1> that makes it much easier to answer the often tricky question of whether a certain syllable is or is not pitch-accented. The corpus analysis of Niebuhr, Schroeder, and Baumann (2015) suggests that listeners make systematic use of all known prominence cues, that is, F0 movement, duration, and acoustic-energy level, when deciding between accent levels <1> and <2>.

8.4.2 Tune Structure and Phrase Boundaries

As illustrated in figure 8.4, a well-formed (i.e., noninterrupted) tune in the KIM consists of an intonation phrase with either four or five different types of syntagmatic slots, each of them having its own paradigmatic oppositions. A single-accent phrase includes a phrase-initial contour element, followed by a pitch accent, a phrase-final element, and a boundary category. When a phrase has more than one pitch accent, the two accents are linked by one of two phonologically different concatenation-contour categories. So unlike in many other intonation models (e.g., the autosegmental-metrical framework), the way in which pitch accents are linked is in the KIM not merely a phonetic consequence of the types of adjacent pitch accents or the temporal interval between them.

The original version of the KIM distinguished two types of phrase boundaries, which could roughly be termed "small/weak boundary" and "large/strong boundary." Kohler (1997) already assumed that this simple binary distinction may be insufficient by admitting that "the linguistically and phonetically relevant categorization of these boundaries is not well understood" (10), and later analysis of spontaneous

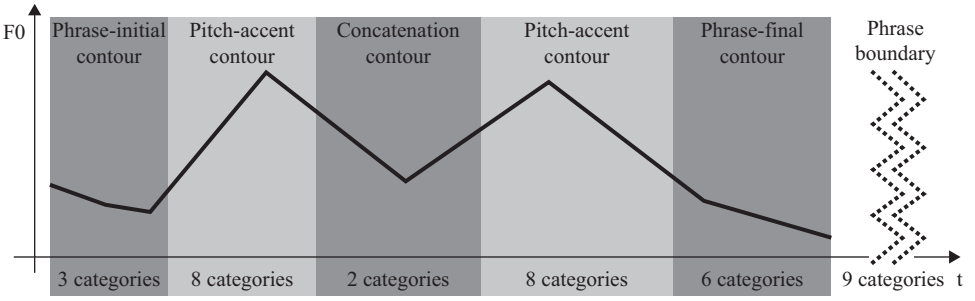


Figure 8.4
The syntagmatic structure of intonation phrases in the KIM. The concatenation contour occurs only in phrases with more than one pitch accent.

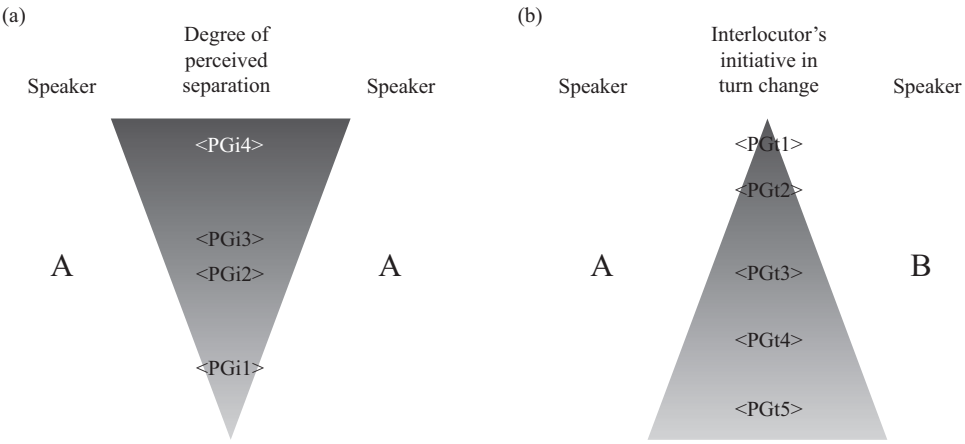


Figure 8.5
The phrase-boundary categories distinguished by the KIM for (a) turn-internal and (b) turn-final prosodic phrases.

speech dialogues and corresponding perception experiments by Peters (2012) led to a much more complex 4 + 5 system. The former number refers to turn-internal and the latter number to turn-final boundary categories. Figure 8.5 provides a summary of this system.

All turn-internal boundaries show similar degrees of phrase-final lengthening and intensity increases or decreases for rising or falling intonation movements, respectively. The distinguishing features are whether they are followed by a pause and end in an intonation movement that reaches into the limits of the speakers individual pitch range. Phrase boundaries without large final intonation movements and without following pauses are the weakest. They often occur unnoticed by speakers and listeners within syntactic constituents. Phrase boundaries without large final intonation movements but with a following pause are typically unavoidable "breathing pauses" within a speaker's line of argument. They can, but need not, occur at syntactic boundaries. A similar degree of perceptual separation is created by phrase boundaries with a large final intonation movement but without a following pause. They predominantly coincide with syntactic boundaries and are used, for example, for listing or contrasting arguments or items. Phrase boundaries that show both large final intonation movements and following pauses are the strongest and often are used to mark topic changes. They always occur at syntactic boundaries. The corresponding PROLAB labels are <PGi1>, <PGi2>, <PGi3>, and <PGi4> (<PG> stands for "Phrasen-Grenze" and <i> for "intern").

Turn-final phrase boundaries are also distinguished by whether the final intonation movement reaches into the limits of the speaker's individual pitch range. In addition, they take into account whether the phrase-final intonation is rising or falling and whether the turns end in overlap with the beginning of the interlocutor's turn. Together, these features reflect the speaker's and the interlocutor's relative contributions to the turn change (figure 8.5b). The two categories labeled <PGt1> and <PGt2> are the strongest turn-final phrase boundaries. They end in a large intonation movement and show no overlap with the interlocutor's turn. Both signal an intended turn-yielding, but <PGt1> ends in a fall, whereas <PGt2> ends in a rise, which in almost

all cases is associated with a question. In contrast, <PGt3> and <PGt4> signal that the interlocutor can, but need not, take the turn. <PGt3> marks the end of a terminal low fall and is used when a speaker has completed her line of argument. The opposite is true for <PGt4>, which is why this phrase boundary can also occur in overlap with the beginning of the interlocutor's turn. Finally, <PGt5> indicates a forced turn-yielding and is characterized by a high final intensity level, no pronounced final F0 movement, and a phrase boundary in overlap with the interlocutor's turn.

8.4.3 Phrase-Initial Contour

The KIM distinguishes three types of phrase-initial contours, based on, for example, the experiments of Kohler (1991a, 2011):

- The *default phrase-initial contour* category rises from a lower pitch level to the level of the F0 maximum of first pitch accent. Its PROLAB symbol is <LP> (low prehead). It is a neutral way of opening a discussion or relating to previous utterances of the speaker or the interlocutor.
- *Falling phrase-initial contours* that start at a similar F0 level as the maximum of the first pitch accent are marked by <HP> (high prehead). Speakers use this phrase-initial category to convey their cooperativeness in the conversation. In addition to formal evidence from perception experiments, this attitudinal meaning is also supported by speech data analyses: <HP> occurs particularly often when speakers continue a line of argument or complete utterances of their dialogue partners.
- *High-falling phrase-initial contours* that set in and start falling at a much higher F0 level than the maximum of the first pitch accent represent a separate category in the KIM and are marked in PROLAB by <HP2>. Unlike <HP>, <HP2> is used to take up and critically revise what has been said before. In this sense, <HP2> conveys the opposite of cooperativeness. For example, in an utterance like "the car, the dog, and the children," an <HP> onset could be paraphrased as, "We have to take these things into account as well," whereas a <HP2> onset means, "This is what you forgot" or "This is why your plan won't work."

8.4.4 Pitch-Accent Contour

The KIM distinguishes two basic classes of pitch-accent contours: rising-falling peaks and (falling-)rising valleys (Niebuhr and Kohler 2004). Both contour classes have four distinctive intonation categories, which results in a total of eight different pitch accents.

The KIM does not make any restrictions concerning the occurrence of these eight categories in prenuclear and nuclear positions. Although Kiel Corpus analyses show that valley contours are more frequent in nuclear than in prenuclear positions, every accent category can basically occur with the same meaning in every position (see, e.g., Niebuhr and Zellers 2012 for pitch-accent peaks).

8.4.4.1 Pitch-accent peaks Synchronization plays a major role in distinguishing pitch accents in the KIM. This is most obvious for the peak categories, which are all defined by their synchronization. So unlike in, for example, autosegmental-metrical models, time or timing is directly phonologically relevant in the KIM instead of being an epiphenomenon of how and what tones are associated with the accented syllable. As for the latter, another important difference between autosegmental-metrical models and the KIM is that the Kiel model refers to the accented-vowel boundaries rather than the accented-syllable boundaries in describing differences between pitch-accent categories.

In its original version, the KIM distinguished three peak categories based on studies of Kohler (1987) and Niebuhr (2007c). However, in dealing with the Kiel Corpus of Spontaneous Speech, a fourth category was introduced, also inspired by the distinction between H* and L+H* in German GToBI (Grice and Baumann 2002). Kohler (2005) provided the perceptual evidence for this fourth category.

- The early peak reaches the F0 maximum before the accented-vowel onset. Thus, F0 is falling through the vowel to a low level. The PROLAB symbol for the early pitch accent is <)>. The early accent expresses that the corresponding piece of information is given and cannot or should not be changed. Depending on the semantic-pragmatic context, the specific meaning of early peaks can range from resignation ("there is nothing we can do about it") to reliability ("what I say is an immutable fact"). Not least for these reasons, the early category often occurs at the end of a speaker's turn.
- The medial peak reaches its F0 maximum within the accented-vowel boundaries, typically in the second half of the vowel. Thus, parts of the rising and falling slope are located inside the vowel. The corresponding PROLAB symbol is <^>. The medial peak is the neutral way of accentuation in German. It marks new pieces of information or is used to put a previously addressed piece of information up for discussion again (which is why even given information can occur with a medial peak; Niebuhr 2007c and Kohler 2005, 2007).
- The late-medial peak is characterized by a rise throughout the vowel. The rising slope starts before the accented-vowel onset and continues until after the accented-vowel offset. The F0 peak maximum is typically reached in the next sonorant sound segment. The late-medial peak is marked in PROLAB by <{>. It conveys that the corresponding piece of information is at odds with the speaker's expectation. The speaker makes "a matter-of-fact statement of this contrast" (Kohler 2005, 90).
- The late peak starts rising after the accented-vowel onset and continues rising until after the accented-vowel offset. The peak maximum is located quite high in the speaker's pitch range and typically reached still later than for the late-medial peak, that is, in the vowel of the following unaccented syllable. However, unlike for the late-medial peak, the rise must immediately change into a steep fall, part of which has to be realized within the accented word (Niebuhr and Zellers 2012). The PROLAB symbol of the late peak is <(>. The late peak also signals a contrast to the speaker's expectation, but in addition to the late-medial peak, the late peak "adds a personal expressive evaluation to the contrast" (Kohler 2005, 90). Depending on the semantic-pragmatic context, this can signal either a positive surprise or incredulity in the sense that the speaker is not willing to simply accept the facts that led to this contrast.

To make the meanings and meaning differences of the four pitch-accent peaks more tangible, we can combine them with the simple particle "no." In this case, their meanings can be paraphrased as follows: "And that's that" (early), "I'm against it, but let's discuss that" (medial), "This comes unexpected, but I can live with that; I can't change it anyway" (late-medial), and "I can't believe that this is happening to me, and I will do something about it" (late).

Figure 8.6a illustrates the synchronization characteristics of the four pitch-accent peaks relative to the accented-vowel boundaries. More recently, it was found that peak

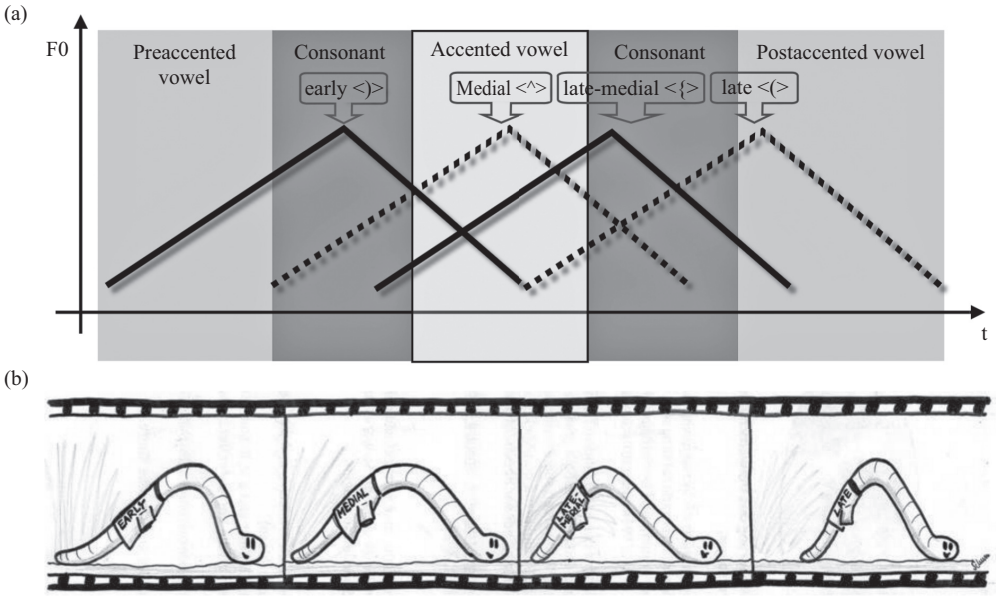


Figure 8.6
The four pitch-accent peaks of the KIM shown (a) in their distinctive synchronization windows and (b) in additional combination with their characteristic peak shapes, which together constitute the "worm paradigm." *Source*: Drawing courtesy of Nathalie Schümchen.

shape also matters, particularly with respect to the steepness of rising and falling slopes. Some shapes are more effective than others in signaling pitch-accent categories, and as a consequence, their synchronization ranges become larger and overlap with those of other categories across the vowel boundaries. For example, a peak with a slowly rising-falling shape is still perceived as being medial, even when the F0 maximum is located well after the accented-vowel offset. A peak with a fast falling shape is still perceived as being early, even when its F0 maximum is clearly reached inside the accented vowel (Niebuhr 2007a, 2007c). In contrast, accent peaks with a plateau-shaped maximum need not rise at all into the accented vowel to be identified as being medial (Niebuhr 2011). This perceptually determined interplay of peak synchronization and peak shape also manifests itself in production, for example, in the form of speaker-specific trade-offs between the two parameters, with "shapers" and "aligners" at the extreme ends of the trade-off continuum (Niebuhr et al. 2011).

When the four pitch-accent categories are depicted in their defined synchronization window and with their most effective peak shape, they look like still images of a worm that successively crawls across the accented vowel (see figure 8.6b). Therefore, the four accent peaks together form a set of contrasts that is also called the "worm paradigm." The worm starts crawling before the accented vowel with a slowly rising, fast-falling shape (early), takes a more symmetrical slowly rising-falling shape inside the vowel (medial), crosses the vowel offset with a fast-rising slow-falling shape (late-medial), and ends his trip through the accented vowel in the post-accented vowel, having a symmetrical fast rising-falling shape (late).

8.4.4.2 Pitch-accent valleys The attitudinal meanings of peak categories are oriented toward a particular piece of information or toward the speakers themselves. In contrast, valleys bring in the addressee. While the meanings of peaks and valleys point broadly in diametrically opposed directions, Kohler stresses that from a phonetic-phonological perspective, "valleys are not the mirror images of peaks, and it is not surprising that they should have different perceptual status" (1991a, 308). In terms of phonological contrasts, this means, for example, that there is only perceptual evidence for a binary distinction based on synchronization (Kohler 1991a; Niebuhr and Kohler 2004; see figure 8.7a).

- Early valleys reach their F0 minimum before the accented-vowel onset. Thus, F0 rises throughout the accented vowel. The corresponding symbol in PROLAB is <]>. Early valleys invite dialogue partners to make a matter-of-fact contribution to the discussion, for example, by answering a question or expressing their opinion about a certain piece of information.
- Late valleys differ from early valleys in that they start rising inside the accented vowel. The PROLAB symbol of late valleys is <[>. Like early valleys, late valleys also extend an invitation to the dialogue partner to contribute to the discussion. However, in the case of late valleys, this invitation is aimed at the interlocutor's feelings about a certain piece of information.

So a discourse marker such as "okay" can call for backchanneling, agreement, or disagreement in combination with an early valley, but means, "How do you feel about it?" or, "Can you live with that?" in combination with a late peak.

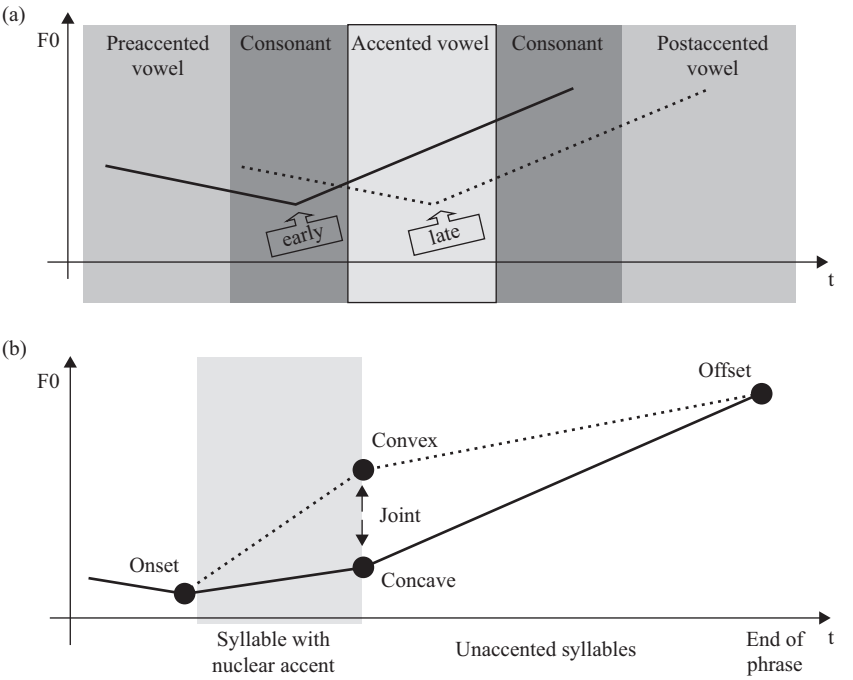


Figure 8.7
The two pitch accent valleys of the KIM shown (a) in their distinctive synchronization windows and (b) in combination with their distinctive feature of a convex or concave rise.

In addition to the phonological distinction based on synchronization, a further distinction is based on shape. This is another aspect in which valleys differ from peaks, for which shape is (as far as we know today) not an independent phonological feature. However, in the case of valleys, F0 rises can take a convex or concave shape (seen from above), for example, based on a hypothetical joint located at the end of the accented syllable (see figure 8.7b):

- Valleys with a convex shape "restrict" dialogue partners in the sense that they reduce their alternatives in the conversation.
- Valleys with a concave shape "activate" dialogue partners and invite them to shape the further discussion at their own discretion.

In other words, from the speaker's point of view, with the convex shape, the speaker remains "at the helm," whereas, with a concave shape, he or she hands the control over to the interlocutor. Depending on the semantic-pragmatic context, this can result in different interpretations. For example, in statements like "Then came Anna," concave shapes ask for conformation or a short comment, whereas convex shapes signal "hold on and let me continue." In the case of questions like "Are you Anna?," the meanings of convex and concave shapes can be paraphrased as "I just ask for your name and for nothing else; I continue speaking afterward" (convex) or "Please also tell me a bit more about you" (concave). Correspondingly, convex valleys occur more often turn-internally and concave valleys turn-finally.

Production and perception experiments by Dombrowski and Niebuhr (2005, 2010) put the convex-concave distinction on solid empirical grounds for the early valley category (cf. figure 8.4b). However, there is also initial evidence from pilot perception experiments that the distinction also applies to the late valley category. As the PROLAB inventory was last updated in 2010, there are no symbols for the convex-concave distinction as yet. However, an obvious solution would be to add an <x> for convex and a <c> for concave right before the respective synchronization label. Hence, <2x]> would, for instance, denote a convex early valley pitch accent at a default prominence level.

8.4.4.3 The minus accent Phenomenologically similar to the concept of phrase accents in GToBI (Grice, Ladd, and Arvaniti 2000) but syntagmatically not restricted to post-nuclear prominences, the KIM takes into account that syllables can be made prominent for rhythmical purposes only. Barry (1981) points out the important guiding function of speech rhythm (in West Germanic languages). If syllables are made prominent for rhythmical purposes, they need not be associated with a pitch accent, that is, with a separate attitudinal meaning. Therefore, the KIM offers the possibility of marking prominent syllables without a pitch-accent label. In this case, the phonological prominence level—often <1>, sometimes <2>—is followed by a minus sign <–>, which is why the F0 course on the corresponding syllable is also called the *minus accent*. Analyzing the annotations in the Kiel Corpus of Spontaneous Speech, Peters, Kohler, and Wesener (2005) found that minus accents are relatively rare, but not exotic (about 7 percent of all prominent syllables). They occur typically in phrase-initial or phrase-final position. In the latter position, they are often related to disfluencies. About 20 percent of all minus accents represent prominent syllables inside "hat patterns" (see section 8.4.5). Not least for this reason, minus accents are typically annotated when F0 runs flat across a prominent syllable.

8.4.5 Concatenation Contour

The separate phonological categorization of the F0 course in between two accented syllables is probably one of the two most noticeable features in which the KIM differs from other models of intonation, next to the KIM's prominence-level distinction. In the original version of the KIM (Kohler 1991a, 1991b), there was a tripartite distinction among the concatenation contours, symbolized in PROLAB by <0.>, <1.>, and <2.> (see figure 8.8a).

The <0.> category means that there is no F0 indentation between two pitch accents, which results in a so-called *hat pattern.* In contrast, <1.> refers to a moderate and <2.> to a strong F0 indentation, the latter reaching down to the bottom of the speaker's individual pitch range. However, a more detailed and comprehensive series of perception experiments summarized in Ambrazaitis and Niebuhr (2008, 2014) was inconsistent with this phonological analysis. First, the experiments showed that hat patterns do not have to be completely flat. Even a slight F0 indentation of about two or three semitones can still have the same communicative function as a completely flat hat. Second, Ambrazaitis and Niebuhr found no empirical support for a distinction between <1.> and <2.>. As is displayed in figure 8.8b, this led to a revision of the concatenation-contour paradigm in favor of a binary phonological distinction: hat pattern versus dip pattern. It is suggested here that the former is marked by <h> and the latter by <d>.

The hat pattern looks like a staple for a stapler, and in fact it has a bracketing function that holds two or more accented pieces of information together, whereas the opposite applies to the dip pattern. So, for example, when produced with the hat pattern, the greeting "Dear ladies and gentlemen" is content-wise the same as "Dear audience," whereas a dip pattern signals that the two groups of listeners are addressed

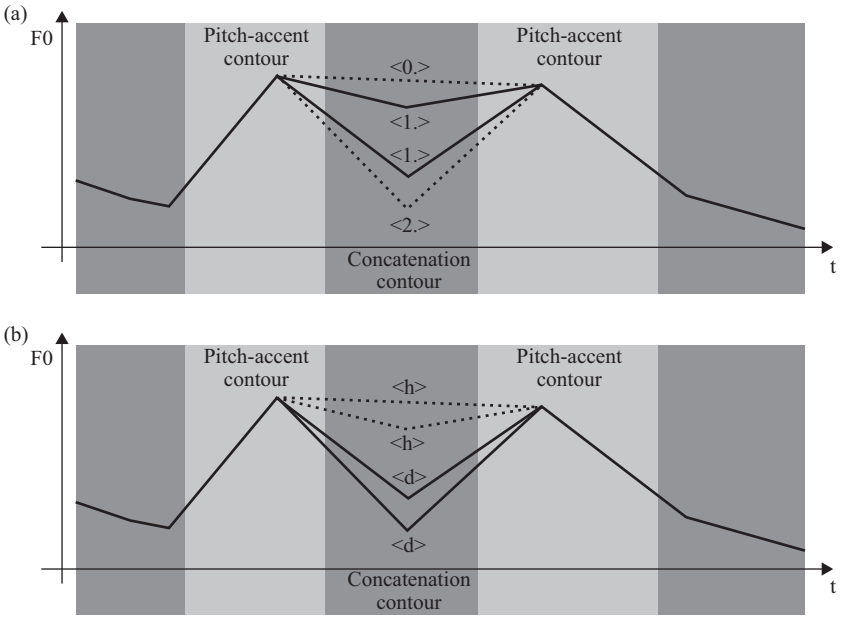


Figure 8.8
The concatenation contours of the KIM. (a) The original paradigm with three categories. (b) The revised paradigm that distinguishes only two categories: hat and dip contours.

separately (as if they were sitting on opposite sides of the room). Likewise, "Would you like tea or coffee?" with a hat pattern asks the interlocutor to choose between a closed set of two alternatives, whereas the same question with a dip pattern becomes an open list of suggestions, equal to "Would you like something to drink?" Also, interlocutors are basically given the option to ask for a hot chocolate, which would not be possible in connection with the hat pattern.

In the context of concatenation phenomena, it must also be noted that the KIM distinguishes between upstep and downstep (Kohler 1991a, 1991b, 1997). This applies to pitch-accent peaks. If the F0 maximum of pitch-accent peak B is higher than that of the previous peak A, then peak B is marked as upstepped by placing the PROLAB label <|> in front of the peak-category label. However, upstep marking requires, per definition, that the two accent peaks have the same phonological prominence level. So <2^ d |2^> is a valid upstep annotation, but <1^ d |2^> is not. This restriction is to avoid redundancies, as higher prominence levels are associated with larger F0 peak ranges anyway.

8.4.6 Phrase-Final Contour

Given the contour differences between pitch-accent peaks and valleys, it is logical that each of these two contour classes has its own phonological paradigm of phrase-final contours (the two paradigms are depicted in figure 8.9a and 8.9b):

- Peak contours can end in a terminal fall until the lower limit of the speaker's individual pitch range. This is marked in PROLAB by <2.>. The categoricalness, dominance, and potential impoliteness conveyed a terminal fall can be attenuated "without appearing to be indecisive and insecure" (Ambrazaitis 2005, 198). This is done by attaching a small, hardly audible rise of about one to two semitones (five to fifteen hertz; Peters 2000) to the terminal fall. The attenuated terminal fall was called "pseudo-terminal" and is symbolized in PROLAB by <2;>. Peters (2000) discovered by analyzing <2;> annotations in the Kiel Corpus of Spontaneous Speech that pseudo-terminal falls are far more often produced by women than by men, probably due to the traditional role model of women in a patriarchal society.
- Falls not reaching into the lower limit of the speaker's pitch range are called "non-terminal." This type of phrase-final peak contour, which is marked by <1.> in PROLAB, typically occurs when speakers reflect on different alternatives. For example, they make a statement, and then they consider it from a different perspective in the subsequent intonation phrase. A prime example are statements such as "We can do it" or "A great idea" at the end of which the interlocutor already hears (due to <1.>) that a "but" will follow.
- Cases in which a peak contour changes at the F0 maximum into a level plateau are indicated by <0.> in PROLAB. Phrase-final <0.> contours typically occur in turn-medial position. So like <1.>, <0.> signals that the speaker will continue. But unlike <1.>, <0.> is used when a statement is continued or further justified rather than contrasted with a different statement in the next phrase.
- Phrase-final contours after valley contours are distinguished by how high they rise. There are two phonological categories of phrase-final rises: a medium rise and a high rise. The original version of the KIM (Kohler 1991a, 1991b) did not specify how exactly the two rise categories are defined. Recent evidence from production and perception (see also Michalsky 2014) led to the following definitions (see figure 8.6b): medium rises do not reach beyond the F0 maximum of the immediately

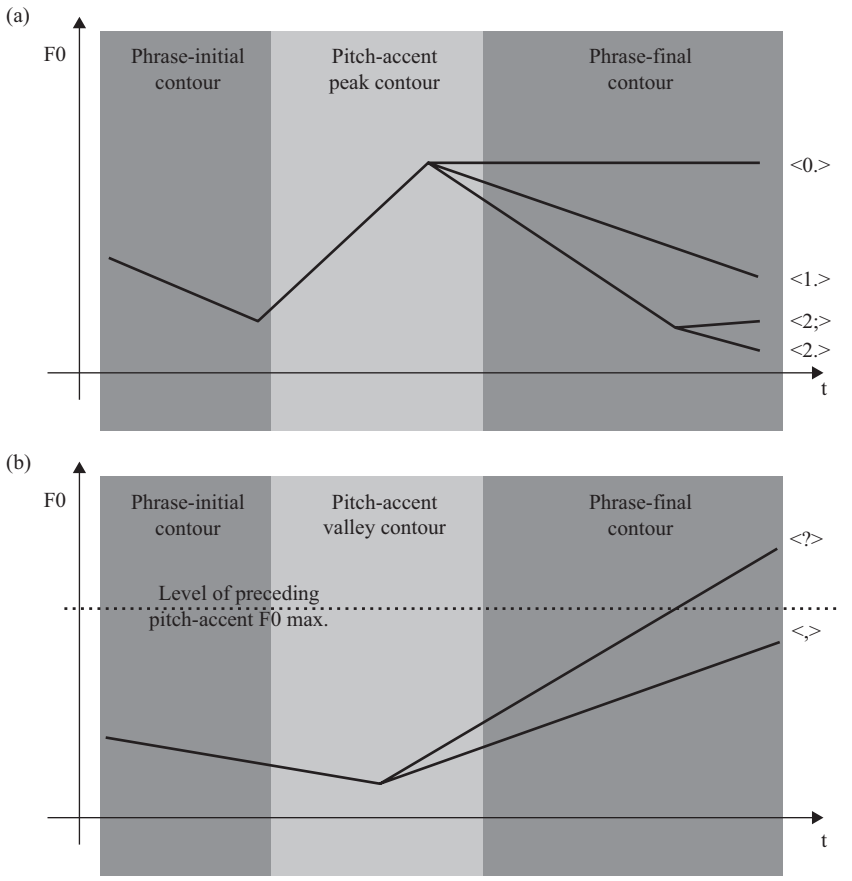


Figure 8.9
The paradigmatic oppositions among the phrase-final contours of the KIM after (a) pitch-accents peaks (four categories) and (b) pitch-accent valleys (two categories).

preceding pitch-accent contour. The opposite applies to high phrase-final rises. That is, they continue beyond the F0 maximum of the immediately preceding pitch-accent contour. Medium phrase-final medium rises are symbolized by <,> and have a similar "continuation function" as <1.> and <0.> after peak contours. However, unlike the latter two categories, the <,> category is listener-oriented rather than information-oriented.

- High phrase-final rises are the counterpart of phrase-final <2.> falls. Thus, they are also called "terminal rises" (Kohler 1991a, 1991b). Their PROLAB symbol is <?>. As is already indicated by the question-mark symbol, high phrase-final rises typically occur at the ends of interrogative utterances or when the turn is for other reasons handed over to the interlocutor.

8.4.7 Additional Categories of Emphatic Intonation

Work with the Kiel Corpus of Spontaneous Speech over many years made it clear that emphatic expressions are anything but rare in informal everyday conversation. Therefore, the KIM started a new line of research around 2006, aimed at discovering the

form-function links of emphatic intonation contours in German and consolidating them into a system of phonological paradigms that supplements the KIM's regular intonational phonology.

Emphasis is "a very broad and general term applied to any kind of prominence attached to a linguistic element" (Trask 2004, 188). Newman (2015, 73) calls it "a convenient phonological catch-all." The use of the term *emphasis* in the literature ranges from function-oriented definitions such as "emotional involvement on the part of the speaker" (Ladd 1996, 129) to signal-oriented definitions in which emphasis is an acoustic feature reflecting the "relative energy in the higher frequency bands" (Heldner 2003, 40). Against this complex terminological background, the KIM provides an explicit definition of emphasis. This definition combines form and function and separates emphasis from the noncontrolled (i.e., physiologically determined) long-term effects of emotion on speech. Emphasis manifests itself as local extra effort, added by the speaker to both articulatory and phonatory patterns with the intention to intensify an aspect of the communication.

For these intensified aspects of communication, figure 8.10 shows that the KIM makes a tripartite distinction, inspired by the communication models of Bühler (1934) and von Thun (1981), as well as the notion of emphasis of British school phoneticians (Jones 1969):

- *Emphasis for intensity* is an accent-level phenomenon that intensifies the semantics of the accented piece of information. Emphatic accents are not just regular pitch accents produced with more articulatory and phonatory effort. Rather, this extra effort adds an expressive evaluation to the meaning of the pitch accent.
- *Emphasis for contact* relates to the intonational nucleus. It intensifies the speaker's connection to the interlocutor—for example, to assess this connection or bind the two dialogue partners closer together.
- *Emphasis for attention* is a phrase-level phenomenon whose intensification is directed toward the communication channel, making the interlocutor more receptive or susceptible to the speaker's message.

A fourth potential category would be *emphasis for contrast*, which relates to the informational or situative context of the communication (Kohler 2006). It includes the late-medial and late pitch-accent categories and the corresponding patterns of

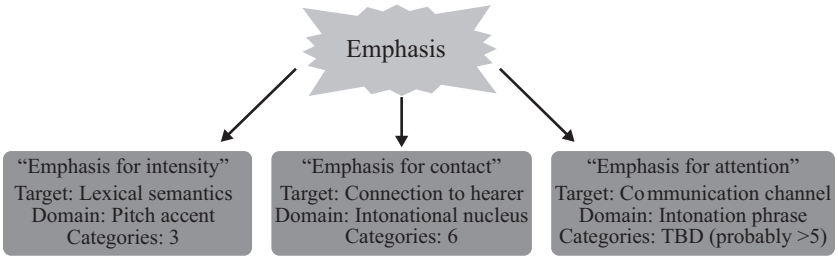


Figure 8.10
The three emphasis classes distinguished in the KIM. Each class includes at least three phonological categories. The exact number of categories is still unknown as the emphasis system is subject of ongoing research. TBD, to be determined.

phonological prominence. It is not entirely clear yet whether these accent patterns meet all criteria of the KIM's emphasis definition (e.g., extra effort in the articulatory and phonatory domain; Görs and Niebuhr 2012). For this reason, "emphasis for contrast" was omitted in figure 8.10.

Overall, evidence from combined production and perception studies have identified more than a dozen emphasis categories. Providing detailed portraits for all of these categories would make this chapter far too long. Therefore, the following sections just give a brief overview of each class of emphasis categories.

8.4.7.1 Emphasis for intensity Emphasis for intensity is the oldest and best investigated class of emphasis phenomena within the KIM. Studies by Kohler (2006) and Niebuhr (2010) suggested distinguishing three categories: positive intensification, negative intensification, and reinforcement. The corresponding acoustic profiles are sketched in figure 8.11.

Positive intensification is phonetically based on the strengthening of sonorous features in the accented syllable. For instance, the accented vowel is strongly lengthened, partly at the expense of the preceding consonant. The F0 rise into the vowel freezes at a high level and creates an extensive plateau that is left around the end of the vowel in a shallow fall. The acoustic energy contour also increases and decreases slowly at the onsets and offsets of the vowel, in this way paralleling the F0 deceleration. In contrast, negative intensification weakens sonorous features. This involves a quickly articulated accented vowel with large and abrupt acoustic energy changes at both edges and a pointed F0 peak that falls steeply to a low level right after the accented-vowel onset. A pressed and irregular voice quality adds to the "barking" impression given by the vowel. Unlike the vowel, the preceding consonant is strongly lengthened. Reinforcement is a kind of hybrid form that shares features of positive and negative intensification. Reinforcement resembles negative intensification in terms of the lengthened consonant at the accented-syllable onset and the high F0 and acoustic energy dynamics inside the accented vowel. However, as for positive intensification, F0 is rising into the vowel; the vowel is not shortened and shows a modal voice quality (Landgraf 2014).

Emphasis for intensity underscores the semantics of the accented word. But while positive and negative intensification target the word's valence, reinforcement is aimed

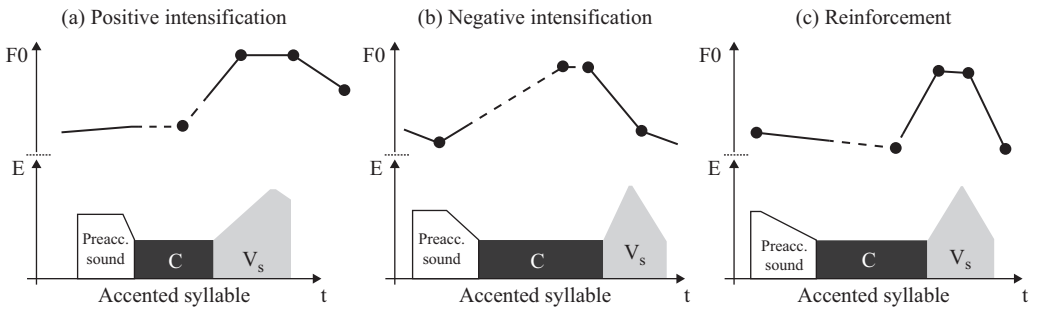


Figure 8.11
Prosodic profiles of the three emphatic categories in the KIM that aim at intensifying lexical semantics, particularly of adjectives: (a) positive intensification, (b) negative intensification, and (c) reinforcement. Preacc = preaccented.

at the truthfulness of the information represented by the accented word. Depending on the emphasis category, a sentence such as "This was a HUGE traffic jam" can additionally convey the following: "That's awesome. I have never seen so many cars before in one place!" (positive intensification); "Oh, crap; this will take forever!" (negative intensification); or "You can trust my assessment. I have seen a lot of traffic jams in life!" (reinforcement). Corpus analyses showed that emphasis for intensity is not restricted to adjectives like *HUGE*, although this is the most frequently used word class. Moreover, all three emphasis categories basically occur with all categories of pitch accent. Still, early peaks and nonearly peaks are more frequently associated with negative or non-negative intensification, respectively.

In PROLAB, emphasis categories are represented by capital letters placed between the phonological prominence level (typically <3>) and the pitch-accent category. In the case of emphasis for intensity, the PROLAB labels are <P> for positive intensification, <N> for negative intensification, and <R> for reinforcement.

8.4.7.2 Emphasis for contact One type of emphasis for contact refers to the pairs of intonational plateaus that are also known as "calling contours" (henceforth *nuclear one-step down contours*). Production and perception studies showed that there are actually three different categories of such nuclear one-step-down contours: they express reluctance, harmony, or disharmony (Niebuhr 2015).

One-step-down contours signaling reluctance are characterized by a medium-sized rise to the first plateau, followed by a tiny downward step of about one or two semitones to the second plateau. The second plateau is much longer than the first. Both plateaus have similarly high acoustic energy levels and start similarly "late" in the vowels of the accented and postaccented syllables. The plateau sequence expressing harmony is also characterized by a medium-sized rise up to the first plateau, but the step down to the second plateau is about twice as large as in the reluctance condition, that is, between three and four semitones. Both plateau durations and acoustic energy levels decrease from first to second plateau. Harmony plateaus start almost simultaneously with the respective vowel onset. Disharmony differs from harmony in that duration and acoustic energy levels increase rather than decrease from first to second plateau. Moreover, compared with harmony, disharmony is associated with larger upward and downward pitch steps to the first and second plateaus. The plateau alignment is intermediate between those of harmony and reluctance.

A second type of emphasis for contact adds another rise to a nuclear peak contour, in this way creating a rising-falling-rising intonational nucleus. Functionally, these nuclear contours are manipulative in the sense that they infringe on the interlocutor's self-determination. This can be done for three different purposes that have clearly distinct prosodic profiles and are thus analyzed as different emphasis categories: threatening, pleading, and confessing. *Threatening* is prosodically characterized by fast and extensive F0, duration, and intensity changes. In brief, syllables with rising or high pitch are long and loud; syllables with falling or low pitch are short and soft. *Pleading* is based on strong lengthening of the accented syllable, a very high reaching phrase-final rise, and a constantly high acoustic energy level. *Confessing* involves a strong lengthening of unaccented syllables and a reduced F0 range at a constantly high level.

Given that emphasis categories are indicated in PROLAB by a capital letter in between accent level and pitch-accent category, an obvious decision was to label the three one-step contours by <W> for reluctance ("<W>iderwille" in German), <H> for harmony, and <D> for disharmony. The three rise-fall-rise contours receive the letters

<T> for threatening, <B> for pleading ("<B>itten" in German), and <C> for confessing. So possible PROLAB annotations would be <3W^> and <3T(>, for example.

8.4.7.3 Emphasis for attention The class of emphasis-for-attention categories is the least well investigated in the KIM. At the current state-of-the-art, it includes two pairs of categories based on the reiteration of either words (e.g., "very very nice!" or "go go!") or pitch-accent patterns (produced in a syllable-by-syllable or foot-by-foot fashion, as in "eve-ry-sin-gle-day!" (see the chanting concept in Auer, Couper-Kuhlen, and Müller 1999). These reiterations are integrated into an overarching, coherent prosodic structure (e.g., Niebuhr et al. 2012). An additional type of phrase-level phenomenon subsumed under emphasis for attention is what Strangert (2003) called "emphasis by pausing." Common to all types and categories of emphasis for attention is that the extra effort is spent on disrupting the fluency or rhythmicity of an utterance. However, as the research on the forms and functions of all these phenomena is still ongoing, it would be premature to specify the exact number of categories and suggest corresponding PROLAB labels here.

8.5 Working with the KIM

Researchers who are interested in working with the KIM can find all relevant details about the syntactic rules of PROLAB in the KIM-Trainer compiled by Peters and Kohler (2004). The trainer also provides instructive examples and practical exercises, and it summarizes previous experiences with PROLAB annotations. Only some very basic points of the trainer should be mentioned here.

8.5.1 The Annotation Cycle

As is suggested by Kügler et al. (2015), annotators should start by determining the location and categories of phrase boundaries. Then the corresponding speech section should be listened to several times to define its prominence structure in terms of the four prominence levels. Each lexically stressed syllable of a word is assigned a number between <0> and <3>. The labels are set at the vowel onsets of the stressed syllables. In the following step, pitch-accent categories should be determined for all stressed syllables with prominence levels above 0. The labels of accent categories are put after the prominence number. In the case of an upstepped pitch accent, <|> is added before the prominence level in this step of the annotation. The final step of the annotation fills the gaps between the pitch accents. That is, phrase-initial, phrase-final, and concatenation contour categories are identified, and the corresponding PROLAB labels are added. The phrase-initial and -final labels are set at the beginning or end of the prosodic phrase, respectively, so phrase-initial/-final contour labels and phrase-boundary labels share the same time stamp. Labels of concatenation contours are placed at the same point in time as the pitch accent at which the concatenation contour ends (the second of two pitch accents or the rightmost one in a sequence of three or more pitch accents).

8.5.2 Potential Problems

Although it was found that the prominence level distinctions offered by the KIM can be reliably annotated, prominence levels are sometimes hard to specify. To solve this problem, it proved helpful to first listen to a longer stretch of the speech material to be annotated. This gives the annotator an impression of how the reduced prominence

level <1> sounds in comparison to levels <0> and <2> for that inidividual speaker. The same also applies to hat patterns in comparison to dip patterns and pseudo-terminal in comparison to terminal and nonterminal phrase-final falls.

Another tricky distinction is that between peak and valley contours at the beginning of a hat pattern. The preliminary convention is to label high-reaching rises, particularly those with a concave shape, as valley contours and all other hat-pattern initial rises as peak contours.

In general, it is highly recommended that annotators familiarize themselves with the categories of the model and also learn to realize them in contrast to one another. On this basis, annotators can imitate what the speaker produced and, moreover, contrast the intonation category in question with its paradigmatic neighbors. Such ad hoc contrasts often facilitate the decision-making process in prosodic annotation. However, learning to produce all intonation categories and contrasts requires considerable supervised training and can easily take several months.

8.6 Perspectives for Further Developments

This section briefly summarizes findings that have been obtained within the KIM framework but which are only insufficiently or not at all taken into account by the current version of the model. The findings come from three different fields of research.

8.6.1 Vowel Boundaries Are Not Directly Relevant

A key assumption of the KIM is that the peak and valley pitch-accent contours are distinguished by the synchronization of contour points relative to onset and/or offset of the accented vowel. However, a growing body of evidence undermines this vowel-related phonological concept. For example, for a stepwise peak shift across the C-V boundary of the accented syllable, pitch-accent identification changes sooner from early to medial, when the intensity difference between consonant (C) and accented vowel (V) is smaller, that is, when the consonant already has almost the same intensity level as the following vowel. Moreover, the identification change from early to medial is more abrupt when the intensity increase into the vowel is faster. Analogous effects of magnitude and slope of the intensity change on pitch-accent identification were found for a stepwise peak shift across the V-C boundary of the accented syllable and the corresponding change from medial to late pitch-accent category (Niebuhr 2006, 2007a).

These findings indicate that it is not the vowel boundary in terms of the change in formant pattern that matters in the synchronization and identification of pitch accents. Rather, what matters are the increases and decreases in intensity caused by the accented vowel and how the pitch-accent contour is positioned (and even shaped) relative to the resulting areas of higher and lower intensity. Accordingly, all other things being equal, different accented-syllable compositions lead to different identification boundaries of pitch-accent categories within the same peak-shift continuum. This fact is not least reflected in the similar effects of syllable structure on pitch-accent alignment that were measured across many languages. Shifting the focus in synchronization from sound segments to their intensity patterns also takes into account that pitch-accent categories can also be realized and distinguished on interjections like “mmm,” which have no separate vowel segment. Speakers produce the prolonged nasal [m:] in “mmm” with a rising-falling intensity pattern, and shape and synchronize the pitch-accent contour relative to this intensity pattern, just as in a CVC syllable (Kirsch 2015).

8.6.2 Intonation beyond F0, Part 1: Pitch-Accent Specific Micro-Rhythms

Pitch accents are not simply put on top of a string of sound segments. Rather, they change the durations and intensity levels in the environment in which they occur, that is, the triplet of preaccented, accented, and postaccented syllable. Pitch accents create an imprint on this triplet in the form of a pitch-accent–specific micro-rhythm. Compared with a medial pitch accent, for which the preaccented and postaccented syllables are rhythmically symmetrical, that is, they have about the same low duration and intensity levels and hence appear similarly nonprominent, the early pitch accent slightly strengthens the preaccented syllable by making it longer and louder and further weakens the postaccented syllable by making it even shorter and softer. The opposite applies to the late pitch accent.

These prominence patterns and the resulting micro-rhythms in the triplet of preaccented, accented, and postaccented syllable also emerged in a finger-tapping task: speakers are asked to produce short test utterances with early, medial, and late peaks while tapping their index finger for each syllable on a device that records length and strength of these finger-tappings (Niebuhr 2019). Niebuhr further showed, in line with Niebuhr and Pfitzinger (2010), that listeners can identify early, medial, and late pitch accents only by hearing their corresponding micro-rhythms. Moreover, combining pitch accent peaks in early, medial, and late synchronization with mismatching micro-rhythms (in resynthesized stimuli), diminishes pitch-accent identification and makes the corresponding stimulus utterances sound less natural.

8.6.3 Intonation beyond F0, Part 2: Segmental Intonation

Voiceless fricative sounds, including postaspiration sections, all convey aperiodic pitch impressions. A number of acoustic analyses across languages and language varieties showed that the spectral patterns of such fricative sounds do not remain constant in different F0 contexts. Rather, the spectral energy distribution of a fricative sound varies such that the aperiodic pitch impression it creates changes in parallel to the adjacent F0 context (Niebuhr 2008, 2009, 2012; Ritter and Röttger 2014; Zygis et al. 2014). Projected onto English, this would mean, for example, that the /ʃ/ in the middle of *mushroom* sounds higher in the context of a high pitch-accent peak and lower in the context of a low pitch-accent valley. Likewise, the post-aspiration at the end of *lack* would sound higher in the context of a phrase-final F0 rise but lower in the context of a phrase-final F0 fall.

In addition, perception experiments provided evidence that listeners can integrate context-adequate aperiodic pitch impressions in their perception of intonation contours, for example, in that they fill in utterance-internal F0 gaps or restore truncated phrase-final F0 movements (Kohler 2011; Mixdorff and Niebuhr 2013). Particularly with respect to this potential perceptual relevance, F0 or pitch-oriented fricative variation in speech was termed *segmental intonation*.

8.6.4 Implications for the Further Development of the KIM

In order to cope with the new empirical findings in sections 8.6.1 to 8.6.3, the KIM must consistently implement the principles on which it is based. The current KIM is inconsistent insofar as it sets up phonological categories based on groups of listeners and the intonational meanings they identify. But it then defines these phonological categories by acoustic rather than perceptual features. Furthermore, the KIM takes an intermediate position in the conceptual issue of contour and tonal-target (level) models. It places contours in the center of the phonology, but at the same time decomposes

these contours into single-movement features or local F0 landmarks, for example, in the case of pitch-accent and concatenation categories. However, this "pattern approach" is not pursued as rigorously as it could be.

So, in the opinion of this author, the further development of the KIM requires two steps:

1. With respect to the model's *paradigmatic* structure, the focus in defining intonational contrasts must be shifted from acoustics to perception.
2. With respect to the model's *syntagmatic* structure, the focus in defining the building blocks of intonation must be shifted from a heterogeneous mixture of contours and local F0 landmarks to Gestalt-like patterns of perceived pitch.

An initial basis for this further development was laid out by Niebuhr (2007b) in the *contrast theory of intonation perception* (C-TIP), see also Barnes et al. (2020). The theory capitalizes on the fact that perception follows the same principles across modalities. Therefore, the construction of intonational categories can be conceived in analogy to the relatively well-understood construction of visual objects. A basic assumption of the contrast theory is that intonational categories consist of two gestalt-like patterns: a pitch gestalt and a prominence gestalt that is associated with the pitch gestalt.

Pitch gestalts are constituted by tonal elements. According to the "spectral constraint" and "tonal movement coding" principles of House (1990, 1996), the F0 course is perceptually broken down by the listeners into a sequence of pitch levels and pitch movements. Pitch levels are directly translated into single tonal elements; pitch movements are represented by two tonal elements, one at each end of the movement. Aperiodic pitch impressions and their respective tonal elements are integrated in the pitch gestalt. The number of perceived movements and level tones and hence the total number of tonal elements within a gestalt can vary (e.g., due to the duration of F0 slopes and the coinciding segment/formant structures), but this variation is phonologically irrelevant as long as the gestalt pattern as a whole remains constant.

In the case of the KIM's three oldest pitch-accent categories (early, medial, and late peak), the pitch gestalt is always the same: rising-falling. Figure 8.12a illustrates the simplest rising-falling pattern, based on three tonal elements. It also shows that the phonological differences among early, medial, and late peaks lie in the prominence gestalt associated with the pitch pattern. This pattern is created in that each tonal element of the pitch gestalt is contrasted with the preceding element (within the same gestalt unit) in terms of its pitch, length, and loudness properties. These multidimensional backward contrasts and the concomitant contrast enhancements and top-down expectations create a specific prominence level for each tonal element and, hence, a specific prominence pattern across all elements.

This key mechanism explains effects of synchronization, shape, and their interplay with syllable structure. For example, changes in F0 peak shape cause changes in the duration of tonal elements, and longer tonal elements are perceived as being more prominent than shorter tonal elements. Even more important from the perspective of the current KIM, synchronization is no longer a direct phonological feature in the C-TIP. Rather, synchronization is merely considered *the most efficient and effective strategy for increasing* the relative prominence of a tonal element in the pitch gestalt by exploiting the given high-intensity area of the accented vowel. In terms of the metaphor displayed in figure 8.12b, synchronizing of F0 peaks (or valleys) relative to the segmental string is about moving tonal elements "into the light" of the (high-intensity)

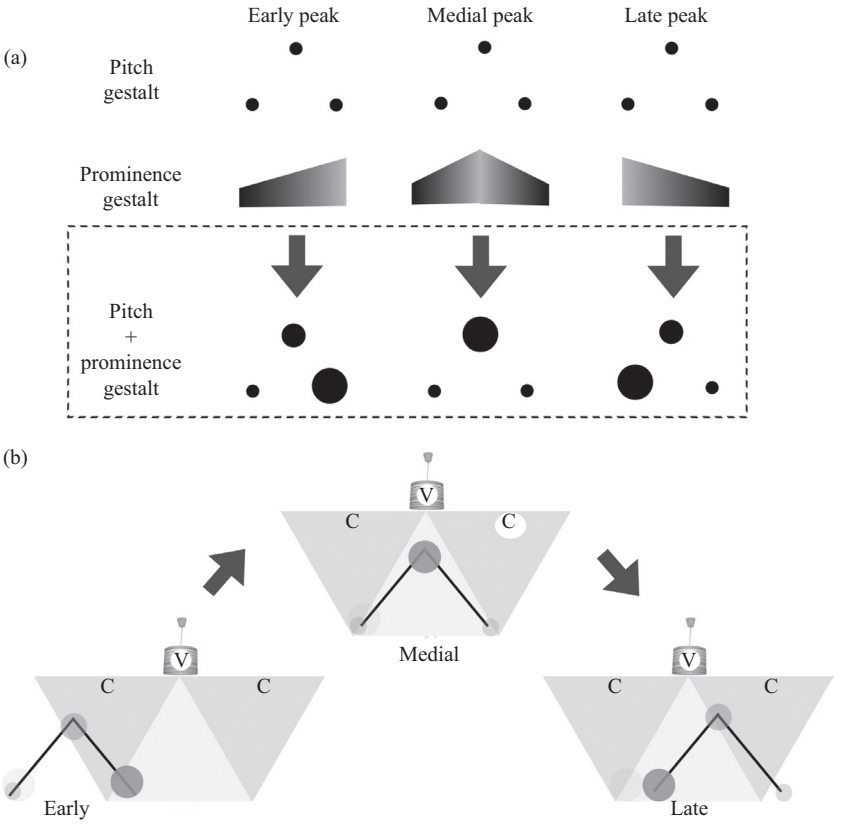


Figure 8.12
(a) Representation of pitch-accent peaks in the contrast theory of intonation perception (Niebuhr 2007b). (b) Implications for the role of synchronization in pitch-accent identification.

vowel section, or out of the light into the "darker" (low-intensity) sections of adjacent consonants.

The metaphor in figure 8.12b also implies, in line with the empirical evidence in section 8.6.1, that the accented-vowel boundaries are important points of reference in intonational phonology. Yet these points of reference are not absolute. There is more leeway for peak/valley synchronization relative to accented-vowel boundaries when the surrounding (preceding or following) consonants are already quite "light," as in the case of approximants. Moreover, pitch accents synchronized closely to the vowel boundaries are identified less clearly when the intensity increases or decreases from or to the surrounding consonants are shallower. Ultimately, speakers can decide not to make use of synchronization at all and instead change the duration and intensity pattern of the segments that coincide with the tonal elements or the shape of the peak. Speakers can also make use of all options at the same time and to different degrees. This individual or context-specific behavior emerges in production data, for example, in the form of a continuum from "F0 shapers" to "F0 aligners" (Niebuhr et al. 2011) or in the alignment changes under different syllable-structure conditions. The assumption of a separate prominence gestalt is also consistent with the creation of pitch-accent specific micro-rhythms (see section 8.6.2) and explains why listeners are able associate

rhythmic sequences of monotonous tones with the corresponding pitch-accent categories of early, medial, and late peak.

8.7 Conclusion

The simple intonational concepts of the KIM, as well as of many other intonation models, appear increasingly outdated against our growing understanding of the actual complexity of intonation. Producing and perceiving intonational categories involves a complex interplay of F0 variables and reaches beyond F0 into other prosodic and even segmental parameters. This complexity calls for a shift in intonational modeling from the acoustic to the perceptual domain, in which the multitude of acoustic factors, their variation, and effects may become understandable as simple trade-offs in patterns of perceived pitch and prominence. This idea will shape the future development of the KIM. It is in many respects similar to the intensity-weighted tonal-center-of-gravity approach developed independently by Barnes et al. (2010, 2012, 2020) from an autosegmental-metrical starting point. Implementing and potentially combining these perception-oriented lines of research is undoubtedly worthwhile, but will be challenging long-term issues, next to establishing and refining the phonological categories of intonation, emphasis, and prominence themselves. Linking phonology and modeling will raise the additional question of whether perception-oriented ideas and concepts are applicable in the same way to all paradigms and categories.

References

Ambrazaitis, G. I. 2005. "Between Fall and Fall-Rise: Substance-Function Relations in German Phrase-Final Intonation Contours." *Phonetica* 62:196–214.

Ambrazaitis, G. I. 2009. *Nuclear Intonation in Swedish: Evidence from Experimental-Phonetic Studies and a Comparison with German* (*Travaux de l'institut de linguistique de Lund* Volume 49). Lund: E-Husets.

Ambrazaitis, G. I., and O. Niebuhr. 2008. "Dip and Hat Pattern: A Phonological Contrast of German?" Paper presented at the Fourth International Conference of Speech Prosody, Campinas, Brazil, May 6–9, 269–272.

Ambrazaitis, G. I., and O. Niebuhr. 2014. "Dip vs. Hat Pattern: Is the Intonation Pattern between Two Pitch Accents a Phonological Feature of German?" Poster presented at Lund Symposium 2014, Lund, Sweden, December 7, 2014.

Arnold, D., P. Wagner, and B. Möbius. 2011. "Evaluating Different Rating Scales for Obtaining Judgments of Syllable Prominence from Naïve Listeners." Paper presented at Proceedings of the Sixteenth International Congress of Phonetic Sciences, August 17–21, 252–255.

Asu, E. L. 2006. "Rising Intonation in Estonian: An Analysis of Map Task Dialogues and Spontaneous Conversations." Paper presented at the Fonetiikan Päivät 2006/The Phonetics Symposium 2006, Helsinki, Finland, August 30–31, 1–8.

Auer, P., E. Couper-Kuhlen, and F. Müller, F. 1999. *Language in Time: The Rhythm and Tempo of Spoken Interaction*. New York: Oxford University Press.

Barnes, J., H. Mixdorff, and O. Niebuhr. 2020. "Phonetic Variation in Tone and Intonation Systems." In *The Oxford Handbook of Language Prosody*, edited by C. Gussenhoven and A. Chen, 125–149. Oxford: Oxford University Press.

Barnes, J., N. Veilleux, A. Brugos, and S. Shattuck-Hufnagel. 2010. "The Effect of Global F0 Contour Shape on the Perception of Tonal Timing Contrasts in American English Intonation." Paper presented at the Fifth International Conference of Speech Prosody, Chicago, IL, USA, May 10–14, 1–4.

Barnes, J., N. A. Veilleux, A. Brugos, and S. Shattuck-Hufnagel. 2012. "Tonal Center of Gravity: A Global Approach to Tonal Implementation in a Level-Based Intonational Phonology." *Laboratory Phonology* 3:337–383.

Barry, W. J. 1981. "Prosodic Functions Revisited Again!" *Phonetica* 38:120–134.

Bloch, B. 1948. "A Set of Postulates for Phonemic Analysis." *Language* 24:3–46.

Boersma, P. 2001. "Praat, a System for Doing Phonetics by Computer." *Glot International* 5:341–345.

Bruce, G. 1977. *Swedish Word Accents in Sentence Perspective*. Lund, Sweden: Gleerup.

Bryant, G. A., and J. Fox Tree 2005. "Is There an Ironic Tone of Voice?" *Language and Speech* 48:257–277.

Bühler, K. 1934. *Sprachtheorie. Die Darstellungsfunktion der Sprache*. Jena, Germany: Fischer.

Carlson, R., B. Granström, and S. Hunnicut, S. 1990. "Multilingual Text-to-Speech Development and Applications." In *Advances in Speech, Hearing, and Language Processing*, edited by W. A. Ainsworth, 269–296. London: JAI Press.

de Saussure, F. 1916. *Cours de linguistique général*. Paris: Payot.

Dombrowski, E. 2003. "Semantic Features of Accent Contours: Effects of F0 Peak Position and F0 Time Shape." Paper presented at the Fifteenth International Congress of Phonetic Sciences, Barcelona, Spain, August 3–9, 1217–1220.

Dombrowski, E., and O. Niebuhr. 2005. "Acoustic Patterns and Communicative Functions of Phrase-Final Rises in German: Activating and Restricting Contours." *Phonetica* 62:176–195.

Dombrowski, E., and O. Niebuhr. 2010. "Shaping Phrase-Final Rising Intonation in German." Paper presented at the Fifth International Conference on Speech Prosody, Chicago, IL, USA, May 10–14, 1–4.

Fant, G., and A. Krukenberg. 1989. "Preliminaries to the Study of Swedish Prose Reading and Reading Style." *Speech Transmission Laboratory—Quarterly Progress and Status Report, Royal Institute of Technology, Stockholm* 2:1–83.

Görs, K., and O. Niebuhr. 2012. "Hocus Focus: What the Elicitation Method Tells Us about Types and Exponents of Contrastive Focus." Paper presented at the Sixth International Conference of Speech Prosody, Shanghai, China, May 22–25, 262–265.

Grice, M., and S. Baumann. 2002. "Deutsche Intonation und GToBI." *Linguistische Berichte* 191:267–298.

Grice, M., D. R. Ladd, and A. Arvaniti. 2000. "On the Place of Phrase Accents in Intonational Phonology." *Phonology* 17:143–185.

Gussenhoven, C. 1999. "Discreteness and Gradience in Intonational Contrasts." *Language and Speech* 42:283–305.

Halliday, M. A. K. 1967. *Intonation and Grammar in British English*. The Hague, the Netherlands: Mouton.

Heldner, M. 2003. "On the Reliability of Overall Intensity and Spectral Emphasis as Acoustic Correlates of Focal Accents in Swedish." *Journal of Phonetics* 31:39–62.

Hirschberg, J. 2002. "The Pragmatics of Intonational Meaning." Paper presented at the First International Conference of Speech Prosody, Aix-en-Provence, France, April 11–13, 65–68.

Hirschberg, J. 2004. "Pragmatics and Intonation." In *Handbook of Pragmatics*, edited by L. R. Horn and G. Ward, 515–537. Oxford: Basil Blackwell.

House, D. 1990. "Tonal Perception in Speech." *Travaux de l'institute de linguistique de Lund* 24:7–163.

House, D. 1996. "Differential Perception of Tonal Contours through the Syllable." Paper presented at the Fourth International Conference of Spoken Language Processing, Philadelphia, PA, USA, October 3–6, 2048–2051.

John, T. 2012. "EMU Speech Database System: Praxisorientierte Weiterentwicklung der Funktionalität, Benutzerfreundlichkeit und Interoperabilität sowie die Aufbereitung des Kiel Corpus als EMU-Sprachdatenbank." PhD diss., Ludwig-Maximilans-Universität Munich.

Jones, D. 1969. *An Outline of English Phonetics*. 9th ed. Cambridge: Heffer.

Kirsch, A. 2015. "Die Produktion und Perzeption früher, mittlerer und später Gipfel über `mmm." BA thesis, Kiel University.

Kleber, F. 2006. "Form and Function of Falling Pitch Contours in English." Paper presented at the Third International Conference of Speech Prosody, Dresden, Germany, May 2–5, 61–64.

Kohler, K. J. 1987. "Categorical Pitch Perception. Paper presented at the Eleventh International Congress of Phonetic Sciences, Tallinn, Estonia, August 1–7, 331–333.

Kohler, K. J. 1991a. "A Model of German Intonation." *Arbeitsberichte des Instituts für Phonetik und Digitale Sprachverarbeitung (AIPUK)* 25:295–360.

Kohler, K. J. 1991b. "Prosody in Speech Synthesis: The Interplay between Basic Research and TTS Application." *Journal of Phonetics* 19:121–138.

Kohler, K. J. 1996. "Labelled Data Bank of Spoken Standard German: The Kiel Corpus of Read/Spontaneous Speech." Paper presented at the Fourth International Conference on Spoken Language Processing, Philadelphia, PA, USA, October 3–6, 1938–1941.

Kohler, K. J. 1997. "Modelling Prosody in Spontaneous Speech." In *Computing Prosody*: *Computational Models for Processing Spontaneous Speech*, edited by Y. Sagisaka, N. Campbell, and N. Higuchi, 187–210. New York: Springer.

Kohler, K. J. 2004. "Pragmatic and Attitudinal Meanings of Pitch Patterns in German Syntactically Marked Questions. In *From Traditional Phonology to Modern Speech Processing*: *Festschrift for Professor Wu Zongji's 95th Birthday*, edited by G. Fant, H. Fujisaki, J. Cao, and Y. Xu, 205–214. Beijing: Foreign Language Teaching and Research Press.

Kohler, K. J. 2005. "Timing and Communicative Functions of Pitch Contours." *Phonetica* 62:88–105.

Kohler, K. J. 2006. "What Is Emphasis and How Is It Coded?" Paper presented at the Third International Conference of Speech Prosody, Dresden, Germany, May 2–5, 748–751.

Kohler, K. J. 2007. "Data Acquisition and Modelling in Speech Communication." Paper presented at the ParaLing07, Sixteenth International Congress of Phonetic Sciences, Saarbrücken, Germany, August 5–10, 1–6.

Kohler, K. J. 2011. "Communicative Functions Integrate Segments in Prosodies and Prosodies in Segments." *Phonetica* 68:26–56.

Kügler, F., B. Smolibocki, S. Baumann, B. Braun, M. Grice, S. Jannedy, O. Niebuhr, et al. 2015. "DIMA: Annotation Guidelines for German Intonation." Paper presented at the Seventeenth International Congress of Phonetic Sciences, Glasgow, Scotland, August 10–14.

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

Landgraf, R. 2014. "Are You Serious? Irony and the Perception of Emphatic Intensification." Paper presented at the Fourth International Symposium on Tonal Aspects of Languages, Nijmegen, The Netherlands, May 13–16, 91–94.

Loevenbruck, H., M. Ben Jannet, M. D'Imperio, M, Spini, and M. Champagne-Lavau. 2013. "Prosodic Cues of Sarcastic Speech in French: Slower, Higher, Wider." Paper presented at the Fourteenth Interspeech Conference, Lyon, France, August 25–29, 1470–1474.

Michalsky, J. 2014. "Scaling of Final Rises in German Questions and Statements." Paper presented at the Seventh International Conference of Speech Prosody, Dublin, Ireland, May 20–23, 978–982.

Mixdorff, H., and O. Niebuhr. 2013. "The Influence of F0 Contour Continuity on Prominence Perception." Paper presented at the Fourteenth Interspeech Conference, Lyon, France, August 25–29, 230–234.

Nash, R., and A. Mulac. 1980. "The Intonation of Verifiability." In *The Melody of Language: Intonation and Prosody*, edited by L. R. Waugh and C. H. van Schooneveld, 219–241. Baltimore: University Park Press.

Newman, D. 2016. "Book Review of Instrumental Studies in Arabic Phonetics." *Phonetica* 73, 79–82.

Niebuhr, O. 2006. "The Role of the Accented-Vowel Onset in the Perception of German Early and Medial Peaks." Paper presented at the Third International Conference of Speech Prosody, Dresden, Germany, May 2–5, 109–112.

Niebuhr, O. 2007a. "Categorical Perception in Intonation: A Matter of Signal Dynamics?" Paper presented at the Eighth Interspeech Conference, Antwerp, Belgium, August 27–31, 109–112.

Niebuhr, O. 2007b. *Perzeption und kognitive Verarbeitung der Sprechmelodie. Theoretische Grundlagen und empirische Untersuchungen.* Berlin: de Gruyter.

Niebuhr, O. 2007c. "The Signalling of German Rising-Falling Intonation Categories: The Interplay of Synchronization, Shape, and Height." *Phonetica* 64:174–193.Niebuhr, O. 2008. "Coding of Intonational Meanings beyond F0: Evidence from Utterance-Final /t/ Aspiration in German." *Journal of the Acoustic Society of America* 142:1252–1263.

Niebuhr, O. 2009. "Intonation Segments and Segmental Intonations. Paper presented at the Tenth Interspeech Conference, Brighton, UK, September 6–10, 2435–2438.

Niebuhr, O. 2011. "Alignment and Pitch-Accent Identification: Implications from F0 Peak and Plateau Contours." *Arbeitsberichte des Instituts für Phonetik und Digitale Sprachverarbeitung (AIPUK)* 38:77–95.

Niebuhr, O. 2012. "At the Edge of Intonation: The Interplay of Utterance-Final F0 Movements and Voiceless Fricative Sounds." *Phonetica* 69:7–27.

Niebuhr, O. 2015. "Stepped Intonation Contours: A New Field of Complexity." In *Tackling the Complexity of Speech*, edited by O. Niebuhr and R. Skarnitzl, 9–74. Prague: Nakladatelství Epocha.

Niebuhr, O. 2019. "Pitch Accents as Multiparametric Configurations of Prosodic Features—Evidence from Pitch-Accent Specific Micro-rhythms in German." In *A Sound Approach to*

Language Matters: In Honor of Ocke-Schwen Bohn, 321–351. AU Library Scholarly Publishing Services. https://doi.org/10.7146/aul.322.218.

Niebuhr, O., M. D'Imperio, B. Gili Fivela, and F. Cangemi. 2011. "Are There 'Shapers' and 'Aligners'? Individual Differences in Signalling Pitch Accent Category." Paper presented at the Seventeenth International Congress of Phonetic Sciences, Hong Kong, China, August 17–21, 120–123.

Niebuhr, O., P. Jarzabkowska, U. Lorenz, C. Schulz, and F. Sodigov. 2012. "Say It Again, Sam! Phonetic Forms and Functions of Emphatic Reduplication in German." Paper presented at the Sixth International Conference of Speech Prosody, Shanghai, China, May 22–25, 258–261.O'Connor, J. D., and J. F. Arnold. 1970. Intonation of Colloquial English. London: Longman.

Niebuhr, O., and K. J. Kohler. 2004. "Perception and Cognitive Processing of Tonal Alignment in German." Paper presented at the First International Symposium on Tonal Aspects of Languages, Nanjing, China, March 28–31, 155–158.

Niebuhr, O., and H. P. Pfitzinger. 2010. "On Pitch-Accent Identification; the Role of Syllable Duration and Intensity." Paper presented at the Fifth International Conference on Speech Prosody, Chicago, IL, USA, May 10–14, 1–4.

Niebuhr, O., B. Schroeder, and S. Baumann. 2015. "Cues to Perceptual Prominence of Prenuclear Pitch Accents: Evidence from German Spontaneous Speech." In *Proceedings of the International Conference on Prominence in Language*, 1–2.

Niebuhr, O., and M. Zellers. 2012. "Late Pitch Accents in Hat and Dip Intonation Patterns." In *Understanding Prosody: The Role of Context, Function and Communication*, edited by O. Niebuhr, 159–186. Berlin: de Gruyter.

Osgood, C. E., G. J. Suci, and P. H. Tannenbaum. 1957. *The Measurement of Meaning*. Champaign: University of Illinois Press.

Peters, B. 2000. "Individuelle und geschlechtsspezifische Unterschiede in der prosodischen Gestaltung deutscher Lese- und Spontansprache." *Sprache und Kultur, Forum Angewandte Linguistik* 38:153–162.

Peters, B. 2005. "The Database: The Kiel Corpus of Spontaneous Speech." *Arbeitsberichte des Instituts für Phonetik und Digitale Sprachverarbeitung (AIPUK)* 35a:1–6.

Peters, B. 2012. *Form und Funktion prosodischer Grenzen im Gespräch: Ein phonetischer Beitrag zur Gesprächsforschung*. Saarbrücken, Germany: Südwestdeutscher Verlag für Hochschulschriften.

Peters, B., and K. J. Kohler. 2004. "Trainingsmaterialien zur prosodischen Etikettierung mit dem Kieler Intonationsmodell KIM." www.ipds.uni-kiel.de/kjk/pub_exx/bpkk2004_1/TrainerA4.pdf.

Peters, B., K. J. Kohler, and T. Wesener. 2005. "Melodische Satzakzentmuster in prosodischen Phrasen deutscher Spontansprache: Statistische Verteilung und sprachliche Funktion." *Arbeitsberichte des Instituts für Phonetik und Digitale Sprachverarbeitung (AIPUK)* 35a:7–54.

Petrone, C., and O. Niebuhr. 2013. "On the Intonation in German Intonation Questions: The Role of the Prenuclear Region." *Language and Speech* 57:108–146.

Rietveld, T., and A. Chen. 2006. "How to Obtain and Process Perceptual Judgements of Intonational Meaning." In Methods in Empirical Prosody Research, edited by S. Sudhoff, D. Lenertová, R. Meyer, S. Pappert, P. Augurzky, I. Mleinek, N. Richter, and J. Schliesser, 283–319. Berlin: Mouton de Gruyter.

Ritter, S., and T. B. Roettger. 2014. "Speakers Modulate Noise-Induced Pitch according to Intonational Context." Paper presented at the Seventh International Conference of Speech Prosody, Dublin, Ireland, May 20–23, 1–5.

Strangert, E. 2003. "Emphasis by Pausing." Paper presented at the Fifteenth International Congress of Phonetic Sciences, Barcelona, Spain, August 3–9, 2477–2480.

Strömbergsson, S., J. Edlund, and D. House. 2012. "Prosodic Measurements and Question Types in the Spontal Corpus of Swedish Dialogues." Paper presented at the Thirteenth Interspeech Conference, Portland, OR, USA, September 9–13, 839–842.

Trask, R. L. 2004. *A Dictionary of Phonetics and Phonology*. New York: Routledge.

Trubetzkoy, N. S. 1939. *Grundzüge der Phonologie*. Travaux du Cercle Linguistique de Prague, vol. 7. Prague: Cercle Linguistique de Prague.

von Thun, F. S. 1981. *Miteinander reden: Störungen und Klärungen. Psychologie der zwischenmenschlichen Kommunikation*. Reinbek, Germany: Rowohlt.

Wahlster, W., and J. Engelkamp. 1992. *Scientific Goals and Networks of Work Packages of the Verbmobil Project*. Saarbrücken, Germany: Deutsches Forschungszentrum für Künstliche Intelligenz.

Winchester, S. 2003. *The Meaning of Everything: The Story of the Oxford English Dictionary*. Oxford: Oxford University Press.

Xu, Y. 2004. "The PENTA Model of Speech Melody: Transmitting Multiple Communicative Functions in Parallel." Paper presented at the International Symposium from Sound to Sense, Boston, MA, USA, June 11–13, 91–96.

Żygis, M., D. Pape, L. M. T. Jesus, and M. Jaskuła. 2014. "Intonation of Statements and Polar Questions in Whispered, Semi-Whispered and Normal Speech Modes." Paper presented at the Seventh International Conference of Speech Prosody, Dublin, Ireland, May 20–23, 1–5.

The MIT Press would like to thank the anonymous peer reviewers who provided comments on drafts of this book. The generous work of academic experts is essential for establishing the authority and quality of our publications. We acknowledge with gratitude the contributions of these otherwise uncredited readers.

This book was set in Stone Serif and Stone Sans by Westchester Publishing Services.

Library of Congress Cataloging-in-Publication Data is available.
Names: Barnes, Jonathan, 1970– editor. | Shattuck-Hufnagel, Stefanie, editor.
Title: Prosodic theory and practice / edited by Jonathan Barnes and Stefanie Shattuck-Hufnagel.
Description: Cambridge, Massachusetts : The MIT Press, 2022. | Includes bibliographical references and index.
Identifiers: LCCN 2021000764 | ISBN 9780262543170 (paperback)
Subjects: LCSH: Prosodic analysis (Linguistics)
Classification: LCC P224 .P739 2022 | DDC 414/.6—dc23
LC record available at https://lccn.loc.gov/2021000764