



Modeling Dimensions of Prosodic Prominence

Simon Roessig* and Doris Mücke

Institut für Linguistik – Phonetik, Universität zu Köln, Cologne, Germany

Detailed modifications both in the laryngeal as well as in the supra-laryngeal domain have been shown to be used by speakers of German to express prosodic prominence. This paper aims to bring the two domains together in a joint analysis and modeling account. We report results on the prosodic marking of focus types from 27 speakers that were recorded acoustically and with electromagnetic articulography. We investigate the intonational patterns (tonal onglide) as well as the articulatory movements during the vowel production (lip aperture and tongue body position). We provide further evidence for categorical and continuous modifications across and within accentuation and sketch a dynamical model that accounts for these modifications on multiple dimensions as the consequence of scaling the same parameter. In this model, the prosodic dimensions contribute differently to the complex shape of the compositional attractor landscape and respond differently to the scaling of the system. The study aims to add to our understanding of the integration of speech sounds in a two-fold manner: the integration of different channels of prosody (laryngeal and supra-laryngeal) as well as the interplay of categorical and continuous aspects of speech.

OPEN ACCESS

Edited by:

Adamantios Gafos,
University of Potsdam, Germany

Reviewed by:

Argyro Katsika,
University of California, Santa Barbara,
United States

Mariapaola D'Imperio,
Aix-Marseille Université, France

*Correspondence:

Simon Roessig
mail@simonroessig.de

Specialty section:

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

Received: 15 March 2019

Accepted: 31 July 2019

Published: 10 September 2019

Citation:

Roessig S and Mücke D (2019)
Modeling Dimensions of Prosodic
Prominence. *Front. Commun.* 4:44.
doi: 10.3389/fcomm.2019.00044

Keywords: prosody, dynamical systems, articulation, intonation, speech production, attractors

INTRODUCTION

In the last decades, a growing body of research has pointed out the dynamical nature of the mind (e.g., Kelso, 1995; van Gelder and Port, 1995; Port, 2002; Spivey, 2007). To overcome limitations imposed by symbolic approaches, researchers from many disciplines have turned to the framework of dynamical systems describing a multitude of different cognitive processes including the production and perception of speech sounds and their cognitive representations (Browman and Goldstein, 1986; Tuller et al., 1994), organization of semantic knowledge (Mirman and Magnuson, 2009) as well as movement coordination (Haken et al., 1985).

In the fields of phonetics and phonology, the dynamical perspective has the potential to shed new light on the question of how the categorical and the continuous aspects of speech are related (Browman and Goldstein, 1986; Hawkins, 1992; Tuller et al., 1994; Port, 2002; Gafos, 2006; Gafos and Benus, 2006; Lancia and Winter, 2013; Roon and Gafos, 2016; Iskarous, 2017; Mücke, 2018). Phonology and phonetics have long been conceptualized as two separate modules with a process of translation to mediate between them. While phonology comprises the categorical representations of speech sounds and computations that operate on them (rules or ranked constraints), phonetics implements speech sounds in a physical representation. Thus, the translation from phonology to phonetics must be a process of transforming a discrete symbolic representation into a continuous signal. The division between phonology and phonetics into a discrete, symbolic domain on the one hand and a continuous, physical domain on the other is based on the observation that speech is

characterized by abstract mental categories and continuous signals at the same time. While this perspective of duality appears to be a plausible motivation for a clear-cut separation of phonology and phonetics at first sight, accumulating evidence shows that the categorical and the continuous sides of speech are deeply intertwined. Crucially, this evidence questions a purely categorical, abstract nature of phonological representations (Pierrehumbert et al., 2000; Port, 2006; Ladd, 2011; Pierrehumbert, 2016). The dynamical perspective of the mind does not posit a strict division between categorical and continuous aspects of speech production and perception. In this view, the mind works in a completely continuous manner—there are no pure, symbolic mental states (Spivey, 2007). While the mind is in constant flux, it gravitates toward relatively stable states, called attractors. These attractors are the analogs to categorical representations in the symbolic computation view. Since attractors are located in a fully continuous space that is not separated into discrete areas, it is sensible to talk of *quasi-categories* in the context of attractors. Crucially, the fact that attractors are part of the continuous state space of the system makes a translation from the categorical to the continuous superfluous. As such, speech sound categories can be represented as stable states on multiple continuous dimensions. While the notion of the attractor reflects the observation that these categories are relatively stable, the continuous nature of the system allows for fine-grained variation around the attractor—induced for example by prosody or by stronger intention to achieve a communicative goal.

One of the potential strengths of the dynamical systems approach is that it can deal with variation in speech production when investigating sound patterns. In a symbolic, modular view, only the discrete end result of a phonological computation—be it by virtue of rules or ranked constraints—is passed on to the phonetic implementation. The phonetic implementation module has no access to the “history” of discrete operations performed and implements symbols into physical signals regardless of the way they were obtained by the phonological module. Incomplete neutralization in German is a classic case that questions the plausibility of this chain: The final obstruents of <Rad> /ʁad/ (“wheel”) and <Rat> /ʁat/ (“advice”) should be completely indistinguishable for the phonetic implementation module after the neutralization rule described for German has turned both forms into [ʁat^h]. Numerous studies demonstrated that this is not the case and that there are indeed systematic acoustic differences between the two words such as voice onset time, closure duration, or the duration of the preceding vowel (see among others Port and O’Dell, 1985, Port and Crawford, 1989, Roettger et al., 2014, and Roettger and Baer-Henney, 2018, for Dutch: Ernestus and Baayen, 2006). In the modular view, the phonetic component should not be able to produce different signals based on the two phonological representations because they are identical. Gafos (2006) and Gafos and Benus (2006) showed how a dynamical perspective can deal with the observed variation. The categories of *voiceless* and *voiced* are conceptualized as two attractors in a continuous space of voicing. At the ends of syllables, the voiceless attractor is the most stable of the two attractors. However, the exact location of the attractor basin can be modulated

by lexical factors and the speaker’s communicative intention, allowing for subtle differences in the acoustic realization of the voiceless obstruent.

Variation also plays an important role in the domain of intonation research. On the one hand, many studies have shown that there is a probabilistic mapping between functions and forms that are described as prosodic categories (Grabe, 2004; Röhr and Baumann, 2010; Yoon, 2010; Baumann et al., 2015; Ritter and Grice, 2015; Cangemi and Grice, 2016). On the other hand, a great deal of variation can be found in the realization of these prosodic categories. For example, the same type of nuclear pitch accent—the part of the pitch contour on and around the most prominent word in the phrase—can be used for different functions. However, the accent’s realization in terms of the *height* of the pitch peak and the temporal *alignment* of the peak to the accented syllable is often systematically varied by speakers (Ladd and Morton, 1997; Kügler and Gollrad, 2015). Grice et al. (2017) investigated the distribution and realization of pitch accents in German focus marking and demonstrated how continuous and categorical variation go hand in hand. The authors compared focus constructions similar to those exemplified in (1–3) (English translations are given below). In all three cases, the word “Jana” in the answer (A) usually receives the nuclear pitch accent. Example (1) illustrates a case of *broad focus*, where the whole sentence is in focus and “Jana” functions as the exponent of the focus domain (Uhmann, 1991). In example (2), “Jana” is the only word in focus, a condition that is often called *narrow focus* (Ladd, 1980). Example (3) is quite similar to (2) but “Jana” contrasts with another word in the immediate context (“Paul” in the question Q)—this condition is called *contrastive focus*.

- (1) Q: Was gibt’s Neues?
What’s up?
A: Melanie will Jana treffen.
Melanie wants to meet Jana.
- (2) Q: Wen will Melanie treffen?
Who does Melanie want to meet?
A: Melanie will Jana treffen.
Melanie wants to meet Jana.
- (3) Q: Will Melanie Paul treffen?
Does Melanie want to meet Paul?
A: Melanie will Jana treffen.
Melanie wants to meet Jana.

As already mentioned, in all three cases, the nuclear pitch accent is usually placed on the last noun, “Jana.” Grice et al. (2017) showed that the distributions and realizations of pitch accent types differs between the focus conditions. However, their results suggest—as already reported in Mücke and Grice (2014)—that the mapping between focus types and pitch accent categories is not one-to-one. There are general tendencies for certain focus types to be more frequently realized with certain pitch accent types, for example broad focus with H+!H* accents, narrow focus with H* accents, and contrastive focus with L+H* accents. But the focus types are also realized with different accent types—for example, there is a considerable number of rising accents in the broad focus productions of some speakers. Crucially,

Grice et al. (2017) and Roessig et al. (2019) demonstrated that variation in the phonetic parameters (peak alignment, target height, tonal onglide) within each pitch accent category is used to signal focus types as well. Moreover, this variation *within* category boundaries seems to mimic the variation *across* category boundaries: Some speakers, for example, use the shallower H* accent in narrow focus primarily and the more rising L+H* accent in contrastive focus. Others use H* for both functions but increase the magnitude of the rising f0 movement from narrow focus to contrastive focus.

While f0 is a strong acoustic parameter in prosody, it is important to acknowledge that speakers exploit many phonetic dimensions to express prosodic structure. This means that prosodic structure is encoded in more than one phonetic exponent, a phenomenon that has recently been discussed in the context of pleiotropy by Gafos et al. (2019). For prosodic prominence, this implies that speakers can use multiple cues in different combinations to express the same degree of prominence. There are several important strategies of the supra-laryngeal system to highlight important prosodic information in the phonetic substance. The first strategy is referred to as *sonority expansion* (Beckman et al., 1992). Sonority expansion enhances the vowel's sonority to strengthen the syntagmatic contrasts between accented and unaccented syllables. Under accent, speakers intend to produce louder and more sonorous syllables by opening the mouth wider. A more open oral cavity allows for a greater radiation of acoustic energy from the mouth. The second strategy is referred to as *localized hyperarticulation* (de Jong, 1995). It is based on the H&H model developed by Lindblom (1990) and follows the observation that signatures of prominence can be identified by a more extreme articulation of the tongue body in vowel productions. The hyperarticulation strategy involves the enhancement of paradigmatic features such as the place feature for a specific vowel. The tongue body position is lower in low vowels such as /a/, while it is more fronted in front vowels such as /i/ and more retracted in back vowels such as /ʊ/ (de Jong et al., 1993; Harrington et al., 2000; Cho and McQueen, 2005).

During the production of low vowels, sonority expansion, and hyperarticulation are non-competing strategies. Lower tongue and jaw positions accompanied by a higher degree of lip opening both increase specifications of manner and place targets. In addition, low vowels are associated with a low degree of coarticulatory resistance, therefore allowing for a high amount of prosodic variation in the temporal and spatial domains. Prosodic strengthening is more complicated in high vowels. While sonority expansion triggers a more open vocal tract to produce louder vowels, localized hyperarticulation induces smaller constriction degrees to increase the vowel's place feature. In addition, high vowels are associated with a high degree of coarticulatory resistance, thus allowing for less prosodic variation at least in the spatial dimension (Mücke and Grice, 2014). However, these highlighting strategies can be combined in the coordination of different articulatory subsystems. While the lingual system is mainly involved in hyperarticulation to increase the place feature in vowels such as /i/ and /ʊ/, the mandibular and the labial system attribute to sonority expansion

by increasing the degree of lip opening. In the acoustic output, this leads to louder and longer syllables with more peripheral formant frequencies (Australian English: see Harrington et al., 2000; American English: see de Jong et al., 1993, as well as Cho, 2005).

Examples (1–3) above illustrate different focus constructions in which the last noun in the sentence (“Jana”) is in the focus domain and receives the nuclear pitch accent. In example (4), the word occurs out of focus, i.e., in the *background*, and as such does not receive the nuclear accent in English and German. Many studies that investigated the above mentioned strategies of prosodic prominence marking concentrated on the distinction between unaccented and accented syllables and compared words in the most divergent conditions, i.e., *background* to words in *contrastive focus* [see example (3)].

- (4) Q: Will Paul Jana treffen?
Does Paul want to meet Jana?
 A: Melanie will Jana treffen.
Melanie wants to meet Jana.

More recently, Mücke and Grice (2014) investigated the adjustments of lip opening gestures within the group of accented words in different focus types (broad vs. narrow vs. contrastive focus) in comparison to adjustments of the lip kinematics between unaccented and accented words (background vs. {broad, narrow, contrastive}). They found the strongest modifications when comparing target words in contrastive focus to target words in the background. During the production of different vowel types, the speakers produced larger, longer, and faster lip opening movements, thus increasing sonority of vowels in prominent positions. However, when comparing background and broad focus, they found only subtle kinematic adjustments. Even though there were tendencies to increase sonority from background to broad focus, the modifications were not systematic. However, when comparing different focus structures within accentuation, i.e., broad, narrow, and contrastive focus, they found larger, longer, and partially faster lip movements from broad focus to contrastive focus, but no clear distinction between narrow focus and contrastive focus. On the basis of their results, Mücke and Grice (2014) concluded that supra-laryngeal articulation may be directly related to focal prominence and not mediated by accentuation itself. These articulatory findings are in line with recent work by Baumann and Winter (2018) who showed that listeners' judgements of prosodic prominence are influenced by a multitude of categorical (pitch accent type and placement) and continuous acoustic factors (e.g., intensity and duration).

In this paper, we investigate the prosody of focus marking in German in both the laryngeal and the supra-laryngeal domain. We analyse acoustic f0 movements in combination with articulatory movements tracked from the lingual and labial system using a 3D Electromagnetic Articulograph (EMA). In our articulatory measurements, we quantify the parameters related to the displacement of lip opening and lowest position of the tongue body in the vowel /a/, between unaccented and accented (out of focus/background vs. broad focus) and within accentuation (broad focus, narrow focus, contrastive focus).

We demonstrate that categorical and continuous adjustments are made by speakers to express focus structure by virtue of prosodic prominence. Finally, we sketch a dynamical system that accounts for the modifications with attractor landscapes that are shaped by the contribution of the different prosodic dimensions under scrutiny. This model is able to account for both categorical and continuous variation as the outcome of the process of scaling the single control parameter of the system. Crucially, we demonstrate how this scaling of the control parameter modulates all prosodic dimensions, laryngeal and supra-laryngeal, at the same time. In this way, the present work attempts to contribute to our understanding of the integration of multiple channels or tiers in speech production.

METHODS

Speakers, Recording Procedure, Speech Material

Twenty-seven monolingual native speakers of German were recorded with 3D Electromagnetic Articulography (EMA) using a Carstens AG501 articulograph and acoustically using a head-mounted condenser microphone. All recordings took place at the IfL Phonetics department of the University of Cologne. To track the movements of the articulators, sensors were placed on the upper and lower lip, tongue tip, tongue blade, and tongue body. Reference sensors were placed on the bridge of the nose and behind the ears to compensate for head movements. A bite plate measure was used to rotate the occlusal plane. The kinematic data were recorded at 1,250 Hz, downsampled to 250 Hz and smoothed with a 3-step floating mean. In this study, we analyse the data from the lip sensors and the tongue body sensor (backmost tongue sensor). The acoustic recordings were carried out with an AKG C520 headset microphone into a computer via a PreSonus AudioBox 22 VSL interface at a sampling rate of 44.1 kHz and a bit depth of 16 bit. At the time of recording, the speakers were aged between 19 and 35. 17 of them were female, 10 were male. None of the subjects had a special training in phonetics, phonology or prosody, or reported any speech or hearing impairments. The participants received compensation for their participation in the study. The actual recording session after the participant had been prepared lasted about 45 min including a training session.

The participants were seated in front of a screen and were involved in an interactive animated game. They were told that the game revolved around two robots working in a factory, in which one of them likes to move around the tools. The other robot, slightly older and technologically outdated, needs the participant's help to retrieve these tools. In each trial, the participant first saw one robot placing the tool on an object in the factory room and leaving the scene. In the next step, the second, older robot entered the scene. This robot did not enter the factory room but stopped in front of the closed door asking a question about the action of the first robot. After the participant's answer, the door opened,

TABLE 1 | Example question-answer-pairs to elicit the focus structures.

Focus structure	Example trigger and target sentence
Background	Q: Hat er die Säge auf die Wohse gelegt? <i>Did he put the saw on the Wohse?</i> A: Er hat [den Hammer]_F auf die Wohse gelegt. <i>He put the hammer on the Wohse.</i>
Broad	Q: Was hat er gemacht? <i>What did he do?</i> A: Er hat [den Hammer auf die Wohse gelegt.]_F <i>He put the hammer on the Wohse.</i>
Narrow	Q: Wo hat er den Hammer hingelegt? <i>Where did he put the hammer?</i> A: Er hat den Hammer [auf die Wohse]_F gelegt. <i>He put the hammer on the Wohse.</i>
Contrastive	Q: Hat er den Hammer auf die Mahse gelegt? <i>Did he put the hammer on the Mahse?</i> A: Er hat den Hammer auf [die Wohse]_F gelegt. <i>He put the hammer on the Wohse.</i>

the second robot entered the room, took the tool and left the scene.

For the robot's questions, natural productions by a male, native German speaker were used. These questions served as triggers for the focus structures of the answers and were chosen such that the target word denoting the object (where the tool is placed) could be in broad focus, narrow focus, contrastive focus, or in background (with a contrastive focus on the direct object). **Table 1** shows examples for such question-answer-pairs with square brackets and subscript F marking the focus domain. Each question was given auditorily and shown as a combination of pictures in a thought bubble above the head of the robot: the question tool on top of the question object in the case of background and contrastive focus; a simple question mark in the case of broad focus; the object and the question word "wo?" ("where?") in the case of narrow focus. The answers that the participant had to produce were always given in written form at the bottom of the screen. Many participants reported that they were able to give the answers without reading them on the screen after some trials. The participants were asked to always produce the answer with the same syntactic structure and to not add any words like "no." None of the participants had any problems with this restriction. Likewise, none of the participants reported that they found the sentences unnatural or difficult.

Twenty German sounding disyllabic nonce words with a C1V1:C2ə structure were chosen as target words. Since it is important to control for the segmental context in EMA experiments, we used nonce words in target positions. This enabled us to also control for the frequency of the target words. The words were designed such that the word stress was on the first syllable and the consonants (C1 and C2) either require movements of the labial system or the tongue tip to avoid influences on the tongue body measures for the vowel. The first consonant was chosen from the set of /n m b l v/, the second consonant from /n m z l v/. The first, accented vowel was either /a:/ or /o:/, the second always schwa. The consonants and vowels were combined such that each first consonant occurred

twice with each first vowel and each second consonant-schwa-combination occurred four times in the whole set. Special care was taken that the words did not overlap with real German words. All words were presented with the female determiner “die” /di:/. All participants pronounced the words as expected. The target words are given in **Table S1** in the Supplementary Materials.

Each target word was associated with a fictitious visual object. This association remained fixed through the whole experiment and across all participants. The participants were presented with all objects and target words in a preparation phase immediately before the experiment and were asked to read the words aloud with the determiner “die” (“die Nohme,” “die Lahse,” etc.). This phase lasted a few minutes and was included to ensure that no participant placed the stress on the second syllable. In fact, all participants placed word stress on the first syllable starting with the first production.

As described above, in each trial, a tool is placed on one of the fictitious objects. Each object was paired with a tool to occur with. The tools are given in **Table S2** in the Supplementary Materials. As there are 10 tools and 20 target words, each tool had to occur twice. Furthermore, for the background condition and the contrastive focus condition, a competitor tool or object was needed, respectively (for the direct object of the question when the target word was in the background: “Did he place X on A?” “He placed Y on A!”; and for the indirect object of the question when the target word was in contrastive focus: “Did he place X on A?” “He placed X on B!”). These combinations were fixed for each participant, yielding 20 quadruples of target object, tool, competitor object, and competitor tool. The competitor object was chosen such that the first consonant or the first vowel did not equal the first vowel or consonant of the target object. The competitor tool was selected such that it differed in the first consonant from the target sentence tool. The 20 quadruples occurred with all four focus conditions, which resulted in a total of 80 trials. Sixteen trials with different object-tool-quadruples preceded the actual experiment session.

The order of trials was randomized for each of the 27 participants. Subsequent trials were not allowed to contain the same target word or tool used in the target sentence. Furthermore, there were no three subsequent trials with the same focus condition. For two subsequent trials with identical focus condition an upper limit was set: In only 15% of the list, two adjacent trials with equal focus conditions occurred.

The scenes, objects, tools, and robots were drawn by a professional book illustrator. The game was developed as an interactive website using HTML and JavaScript with jQuery for animation (e.g., robots’ arm and mouth movement, the door opening, and closing). The experimenter, sitting behind the participant, pressed a key on the keyboard to make the robot move toward the tool and proceed to the next trial. There was a “rescue key” to repeat the trial in case something went wrong. Between trials, the scenery disappeared for 4 s and the screen transitioned through a series of light, muted colors. This was done to detach the trials from one another to make sure that the focus structure of the target sentence made reference to the current trial only. Points were counted for each complete trial in the lower right corner of the screen to make the task more

game-like. **Figure S1** in the Supplementary Materials shows an example of the experiment screen, where the second robot has just asked his question and is waiting for the answer. The code of the experiment app is available for download: <http://doi.org/10.5281/zenodo.2611287>.

Measures

In this paper, only a subset of the data is reported on. Since the vowel /o/ involves lip rounding, lip aperture in syllables with /o/ cannot be compared to values of syllables with /a/. We decided to restrict our analysis to the target words with /a/ in the stressed syllable. From all 1,080 productions (27 speakers × 4 focus conditions × 10 target words), a minority of cases (3.7%) had to be excluded due to mispronunciations, strong disfluencies, or technical problems during the recording session. The used data set comprises 1,040 tokens and is available for download: <https://osf.io/jx8cn>.

One trained annotator labeled the beginning and end of the accented syllable of each target word using the waveform and the spectrogram in the emuR speech database system (Winkelmann et al., 2018). Within the boundaries of the syllable, lip aperture was evaluated as the Euclidean distance between the lips (Byrd, 2000) as given in Equation 1. An automatic procedure was used to retrieve the maximum of the trajectory within the boundaries of the labeled acoustic syllable. The maximal lip aperture represents the widest opening of the lips during the production of the vowel /a/. In addition, the lowest point of the tongue body during the production of /a/ was measured by finding the minimum of the recorded vertical trajectory within the boundaries of the labeled acoustic syllable. All values (lip aperture and tongue body position) were z-scored for each speaker. **Figure 1A** shows schematic depictions of the articulatory measures.

$$\text{lip aperture } x = \text{upper lip } x - \text{lower lip } x$$

$$\text{lip aperture } y = \text{upper lip } y - \text{lower lip } y$$

$$\text{lip aperture} = \sqrt{(\text{lip aperture } x)^2 + (\text{lip aperture } y)^2} \quad (1)$$

To assess the differences in the f0 contours, we measured the tonal onglide of each nuclear pitch accent. **Figure 1B** provides a schematic depiction of the tonal onglide measure. Tonal onglide characterizes the portion of the f0 movement toward the main tonal target of the pitch accent (Ritter and Grice, 2015; Roessig et al., 2019). In terms of an autosegmental-metrical analysis, like GToBI (Grice et al., 2005), L+H*, and H* pitch accent types are described by a rising movement and result in positive onglide values. In contrast, the accent types H+L* or H+!H* are described by a falling movement from the initial high portion of the accent down to the L* or !H* on the accented syllable and result in negative onglide values. In addition to capturing the direction of the tonal movement (“is it rising or falling?”), the tonal onglide reflects the magnitude of the rise or fall in semitones (“how much does it rise or fall?”). It should be emphasized here that pitch accent categories are multi-dimensional and thus best described by multiple variables. Tonal onglide is a continuous variable that represents both the direction of the pitch movement as well as the magnitude of this movement, but it does not

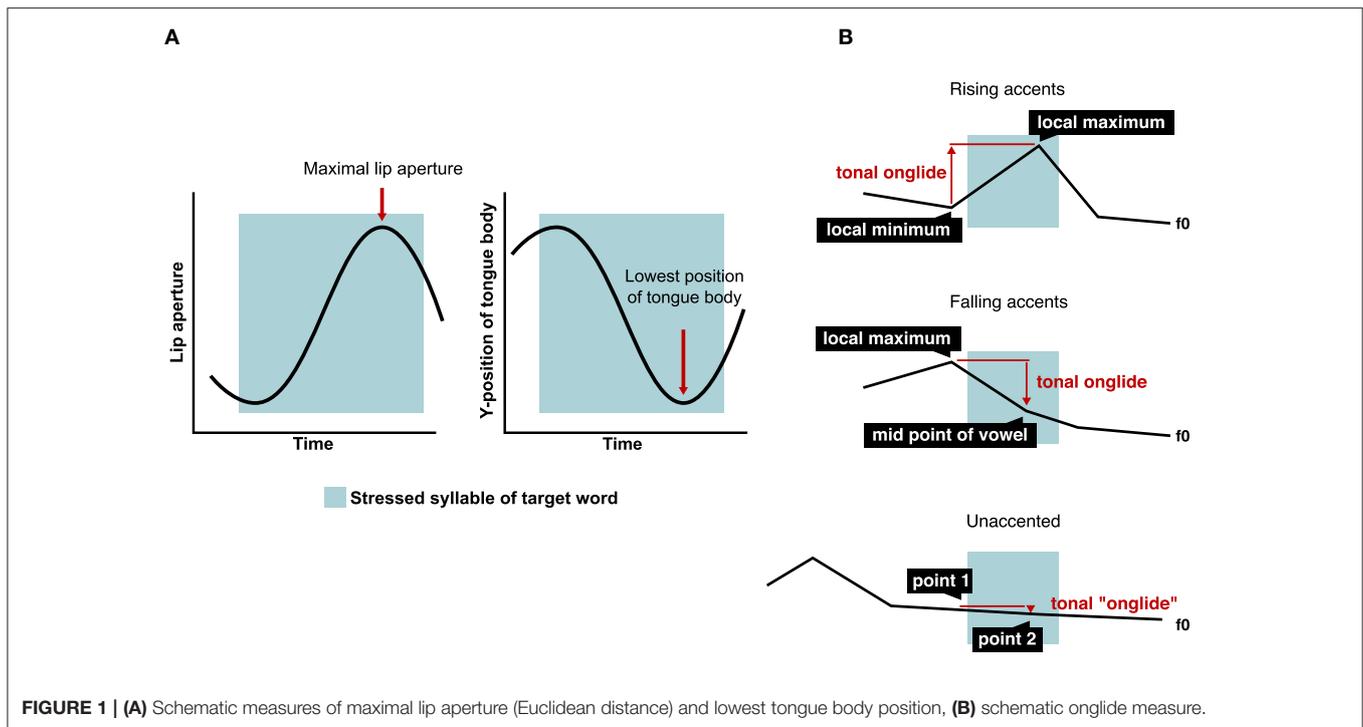


FIGURE 1 | (A) Schematic measures of maximal lip aperture (Euclidean distance) and lowest tongue body position, **(B)** schematic onglide measure.

capture all relevant details of pitch accents (see Grice et al., 2017 for an investigation of the characteristics of pitch accents in terms of tonal onglide and its relation to other parameters). Nevertheless, it has been shown that the tonal onglide movement is a perceptually relevant parameter of pitch accents in German (Baumann and Röhr, 2015; Ritter and Grice, 2015).

Two labelers with training in prosody annotated the f_0 movements with a simple labeling scheme without having access to the intended focus structures of the sentences: First, the labelers identified all utterances in which the speaker did not place the nuclear pitch accent on the object. Second, the labelers judged perceptually whether the nuclear pitch accent was falling or rising. Third, the labelers identified the beginning and the end of the onglide movement manually within a window of three syllables including the accented syllable in the center, the syllable before and the syllable after.

For rising accents, a local minimum just before the rising movement was annotated in the pre-accented syllable or the accented syllable itself as the beginning of the onglide movement. A local maximum at the end of the rise was labeled in the accented syllable or the post-accented syllable as the end of the movement. For falling accents, a relatively high point at the start of the fall was labeled in the pre-accented syllable or the accented syllable itself as the beginning of the onglide movement. Since the f_0 is usually falling throughout the syllable in a falling accent and hence a tonal target is virtually impossible to determine, the midpoint of the vowel of the accented syllable was marked as the end of the accentual movement.

If the nuclear accent was not placed on the target word, it is placed on the direct object of the sentence. In this case, the part of the phrase containing the target word and the following verb is characterized by a low stretch of f_0 . This situation was found in

almost all cases of the background condition and in a minority of cases of the other conditions. When this deaccentuation of the target word occurred, an “onglide” measure was done with fixed time points (5 ms before the start and 50 ms before the end of the stressed syllable) since it is not possible to identify the beginning and the end of a tonal movement. We cannot speak of a real onglide here since there is no movement of a pitch accent. However, this measure makes it possible to compare and model the intonation of all utterances, with accented and unaccented target words, and to relate the intonational and articulatory modifications used to express focus structure across all experimental conditions.

Although using the semitones scale already eliminates a great deal of variation between speakers, normalization is needed to make the speakers more comparable. To do so, we divided each rising onglide value by the mean of the speaker’s rising onglides, and each falling onglide value by the mean of the speaker’s falling onglides. It is plausible that a rise is best interpreted in relation to other rises, while a fall is best interpreted in relation to other falls of the same speaker. For example, a raw onglide value of +6 semitones might be quite extreme for a speaker with a mean of +4 semitones for rises compared to a speaker with a mean of +6 semitones for rises. For the unaccented cases, where we cannot speak of rises and falls, we used the overall mean of the absolute onglide values for each speaker.

RESULTS

Intonation

Before presenting the quantitative results, we turn to some examples of the main intonational modifications in **Figure 2**. The informative value of these examples is of course limited since they

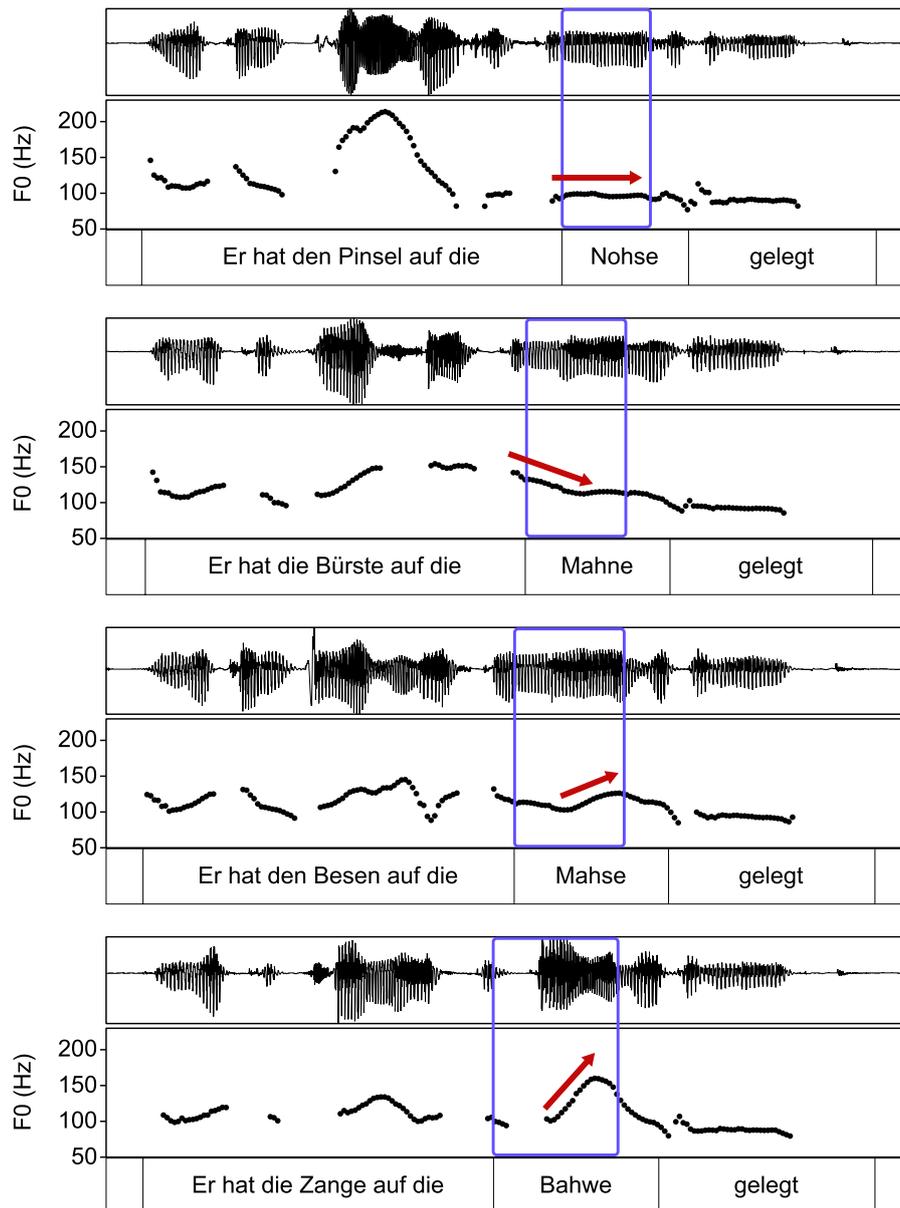
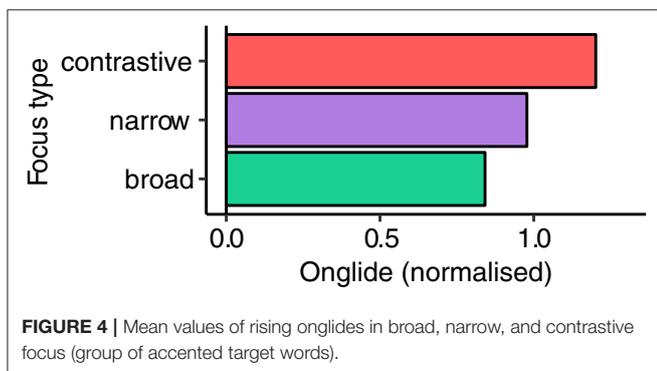
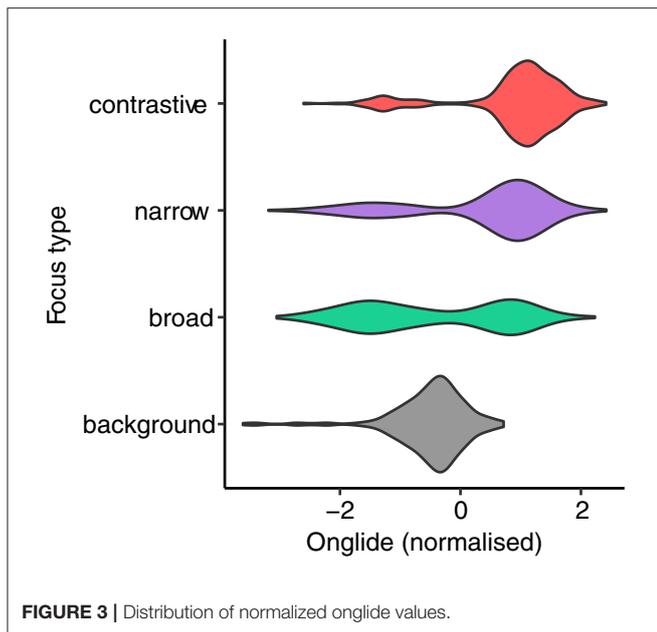


FIGURE 2 | Examples from one male speaker producing background, broad focus, narrow focus, and contrastive focus (from top to bottom). The stressed syllable of the target word is marked by the blue box. [The sentences are of the form "He placed the <tool> on the <target>," literally: "He has the <tool> on the <target> placed," see the methods section for more information on the speech material].

only represent individual utterances. However, accompanying the quantitative results, they help to give a thorough insight into the data. The figure shows examples from one male speaker producing the conditions background, broad focus, narrow focus, and contrastive focus (from top to bottom). The stressed syllable of the target word is marked by the blue box, the arrows illustrate roughly the f_0 movement that is captured by the onglide measure. This speaker uses a flat f_0 stretch on the target word in the background condition (the target word is unaccented), a

falling accent in broad focus and rising accents in narrow focus and contrastive focus. Comparing these last two conditions, a larger magnitude of the rise can be attested in contrastive focus.

Figure 3 presents the normalized onglide values of all speakers for the four focus types in a violin plot. In the background condition, the data show a single mode located slightly below zero. For broad focus, we can observe a bimodal shape of the distribution, with almost equal numbers of falling and rising onglides. In narrow and contrastive focus, the right mode is



more pronounced. Since rising accents dominate the data, we look at the means of rising accents in **Figure 4**. In addition to the increase in the number of accents with a rising onglide, the magnitude of the onglides become larger, as reflected in the stepwise growth of the mean from broad focus to narrow focus, and from narrow focus to contrastive focus. Note that we treat all rises as one group. Many autosegmental-metrical systems like GToBI (Grice and Baumann, 2002) posit two rather similar rising accents, H^* and $L+H^*$. While we do not deny the existence of the two types of pitch accents, our analysis is not intended to be an autosegmental-metrical analysis. As outlined in the methods section, the labelers did not classify each accent beyond deciding whether it is a rise or a fall.

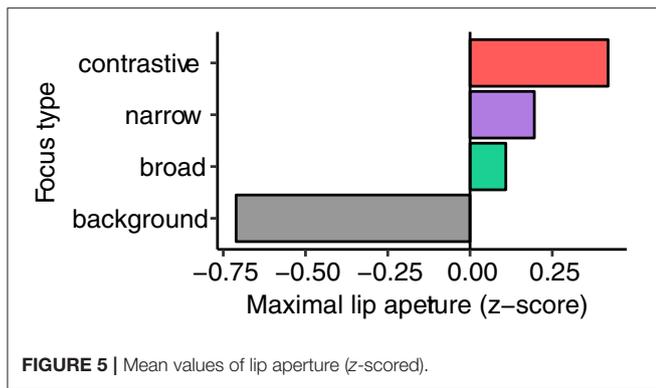
We analyse the results using a Bayesian linear mixed model in R (R Core Team, 2018) with the package brms (Bürkner, 2018) that implements an interface to Bayesian inference with MCMC sampling in Stan (Carpenter et al., 2017). We report the estimated differences between focus conditions in terms of posterior means, 95% credible intervals, and the probability of the estimate being greater than zero.

Given the data and the model, the 95% credible intervals indicate the range in which one can be certain with a probability of 0.95 that the difference between estimates can be found. To calculate the differences between focus types, we subtract the posterior samples for background from broad focus (broad–background), broad focus from narrow focus (narrow–broad), narrow focus from contrastive focus (contrastive–narrow), and broad focus from contrastive focus (contrastive–broad).

The model includes normalized onglide as the dependent variable, focus type as a fixed effect, and random intercepts for speakers and target words as well as by-speaker and by-target-word slopes for the effect of focus type. Since the distribution of the dependent variable is bimodal, we use a prior for the predictor that is characterized by a mixture of two Gaussian distributions centered around -0.5 and 0.5 respectively. The model estimates the parameter θ that represents the extent to which the two Gaussian distributions are mixed. For this parameter, we use a prior centered around zero. Differences in θ indicate the differences in the proportions of the two modes in the onglide data. The model runs with four sampling chains of 5,000 iterations each, preceded by a warm-up period of 3,000 iterations.

We start with the results for the mixing parameter. Given the model and the data, the analysis yields strong evidence for differences in the posterior probabilities for the mixing parameter θ between broad focus and narrow focus ($\hat{\beta} = 1.35$, 95% CI = [0.09, 2.49], $Pr(\hat{\beta} > 0) = 0.98$), narrow focus and contrastive focus ($\hat{\beta} = 1.74$, 95% CI = [0.28, 3.37], $Pr(\hat{\beta} > 0) = 0.99$), as well as broad focus and contrastive focus ($\hat{\beta} = 3.09$, 95% CI = [1.24, 4.95], $Pr(\hat{\beta} > 0) = 1$), i.e., within the group of accented target words. In all cases, the differences are positive indicating a growth of the right mode from broad to narrow focus, and from narrow to contrastive focus. As to the difference between background and broad, the model also suggests that the mixing proportion of the two modes is different ($\hat{\beta} = -2.49$, 95% CI = [-3.82, -1.14], $Pr(\hat{\beta} > 0) = 0$). This comes as no surprise since the distribution of background is unimodal whereas the distribution of broad is bimodal. However, the model calculates a negative difference. This is due to the fact that the model takes the right mode of the prior mixture to capture the unimodal distribution of background. The mixing parameter we report here is higher when the right mode is stronger and the left mode is weaker (note that the model can also estimate the mixing parameter that describes the exact opposite situation but the direction of differences is mirrored in the same way regardless; both parameters cannot be estimated at the same time). Thus, it makes sense—for the sake of completeness—to report the probability of the difference between background and broad focus in the mixing parameter to be lower than zero: $Pr(\hat{\beta} > 0) = 1$.

To assess the differences between the focus conditions regarding the rising distributions, we investigate the mean estimates of the right Gaussian sub-distribution. We only look at broad focus, narrow focus, and contrastive focus since we can only speak of a rising accent in these conditions. The model



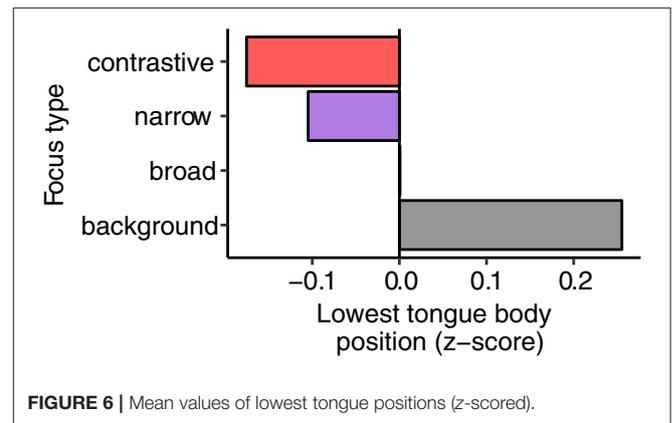
provides evidence for differences in the posterior probabilities between broad focus and narrow focus ($\hat{\beta} = 0.16$, 95% CI = $[-0.02, 0.35]$, $Pr(\hat{\beta} > 0) = 0.96$), narrow focus, and contrastive focus ($\hat{\beta} = 0.23$, 95% CI = $[0.10, 0.36]$, $Pr(\hat{\beta} > 0) = 1$) as well as broad focus and contrastive focus ($\hat{\beta} = 0.39$, 95% CI = $[0.21, 0.58]$, $Pr(\hat{\beta} > 0) = 1$). In all cases, the differences are positive, indicating that the model estimates the rises to become increasingly large from broad focus to narrow focus, and from narrow focus to contrastive focus.

Supra-Laryngeal Articulation

We now turn to the results of the supra-laryngeal parameters. **Figure 5** gives the mean values of the maximal lip aperture for all speakers and focus types (the raw distributions are shown in **Figure S2** in the Supplementary Materials). There is a clear jump from background to broad, with larger distances between the lips for broad focus. The differences between broad focus and narrow focus, as well as between narrow focus and contrastive focus are more subtle, especially between broad and narrow focus. In sum, these results show a modification of the lip opening gesture between unaccented and accented target words as well as within the group of accented words with a ranking from broad to contrastive: background < broad focus < narrow focus < contrastive focus.

Figure 6 presents the mean values of the lowest tongue positions for all speakers and focus types (the raw distributions are shown in **Figure S3** in the Supplementary Materials). As with lip aperture, a larger jump from background to broad focus can be found, i.e., between unaccented and accented words. But there are also differences between broad focus and narrow focus and narrow focus and contrastive focus, i.e., within the group of accented words. Overall, the same ranking as for lip aperture can be attested for the lowest tongue body position: background > broad focus > narrow focus > contrastive focus (reversed because the tongue position is lowered and the values thus decrease).

Analogously to the tonal onglide analysis in Intonation, we analyse the results using Bayesian linear mixed models in R (R Core Team, 2018) with the package brms (Bürkner, 2018). We report the estimated differences between focus conditions in terms of posterior means, 95% credible intervals. Given the data and the model, the 95% credible intervals indicate the



range in which one can be certain with a probability of 0.95 that the difference between estimates can be found. To calculate the differences between focus types, we subtract the posterior samples for background from broad focus (broad–background), broad focus from narrow focus (narrow–broad), narrow focus from contrastive focus (contrastive–narrow), and broad focus from contrastive focus (contrastive–broad). In the case of the maximal lip aperture, we report the probability of the estimate being greater than zero because we are interested in whether the lip aperture increases from one focus type to another. In the case of the lowest tongue position, we report the probability of the difference being smaller than zero, because we are interested in whether the tongue position is lower, i.e., the values decrease, from one focus type to another.

The models include either the z-scored maximal lip aperture or the z-scored lowest tongue positions as the dependent variable. In both models, focus type is a fixed effect, and random intercepts for speakers and target words as well as by-speaker and by-target-word slopes for the effect of focus type are included. We use regularizing priors centered around zero. The models run with four sampling chains of 5,000 iterations each, preceded by a warm-up period of 3,000 iterations.

We start with the modeling results for the maximal lip aperture. Given the model and the data, the analysis yields clear differences in the posterior probabilities between background and broad focus ($\hat{\beta} = 0.81$, 95% CI = $[0.65, 0.97]$, $Pr(\hat{\beta} > 0) = 1$), narrow focus and contrastive focus ($\hat{\beta} = 0.22$, 95% CI = $[0.04, 0.40]$, $Pr(\hat{\beta} > 0) = 0.99$), as well as broad focus and contrastive focus ($\hat{\beta} = 0.30$, 95% CI = $[0.12, 0.48]$, $Pr(\hat{\beta} > 0) = 1$). For broad focus and narrow focus, the model provides evidence for a positive difference which is, however, weaker than in the other cases ($\hat{\beta} = 0.09$, 95% CI = $[-0.10, 0.25]$, $Pr(\hat{\beta} > 0) = 0.84$). In sum, there is a clear increase in the maximal lip aperture from background to broad focus, i.e., from unaccented to accented. Within the group of accented target words, overall, the maximal lip aperture increases. Narrow focus seems to be closer to broad focus although the model still yields evidence for a difference between the two.

We now turn to the results for the lowest tongue position. Given the model and the data, the analysis yields clear differences in the posterior probabilities between background and broad

focus ($\hat{\beta} = -0.25$, 95% CI = $[-0.44, -0.07]$, $Pr(\hat{\beta} > 0) = 1$). This shows that when going from unaccented to accented, the tongue position for the low vowel /a/ is lowered. For the oppositions of broad focus and narrow focus ($\hat{\beta} = -0.11$, 95% CI = $[-0.31, 0.10]$, $Pr(\hat{\beta} < 0) = 0.85$) as well as narrow focus and contrastive focus ($\hat{\beta} = -0.07$, 95% CI = $[-0.28, 0.15]$, $Pr(\hat{\beta} < 0) = 0.75$), the model also provides evidence for differences, although they are not as strong as between background and broad, with 0.85 and 0.75, respectively. When comparing broad focus and contrastive focus, however, the evidence for the difference is stronger again ($\hat{\beta} = -0.18$, 95% CI = $[-0.40, 0.04]$, $Pr(\hat{\beta} < 0) = 0.95$), indicating that there is a substantial decrease in the lowest tongue position within the group of accented focus types.

DYNAMICAL MODEL

The results presented in the previous section show the following pattern: On the tonal tier, when going from background to broad focus, i.e., unaccented to accented, the distribution of flat f0 is split into a bimodal distribution. This bimodal distribution reflects that, when a pitch accent is placed, this accent can be either falling or rising. Both falling and rising accents are found in productions of broad focus, a result that is in line with Mücke and Grice (2014) and Grice et al. (2017). When going from broad focus to narrow focus, the number of rising accents increases while the number of falling accents decreases. This trend continues from narrow focus to contrastive focus. In addition, the magnitude of the rising movements increases between broad and narrow focus and between narrow focus and contrastive focus. The dominance of rising accents as well as the increase in magnitude of the tonal onglide of these rises help to make the accent more prominent.

On the articulatory tier, there is a continuous increase in the lip aperture and a lowering in the tongue body position from background to contrastive focus related to prosodic strengthening strategies during the production of the vowel in the target syllables. The increase in lip aperture can be attributed to sonority expansion, i.e., the speaker produces a louder vowel in the accented syllable (Beckman et al., 1992; Harrington et al., 2000). More energy radiates from the mouth, strengthening the syntagmatic contrast between accented and unaccented syllables in the utterance. The lowering of the tongue during the low vowel /a/ can be related to the strategy of localized hyperarticulation, i.e., the speaker intends to increase the paradigmatic contrast between the low vowel /a/ and any other vowel that could have occurred in the target syllable. The hyperarticulation of the vowel's place target [+low] is related to feature enhancement (de Jong, 1995; Cho, 2006; Mücke and Grice, 2014). Note that in this case the lowering of the tongue also contributes to sonority expansion. Both types of modifications can be seen as strategies to enhance the prominence of the target word from background to contrastive focus with intermediate steps for broad and narrow focus. In this section, we propose a dynamical system that models the tonal and articulatory modifications as the result of the scaling of one control parameter. Before turning to the actual

model, we introduce some of the concepts of dynamical systems that are important for the present work.

The dynamical perspective of the mind, as explained in the introduction, views the mind not as a machine that manipulates symbols with discrete operations. Rather, it is conceptualized as a continuous system that is constantly in flux. This dynamical system follows predictable patterns of behavior in gravitating toward attractors, stable states in its space of possible states. To describe this evolution of the system through the state space over time, the language of differential equations can be employed (Iskarous, 2017). In this formal language, one way of formulating a dynamical system is by giving its potential energy function and its force function—the negative derivative of the potential energy function. The graph of the potential energy curve can give a good impression of the attractors present in the system, the attractor landscape. Consider the black lines in **Figure 7** presenting the potential energy curves of a system with two attractors (left) and another system with one attractor (right). On the x-axis, the state space is shown. This is the space of all possible states of the system, and crucially it is continuous. However, the system is moving toward local minima in the potential energy which are the attractors of the system.

The functions corresponding to the graphs are given in Equation 2 (two attractors) and 3 (one attractor). Both equations include a parameter k , called the control parameter of the system. By scaling this parameter, the system is “moved” through its possible patterns of behavior (Kelso, 2013). As a consequence, the attractor landscape can change when the parameter value is modulated. The black lines of **Figure 7** show the attractor landscape when the control parameter k is 0. The blue lines demonstrate how the system changes if the control parameter is increased to 0.5. In the case of the two-attractor landscape, the right attractor has become deeper than the left attractor and its deepest point also moved slightly to the right on the x-axis (the state space). In the case of the one-attractor landscape, the attractor also moved toward the right on the x-axis.

$$V(x) = \frac{x^4}{4} - kx - \frac{x^2}{2} \quad (2)$$

$$V(x) = \frac{(x - k)^2}{2} \quad (3)$$

A useful metaphor to illustrate how noise works in a dynamical system is to imagine a ball rolling through an attractor landscape like the one in **Figure 7** (left). When the ball is put into the attractor landscape at some random point, it will roll down into one of the two attractor valleys. We can enrich this metaphor by adding wind to the system that represents the notion of noise—a very important component in dynamical systems (Haken, 1977). In this scenario, the ball is pushed away from its original trajectory from time to time. Sometimes these gusts of wind are strong and the ball is pushed far away, sometimes they are weak and it is only perturbed slightly. When the control parameter k is 0, and the two attractors of the system are symmetrical, it takes the same strength of wind gusts to push the ball out of both attractors. But if $k \neq 0$, one of the attractor basins is deeper. For

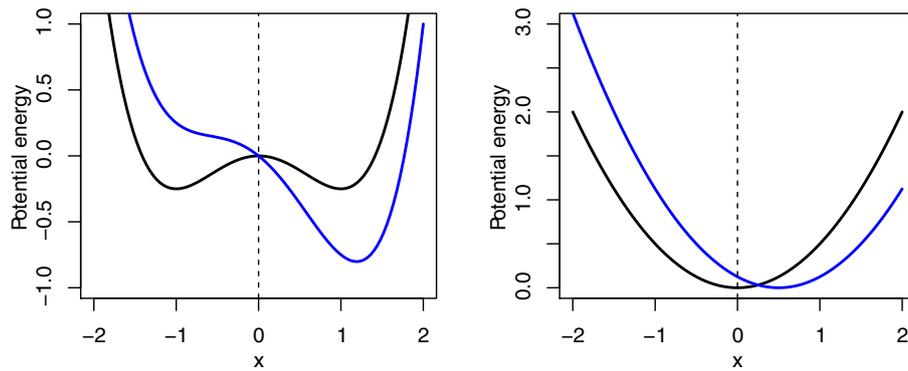


FIGURE 7 | Potential energy functions of system with two attractors (left) and system with one attractor (right). Black line: control parameter $k = 0$. Blue line: control parameter $k = 0.5$.

this deeper attractor, it will take stronger gusts of wind to push the ball out of it. Thus, this attractor is more stable than the other.

Another crucial feature of dynamical systems is that they can exhibit qualitative changes as a control parameter is scaled continuously, also called *bifurcations* (Gafos and Benus, 2006; Kelso, 2013). The model of Haken et al. (1985), for example, describes the shift between anti-phase and in-phase coordination of finger movements as an abrupt change in an attractor landscape that occurs when the tempo of the movement is scaled up continuously (anti-phase: 180° phase transition; in-phase: 0° phase transition). Starting at anti-phase coordination and scaling the tempo up, the mode of coordination remains anti-phase for some time but “breaks down” and changes to in-phase at a certain upper threshold. In the lower tempo ranges, two coordination patterns are possible (in-phase and anti-phase) while beyond the critical boundary, only one coordination pattern, in-phase, is possible. To model this phenomenon, Haken et al. (1985) proposed a dynamical system with two attractors for the lower range of tempo values (one attractor for in-phase and one attractor for anti-phase). For higher tempo values, the model exhibits a simpler landscape with a sole attractor for in-phase.

Equation 4 gives another example system. In **Figure 8**, the consequences of scaling of this system’s control parameter k can be observed: As long as k has a value below 0, the system is characterized by a mono-stable attractor landscape (one attractor). As the parameter k passes 0, the landscape becomes bistable (two attractors).

$$V(x) = \frac{x^4}{4} - k \frac{x^2}{2} \quad (4)$$

Modeling the Tonal Onglide

The part of the model dealing with the intonation side of our data is based on three observations: First, the proportion of falling and rising accents changes from broad to narrow focus, and from narrow to contrastive focus such that the number of rises increases. Second, the magnitude of the rises shifts subtly toward more extreme values, i.e., the rises become increasingly large from broad to narrow focus, and from narrow

to contrastive focus. Third, the shape of the distribution changes from unimodal (“flat”) to bimodal (“rising” vs. “falling”) when going from background to broad focus.

In the two examples of dynamical models above we have laid out the foundations of how we can incorporate these observations into our model. The presence of two modes in the tonal onglide data for broad, narrow and contrastive focus but only one mode for background requires that we use a model with a bistable attractor landscape for a certain range of control parameter values and a monostable attractor landscape for a different range of control parameter values. Within the range of bistability, a change in the control parameter should cause a tilt to the rising side of the attractor landscape. This tilt must go hand in hand with a slight shift of the location of the deepest point of the attractor toward higher values of the state space (the x axis in the graphs of the potential energy function).

One possible model is given by the potential energy function $V(x)$ in Equation 5. **Figure 9** illustrates the consequences of changing the control parameter k : When k is smaller than zero, the system has a single attractor. As it passes zero, it becomes bistable. When k is scaled further, the system tilts to the right giving the right attractor more stability.

$$V(x) = \frac{x^4}{4} - \left(1 - e^{-k}\right) \frac{x^2}{2} - |k| (k - 1) \frac{x}{4} \quad (5)$$

We take the system expressed by Equation 5 as a model for our onglide data and use simulations to evaluate predictions of the system to assess how well it can account for the structure of our observational data. We use a simulation method inspired by the software accompanying Gafos (2006), reimplemented and modified for our purposes. The code is available for download: <https://osf.io/jx8cn>.

The simulation operates on the force function, the negative derivative of the potential energy function. It starts at a random initial state and estimates the solution to the corresponding stochastic differential equation (Brown et al., 2006). The method calculates the change of the system at the current state and adds it to the current state to get to the next state. For the sake of simplicity, the simulation implements a time window that

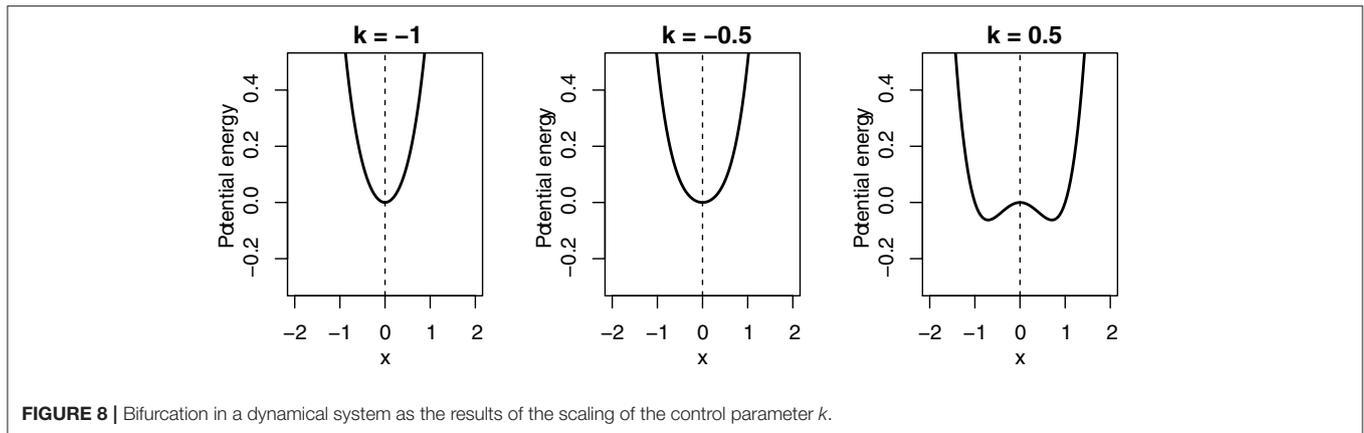


FIGURE 8 | Bifurcation in a dynamical system as the results of the scaling of the control parameter k .

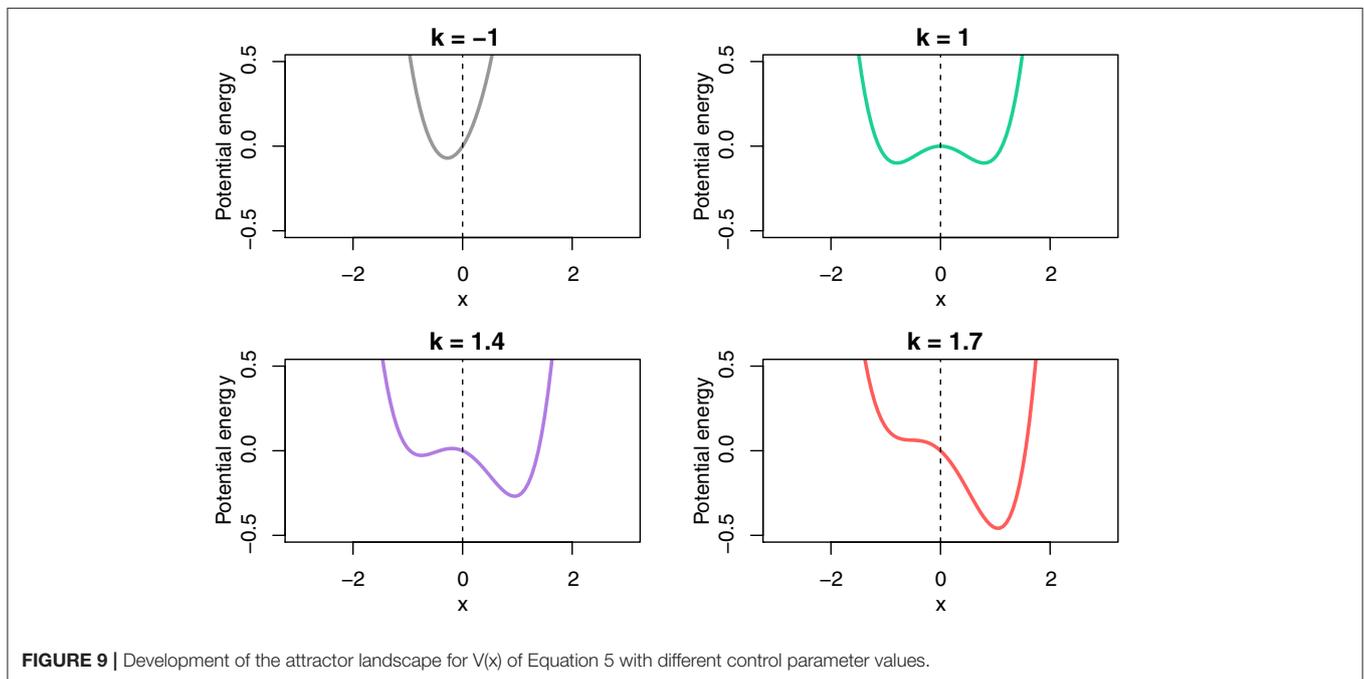
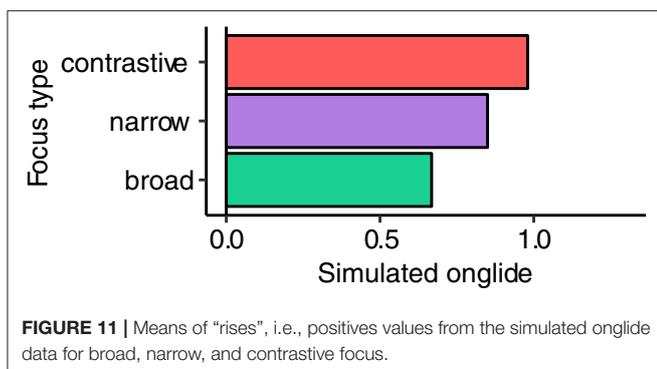
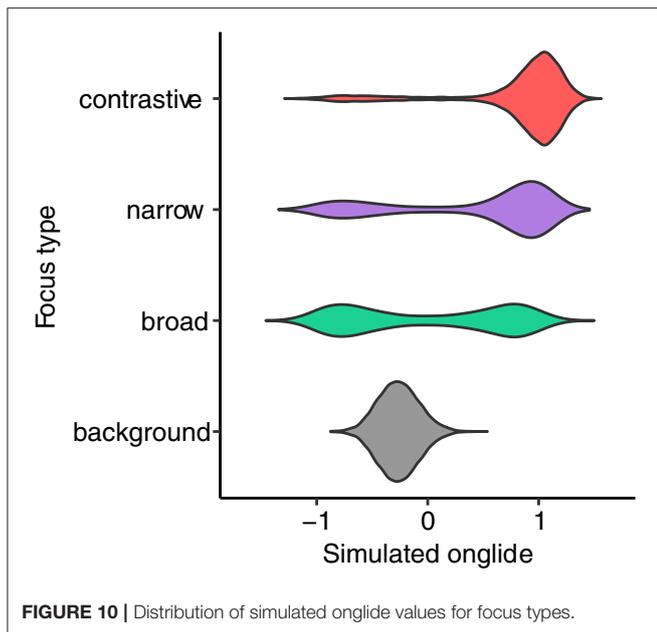


FIGURE 9 | Development of the attractor landscape for $V(x)$ of Equation 5 with different control parameter values.

always has the same length. Thus, after a fixed period of time, i.e., a fixed number of small time steps, in our case 10,000, a single simulation run stops and the current state is registered as the result. Crucially, during each step of the simulation, Gaussian noise is added to the current state. By adding noise, the simulation results are able to reflect the patterns of relative stability of the attractors: Noise pushes the system away from its current state, but the more stable an attractor, the smaller the influence of noise on this state. In other words, when the system is close to a more stable attractor, the probability is higher that it will stay in the basin of the attractor despite the noise. On the contrary, when the system is near a less stable attractor, it is more likely to be pushed away from the attractor basin—eventually ending up in the vicinity of the more stable attractor. The simulation is run 10,000 times (i.e., 10,000 data points with 10,000 time steps each). We can conceive of a single simulation run as one production of an intonation contour.

We use the k values exemplified by the corresponding attractor landscapes in **Figure 9** for the four focus types. Background is modeled with $k = -1$, broad focus is modeled with $k = 1$, narrow focus is modeled with $k = 1.4$, contrastive focus is modeled with $k = 1.7$. The results of the simulations are shown in **Figure 10**. The same pattern as in the results for the tonal onglide can be observed here: the system produces a unimodal distribution slightly below zero for background. The distribution for broad focus is symmetrical. In narrow and contrastive focus, the right mode (rising) becomes increasingly strong. The mean values of the rising distributions also show essentially the same stepwise increase for the “accented” focus types (broad, narrow and contrastive focus), as presented in **Figure 11**. This shows that the attractor basin moves on the dimension of possible states toward more extreme values when the control parameter value is increased and the attractor landscape tilts to the right side.



Enriching the Model

As outlined in the results section, not only the proportion and the scaling of accents are modified by speakers to express focus types, but the lip and tongue body kinematics of the vowel /a/ are also affected. The lips are opened wider, the tongue body position is lower. We can view these modifications as the outcome of a multi-dimensional system of prosody to signal information structure. In this system, the control parameter is used to scale the attractor landscape on many dimensions to achieve the bundle of prosodic modifications. The attractors of the landscape are the result of the combination of these multiple dimensions. The way in which the dimensions shape the multi-dimensional attractor landscape will, however, be different: Some of the dimensions will contribute a rather complex shape, like the tonal onglide with its two stable states for falling and rising—a dimension of the system that can be described well with the two-attractor landscape. Other dimensions will contribute a simpler shape, like the lip and tongue body movements, that can be described with a monostable attractor landscape.

Figure 12 attempts to give an impression of a system with more than one dimension. It combines the landscape for the tonal onglide defined in the previous section with a parabolic landscape for the Euclidian distance of the lips, that could be modeled by a potential energy function as the one given in Equation 3 above. This results in the potential energy function given in Equation 6 which models the tonal onglide as the state of the variable x , and the lip aperture as the state of the variable y . In this function, the control parameter k affects both dimensions.

$$V(x, y) = \frac{x^4}{4} - (1 - e^{-k}) \frac{x^2}{2} - |k|(k-1) \frac{x}{4} + \frac{(y-k)^2}{2} \quad (6)$$

Like in the one-dimensional illustrations above, the potential energy of the system is drawn on the vertical axis. On the left, it is shown what the attractor landscape looks like when the control parameter k is 1. In the tonal onglide dimension, both falling and rising onglides are equally possible. On the right, it is illustrated what the attractor landscape looks like when the control parameter k is increased to 1.4. Now, on the tonal onglide dimension, the right attractor has gained more stability. This leads to more instances of this pitch accent category (e.g., rising) and larger rises. In addition, this attractor has moved toward more extreme values. On the lip aperture dimension, the deepest point of the parabolic shaped attractor drifted toward more extreme values, too. Although we can only visualize two dimensions here, we can imagine that more than two dimensions can shape the attractor landscape. And in fact, it seems plausible to assume that even more than the three dimensions investigated in this paper contribute to the prosodic marking of focus.

The probability density function of a non-deterministic, first-order dynamical system can be found as a stationary solution to the Fokker-Planck equation for the system (Haken, 1977; Gafos and Benus, 2006). In **Figure 13**, the graphs of probability functions are given for the system with two dimensions and the control parameter values used in the previous section to model the focus types (background: $k = -1$, broad focus: $k = 1$, narrow focus: $k = 1.4$, contrastive focus: $k = 1.7$). In **Figure 14**, the same distributions are given from a different perspective to make it easier to grasp the change on the lip aperture dimension. While the tonal onglide becomes bistable as the parameter k is scaled from -1 to 1 and then gains more and more stability on the right mode, the attractors also move on the dimension of lip aperture. First with a big step, from background to broad, and then subtly when going from broad focus to narrow focus, and from narrow focus to contrastive focus. Note that on this dimension the change is similar to what happens to the rising accents of the tonal onglide: While the probability on this dimension remains characterized by a single mode, this mode moves toward more extreme values when k is scaled.

The dimension of the tongue position also contributes a single attractor that is very similar to the one for the lip aperture, except that an increase in k makes it move toward lower values (the

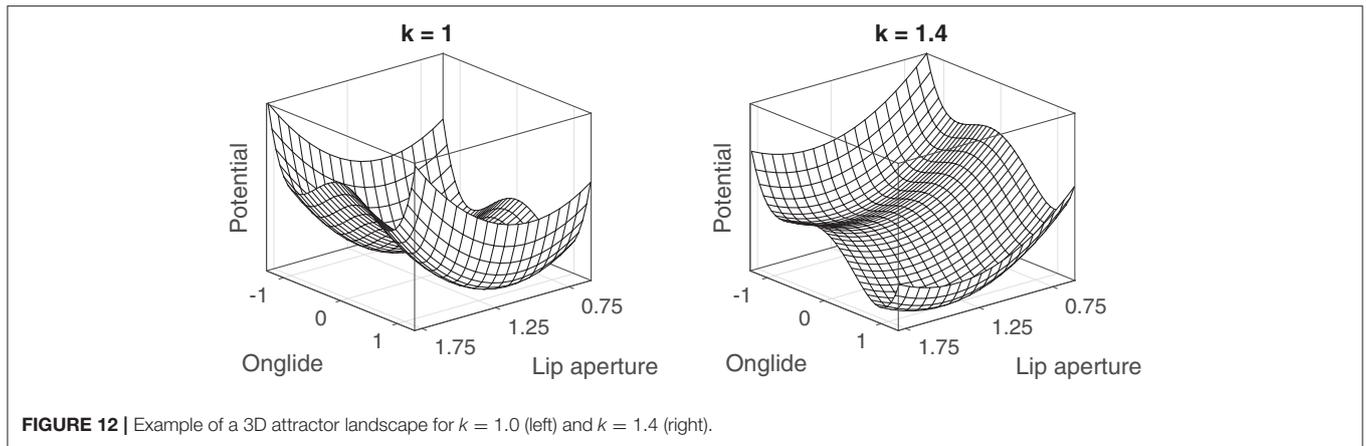


FIGURE 12 | Example of a 3D attractor landscape for $k = 1.0$ (left) and $k = 1.4$ (right).

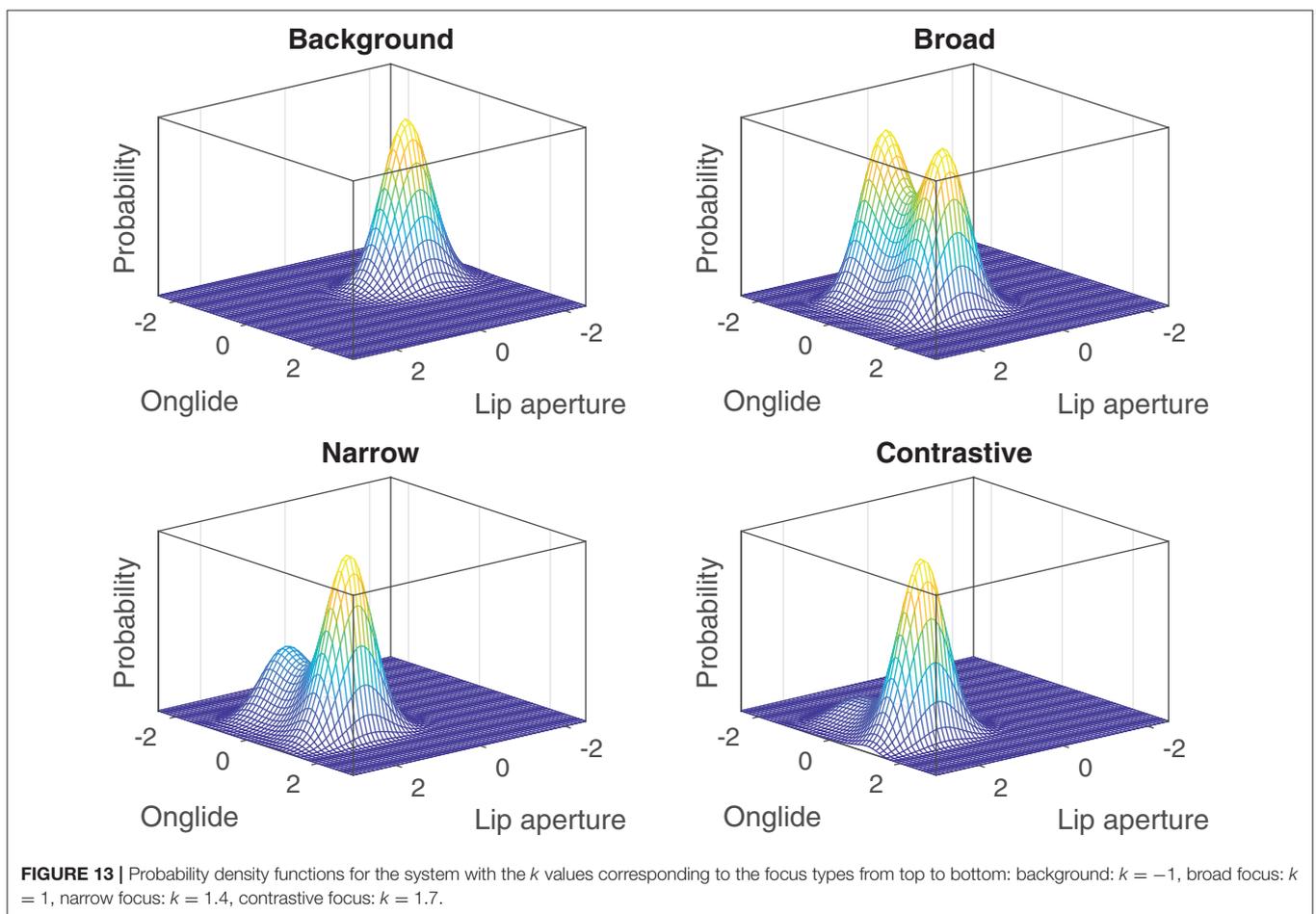
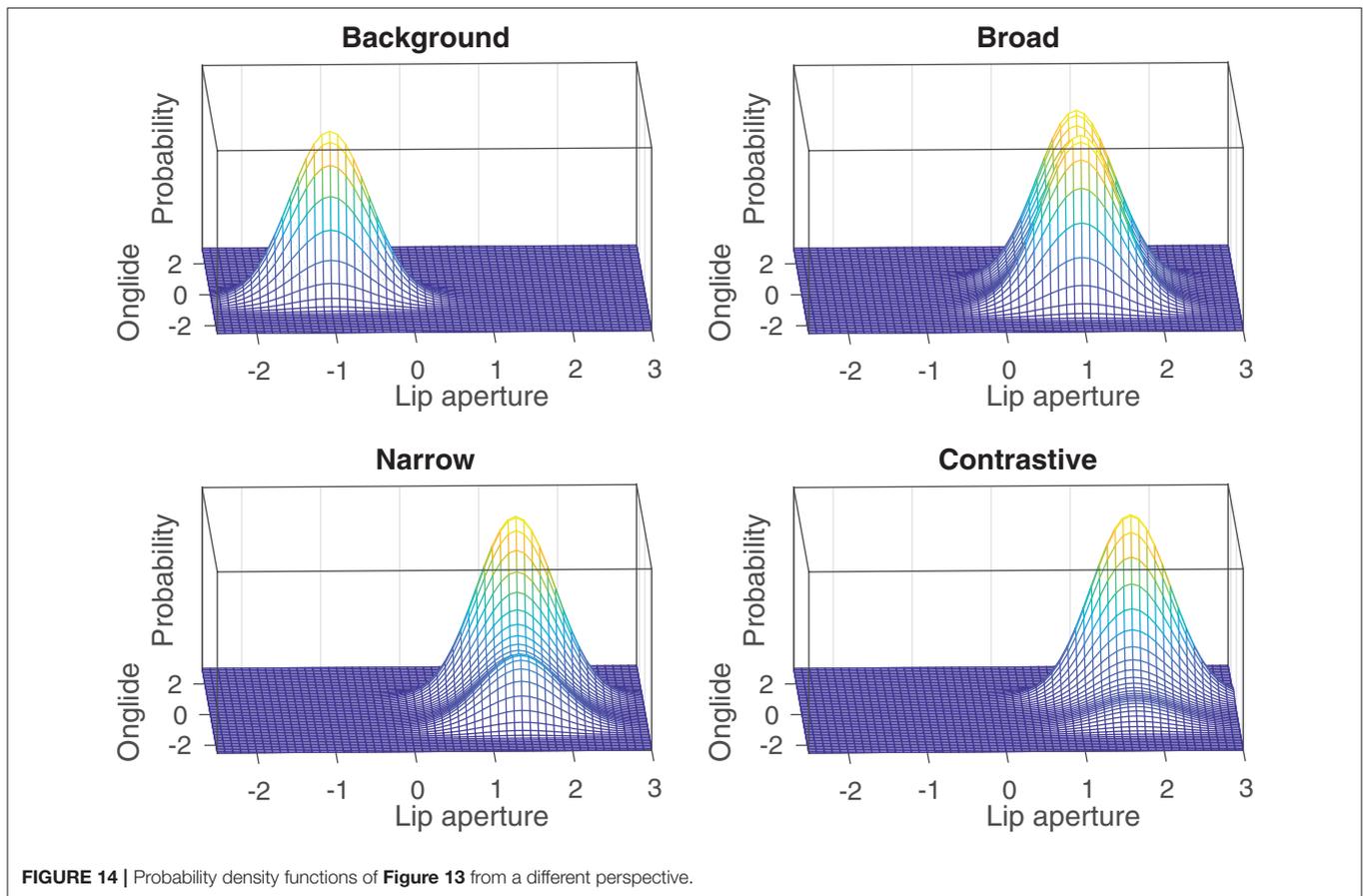


FIGURE 13 | Probability density functions for the system with the k values corresponding to the focus types from top to bottom: background: $k = -1$, broad focus: $k = 1$, narrow focus: $k = 1.4$, contrastive focus: $k = 1.7$.

tongue body is lowered). Equation 7 represents an attempt to sketch how such a system could be described with a potential energy function of three variables.

$$V(x, y, z) = \frac{x^4}{4} - (1 - e^{-k}) \frac{x^2}{2} - |k|(k - 1) \frac{x}{4} + \frac{(y - k)^2}{2} + \frac{(z + k)^2}{2} \quad (7)$$

It should be noted that none of the functions given here reproduces the measured values exactly. We have focussed on the qualitative correspondence of the experimental observations and the theoretical model (which the presented system is able to capture). The coefficients for the model are chosen for presentation purposes here. For example, the differences between the values for the focus types with regard to the lip aperture are greater compared to the tongue movement



as different articulators naturally produce different magnitudes of movements. This fact is not reflected in the system. The system only provides a scheme of how we can picture the score of prosodic dimensions in a single system with one control parameter.

DISCUSSION

In this study, we have presented data on the prosodic marking of focus in German from 27 speakers. These data contribute to the increasing evidence of the systematic use of continuous variation in speech and the deep intertwining of this continuous variation with categorical variation. Moreover, the data show how speakers use a combination of cues related to the laryngeal and supra-laryngeal tiers to enhance prosodic prominence. This combination of prosodic dimensions is taken up by our dynamical model.

With regard to the intonation results, our analysis shows that there is no one-to-one mapping between focus types and accent types. However, there are probabilistic tendencies that can be described as patterns of relative stability between the quasi-categories represented by the attractors. With regard to the articulatory results, the study adds evidence to the finding that prosodic prominence is expressed gradually: There are not only modifications in terms of prosodic strengthening between unaccented and accented, but also within the group

of accented targets to make the word more prominent. The increase in lip aperture during vowel production can be viewed as sonority expansion, while the corresponding lowering of the tongue body position can be interpreted as hyperarticulation of the vowel /a/ by enhancing the vowel's place feature [+low] and an increase in sonority at the same time. Since the vowel is low, the strategies of localized hyperarticulation and sonority expansion are compatible. The speakers intend to produce louder and more peripheral vowels (de Jong, 1995; Harrington et al., 2000; Cho, 2006; Mücke and Grice, 2014). Our results are generally in line with the findings of Mücke and Grice (2014) for German. The data support the assumption that prosodic strengthening in the articulatory domain is not just a concomitant of accentuation but is directly controlled to express different degrees of prominence. However, the modifications between target words in background and broad focus reported in the present study are stronger than those reported in Mücke and Grice (2014) who did not find systematic differences between background and broad. This might be attributed to the fact that the data set of the present paper (27 speakers) is considerably larger than in the study by Mücke and Grice (5 speakers) and therefore less sensitive to speaker-specific variation.

The results of the present study underscore that it is fruitful to analyse categorical and continuous aspects jointly and that theoretical devices that treat phonology and phonetics as a single

system are needed. The dynamical perspective of the mind as endorsed by many researchers within the fields of phonology and phonetics (Browman and Goldstein, 1986; Tuller et al., 1994; Port, 2002; Gafos and Benus, 2006; Nava, 2010; Mücke, 2018) and beyond (Haken et al., 1985; Thelen and Smith, 1994; Kelso, 1995; Smith and Thelen, 2003; Spivey and Dale, 2006; Spivey, 2007) is well-suited to provide a view on the sound patterns of language without the need for a translation process between categorical and continuous aspects.

With respect to this intertwining of categorical and continuous aspects of prosodic prominence, it is worthwhile to take a short look at how the current approach relates to the widespread view of prosodic prominence as a characteristic of a hierarchically organized structure. In the literature, different hierarchies of prosodic structure have been proposed (Nespor and Vogel, 1986; Pierrehumbert and Beckman, 1988; Hayes, 1989; Selkirk, 1996; Shattuck-Hufnagel and Turk, 1996). Although the proposals disagree as to the existence of some levels, they all share the assumption that utterances can be decomposed into hierarchically organized constituents. A minimal structure that most researchers in the field agree upon can be outlined as follows (Grice, 2006): An utterance consists of one or more intonational phrases which contain one or more smaller phrases (e.g., an *intermediate phrase*). A constituent on the smallest level of phrasing contains one or more words, a word contains one or more feet, and a foot contains one or more syllables. Regarding the results of the current study, it is interesting to look at how this prosodic hierarchy has been related to prosodic prominence. One approach is to assume that the levels in the hierarchy are headed by prominences (Beckman and Edwards, 1994; Shattuck-Hufnagel and Turk, 1996). For example, a nuclear pitch accented syllable is the head of an intermediate phrase. Applying this view to the productions of the current corpus, this theory would interpret the increase of supra-laryngeal articulatory effort in the target word's stressed vowel as a correlate of the reorganization in the prosodic prominence structure as the nucleus is placed on the target word and hence the head status is moved from the stressed syllable of the direct object (the tool) to the stressed syllable of the target word. In our model of the production of prosodic patterns, the attractor basin situated on the continua of the articulatory dimensions moves toward more extreme values. In the tonal domain, controlled by the laryngeal system, we model this reorganization as a bifurcation on the dimension of onglide such that the system evolves from monostability (flat f_0) toward bistability to reflect that the newly assigned nuclear pitch accent can be falling or rising.

However, the findings of the current study go beyond what we can conceptualize as a reorganization of the head-assignment in the prosodic hierarchy. They contribute to an understanding of prosodic prominence that is sensitive to both categorical and more fine-grained, continuous phenomena. When we look at the productions with the nuclear pitch accent in the same position, i.e., the same assignment of the head status, we observe that the change of the focus type (broad focus \rightarrow narrow focus \rightarrow contrastive focus) leads to an additional increase in prominence with an increase in articulatory effort, a higher probability of rising accents, and larger tonal onglides. In the modeling

approach, this is reflected by an increase in the continuous control parameter.

Support for the idea that the structure of prosodic prominence in the phrase can be modified even in cases where the nuclear pitch accent is not reassigned, i.e., the nuclear pitch accent remains on the target word, comes from work on the perceived prominence of pitch accent types by Baumann and Röhr (2015). Their study showed that, in general, rising accents are perceived as more prominent than falling accents. Beyond the level of reorganization of the prosodic hierarchy, the choice and realization of the nuclear pitch accent work on the assignment of prosodic prominence. In our view, all these processes are the result of a non-linear dynamical system that does not assume a separation of the categorical, phonological, and the continuous, phonetic level.

In the modeling section of the present work, we have sketched a system that brings together different dimensions of prosodic prominence. The dimensions contribute to the shared attractor landscape in different manners. In the most complex dimension, the tonal onglide, we can see how the continuous scaling of a control parameter can lead to qualitative changes: The landscape goes from monostable (unaccented) to bistable (accented). The bistable landscape is then able to account for the proportions of falling and rising accents (categorical variation) as well as the increase in rising onglides (continuous variation). We have demonstrated a scenario in which one control parameter can account for changes in a multidimensional space including intonation and articulation. As already mentioned, the model does not attempt to exactly reproduce the values obtained from the phonetic analyses. It is rather seen as a proof of concept to demonstrate how we can think of prosody in a dynamical systems framework. The results presented in this paper concentrate on a subset of phonetic dimensions that play an important role for prosodic prominence. And so the model outlined on the basis of these results is restricted. In fact, the state space of a full model would include all relevant parameters including dimensions related to duration and relative timing. For example, in the articulatory domain, the duration of the lip and tongue movements is expected to be longer in prominent syllables. But even with a more complex model—one that could also include more than one control parameter—the main idea persists: the same mechanism that modulates the tonal domain also leads to changes in the articulatory domain. The domains with their multiple dimensions form a bundle to be used by the speaker to express prosodic prominence. These bundles might vary between languages, the attractor landscapes are conceptualized as part the speaker's knowledge of phonetics and phonology.

The concept of a multi-dimensional attractor landscape can in principle be extended to any number of dimensions, and is in line with the finding that phonological entities are characterized by many dimensions (Lisker, 1986; Coleman, 2003; Winter, 2014; Mücke, 2018) and that intonational categories are no exception (for Italian and German: Niebuhr et al., 2011; for German: Cangemi et al., 2015; for Italian: Cangemi and Grice, 2016; for English: Barnes et al., 2012, *inter alia*). Furthermore, in this work, we have conceptualized the dimensions to be orthogonal. Future research should investigate how the different dimensions interact. In addition, the model proposed in the current work

takes into account the patterns of all speakers pooled together. In Roessig et al. (2019), we take a closer look at the intonation patterns of different speaker groups. We demonstrate that it is possible to conceptualize the different speaker-specific patterns as different uses, or scaling strategies, of the same system. For the unidimensional system presented in that study, it seems to be sufficient to assume that speakers use different ranges of values for the control parameter. For a more complex system, it might be necessary to assign more weight to one or more dimension in order to reflect the fact that speakers might not exploit all phonetic dimensions to the same degree.

The model presented in the current paper is a model of the production of prosodic patterns. We can, however, speculate that the perception of prosodic patterns can be modeled in a similar fashion. Attractors offer a flexible framework to model stability and variability in systems of different kinds and different environments. As such, they are also applicable to speech perception. In fact, similar models have been employed to account for phenomena in the perception of speech sound or lexical access (Tuller et al., 1994; Spivey et al., 2005). In addition, we might speculate that there is a strong connection between the attractor landscapes for production and those for perception, including a huge variety of acoustic and articulatory cues (Baumann and Winter, 2018; Gafos et al., 2019)—but this topic is beyond the scope of the current study and has to be left open for future research.

DATA AVAILABILITY

The datasets generated for this study are available for download: <https://osf.io/jx8cn/>.

ETHICS STATEMENT

This study was carried out in accordance with the recommendations of the Local Ethics Committee of the University of Cologne with written informed consent from

REFERENCES

- Barnes, J., Veilleux, N., Brugos, A., and Shattuck-Hufnagel, S. (2012). Tonal center of gravity: a global approach to tonal implementation in a level-based intonational phonology. *Lab. Phonol.* 3, 337–383. doi: 10.1515/lp-2012-0017
- Baumann, S., and Röhr, C. (2015). “The perceptual prominence of pitch accent types in German,” in *Proceedings 18th ICPhS* (Glasgow: University of Glasgow), 298.
- Baumann, S., Röhr, C., and Grice, M. (2015). Prosodische (De-)Kodierung des Informationsstatus im Deutschen. *Zeitschr. Sprachwiss.* 34, 1–42. doi: 10.1515/zfs-2015-0001
- Baumann, S., and Winter, B. (2018). What makes a word prominent? Predicting untrained German listeners’ perceptual judgments. *J. Phonet.* 70, 20–38. doi: 10.1016/j.wocn.2018.05.004
- Beckman, M., and Edwards, J. (1994). Articulatory evidence for differentiating stress categories. *Papers in Laboratory Phonology*, p. 3.
- Beckman, M., Edwards, J., and Fletcher, J. (1992). “Prosodic structure and tempo in a sonority model of articulatory dynamics,” in *Papers in Laboratory Phonology II: Segment, Gesture, Prosody* (Cambridge), 68–86.

all subjects. All subjects gave written informed consent in accordance with the Declaration of Helsinki. The protocol was approved by the Local Ethics Committee of the University of Cologne (application 16–404).

AUTHOR CONTRIBUTIONS

SR and DM: substantial contributions to the conception and design of the work as well as the acquisition, analysis, and interpretation of data for the work.

FUNDING

This work was supported by the German Research Foundation (DFG) as part of the SFB1252 Prominence in Language in the project A04 Dynamic modeling of prosodic prominence at the University of Cologne.

ACKNOWLEDGMENTS

The authors thank Timo B. Roettger and Bastian Auris for their advice on the statistical analyses, Stefan Baumann for discussions about prosody, as well as the reviewers for their helpful comments. All errors are ours.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fcomm.2019.00044/full#supplementary-material>

Figure S1 | Example screen from experiment during a trial with contrastive focus condition.

Figure S2 | Distributions of the maximal lip aperture (z-scored) for all speakers.

Figure S3 | Distributions of the lowest tongue body position (z-scored) for all speakers.

Table S1 | Target words (all nonce words).

Table S2 | Tools with English translation.

- Browman, C. P., and Goldstein, L. (1986). Articulatory phonology: an overview. *Phonol. Yearbook* 3, 219–252. doi: 10.1017/S0952675700000658
- Brown, S. D., Ratcliff, R., and Smith, P. L. (2006). Evaluating methods for approximating stochastic differential equations. *J. Mathemat. Psychol.* 50, 402–410. doi: 10.1016/j.jmp.2006.03.004
- Bürkner, P.-C. (2018). Advanced bayesian multilevel modeling with the R package brms. *R J.* 10, 395–411. doi: 10.32614/RJ-2018-017
- Byrd, D. (2000). Articulatory vowel lengthening and coordination at phrasal junctures. *Phonetica* 57, 3–16. doi: 10.1159/000028456
- Cangemi, F., and Grice, M. (2016). The importance of a distributional approach to categoricity in autosegmental-metrical accounts of intonation. *Lab. Phonol.* 7, 1–20. doi: 10.5334/labphon.28
- Cangemi, F., Krüger, M., and Grice, M. (2015). “Listener-specific perception of speaker-specific productions in intonation,” in *Individual Differences in Speech Production and Perception*, eds S. Fuchs, D. Pape, C. Petrone, and P. Perrier (Frankfurt: Peter Lang), p. 123–145.
- Carpenter, B., Gelman, A., Hoffman, M., Lee, D., Goodrich, B., Betancourt, M., et al. (2017). Stan: a probabilistic programming language. *J. Stat. Softw.* 76, 1–32. doi: 10.18637/jss.v076.i01

- Cho, T. (2005). Prosodic strengthening and featural enhancement: evidence from acoustic and articulatory realizations of /a,i/ in English. *J. Acoust. Soc. Am.* 117, 3867–3878. doi: 10.1121/1.1861893
- Cho, T. (2006). “Manifestation of prosodic structure in articulatory variation: Evidence from lip kinematics in English,” in *Laboratory Phonology 8*, eds L. Goldstein, D. H. Whalen and C. T. Best (Berlin; Boston, MA: De Gruyter Mouton), 519–548.
- Cho, T., and McQueen, J. (2005). Prosodic influences on consonant production in Dutch: effects of prosodic boundaries, phrasal accent and lexical stress. *J. Phon.* 33, 121–157. doi: 10.1016/j.wocn.2005.01.001
- Coleman, J. (2003). Discovering the acoustic correlates of phonological contrasts. *J. Phon.* 31, 351–372. doi: 10.1016/j.wocn.2003.10.001
- de Jong, K. (1995). The supraglottal articulation of prominence in English: linguistic stress as localized hyperarticulation. *J. Acoust. Soc. Am.* 97, 491–504. doi: 10.1121/1.412275
- de Jong, K., Beckman, M. E., and Edwards, J. (1993). The interplay between prosodic structure and coarticulation. *Lang Speech* 36, 197–212. doi: 10.1177/002383099303600305
- Ernestus, M., and Baayen, H. (2006). “The functionality of incomplete neutralization in Dutch: the case of past-tense formation,” in *Laboratory Phonology 8: Varieties of Phonological Competence Vol. 8*, editors M. L. Goldstein, D. H. Whalen, and C. Best (Berlin; New York, NY: Mouton de Gruyter), 27–49.
- Gafos, A. I. (2006). “Dynamics in Grammar,” in *Laboratory Phonology 8: Varieties of Phonological Competence*, editors M. L. Goldstein, D. H. Whalen, and C. Best (Berlin, New York, NY: Mouton de Gruyter), 51–79.
- Gafos, A. I., and Benus, S. (2006). Dynamics of phonological cognition. *Cogn. Sci.* 30, 905–943. doi: 10.1207/s15516709cog0000_80
- Gafos, A. I., Roeser, J., Sotiropoulou, S., Hoole, P., and Zeroual, C. (2019). Structure in mind, structure in vocal tract. *Nat Lang Linguist Theory*. doi: 10.1007/s11049-019-09445-y
- Grabe, E. (2004). Pitch accent realization in English and German. *J. Phon.* 26, 129–143. doi: 10.1006/jpho.1997.0072
- Grice, M. (2006). “Intonation,” in *Encyclopedia of Language and Linguistics, 2nd Edn, Vol. 5* ed K. Brown (Oxford: Elsevier), 778–788. doi: 10.1016/B0-08-044854-2/00045-6
- Grice, M., and Baumann, S. (2002). Deutsche intonation und GToBI. *Linguist. Berichte* 191, 267–298.
- Grice, M., Baumann, S., and Benzmüller, R. (2005). “German intonation in autosegmental-metrical phonology,” in *Prosodic Typology: The Phonology of Intonation and Phrasing Sun-Ah Jun* (Oxford: Oxford University Press), 55–83. doi: 10.1093/acprof:oso/9780192249633.003.0003
- Grice, M., Ritter, S., Niemann, H., and Roettger, T. B. (2017). Integrating the discreteness and continuity of intonational categories. *J. Phon.* 64, 90–107. doi: 10.1016/j.wocn.2017.03.003
- Haken, H. (1977). *Synergetics*. Berlin: Springer.
- Haken, H., Kelso, J. A. S., and Bunz, H. (1985). A theoretical model of phase transitions in human hand movements. *Biol. Cybern.* 51, 347–356. doi: 10.1007/BF00336922
- Harrington, J., Fletcher, J., and Beckman, M. E. (2000). “Manner and place conflicts in the articulation of accent in Australian English,” in *Papers in Laboratory Phonology V: Acquisition and the Lexicon*, ed M. Broe (Cambridge: Cambridge University Press), 40–51.
- Hawkins, S. (1992). “An introduction to task dynamics,” in *Gesture, Segment, Prosody*, eds D. Robert Ladd and G. J. Docherty (Cambridge: Cambridge University Press), 9–25.
- Hayes, B. (1989). “The prosodic hierarchy in meter,” in *Rhythm and Meter*, editors P. Kiparsky and G. Youmans (San Diego, CA: Academic Press), 201–260. doi: 10.1016/B978-0-12-409340-9.50013-9
- Iskarous, K. (2017). The relation between the continuous and the discrete: a note on the first principles of speech dynamics. *J. Phon.* 64, 8–20. doi: 10.1016/j.wocn.2017.05.003
- Kelso, J. A. S. (1995). *Dynamic Patterns: The Self-Organization of Brain and Behavior*. Cambridge: MIT Press.
- Kelso, J. A. S. (2013). “Coordination dynamics,” in *Encyclopedia of Complexity and Systems Science*, ed R. A. Meyers (Heidelberg: Springer), 1537–1564. doi: 10.1007/978-0-387-30440-3_101
- Kügler, F., and Gollrad, A. (2015). Production and perception of contrast: the case of the rise-fall contour in German. *Front. Psychol.* 6:1254. doi: 10.3389/fpsyg.2015.01254
- Ladd, D. R. (1980). *The Structure of Intonational Meaning: Evidence from English*. Bloomington, IN: Indiana University Press.
- Ladd, D. R. (2011). “Phonetics in phonology,” in *The Handbook of Phonological Theory*, eds J. Goldsmith, J. Riggle, and A. C. L. Yu (Malden, MA: Blackwell), 348–373. doi: 10.1002/9781444343069.ch11
- Ladd, D. R., and Morton, R. (1997). The perception of intonational emphasis: continuous or categorical? *J. Phon.* 25, 313–342. doi: 10.1006/jpho.1997.0046
- Lancia, L., and Winter, B. (2013). The interaction between competition, learning, and habituation dynamics in speech perception. *Lab. Phonol.* 4, 221–258. doi: 10.1515/lp-2013-0009
- Lindblom, B. (1990). “Explaining phonetic variation: a sketch of the hand theory,” in *Speech Production and Speech Modeling*, eds W. J. Hardcastle and A. Marchal (Dordrecht: Kluwer Aca), 403–439. doi: 10.1007/978-94-009-2037-8_16
- Lisker, L. (1986). “Voicing” in english: a catalogue of acoustic features signaling /b/ Versus /p/ in trochees. *Lang. Speech* 29, 3–11. doi: 10.1177/002383098602900102
- Mirman, D., and Magnuson, J. S. (2009). Dynamics of activation of semantically similar concepts during spoken word recognition. *Memory Cognit.* 37, 1026–1039. doi: 10.3758/MC.37.7.1026
- Mücke, D. (2018). *Dynamische Modellierung von Artikulation und prosodischer Struktur: Eine Einführung in die Artikulatorische Phonologie*. Berlin: Language Science Press.
- Mücke, D., and Grice, M. (2014). The effect of focus marking on supralaryngeal articulation - is it mediated by accentuation? *J. Phon.* 44, 47–61. doi: 10.1016/j.wocn.2014.02.003
- Nava, E. (2010). *Connecting Phrasal and Rhythmic Events: Evidence from Second Language Acquisition*. Ph.D. dissertation, University of Southern California.
- Nespor, M., and Vogel, I. (1986). *Prosodic Phonology*. Dordrecht: Foris Publ.
- Niebuhr, O., D’Imperio, M., Gili Fivela, B., and Cangemi, F. (2011). “Are there “shapers” and “aligners”? individual differences in signalling pitch accent category,” in *Proceedings of the 17th ICPHS* (Hong Kong), 120–123.
- Pierrehumbert, J. (2016). Phonological representation: beyond abstract versus episodic. *Annu. Rev. Linguist.* 2, 33–52. doi: 10.1146/annurev-linguistics-030514-125050
- Pierrehumbert, J., and Beckman, M. E. (1988). *Japanese Tone Structure*. Cambridge: MIT Press.
- Pierrehumbert, J., Beckman, M. E., and Ladd, D. R. (2000). “Conceptual foundations of phonology as a laboratory science,” in *Phonological Knowledge: Conceptual and Empirical Issues*, eds N. Burton-Roberts, P. Carr, and G. Docherty (Oxford: Oxford University Press), 273–304.
- Port, R. (2002). “Dynamical systems hypothesis in cognitive science,” in *Encyclopedia of Cognitive Science*, ed L. Nadel (London: Nature Publishing Group), 1027–1032.
- Port, R. (2006). “The graphical basis of phones and phonemes,” in *Second Language Speech Learning: The Role of Language Experience in Speech Perception and Production*, eds M. Munro and O. S. Bohn (Amsterdam: John Benjamins), 349–365. doi: 10.1075/llt.17.29por
- Port, R., and Crawford, P. (1989). Incomplete neutralization and pragmatics in German. *J. Phon.* 17, 257–282.
- Port, R., and O’Dell, M. (1985). Neutralization of syllable-final voicing in German. *J. Phon.* 13, 455–471.
- R Core Team (2018). *R: A Language and Environment for Statistical Computing. Vienna, Austria*. Retrieved from: <http://www.r-project.org/>
- Ritter, S., and Grice, M. (2015). The role of tonal onglides in german nuclear pitch accents. *Lang. Speech* 58, 114–128. doi: 10.1177/0023830914565688
- Roessig, S., Mücke, D., and Grice, M. (2019). The dynamics of intonation: categorical and continuous variation in an attractor-based model. *PLoS ONE*. 14:216859. doi: 10.1371/journal.pone.0216859
- Roettger, T. B., and Baer-Henney, D. (2018). Towards a replication culture in phonetic research: speech production research in the classroom. *PsyArXiv [Preprint]*. doi: 10.31234/osf.io/q9t7c

- Roettger, T. B., Winter, B., Grawunder, S., Kirby, J., and Grice, M. (2014). Assessing incomplete neutralization of final devoicing in German. *J. Phon.* 43, 11–25. doi: 10.1016/j.wocn.2014.01.002
- Röhr, C., and Baumann, S. (2010). “Prosodic marking of information status in German,” in *Proceedings of the Fifth International Conference on Speech Prosody* (Chicago, IL).
- Roon, K. D., and Gafos, A. I. (2016). Perceiving while producing: modeling the dynamics of phonological planning. *J. Mem. Lang.* 89, 222–243. doi: 10.1016/j.jml.2016.01.005
- Selkirk, E. (1996). “The prosodic structure of function words,” in *Signal to Syntax: Bootstrapping from Speech to Grammar in Early Acquisition*, eds J. Morgan and K. Demuth (Mahwah, NJ: Lawrence Erlbaum), 187–214.
- Shattuck-Hufnagel, S., and Turk, A. E. (1996). A prosody tutorial for investigators of auditory sentence processing. *J. Psycholinguist. Res.* 25, 193–247. doi: 10.1007/BF01708572
- Smith, L. B., and Thelen, E. (2003). Development as a dynamic system. *Trends Cogn. Sci.* 7, 343–348. doi: 10.1016/S1364-6613(03)00156-6
- Spivey, M. (2007). *The Continuity of Mind*. New York, NY: Oxford University Press.
- Spivey, M., and Dale, R. (2006). Continuous dynamics in real-time cognition. *Curr. Dir. Psychol. Sci.* 15, 207–211. doi: 10.1111/j.1467-8721.2006.00437.x
- Spivey, M., Grosjean, M., and Knoblich, G. (2005). Continuous attraction toward phonological competitors. *Proc. Natl. Acad. Sci. U.S.A.* 102, 10393–10398. doi: 10.1073/pnas.0503903102
- Thelen, E., and Smith, L. B. (1994). *A Dynamic Systems Approach to the Development of Cognition and Action*. Cambridge: The MIT Press.
- Tuller, B., Case, P., Ding, M., and Kelso, J. A. S. (1994). The nonlinear dynamics of speech categorization. *J. Exp. Psychol.* 20, 3–16. doi: 10.1037//0096-1523.20.1.3
- Uhmann, S. (1991). *Fokusphonologie*. Tübingen: Niemeyer.
- van Gelder, T., and Port, R. (1995). *Mind as Motion*. MIT Press, Cambridge, MA.
- Winkelmann, R., Jaensch, K., Cassidy, S., and Harrington, J. (2018). *emuR: Main Package of the EMU Speech Database Management System*.
- Winter, B. (2014). Prospects and overviews spoken language achieves robustness and evolvability by exploiting degeneracy and neutrality. *Bioessays* 36, 960–967. doi: 10.1002/bies.201400028
- Yoon, T.-J. (2010). “Speaker consistency in the realization of prosodic prominence in the Boston University Radio Speech Corpus,” in *Proceedings of Speech Prosody* (Chicago, IL).

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

Copyright © 2019 Roessig and Mücke. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.