Categoriality and continuity in prosodic prominence

Simon Roessig

language
science
press

Studies in Laboratory Phonology

Chief Editor: Martine Grice
Editors: Doris Mücke, Taehong Cho

In this series:

1. Cangemi, Francesco. Prosodic detail in Neapolitan Italian.

2. Drager, Katie. Linguistic variation, identity construction, and cognition.

3. Roettger, Timo B. Tonal placement in Tashlhiyt: How an intonation system accommodates to adverse phonological environments.

4. Mücke, Doris. Dynamische Modellierung von Artikulation und prosodischer Struktur: Eine Einführung in die Artikulatorische Phonologie.

5. Bergmann, Pia. Morphologisch komplexe Wörter im Deutschen: Prosodische Struktur und phonetische Realisierung.

6. Feldhausen, Ingo & Fliessbach, Jan & Maria del Mar Vanrell. Methods in prosody: A Romance language perspective.

7. Tilsen, Sam. Syntax with oscillators and energy levels.

8. Ben Hedia, Sonia. Gemination and degemination in English affixation: Investigating the interplay between morphology, phonology and phonetics.

9. Easterday, Shelece. Highly complex syllable structure: A typological and diachronic study.

10. Roessig, Simon. Categoriality and continuity in prosodic prominence

Categoriality and continuity in prosodic prominence

Simon Roessig

Freie Universität Berlin

für Lina und Julius

Contents

1 Introduction

Prosody has been characterised as a "half-tamed savage" (Bolinger 1978: 475) being shaped by both categorical and continuous aspects. According to this view, the categorical, "tamed" side of prosody represents those aspects that are grammaticalised and as such are part of a phonological, symbolic system. The continuous, "untamed" side represents the "unusually generous scope that speakers have [...] in the phonetic implementation" of prosodic categories (Gussenhoven 2004: 49).[1]

The objective of many approaches to prosody has been to "draw a sharp dividing line between the tamed half and the untamed half" (Gussenhoven 2004: 49). Interestingly, this aim resonates with a more general, long-standing debate in linguistics revolving around the question how phonology with its categorical representations and phonetics with its continuous signals are related. The prevalent view has long been that phonological knowledge, the mental representations of speech sounds, is best conceptualised as symbols and discrete rules or constraints that operate on these symbols in an abstract system. The result of the discrete computations has to be translated into a continuous, phonetic signal (Ladd 2006), an acoustic output produced via articulatory movements. The relation between phonology and phonetics is thus characterised by a translation of categorical to continuous, involving two fundamentally different "formal languages" (Gafos & Benus 2006: 906). Fruitful proposals to solve problems arising from the disparity of these representations are rooted in the framework of nonlinear dynamical systems. The framework has gained increasing attention in modelling phenomena in cognition (among others Kelso 1995; van Gelder & Port 1995; Gafos 2006; Gafos & Benus 2006; Port 2002; Spivey 2007; Thelen & Smith 1994; Tuller et al. 1994) because it can provide one formal language to capture categorical and continuous aspects of cognition at the same time.

The present work aims to shed light on the relation between the categoriality and the continuity of prosodic prominence. Crucially, it argues that it is often difficult to draw a sharp dividing line between the tamed and the untamed

---

[1]In this view, the continuous aspects of prosody can become grammaticalised in the course of language change and thus develop to be used in a categorical fashion.

sides of prosody. It demonstrates that what has been termed phonetic or "untamed" seems to form synergies with what has been described as phonological or "tamed", and both work jointly towards the same communicative goals. In particular, the present work investigates recordings of 27 native speakers of German marking focus types in an interactive task. The analysis thereby pursues *integration* in a two-fold manner. On the one hand, the integration of categorical and continuous aspects of pitch accents is examined, revealing that the probabilistic mapping of focus types to pitch accent categories is mimicked by the continuous parameters of the pitch accents. For instance, a higher probability of rising accents is accompanied by larger pitch excursions of these rising accents. On the other hand, the present work integrates multiple dimensions of prosodic focus marking by combining the tonal analyses with investigations of articulatory movements of the lips and the tongue, showing that speakers make use of a rich set of parameters.

These results are incorporated into a dynamical approach that models the discreteness of phonological categories and the continuous nature of phonetic substance as well as the multi-dimensionality of prosodic patterns. The present work thereby emphasises the synergies of categorical and continuous aspects of prosody and questions the need to be able to separate the "tamed" and the "untamed" sides in a theoretical approach. The aim is to contribute to a larger understanding of how "a symbiosis of the symbolic and subsymbolic paradigms" (Smolensky 1988: 19) can be developed. On the one hand, this symbiosis is desirable in order to bridge gaps between the disciplines of phonetics and phonology. On the other hand, the symbiosis is needed to form analytical synergies that can cope with a growing body of findings demonstrating that the sound patterns of language are characterised by a wide array of variability, gradient phenomena and systematic fine-grained details.

The book is structured as follows:

CHAPTER 2 sheds light on the relation of phonetics and phonology, or how the relation of the two has been conceptualised in theoretical frameworks. After very briefly tracing the history of some of the most important ideas that led to today's understanding of phonetics and phonology, the chapter turns to problems that arise from a strict separation of phonetics and phonology, or a purely symbolic phonology. In this context, phenomena like (in)variance of sound categories in the world's languages, assimilation, vowel harmony, and incomplete neutralisation are discussed. Special attention is paid to solutions that are provided by models to cope with these phenomena, such as optional rules, the introduction of stochasticity and scalar values in Optimality Theory, gestural overlap in Articulatory phonology, phonetic implementation rules, and concepts of exemplar theory.

CHAPTER 3 introduces the framework of dynamical systems. This chapter has two parts. In the first part, the basic concepts of dynamical systems and attractors are presented. The most important features of dynamical systems are illustrated using the logistic map, and differential equations and multi-stability are introduced. In the second part of the chapter, applications of dynamical models in phonetics, phonology and beyond are investigated in more detail: the Haken et al. (1985) model of inter-limb coordination patterns, the harmonic oscillator of Articulatory phonology, the coupled oscillator model for the coordination of speech gestures, the categorical perception model of Tuller et al. 1994, and two models by Gafos & Benus (2006) for incomplete neutralisation and Hungarian vowel harmony. This second part takes up the problems of the relation of phonetics and phonology outlined in Chapter 2. It describes how models based on dynamical systems can help to learn more about the relation of phonetics and phonology, and how attractors in dynamical systems relate to the conception of linguistic categories in a traditional sense. Throughout the chapter, the models are illustrated using MATLAB code that accompanies this book and is available for download.

CHAPTER 4 deals with the topic of prosody and prosodic prominence. The chapter sketches some of the concepts that are fundamental to the study presented later in the book, such as pitch accents, prosodic structure and prosodic strengthening. Special attention is paid to what has been described as categorical and what has been described as continuous in prosody research. This chapter introduces the prosody of focus marking – a field of research that the present work attempts to contribute to. The chapter takes three perspectives on focus marking by incrementally adding bits of evidence. It reviews what is known about the tonal and articulatory patterns of prosodic focus marking and refines the view on categorical and continuous phenomena in these patterns. In doing so, it narrows down the subject of the empirical and modelling part of the book.

CHAPTER 5 introduces the objectives of the empirical part of this work and describes the experimental methods used to collect the data that are analysed in the next two chapters. The corpus of collected data described here comprises productions of 27 native speakers of German marking different focus structures (background, broad focus, narrow focus, and contrastive focus) by means of prosody in a controlled experimental environment.

CHAPTER 6 presents the results of the F0 measures of a subset of productions of the corpus. This subset contains all utterances in which the nuclear accent is placed on the target word, i.e. the indirect object. The analysis demonstrates that speakers use categorical and continuous modulations of F0 to mark focus types and that the two types of modulation form a symbiosis. A first dynamical

model for pitch accents is sketched in the second part of this chapter. This model represents an account that reconciles the categoriality and the continuity of pitch accents found in the data. As a result, it provides a first step towards an approach that dispenses with a strict division of abstract representations of pitch accents on the one hand, and their phonetic implementation in terms of F0 on the other.

CHAPTER 7 extends the analysis of F0 patterns including the utterances without nuclear pitch accent on the target word and adds measures of articulatory movements of the lips and the tongue. The chapter sketches a second model that enriches the account of Chapter 6 in two ways: First, the transition from unaccented to accented is conceptualised as a bifurcation, a qualitative change, in the dynamical system. Second, the attractor landscapes are understood as multidimensional constructs in which many dimensions – both laryngeal and supralaryngeal – contribute to a complex, flexible bundle of prosodic prominence.

CHAPTER 8 completes the book with a general discussion.

2 Categoriality and continuity in the sounds of language

A central topic of debate in recent years has concentrated on the question of how phonology, with its categorical representations, and phonetics, with its continuous signals, relate to each other. Although concerned with the same object of study – the sounds of speech and how they are used in communication – phonetics and phonology share a long tradition of separation.

This separation of the two fields at least partly originated from the fact that speech is indeed found to be concrete and continuous, yet at the same time abstract and categorical. Humans produce the acoustic signals we consider as speech sounds by moving their articulators. Both motion and acoustic output can be characterised as continuous. But the sounds of speech have been ascribed an abstract, symbolic nature in the function they serve as part of language (Ladd 2014; Pierrehumbert 1990). This dual nature of speech sounds is at the root of the observation that it is possible for "two physical (acoustic) events that are objectively quite different to count as instances of the same category in the symbolic system, and for two physical events that are objectively very similar to count as instances of two different categories" (Ladd 2014: 29). Many researchers have acknowledged throughout the history of linguistics that the search for an adequate characterisation of speech is not achieved with the mere description of the physical aspects of the sounds of language but has to include the mental or cognitive representations of sounds. Likewise, the description in terms of a purely abstract system void of reference to the phonetic substance proves unpromising as well, since "phonological units and processes are what they are largely due to the physical and physiological structure of the speech mechanism" (Ohala 1990: 155).

This chapter is concerned with how phonetics, phonology and their relation have been characterised by taking a brief look at some history of the fields and discussing several problems arising through a strict separation of categorical, symbolic from continuous, physical aspects that has been assumed by many models. The chapter also concentrates on which solutions have been provided more recently in the literature.

2.1 Phonetics and phonology

The physical, continuous nature of speech is traditionally treated as the domain of *phonetics*. The cognitive, categorical nature of speech is attributed to the field of *phonology*. While scholars working on various aspects of speech sounds did not assume a strong division between the two domains for a long time, the 20th century brought a "split between phonetics and phonology" (Ohala 1997: 680). Ferdinand de Saussure's work, in introducing the distinction between *parole* and *langue*, played an important role in laying the foundations of the division between phonetics and phonology. This division was spelled out more clearly by Trubetzkoy 1958 – although Durand & Laks 2002 note that Saussure's use of the terms "phonetics" and "phonology" was slightly different from Trubetzkoy's. Trubetzkoy (1958: 7) proposed a division between two disciplines ("Lautlehren") each with its own name, object of study and methodology. While phonetics in his view is concerned with the physical form of speech and consequently has to use the methods of the natural sciences, phonology is concerned with the meaning of speech sounds in language, their symbolic character and has to use the methods of the humanities and social sciences. In Saussurean terms, the former discipline studies the sound pattern of parole while the latter studies the sound pattern of langue (Culler 1986; Ladd 2014). This division had great impact on the development of linguistics, the fields of phonetics and phonology and on the way scholars have since viewed language and speech in general. While the division seems natural to many and is commonly taught in introductory courses in linguistics, it has also been the subject of one of the most important debates in the study of language in recent years.

One of the major building blocks for a clear-cut separation of phonetics and phonology is the idealisation of speech as a linearly ordered string of discrete units (Ladd 2014; Durand & Laks 2002). Although the idea that speech can be, in some form, described as a sequence of symbolic units had been present in alphabetic writing systems for millennia, the rise of the *phonemic principle* at the end of the nineteenth century marked a turning point in the history of linguistics. Representations in terms of *phonemes* take the categorical nature of speech into account: Sufficiently similar sounds that do not contrast, called *allophones*, are grouped into one category, one phoneme. For example, whether an /l/ is realised as a voiced lateral [l] or as a the devoiced version [l̥], like in English after voiceless stops, does not influence the categorisation of the sound as the phoneme /l/ according to the principle. In this case, the realisation as a devoiced version is simply attributed to contextual variation. While for many linguists phonemes and phonemic transcriptions offered a useful toolbox to capture what they considered

the most important information when describing a language, the phoneme was also ascribed an important psychological function. Baudoin de Courtenay 1963 considered the phoneme as a "living psychological phonetic unit" that constitutes the "psychological equivalent of sound" (translations: Durand & Laks 2002: 18). In this view, words and sentences are conceptualised as being composed of segment-sized units not only for descriptive purposes, they are represented in the language user's mind as strings of discrete units. These discrete units form the minimal building blocks of the cognitive representation of speech.

Although the concept of the phoneme and its implications were sharply criticised in early generative work (Chomsky 1964) and phonological knowledge began to be represented in terms of distinctive features (Jacobson et al. 1952; Chomsky & Halle 1968), the division between categorical, abstract units and a continuous, physical layer was maintained. Phonological representations of sounds were decomposed and conceptualised as bundles of binary features that make reference to phonetic characteristics. In this way, the rise of distinctive features tied phonetics and phonology closer together. However, the idea of distinctiveness as a *gold standard* for the description of sound patterns as well as the categorical, symbolic nature of the primitives of phonology continued to be central. Moreover, the representations of sound categories in the mind as bundles of features remained faithful to the idea that words are linearly ordered sequences of discrete units and that abstract phonological representations have to be mapped onto a universal set of phonetic entities.

In fact, many theories in the twentieth century have implicitly incorporated the view argued for by Saussure and Trubetzkoy. For example, one of the largest endeavours in linguistics, the development and maintenance of a standardised set of symbols for the description of the world's languages, the *International Phonetic Alphabet* (IPA), builds on the assumption that the phonetic sounds are not part of langue. As such, they are viewed as *universal* categories that exist regardless of the languages they are used in (IPA 1999; Ladd 2014). This assumption has been attacked by many scholars in recent years. Pierrehumbert et al. (2000: 30f), for example, state that "there are no two languages in which the implementation of analogous phonemes is exactly the same" and that "phonological inventories only exhibit strong analogies". A remarkable example is provided by data of Cho & Ladefoged 1999 on voice onset time (VOT) in eighteen languages. VOT ist defined as the duration between the release of a stop and the beginning of vocal fold vibration following the stop and is regarded as one of the most important acoustic parameters to distinguish voicing categories of stop consonants (Lisker 1986). The results of Cho & Ladefoged 1999 are shown in Figure 2.1. Ordered from low to high VOT, it appears that there are hardly any

clear boundaries between the categories that the authors imposed on the data as an interpretation, like "unaspirated", "slightly aspirated", "aspirated" and "highly aspirated". Some of the within-category differences appear at least equal to the between-category differences. This example demonstrates that the existence of a universal, language-independent set of phonetic categories may be problematic. Therefore, the data are hard to reconcile with the stance taken in the prevalent models of phonetics and phonology – a point of view that is well summarised with the words of Chomsky & Halle (1968: 4f):

> [G]eneral linguistic theory might propose, as substantive universals, that [...] phonetic transcriptions must make use of a particular, fixed set of phonetic features. [...] We will be concerned with the theory of "universal phonetics," that part of general linguistics that specifies the class of "possible phonetic representations" of sentences by determining the universal set of phonetic features and the conditions on their possible combinations. The form of each sentence in each language is drawn from this class of possible phonetic representations.

To illustrate this point, if language drew on a fixed set of phonetic categories, the picture obtained by Cho & Ladefoged 1999 should look more like the simulated data shown in Figure 2.2. Compared to this figure, the original picture resembles a continuous increase of VOT lacking clear jumps between the putative categories. Although this remains pure speculation, it is in line with Ladd (2014: 42) who concludes that "any apparent discontinuities in the gradual increase from one end of the VOT scale to the other would disappear" when more data were added to the picture of Cho and Ladefoged.

The idea of a universal set of phonetic categories is an integral part of a *modular* view of phonetics and phonology. As mentioned above, the phoneme was seen as a psychological unit and supposed a division of the abstract representation in the mind and the physical realisation. In this way, phonetics and phonology are conceptualised as two modules from a cognitive perspective. More recent views that dispense with the phoneme and instead assume other representations like feature bundles adhere to a perspective in which modularity plays an important role.

In this perspective of modularity, phonological entities are stored and processed in one module and then passed on to a separate phonetic module to prepare and realise the implementation. Thus, phonetics and phonology are not only separated as scientific disciplines but also viewed as two disparate cognitive domains. Importantly, the separation into clear-cut cognitive modules necessitates

Figure 2.1: Mean VOT values for velar stops from eighteen languages in the study of Cho & Ladefoged 1999. Dashed lines indicate the category boundaries assumed by the authors.

translations from symbolic representations to continuous properties (Ohala 1990; Gafos & Benus 2006) at the *interface between phonetics and phonology* (Keating 1988) – a term extensively discussed in the last decades. Because the domains are encapsulated and administer their own representations and carry out their own computations before passing the result to the next module, the receiving module cannot access the history of steps that lead to this particular representation. As Pierrehumbert (2002: 102) notes, the mainstream view of phonetics, phonology and their interface is architectured as a *feed-forward* system "because no arrows go backwards, from articulatory plans to phonological encoding".

2.2 Gradience

In the above outlined perspectives of phonology, its representations must only entail those features of sounds that cause differences in terms of lexical meaning. The continuous aspects of speech live only on the level of phonetic representations and come into being mainly as a consequence of universal characteristics

Figure 2.2: Simulated VOT values under the assumption of a universal set of phonetic categories.

of articulation, acoustics and audition. The phonological representations are discrete symbols that can be changed by applying discrete rules. An example for such a rule is the assimilation rule given in 2.1 as taken from Nolan (1992: 262). In this example, a unit characterised by the features [+coronal], [+anterior], and [−continuant] inherits the values of these particular features from the following unit. If this rule is applied to the sequence /leɪt kɔːlz/ (*late calls*), the result would be [leɪk kɔːlz].

$$
\begin{bmatrix} +\text{coronal} \\ +\text{anterior} \\ -\text{continuant} \end{bmatrix} \rightarrow \begin{bmatrix} \alpha\ \text{coronal} \\ \beta\ \text{anterior} \\ \gamma\ \text{continuant} \end{bmatrix} / \underline{\hspace{1cm}} \begin{bmatrix} \alpha\ \text{coronal} \\ \beta\ \text{anterior} \\ \gamma\ \text{continuant} \end{bmatrix} \quad (2.1)
$$

Note how the /t/ turned into a [k] – one discrete entity was transformed into another discrete entity. Nolan (1992) demonstrates, using electropalatography (EPG), that the process of assimilation is in fact a gradient process. This means that there are instances of /t/ for which it is possible to find the residual of an alveolar closure during the phase of the plosive. In some instances, there is no complete alveolar closure, but the tongue still has contact to the front parts of

the gum. In other instances, the assimilation is complete and the contact is only velar. In these instances, the contacts are comparable to a /k k/ sequence as in /meɪk kɔːlz/ (*make calls*). In addition to showing that assimilation is a gradient process rather than a discrete process in speech production, Nolan also provides evidence that listeners are able to use this gradient information in perception.

A framework that is able to capture gradient phenomena like this is *Articulatory phonology* (Browman & Goldstein 1986; 1992). In Articulatory phonology, speech sounds are not represented as symbolic units. Instead, the primitives of phonology are gestures. Patterns of speech sounds come into existence through the orchestration of multiple gestures in *gestural scores*. These scores build higher forms, like syllables and words. Crucially, a gesture is viewed as a continuous dynamical system. This makes it possible to describe fine-grained differences in the spatial and temporal manifestations. Assimilation of a /t/ towards a /k/, like in the study of Nolan discussed above, is ascribed to an overlap of the tongue tip gesture for the alveolar closure and the tongue back gesture for the velar closure. The overlap of the two gestures can take any value on a continuous scale from no overlap to complete overlap. In this view, the timing relation of the two gestures is gradient and so is the phenomenon of assimilation. To illustrate how assimilation is modelled in Articulatory phonology, Figure 2.3 shows gestural scores for two instances of *late calls*.[1]

A gestural score is a tabular visualisation of the gestural activations during the production of a word or phrase. The vertical axis shows the *tract variables* that correspond to the recruitment of the articulators. In the cases shown here, tongue tip (TT), tongue body (TB) and glottis (GLO) are displayed. There are many more tract variables available in the Articulatory phonology modelling approach, for an overview see Browman & Goldstein 1992 and Mücke 2018, as well as the next chapter. The horizontal axis of the gestural score displays *time* such that if an interval starts to the right of another interval, it is said to start later. The boxes give the description of the *location* and *degree of constriction* in a tract variable. The constriction locations used in the score in Figure 2.3 are alveolar (alv), palatal, velar and uvular. The constriction degrees used in the scores comprise two types of closure: clo, denoting a full closure, and clo*, denoting a partial closure with a lateral opening. In addition, critical (crit) stands for a very narrow constriction resulting in friction noise. Finally, the degrees of constriction for the vowels used

---

[1]The scores shown here might not provide a full detailed account of the gestural activations for this phrase, as for example the /l/ in English in word-medial positions is often described as "dark". Therefore, /l/ in English may be better described by a tongue tip and a tongue back gesture that vary in the degree of overlap.

here are narrow and mid. For the glottis, wide simply indicates that the glottis is open for voiceless, it is closed for voiced otherwise.

The example shows how the gestural organisation of Articulatory phonology models the gradience of a process like assimilation as continuous variation in the overlap of gestures. In the score at the top of Figure 2.3, there is no overlap between the alveolar closure gesture of the tongue tip (alv clo in the TT row) and the velar closure gesture of the tongue body (velar clo in the TB row). This score corresponds to a rather careful, clear rendition of *late calls* that results in a clear differentiation of the velar and the alveolar stop. In the score at the bottom of the figure, the two gestures overlap to a large degree. In the acoustic signal corresponding to this score, the two stop sounds would not be differentiated well and hence assimilation would be recorded in a symbolic transcription. However, there is no manipulation or transformation of the set of phonological units involved here as in the rule of 2.1 where one symbolic representation was replaced by another. The framework models the process as a gradient change in a continuous phonological representation. Although the boxes and labels (e.g. mid vs. wide) in the gestural score make the modelling approach appear discrete, one has to bear in mind that the boxes and labels are just "shortcuts". A gesture is defined as a continuous dynamical system (Browman & Goldstein 1992; Hawkins 1992), constriction place and degree can vary continuously in this system. In
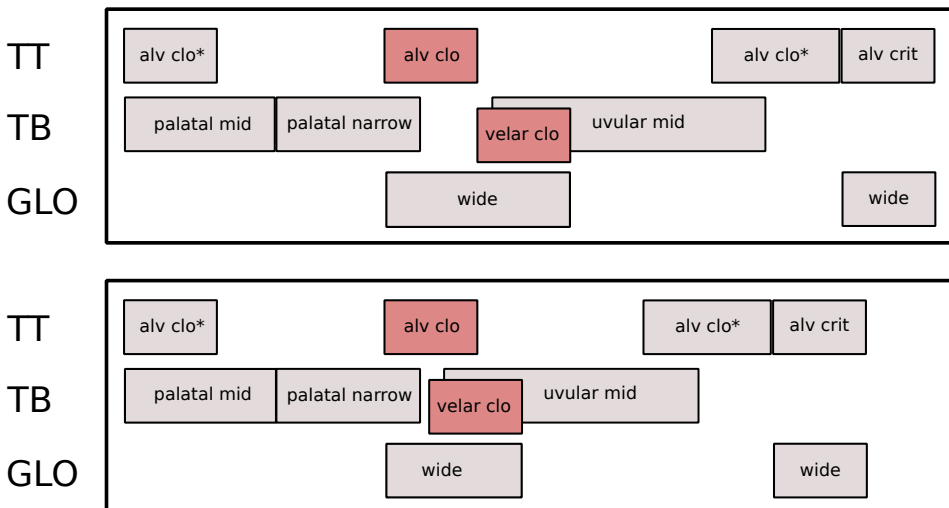

Figure 2.3: Gestural scores of the utterance *late calls* with no overlap of the two stop sounds (top) and high degree of overlap of the two stop sounds (bottom).

consequence, a label like "mid" stands for a scalar value on a continuous scale of constriction degree and not for a discrete symbol from a finite set.

The case of assimilation as described above is an example of *gradience* in speech – a concept that has gained growing attention in recent research in phonetics and phonology (Cohn 2006). Since it is used to denote similar but different phenomena, the term gradience has to be differentiated here. Ladd 2014 referring to Bolinger 1961 distinguishes *physical* from *statistical* gradience. While physical gradience refers to detailed variation on a continuous scale, statistical gradience denotes variation in the statistical patterns of occurrence of an event that can be described as categorical. This differentiation will be referred to in the further course of this work. However, to avoid confusion with other uses of gradience, I will adopt a different terminology. Many scholars use the word gradience to exclusively describe what Ladd calls physical gradience, not statistical gradience. Hence, I will adopt the term *variation* instead. Because statistical gradience or variation refers to the occurrence of categorical events or entities, the term *categorical variation* will be used. Further, I will use the term *continuous variation* to refer to what Ladd 2014 calls physical gradience.

Various theoretical frameworks have dealt with the development of an accurate description of variation. Articulatory phonology, as demonstrated above, provides a means to account for continuous variation as in the case of assimilation. Of course, using the ends of the continuum, it is also suited to describe categorical variation (e.g. no overlap vs. complete overlap). But most theories have focussed mainly on a categorical description of variable patterns in speech sounds. In turning to patterns of language usage, scholars in the variationist tradition of sociolinguistics decades ago acknowledged that multiple forms of a word may coexist. As a consequence, the rule-based framework endorsed by generative phonology was extended to entail variable rules (Labov 1969; Cedergren & Sankoff 1974; Anttila 2007). A variable rule is applied optionally with a specific quantity that denotes how often the rule will be applied. Such a rule may be written in the form X → (Y) / A_B with the parentheses indicating that the outcome of the rule is not generated in all cases ("X optionally turns to Y in the context between A and B").

Other modelling approaches that are able to describe patterns of categorical variation are stochastic extensions of *Optimality Theory* (Prince & Smolensky 2004; henceforth: OT). OT is a theoretical framework that works with symbolic representations but without rules. Instead, it uses ranked *constraints* to evaluate the "winner" from a set of possible *output form* candidates for a given *input form* (Gussenhoven 2004). The winning output form is said to optimally satisfy the constraints and is delivered to phonetic implementation. The tableau in 2.2 gives

a simplified example for computation of the plural form of the English word *kiss* from Gussenhoven & Jacobs 2011. It uses the constraint *SɪʙSɪʙ that is violated if the word contains two adjacent sibilants, the constraint Dᴇᴘ-IO that is violated if a segment not present in the input form is inserted, and the constraint *αᴠᴏɪᴄᴇ-αᴠᴏɪᴄᴇ that is violated if two adjacent sibilants do not share the same quality for the voicing feature (e.g. voiceless may not be followed by voiced). The constraints are ranked from left to right: *SɪʙSɪʙ outranks Dᴇᴘ-IO, Dᴇᴘ-IO outranks *αᴠᴏɪᴄᴇ-αᴠᴏɪᴄᴇ. The symbol * denotes one violation of a constraint for the candidate in that row. The symbol ! indicates whether this violation is "fatal" which means that this violation leads to the "defeat" of this candidate. In the example in 2.2, candidate b is the only candidate that does not violate the constraint of the highest rank, *SɪʙSɪʙ, and thus is the optimal form, marked with the symbol ☞. Of the two violations for candidate a, the violation of *SɪʙSɪʙ is fatal because it already renders it a loser of the competition regardless of its violation of *αᴠᴏɪᴄᴇ-αᴠᴏɪᴄᴇ.

/kɪsz/	*SɪʙSɪʙ	Dᴇᴘ-IO	*αᴠᴏɪᴄᴇ-αᴠᴏɪᴄᴇ
a. kɪsz	*!		*
☞ b. kɪsɪz		*	
c. kɪzz	*!		
d. kɪss	*!		

(2.2)

As can be seen in this simple example, a basic OT approach determines a single optimal form – and the same optimal form will be evaluated by the grammar every time unless the constraint ranking is changed. To implement categorical variation, a probabilistic mapping of input and output form, this account has to be extended. To illustrate the process, one case that will be treated more in-depth here is the probabilistic choice of suffixes in Hungarian *vowel harmony* (Hayes & Londe 2006). Vowel harmony has been described as a phenomenon in which the vowels within a word agree with regard to some phonetic property, like the place feature [±back] (Gafos & Benus 2006). In Hungarian, for many stems, the quality [±back] of the stem vowel determines the choice of the suffix such that the suffix agrees with the stem in its [±back] property. For example, the stem *ablak* /ɔblɔk/ ('window') in which the last vowel is [+back] takes the suffix *nak* /nɔk/ with a back vowel while the stem *üst* /yʃt/ ('cauldron') with a front vowel takes the suffix *nek* /nɛk/ with a front vowel (Hayes & Londe 2006: 62). In addition, front unrounded vowels function as *transparent or neutral vowels* (Hayes & Londe 2006). These vowels can occur between the vowel triggering the vowel harmony and the target of the vowel harmony, e.g. the suffix vowels in the examples above, but do not affect the process of vowel harmony – even if their

quality of the feature [±back] is opposing to the triggering vowel. For instance, the stem *kávé* /kaːveː/ ('coffee') takes the suffix *nak* /nɔk/ – the back vowel of the stem determines the vowel of the suffix regardless of the intervening unrounded front vowel.

Interestingly, stems with a back vowel and one or two transparent vowels have been observed to be able to take back and front suffixes, with statistical preferences for one or the other. For instance, Hayes & Londe (2006) observed that the stem /aːɲiveːl/ occurs with the suffix *nek* /nɛk/ in 83.6% of cases and with the suffix *nak* /nɔk/ in the remaining 16.4% of cases. The authors use stochastic OT (Boersma 1997; Boersma & Hayes 2001) to model this probabilistic suffix alternation. To give a full account is beyond the scope of this chapter. The following description is restricted to a short overview to exemplify how stochastic OT can account for categorical variation. The constraint rankings are not strict in this approach, rather the constraints are assigned ranking strength probabilities. The tableau in 2.3 uses three constraints: Local[NN] which is violated when a stem with two neutral vowels is followed by a back vowel, Local[eː] which is violated when the closest vowel following [eː] is a [+back] vowel, and Distal[B] which is violated when a [+back] vowel is followed by a [−back] vowel somewhere in the word.[2] The fact that they are not ranked strictly (or statically) as in the previous example is expressed by the dashed separation lines between the columns. The ranking strengths of the constraints are given by the probability density functions shown in Figure 2.4. Local[NN] has a mean ranking strength of 101.802, Local[eː] has a mean ranking strength of 100.894, Distal[B] has a mean ranking strength 100.000. The standard deviation is 2 in all cases. From these ranking strength distributions, the probability for a given output form can be calculated. Candidate a in tableau 2.3 from Hayes & Londe (2006: 81) violates Distal[B] three times, candidate b violates the other two constraints but Distal[B] only twice. For candidate b to win, the constraint Distal[B] has to be outranked by both Local[NN] and Local[eː]. The question is: If one ranking strength sample is taken from each of the three distributions presented in Figure 2.4, how often will the two samples for Local[NN] and Local[eː] both be smaller than that for Distal[B]? The answer is: In 16.4% of all cases. In all other cases, candidate a will win. The probabilities are given in the first column, to the left of the candidates.

/aːɲiveːl-nAk/	Local[NN]	Local[eː]	Distal[B]	
0.836 ☞ a. aːɲiveːl-nɛk			***(!)	(2.3)
0.164 ☞ b. aːɲiveːl-nɔk	*(!)	*(!)	**	

[2]This is not the complete set of constraints used by Hayes & Londe (2006) to model the data, it represents a minimal example to illustrate the point of probabilistic constraint rankings.
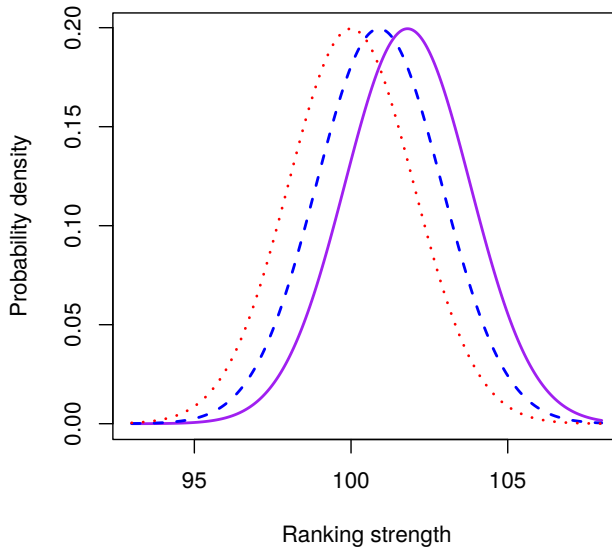
Figure 2.4: Ranking strength probabilities for the constraints based on footnote 15 of Hayes & Londe (2006). Local[NN] (purple, solid line), Local[e:] (blue, dashed line), Distal[B] (red, dotted line).

As becomes clear from this example, stochastic OT is able to cope with statistically gradient patterns, or categorical variation, i.e. the statistical preference for one or another category. Continuous variation is not captured by this framework. In the next chapter, a dynamical systems approach to the problem of the probabilistic suffix choice in Hungarian will be outlined that focusses on continuous sub-symbolic variation (Gafos & Benus 2006). This model builds on articulatory details and shows how continuous variation may contribute to categorical variation.

As outlined in the introduction to this book (Chapter 1), it is one of the aims of the present work to show that categorical and continuous variation often go hand in hand. This view is in line with Ladd (2014: 88) who notes that "[i]n many situations, of course, the two types of variation are likely to interact and reinforce one another." Motivated by the fact that categorical and continuous phenomena in speech often show a high degree of parallelism, Flemming (2001) presents a model rooted in OT that combines continuous and categorical aspects and intends to reconcile phonetic and phonological representations in a formal approach. The main idea revolves around a trade-off of constraints rather than a categorical ranking of constraints. This trade-off can be determined by expressing the violation of each constraint as a scalar quantity. As a result, each possible candidate of the optimalisation process acquires a cost and the candidate with

the lowest cost is selected as the optimal form. This approach will be outlined here in more detail using the example of assimilation of a back vowel to a coronal consonant.

It has been observed that in some languages the contrast between front and back rounded vowels is neutralised in the position between coronal consonants. For example, in Cantonese Chinese the two forms /kʰyt/ ('decide') and /kʰut/ ('bracket') exist, as well as /tʰyt/ ('to take off'). There is, however, no form */tʰut/ (Flemming 2001). This phenomenon has been described as a case of categorical assimilation, i.e. as the outcome of a phonological computation in which the [+back] specification of the vowel is changed to [+front]. In other languages, a parallel process, the co-articulatory fronting of back vowels in the context of coronal consonants has been described as purely phonetic (and as such continuous). In English, for example, /u/ is produced more fronted in *toot* /tut/ compared to *coo* /ku/.

The modelling idea of Flemming (2001) involves the formulation of the following constraints based on a quantification of the second formant (F2) of both the consonants and the vowel. The constraint IDENT(C) requires the target F2 of a consonant to be maintained. Its violation cost is the weighted squared difference between the realised F2 of the consonant and the target F2 of the consonant: $weight_c \cdot (F2(C) - L)^2$, where $F2(C)$ is the realised F2 of the consonant and L is the target F2 of the consonant.

The constraint MINIMISEEFFORT requires that the speaker reduces articulatory effort. It tries to keep changes in F2 from C to V as small as possible. Its violation cost is the squared difference between the realised F2 of the consonant and the realised F2 of the vowel: $weight_e \cdot (F2(C) - F2(V))^2$, where $F2(C)$ is the realised F2 of the consonant and $F2(V)$ is the realised F2 of the vowel.

The constraint MINDIST = Δ requires that the distance between the F2 of the vowel /u/ and its nearest neighbour /y/ is above a certain threshold. Its violation cost is the weighted difference between the distance of /u/ and /y/ in terms of F2 and the threshold: $weight_v \cdot (|F2(y) - F2(u)| - \Delta)^2$, where $F2(y)$ is the F2 of the vowel /y/, $F2(u)$ is the F2 of the vowel /u/, and Δ is the minimum distance between the two, the threshold. This constraint only applies if the contrast is maintained. Its violation cost is not calculated if the contrast is neutralised, i.e. the assimilation is categorical.

In addition, the model uses the quantity MAXIMISECONTRAST. This value represents the benefit of preserving a contrast. While the other three constraints acquire positive costs as they are violated, MAXIMISECONTRAST is subtracted as a negative cost.

The candidates of the optimisation process are inventories of contrasting syllables. First, the total violation cost for a candidate is calculated as the sum of all single weighted violation costs for the constraints. Second, the benefit of maintaining the contrast, MaximiseContrast, is subtracted. After these steps, the inventory with the lowest cost is selected as optimal. Thus, if the benefit of maintaining the contrast is exceeded by the costs obtained by the distinctiveness and effort constraints (Ident(C), MinDist = Δ, and MinimiseEffort) in the realisation of /tut/, it is optimal to neutralise the contrast between /u/ and /y/. The consequence is *categorical assimilation.* On the contrary, if the combined costs for the violation of the constraints Ident(C), MinDist = Δ, and MinimiseEffort do not exceed MaximiseContrast, neutralisation in the form of categorical assimilation is not optimal. However, as a trade-off between Ident(C) and MinimiseEffort, *co-articulatory assimilation* with varying degrees follows. In this model, the difference between languages can be modelled by using the same sets of constraints and modulating the value for MaximiseContrast as well as the weights for the violation costs of the constraints as scalar values. Depending on how the weights are set, the costs for distinctiveness and effort might exceed the benefit of maintaining a contrast or not (Flemming 2001).

Similar to Articulatory phonology, Flemming's approach models degrees of assimilation in the same formal system and does not distinguish between categorical, phonological and continuous, phonetic processes. In Articulatory phonology, assimilation is a fundamentally continuous process as the organisation of gestural activation varies on the continuous dimension of time – affecting the temporal and spatial properties of the phonetic outcome. Categorical behaviour can be found at the ends of the continuum. In Flemming's model, the trade-off between the benefit of contrast preservation and the costs connected to effort and distinctiveness of sound pattern explains the outcome as categorical or continuous assimilation. In both modelling frameworks, categorical variation on the "macro-level" is seen as the result of the interaction of variation on continuous dimensions on the "micro-level".

2.3 Tiny differences, rich memory

Another famous example that poses a problem for the modular view of phonetics and phonology is the phenomenon known as *incomplete neutralisation.* It describes the finding that the neutralisation of the voicing contrast of syllable-final obstruents present in some languages as the result of *final devoicing* is indeed incomplete. Final devoicing or *Auslautverhärtung* is a classic textbook example

showing that voiced obstruents at the end of syllables turn into voiceless obstruents in German. Thus, the contrast between voiced and voiceless is said to be *neutralised* in this context. Similar phenomena have been observed in other languages as well (Bloomfield 1933). In a modular, symbolic account of the case, the voiced obstruent is transformed into a voiceless obstruent at the end of syllables by virtue of a phonological rule like the one given in 2.4 (where $ stands for the end of a syllable). This rule is suited to turn a structure like /ʁad/ into [ʁatʰ] and thus makes the difference between the phonetic outputs of the words *Rad* ('wheel') and *Rat* ('advice') disappear completely.

$$[\text{+voiced}] \rightarrow [\text{−voiced}]/\underline{\hphantom{xxxx}}\$ \qquad (2.4)$$

   Contrary to the prediction of this rule, a considerable number of studies found that the acoustic signals of words like *Rat* and *Rad* are different (Dinnsen & Garcia-Zamor 1971; Port & O'Dell 1985; Charles-Luce 1985; Port & Crawford 1989; Ernestus & Baayen 2006; Roettger et al. 2014). In general, the voicing contrast can be encoded by different phonetic cues like glottal pulsing in the closure of the stop, closure duration, voice onset time, but also duration of the preceding vowel (Lisker 1986). Studies on the incompleteness of the voicing contrast demonstrated that the differences between the acoustic signals regarding these parameters go in the direction of a voicing contrast in inter-vocalic position as in *Räder* vs. *Räte*, although the differences are much smaller. Recently, Roettger & Baer-Henney 2019 added convincing empirical evidence for the robustness of the incompleteness of German final devoicing using a large, diverse data set.

   An approach based on discrete symbolic representations and rules, like the one outlined in 2.4, is clearly not able to account for the continuous, subtle variation reported by the studies above. The question arises how the difference between the two obstruents can be best captured in a model of phonetics and phonology. An early proposal by Port & O'Dell 1985 is to employ *phonetic implementation rules*. While phonetic implementation rules were already implicitly assumed by earlier models (like Chomsky & Halle 1968; Jacobson et al. 1952), they were not considered linguistic in a strict sense, i.e. not part of langue. Port & O'Dell 1985 extend the classic model of separated phonetics and phonology by considering language-specific phonetic implementation rules that belong to the speaker's knowledge of the language. To account for the incompleteness of final devoicing in German, the authors use a rule that phonetically implements the syllable. It is the duty of this rule to devoice a voiced obstruent at the end of syllables. Thus, the voiced obstruent retains its phonological specification [+voiced] during the stage of phonological derivation. Only later, at the stage of phonetic implementa-

tion, a gesture is activated that resembles the one found with sounds that possess the phonological quality [−voiced].

While the inclusion of a phonetic rule that is in fact part of the language "makes the relationship between phonetics and phonology closer by permitting the phonetic implementation system to directly execute the macrostructures in the phonology" (Port & O'Dell 1985: 468–469), it strictly adheres to the modular view of phonetics and phonology. In fact, it emphasises the separation between the two modules despite making them more similar. In a footnote, Port & O'Dell 1985 entertain the interesting idea that *rules* might not be the right devices in general to account for the cognitive underpinnings of language and speech. In the next chapter (Chapter 3), a dynamical model for the subtle differences observed in the syllable-final obstruents in German is introduced (Gafos 2006; Gafos & Benus 2006). This model does not assume rules or a separation of categorical and continuous aspects. As will become clear in this chapter, despite using a continuous formalism, the model is nevertheless able to explain the categorical nature of the [±voiced] feature.

Other accounts have modelled the phenomenon as an artefact of lexical co-activation, also known as *phonetic paradigm uniformity*, without the need for the invention of a new kind of rules (e.g. Ernestus & Baayen 2006; Goldrick et al. 2010; Kleber et al. 2010; Winter & Roettger 2011; Roettger et al. 2014; Seyfarth et al. 2019). The basic idea behind these approaches is that the mental lexicon stores a collection of rich auditory forms of words, like *Rad*, *Räder*, and *Rades*, rather than abstract representations and rules to produce derived forms. During the activation of a word like *Rad*, close lexical neighbours like *Räder* and *Rades* are co-activated. These co-activated forms enhance the probability that the final obstruent of *Rad* is realised with phonetic characteristics slightly pushed in the direction of a voiced obstruent (Ernestus & Baayen 2006; Roettger et al. 2014).

These approaches are rooted in a framework known as *exemplar theory*. Exemplar theory, originally proposed in psychology for the classification of multi-dimensional stimuli more generally (Nosofsky 1986; Hintzman 1986), has gained a lot of attention in phonetics and phonology. It postulates detailed, episodic memory of acoustic traces of words as the basis of speech production and perception (Goldinger 1996; Johnson 1997; Pierrehumbert 2001; Bybee 2001; Pierrehumbert 2016: among others). Exemplar theory is thus fundamentally different from rule-based or constraint-based theories of phonology which, as outlined above, assume abstract units to achieve "maximally simple, redundancy-free representations" (Gahl & Yu 2006: 213). Exemplar theory postulates that the detailed experiences of language use shape the cognitive representation of speech sounds through memorisation of the particular signal (or at least parts of it). Every time

a speaker hears a word – be it spoken by another person or herself – a new acoustic trace called *exemplar* is stored in her memory near the location of the existing exemplars. During the perception and the production of speech sounds, clouds of stored exemplars are activated. In the case of categorisation, the incoming stimulus is compared against the rich inventory of exemplars. The stimulus will be categorised as belonging to the group of stored exemplars that are most similar. In the case of production, the stored clouds of exemplars serve as production targets. In both perception and production, newer exemplars have more influence since memory traces fade away over time (Schweitzer 2012).

Exemplar models are able to account for experimental findings demonstrating the importance of fine-grained phonetic detail in both perception and production (Pierrehumbert 2016). For example, Wright 1979 as well as Jurafsky et al. 2002 showed that more frequent words are produced faster and more reduced compared to less frequent words. This finding also applies to words that are more predictable from the context (Seyfarth 2014; Aylett & Turk 2004; Hall et al. 2018). From an exemplar-based perspective, articulatory effort can be seen in relation to the likelihood of activation. A lexical item with a large cloud of exemplars is activated more easily than a lexical item with a smaller cloud. As outlined above, each time a word is perceived, its acoustic trace is added to the existing exemplars. Hence, less articulatory effort is needed for the transmission of words that are heard very frequently since their exemplar clouds generally accumulate more traces.

In addition to frequency, lexical access can also be facilitated by many other factors, like indexical information about the speaker producing the word: Walker & Hay 2011 showed that listeners were faster and more accurate in lexical decision for words that are more frequently heard in real life said by older speakers, like *typist*, when these words are produced by an older voice in the experiment. Vice versa, words that are more frequently heard in real life from younger speakers, like *checkout*, were more easily processed in the experiment when produced by a younger voice.

Exemplar theory emphasises the importance of phonetic detail for our understanding of speech and offers a completely different view on the relation between phonetics and phonology. While abstractionist frameworks like the rule-based or constraint-based models presented in this chapter view phonology as a reduced, abstract structure that is efficient in terms of memory consumption, exemplar theory posits that the speaker stores and uses large pools of real, experienced "data". While abstractionist models operate on segment-sized units that in some sense adhere to the phonemic principle, exemplar theory uses larger chunks like words. Scholars in the abstractionist tradition have often highlighted

that segment-based approaches are good at explaining the regularities in sound change throughout the whole phonological system, like *chain shifts* described by Grimm's law (Guy 2014). Exemplar theorists on their part have argued that sound changes may not necessarily be regular. While Middle English long /o/ in today's English turned into /u/ in words like *root* and *food*, it also developed into /ʊ/ in words like *good* and *book* or schwa as in *flood* (Guy 2014). However, pure exemplar approaches have difficulties dealing with regularities in sound change – even if these regularities might not complete (Pierrehumbert 2016). Further critique towards purely "data-driven" exemplar models includes evidence that in addition to *token frequency* (accumulated traces of perceived words), *type frequency* of a lexical unit plays a role for the productivity of this pattern (Hay et al. 2004).

As a consequence, exemplar-based approaches have been extended to *hybrid models* which argue that abstract phonological representations also play a role. One of these models is delivered by Pierrehumbert 2002 positing that abstract generalisations and exemplar clouds can be associated with phonological units like phonemes, sequences of phonemes, or words (see also German et al. 2013). The model makes a distinction between production and perception. In production, phonological units play a major role but exemplars bias the production goals of the abstract units. On the contrary, in perception, exemplars play the leading role. The related *Polysp* model (Hawkins & Smith 2001; Hawkins 2003) emphasises that identification of meaning in communication "takes place probabilistically, using all possible available information in parallel to flesh out linguistic structure at all levels" (Hawkins 2003: 391). Following this model, a listener might analyse a given acoustic signal into abstract linguistic units and match the signal with exemplars (Ernestus 2011). Notably, in *Polysp*, exemplars can include non-acoustic memory like visual information. Sumner et al. (2014: 8) endorse a similar view stating that "[l]isteners simultaneously extract linguistic and social information from speech". While hybrid approaches offer promising explanations and testable predictions, Ernestus 2011 notes that a full computational implementation of these models is still in progress.

2.4 Summary

In this chapter, the relation of phonetics and phonology was explored. It was outlined that the 20th century, with the rise of the phonemic principle, introduced a rather sharp separation between continuous, physical phonetics and categorical, symbolic phonology – a distinction that was weaved into the fundamentals

of linguistics. Even models in the tradition of Chomsky & Halle (1968) that dispense with the concept of the phoneme in a strict sense maintain the split as an important assumption. The chapter reviewed some of the problems which challenge a clear-cut separation and highlighted that many scholars nowadays assign a significant role to fine phonetic detail in cognitive representations of sound patterns as well as storage in memory. In this context, it becomes increasingly important to investigate how theoretical approaches are able to deal with continuous and categorical variation in sound patterns and linguistic structure. Models rooted in the framework of dynamical systems offer promising solutions for reconciling the continuous and categorical aspects in one formal language. Since the focus of this book is on modelling approaches within this framework and what they have to offer for an understanding of the relation of categorical and continuous aspects of speech, the next chapter introduces the foundations of dynamical systems and reviews important applications.

3 Dynamical Systems

The previous chapter reviewed models of phonetics and phonology and high-lighted the importance of the ability to capture both categorical as well as con-tinuous aspects of speech. Many problems that arise in modelling the relation of these aspects are connected to the use of fundamentally different representations. While phonology is modelled as a system of discrete computations, phonetics is conceptualised as essentially continuous (Gafos 2006). Hence, in a many wide-spread models, "a mapping between a categorical symbolic representation and a quantitative physical signal" (Ladd 2006: 8) is necessary. In contrast to these translational approaches, dynamical systems offer an alternative by employing a single formal language to capture both categorical and continuous aspects of speech. The potential of dynamical systems can be attributed to the fact that categorical behaviour emerges from continuous changes of parameters in a con-tinuous space of possible states.

The last decades have seen a growing body of research that has pointed out the dynamical nature of the mind (e.g. Kelso 1995; van Gelder & Port 1995; Port 2002; Spivey 2007; Kelso 2013). In order to overcome limitations imposed by purely symbolic approaches, researchers from many disciplines have turned to the framework of dynamical systems describing a multitude of different cogni-tive processes. Dynamical models for action and perception in cognition, includ-ing language and speech, emphasise the idea that the mind "travels" through a continuous, many-dimensional space towards stable states (Spivey 2007: 4). This idea is in sharp contrast to the traditional conception of the mind working like a computer that manipulates and replaces symbolic representations (Fodor 1975; Fodor & Pylyshyn 1981; Harnad 1990; Newell & Simon 1976). In a continuous, dynamical conception of cognition, the mind smoothly passes through multiple states during the process of settling in one stable state, rather than abruptly ex-changing one symbol for the other. This perspective aims to shed light on the unfolding of a cognitive process over time and the relative stability of what can be described as a category in relation to other categories (Port 2002; Gafos & Benus 2006; Spivey 2007).

The present chapter concentrates on the basic concepts of the mathematics of dynamical systems and their applications to the description of patterns in speech.

It attempts to be as illustrative as possible and serves as a background for the modelling approach of the present book for readers who are not familiar with the topic. The chapter is accompanied by MATLAB scripts that run the simulations and produce the plots shown alongside the text. The code can be retrieved here: https://osf.io/4g6s2/. Details about which scripts are used can be found at the end of each section.

3.1 The fundamentals of dynamical systems

Complex, dynamical systems are found in all aspects of the world. Importantly, in such systems, the patterns of behaviour of the system emerge from the interaction of the parts of the system (Fuchs 2013; de Boer 2001). This feature distinguishes them from other systems in which the behaviour is determined by a hierarchical structure, for instance structure that is built-in by design as in many engineered systems. A striking example is the formation of a traveling wave, called *la ola* or *Mexican wave*, through a crowd of people in a stadium – a phenomenon that has been scientifically studied by Farkas et al. (2002). To form the wave, individuals successively stand up and raise their arms. Crucially, this collective, coherent behaviour can neither be triggered, killed, slowed down or speeded up by a single individual (Fuchs 2013). It arises under certain conditions when a small, critical mass of people initiates the movement. For example, since the level of excitement has to cross a certain threshold, it does not arise when the home team is losing. If it starts, the wave can spread over many thousands of people through the local interaction of individuals: Active individuals activate near-by individuals to stand up and raise their arms. Thus, the global near-linear shape of the wave emerges as the result of an interaction of the single parts of the system over time (Farkas et al. 2002).

To understand dynamical systems and their application to describe phenomena in the speech, language and cognition sciences, it is useful to look at some of the basic concepts of dynamical systems. This section will concentrate on these basics without providing a full introduction to the topic. Interested readers are referred to Fuchs (2013), Kaplan & Glass (1995) and Iskarous (2017) among many other great fully-fledged introductions to dynamical systems.

3.1.1 Order and chaos

The aim of the theory of dynamical systems theory is to create compact mathematical descriptions of the behaviour of complex systems like the *la ola*. In doing

so, dynamical systems focus on how a system changes over time based on the state that the system is currently in. To get a general understanding, it is helpful to study the *logistic map*, a system formulated to describe the development of populations that was made popular by biologist May (1976) as a discretised version of the demographic model proposed by Belgian mathematician Verhulst in the mid 19th century. The logistic map is given in Equation 3.1.

$$x_{t+1} = kx_t(1 - x_t) \tag{3.1}$$

The formula defines how the state x of the system at a time point $t + 1$ is calculated. Crucially, this future value of the state variable x depends on the current state at t. In addition, the system has a parameter k that represents the *growth rate*. For example, if $k = 0.7$ and $x_1 = 0.5$ at the present time point, the system will predict $x_2 = 0.7 \cdot 0.5 \cdot (1 - 0.5) = 0.175$ at the next time point. Figure 3.1 shows how the evolution of the system is continued over 19 additional time steps. The graph shows that the points gradually approach zero. In terms of a population model this means extinction of the population. As there is no member of the population left to reproduce, the system will stay in the state with the value zero forever. This state is called the *attractor* of the system. Regardless of the state of the system in the beginning, the system will end up in this state.
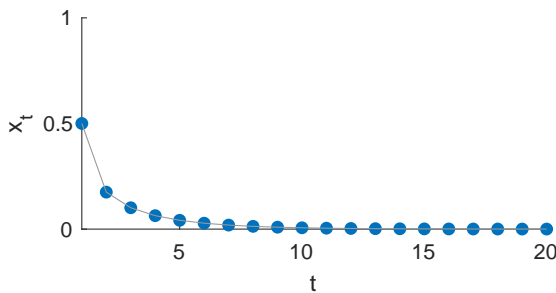

Figure 3.1: Example for the evolution of the logistic map with $k = 0.7$ and $x_1 = 0.5$.

Depending on how the growth rate is chosen, the system can exhibit a variety of patterns. Figure 3.2 gives examples for the logistic map with different values for the growth rate k. In all cases, x_1, the initial state, is 0.42. In the case of $k = 1.2$ (top left), the system monotonically approaches one value. This type of attractor is called *point attractor*, it is the same kind of attractor as the one in the illustration of Figure 3.1. In the case of $k = 2.9$ (top right), the system also approaches one steady state, but while approaching the attractor, it alternates from one side to

the other. When $k = 3.3$ (bottom left), the attractor of the system is not a steady state. Instead, the system oscillates between two values. This is type of attractor is called a *limit cycle*. Another interesting case is $k = 4$ in which the system also oscillates but not in a periodic manner. This behaviour is called *chaos* (Kaplan & Glass 1995). When exhibiting chaotic behaviour, the system will never settle in a steady state or limit cycle but oscillate forever without repeating itself.
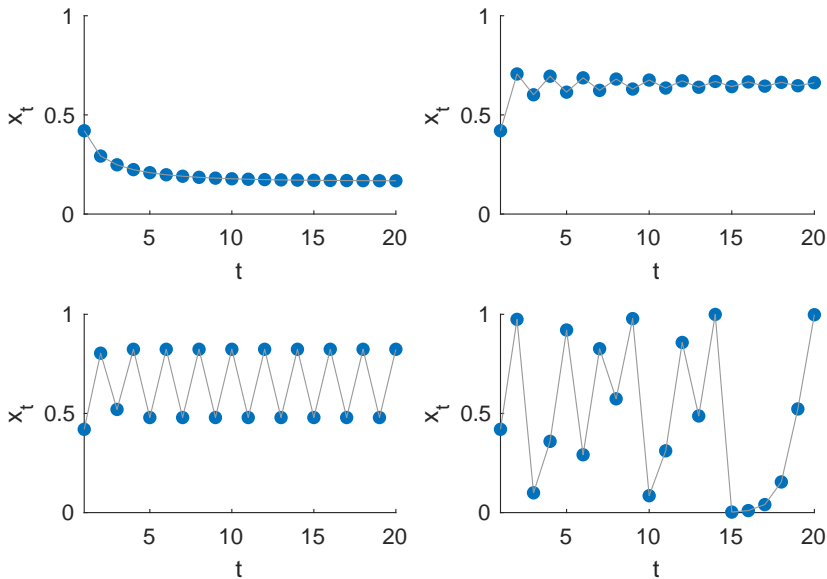

Figure 3.2: Example for the evolution of the logistic map with different values for the growth parameter (top left: $k = 1.2$, top right: $k = 2.9$, bottom left: $k = 3.3$, bottom right: $k = 4$).

Figure 3.3 illustrates the possible patterns of the logistic map as the growth parameter is changed. The plot, also known as a *bifurcation plot* (Feigenbaum 1978), was created by running the logistic map for 2000 iterations for all growth parameter values between 1 and 4 with a step size of 0.01 (i.e. the simulation was first run with $k = 1$, then $k = 1.01$, then $k = 1.02$, and so on). The initial state is 0.42 in all simulations. The axis for the parameter value is the x-axis. Of the 2000 iterations for each parameter value, the last 100 values are plotted on the y-axis as single tiny dots. For $k < 3$, all dots are plotted on top of each other because after a few iterations the system reached a steady state. As the parameter k is increased, the system's attractor is a limit cycle that oscillates between two values. The cycle frequency is then increased to 4, 8, 16 until the

system eventually exhibits aperiodic behaviour as described above. There are, however, bands of growth parameter values in between for which the system moves back to periodic cycles (visible as white stripes in the higher regions on the x-axis) (Kaplan & Glass 1995; Spivey 2007). The behaviour of the system as *k* is scaled beyond 3.5 can be observed in more detail in Figure 3.4. This figure shows the same plot as Figure 3.3 with the x-axis zoomed in to the range of *k* values from 3.5 to 4.
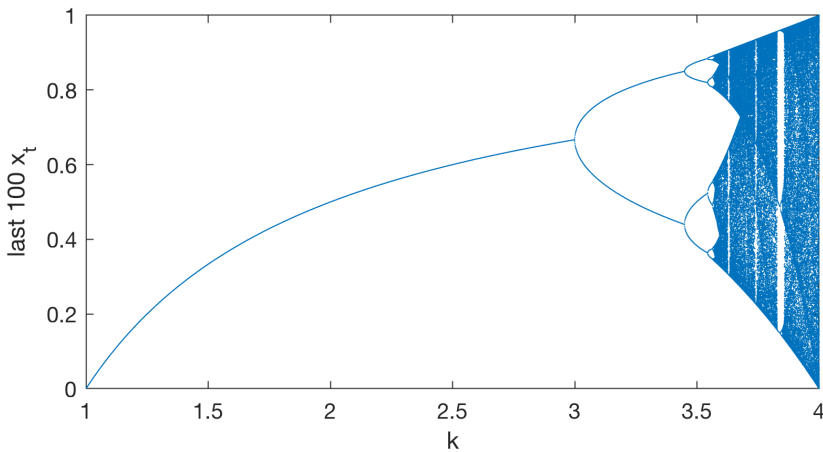

Figure 3.3: Bifurcation plot of the logistic map.

The examples presented around the logistic map demonstrate that an attractor can be a single value or a set of values. For the present work, point attractors will be most important. To change the behaviour of the system, the examples adjusted the growth parameter value *k*. A parameter like *k* is called a *control parameter* – a very important concept for the understanding of dynamical systems. Control parameters can be thought of as the parameters that "move the system through its repertoire of patterns and cause them to change" (Kelso 2013: 1538). They often represent environmental conditions, like the growth parameter in the logistic map example, but are not restricted to this role (Kelso 2013).

An essential property of dynamical systems is that they often exhibit *qualitative change* as the control parameter is scaled. This behaviour is referred to as *bifurcation* – a term used above already with regard to Figure 3.3 and 3.4. As the name of the chart displayed in the two figures, bifurcation plot, already suggested, it shows how the logistic map undergoes bifurcation as the control parameter (the growth parameter) is increased. While the system starts with a
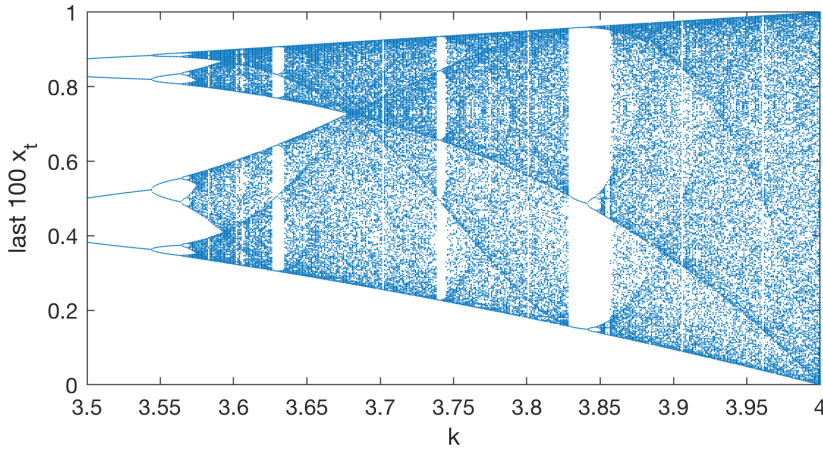
Figure 3.4: Bifurcation plot of the logistic map for values of $k \geq 3.5$.

point attractor, it changes into a pattern with a limit cycle as the control parameter passes a critical threshold. Subsequently, for other control parameter thresholds, the periods of the cycles change before the system turns to aperiodic behaviour. It then changes between aperiodic and periodic patterns for different ranges of control parameter values. All these transitions are qualitative changes, i.e. bifurcations, in the dynamical system.

Code used in this section: logistic_map_evolution.m, logistic_map_bifurcation.m

3.1.2 The use of differential equations in dynamical systems

In the logistic map, time progresses in discrete steps (or generations). When solving the equation, one considers the time points $t_1, t_2, t_3, ..., t_n$. In reality, of course, time is best characterised as continuous. A very common way to describe a dynamical system with reference to continuous time is by using *differential equations*. The equation in 3.2 gives a simple example for a differential equation. It uses a very common notation with a dot above the x which denotes that the function is a derivate with respect to time, $\dot{x} = \dfrac{dx}{dt}$ (Fuchs 2013). The basic idea in this system is very similar to the logistic map: We observe the behaviour of the state variable x. To learn about the change of the system at each state of x, the differential equation of the system is solved.

$$\dot{x} = kx \tag{3.2}$$

To get an impression of what this simple system does, it is useful to start at some randomly chosen state and look at the evolution of the state variable x over time. One way is to use a method known as the *Euler method*, named after Swiss mathematician Leonhard Euler.[1] Simply speaking, this method takes the change at the current state – calculated by solving the differential equation of the system – and adds it to the current state. Given that the state at some time point t is known, the state at time point $t + h$ can be calculated easily as $x_t + hkx_t$ (Fuchs 2013) where h denotes the size of the step that is taken in time (also expressed as Δt). As noted above, time is conceptualised as continuous in differential equations with regard to time, so using a discretising solution seems somehow paradoxical. However, it provides an easily implemented computation method for the solution of differential equations and the step size h can be chosen very small to account for the fact that in continuous time the difference between t and $t + 1$ converges to zero. For the present work, the approximation of continuous time with sufficiently small time steps will be accepted as it will not disturb the results of the account. Figure 3.5 presents the evolution of the system of 3.2 with a k of -0.5 starting at the initial state $x = 0.42$. Similar to the example of 3.1, this system approaches an attractor at zero.
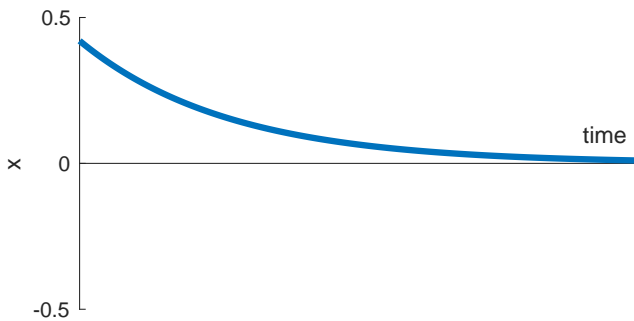

Figure 3.5: Evolution of the system given by the differential equation $\dot{x} = kx$ for $k = -0.5$.

Figure 3.6 demonstrates the effect of scaling the control parameter. For all values of k, the system stays at zero if the initial state is zero (yellow lines). This is because the change in the system is given by $\dot{x} = kx$ and if x is zero, the

---

[1]There are also other methods to solve differential equations, like the Runge-Kutta method. For some systems, like the one of Equation 3.2, it is also possible to give an exact analytical solution. For the exposition of systems considered in this work, subtle differences between the methods do not play a role.

change will be zero as well. For all other initial values, the following pattern can be observed: k values smaller than zero make the system approach the attractor at zero. For k values greater than zero, the system does not converge to any attractor. With $k = 0$ the system does not change at all but stays in its initial state.
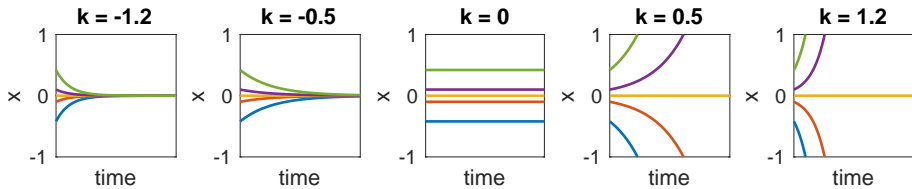

Figure 3.6: Evolution of the system given by the differential equation $\dot{x} = kx$ for different values of the control parameter k. Colours indicate initial states: green = 0.42, purple = 0.1, yellow = 0, red = −0.1, blue = −0.42.

To learn about the patterns of the system without calculating solutions, a *phase space plot* can be a useful tool (Fuchs 2013). In a phase space plot it is particularly interesting to look for *fixed points* – points where the differential equation is zero, i.e. $\dot{x} = 0$. Figure 3.7 shows the phase space plots for $\dot{x} = kx$, for $k < 0$ and $k > 0$. The solutions for the system showed that nothing happens when $k = 0$, the system remains in its initial state. Hence, this case will not be discussed any further here and it will only be investigated what happens for $k < 0$ and $k > 0$. First, the left plot of Figure 3.7 is discussed where the control parameter k is below zero. The function $\dot{x}$ has positive values for negative x values, and negative values for positive x values. Whenever the system is in a negative state (negative x), the change given by $\dot{x}$ is positive, so the system moves towards zero from the left side. Whenever the system is in a positive state (positive x), the change given by $\dot{x}$ is negative, so the system moves towards zero from the right side. This direction is indicated by the arrows on the x-axis. The further away the value of x from zero, the greater the change – this is shown by the size of the arrows. In addition, it is clear that $\dot{x} = 0$ for $x = 0$. In sum, for $k < 0$, when the system is at zero, it stays there; when it is in some other state (be it negative or positive), it gravitates towards zero. The black filled circle at zero indicates that zero is a *stable fixed point* of the system, an attractor.

In the right plot, the space phase plot is shown for cases in which the control parameter k is greater than zero. Whenever the system is in a negative state (negative x), the change given by $\dot{x}$ is negative, so the system moves further away from zero towards $-\infty$. Whenever the system is in a positive state (positive x), the change given by $\dot{x}$ is positive, so the system moves further away from
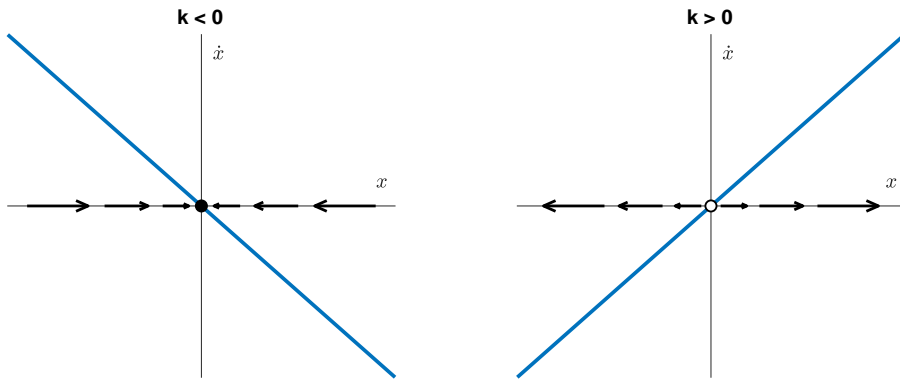
Figure 3.7: Phase space plot of $\dot{x} = kx$. Arrows indicate the direction of change in the system. The filled circle stands for attractor, the empty circle stands for repeller.

zero towards ∞. When the system starts at $x = 0$, it will stay there. But this is the only possibility to be in the state of zero for $k > 0$. All other initial states will lead away from zero, even states that are very close to zero, as for example $x = \frac{1}{10000}$. This is again indicated by the arrows, in this case pointing away from zero. Zero in this scenario is called an *unstable fixed point* or *repeller*, represented by an empty circle.

Code used in this section: ds_straight_line.m

3.1.3 Multistability

Dynamical systems can have more than one attractor – a situation called *multistability*. Multistability is different to the ranges of control parameter values in the logistic map that exhibit a limit cycle attractor. In the case of the limit cycle, all values in the set of the attractor will be visited – the system oscillates between these values. The whole set is one attractor. In the case of a system with more than one attractor, the initial state is crucial in determining in which of the attractors the system will settle. At the end of this chapter, the concept of the stochastic dynamical system will be presented. In this case, random fluctuations play an important part and the role of the initial state in a multistable system is diminished.

The system that will be used here to illustrate the situation in which more than one attractor is present is the one defined by the differential equation in 3.3, a

system that will play a major role in this work. Figure 3.8 gives the solutions for some values of k and some intitial states. Two aspects are important here: First, for $k = 0$ and $k = \pm 0.2$, some trajectories settle in a positive attractor, others settle in a negative attractor. The attractor in which the system settles depends on the initial state. Second, for $k = \pm 0.7$, the situation is different, the system only settles in one attractor regardless of the initial state.
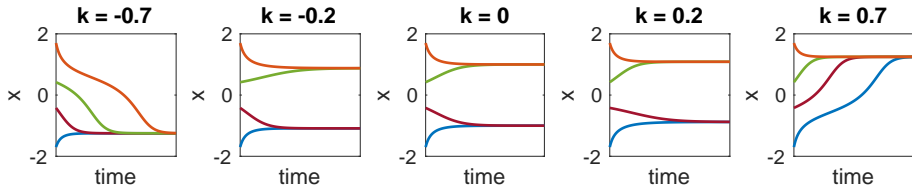
$$\dot{x} = k + x - x^3 \tag{3.3}$$

Figure 3.8: Evolution of the system given by the differential equation $\dot{x} = k + x - x^3$ with different control parameter k values and different initial conditions. Colours indicate initial states: blue = -1.7; red = -0.42; green = 0.42; orange = 1.7.

The behaviour of this system can again be illustrated by the phase space plots (Fuchs 2013), see Figure 3.9. As before, an attractor is drawn as a full circle, a repeller is drawn as an empty circle. When the control parameter is zero, i.e. $k = 0$ (plot in the centre), two attractors and a repeller in the middle exist. When the control parameter is decreased (top row) or increased (bottom row), the attractor layout first remains as described until critical thresholds of the control parameter on both sides (increasing and decreasing) are reached. The symbols $-k_c$ and $+k_c$ stand for these critical values of the control parameter. When k is smaller than $-k_c$, only one attractor exists, and this attractor is located in the negative range of the state axis. When k greater than k_c, only one attractor exists in the system, and this attractor is located in the positive range of the state axis.

When k equals $-k_c$ or k_c, a *half stable point* exists on the opposite side of the attractor, drawn as a square. This point acts like an attractor from one side but like a repeller from the other side. For example, when $k = -k_c$ (second plot in top row), the system will settle in this point when starting with an initial value that is greater than the location of this half stable point, i.e. coming from far right on the x-axis. On the contrary, when the initial state is a tiny bit to the left of the half stable point, the system will converge towards the attractor in the negative range on the x-axis.

In sum, this system exhibits bistability – the presence of two attractors – when k is in the range between $-k_c$ and k_c (i.e. $-k_c < k < k_c$). When the control parameter reaches the critical value on one or the other side ($-k_c$ or k_c), bifurcation occurs and the bistability breaks down: One attractor turns into a half-stable point at the respective critical threshold; beyond the critical values, even this half-stable point vanishes.

So far, differential equations have been used to represent a dynamical system. An additional, tightly connected way to define a dynamical system is by its *po-*
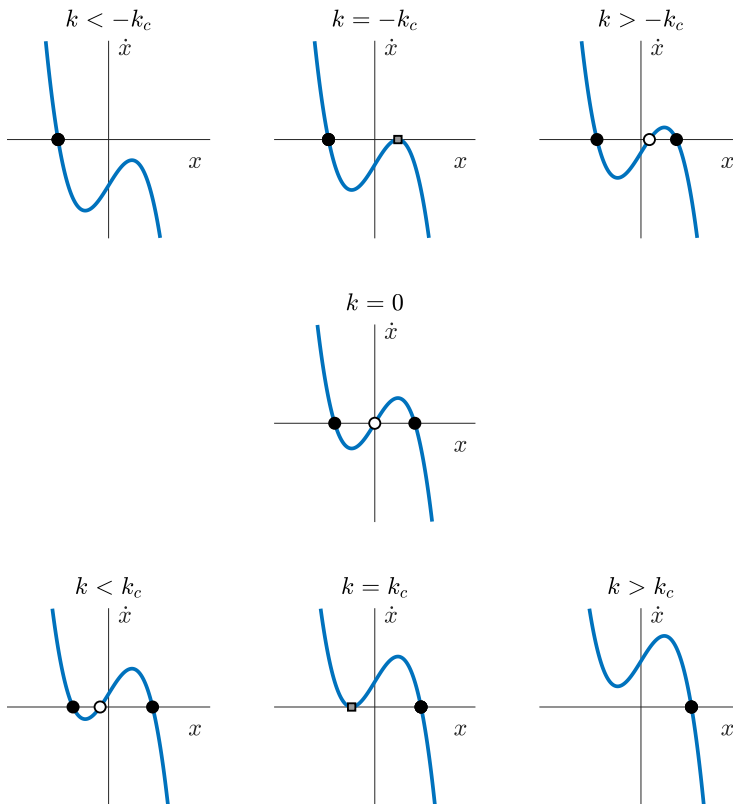

Figure 3.9: Phase space plot for $\dot{x} = k + x - x^3$. The filled circle stands for an attractor, the empty circle stands for a repeller, the square stands for a half-stable point.

tential function. The potential is given by the negative integral of the differential equation. Thus, the relation of the differential equation, which is also often called the *force function*, and the potential can be expressed as $\dot{x} = F(x) = -\frac{dV(x)}{dx}$, where $V(x)$ is the potential and $F(x)$ is the force. The graph of the potential function visualises the *attractor landscape* of the dynamical system. Attractors appear as "valleys", i.e. local minima in the graph. Figure 3.10 gives the potential for the system defined by its differential in Equation 3.3, the potential function itself is given in Equation 3.4. A commonly used metaphor for making the behaviour of the system understandable more intuitively is to imagine a ball or a marble rolling through the attractor landscape given by the potential (Haken & Levi 2012), since "[t]he temporal evolution of a one-dimensional dynamical system corresponds to an overdamped motion of a particle in the landscape of its potential" (Fuchs 2013: 22). Crucially, one has to think of the motion of the ball in "thick or viscous fluid like honey" such that when "it reaches a minimum it will stick there and will not oscillate back and forth" (Fuchs 2013: 22).

$$V(x) = -kx - \frac{x^2}{2} + \frac{x^4}{4} \tag{3.4}$$

Compare Figure 3.10 (potential) to Figure 3.9 (force). When k is zero, two symmetrical valleys are present in the potential. Because of this shape it is often called *double-well potential.* Depending on where the ball starts, it rolls down either one of the valleys. When k is decreased or increased, the function tilts to one side, making one valley deeper and the opposite valley shallower. By passing the critical value of k ($-k_c$ or k_c), one of the attractor valleys disappears – in other words, the attractor becomes unstable and that state of the system ceases to be an attractor. Regardless of where the ball starts now, it will always roll into the remaining attractor. At the critical values, the "fading attractor" takes an elbow shape. The ball metaphor can help to understand the fact that this point is half-stable: Suppose for instance that $k = -k_c$, if the ball starts on the right side of the half-stable point, it will roll down the potential to come to a halt at that point. If, on the contrary, it starts on the left side of the point, it will roll down to the attractor in the negative part of the x-axis.

Code used in this section: double_well.m

3.2 Applications of dynamical systems

This chapter so far has outlined some of the basics of dynamical systems. In what follows, applications in the realm of human movement, language and speech will

be discussed. This section presents a non-exhaustive collection of applications. The models are explained in some detail as they provide inspiration for the modelling approach presented in the second part of the book.

3.2.1 Modelling coordination and speech dynamics

Patterns of human motion have been fruitfully described in terms of dynamical systems. Since speech production represents a highly intricate system of coordinated movements, a large and growing body of research has applied nonlinear
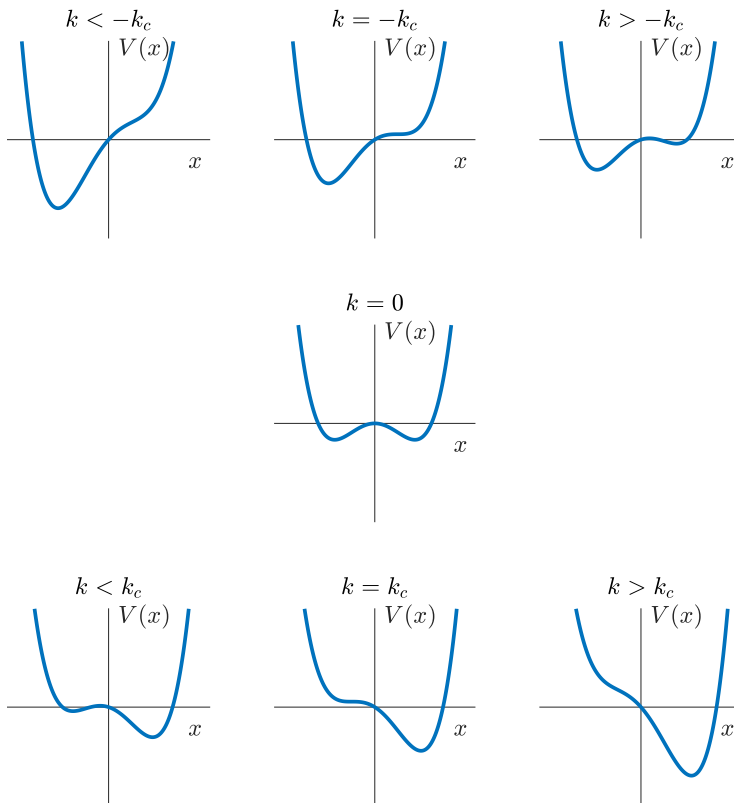

Figure 3.10: Potential $V(x) = -kx - \dfrac{x^2}{2} + \dfrac{x^4}{4}$ corresponding to the phase space plots of the force $\dot{x} = k + x - x^3$.

dynamics to speech. The present subsection introduces and highlights some interesting points of this research.

3.2.1.1 Inter-limb coordination

An influential application of dynamical systems in the domain of human behaviour is the model of Haken et al. (1985). This model presents a mathematical description of interesting observations on the coordination patterns of finger movements made in earlier research by one of the authors. In the study of Kelso (1981), the subjects moved the index fingers of both hands simultaneously at varying frequencies. One remarkable finding was that subjects were able to coordinate the movements of their fingers in two different stable modes: anti-symmetrically, a coordination pattern also known as *anti-phase*, or symmetrically, a coordination pattern known as *in-phase*. These two modes are illustrated in Figure 3.11.

Anti-phase   In-phase

Figure 3.11: Stable coordination patterns found in Kelso (1981) and used for the modelling approach in Haken et al. (1985): anti-phase vs. in-phase.

In terms of a dynamical model, the movement of each finger can be idealised as a harmonic oscillation and each oscillation can be described by its phase angle ϕ_i (Haken et al. 1985). When modelling the coordination patterns introduced in the last paragraph, the *relative phase* is of primary concern. Let ϕ_1 and ϕ_2 be the phase angles of the respective fingers, then $\phi = \phi_2 - \phi_1$ is the relative phase. In the right panel of Figure 3.12, the oscillations of the two fingers in anti-phase (top) and in-phase (bottom) are shown.[2] The dashed line marks one time point t_1. The left panel gives the phase angles of the oscillations at this time point t_1. When the oscillations of the fingers are in anti-phase mode (top), the phase angles are displaced by 180°, i.e. the relative phase ϕ is 180°. At t_1, ϕ_1 is 40° and ϕ_2 is 220° and consequently $\phi = \phi_2 - \phi_1 = 180°$. The relative phase is visualised by the grey shaded circular sector in the phase angle plot on the left. When the oscillations

---

[2]The oscillations for both fingers have identical frequencies and amplitudes in this example.
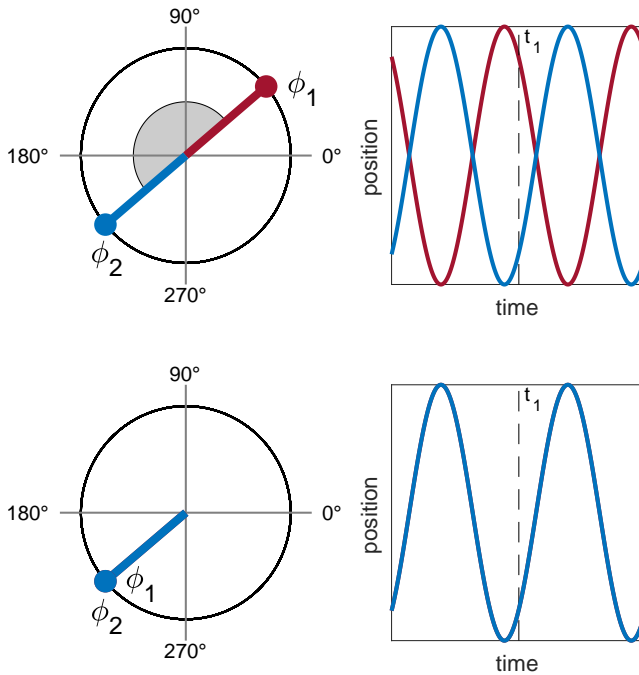
Figure 3.12: Phase angles and evolution of the oscillatory finger movements as modelled by Haken et al. (1985).

of the fingers are in in-phase mode (bottom), their phase angles are identical and the relative phase is 0°. At t_1 in the lower panel, both ϕ_1 and ϕ_2 are 220°, hence $\phi = \phi_2 - \phi_1 = 0°$. When plotted on top of each other, only one of the oscillations is visible in this case.

An additional result of Kelso (1981) was that when subjects started to move their index fingers in anti-phase (as in the top panel of 3.12) and the frequency of this movement was increased, the subjects' finger movements changed abruptly to an in-phase coordination (as in the bottom panel of 3.12) at a critical frequency boundary. To capture these findings, Haken et al. (1985) proposed a model that has attractors for anti-phase and in-phase coordination for certain ranges of the control parameter and a single attractor for in-phase coordination when the control parameter is scaled past a critical threshold. The model is given by its potential energy function in Equation 3.5. The control parameter is the ratio of b and a, i.e $\frac{b}{a}$.

The potential is in fact periodic and so the shape of the attractor landscape repeats over and over again with $V(\phi + 2\pi) = V(\phi)$.

$$V(\phi) = -a\cos(\phi) - b\cos(2\phi) \tag{3.5}$$

In the model, increasing frequency of the movements is conceptualised as a decrease in the ratio $\frac{b}{a}$. Figure 3.13 illustrates the attractor landscapes for different values of the $\frac{b}{a}$. To interpret the effect of scaling $\frac{b}{a}$ towards zero, the metaphor of a ball rolling down the attractor landscape can be employed again. In the figure, the ball starts in one of the anti-phase attractors with $\frac{b}{a} = 1$ (top left). This attractor
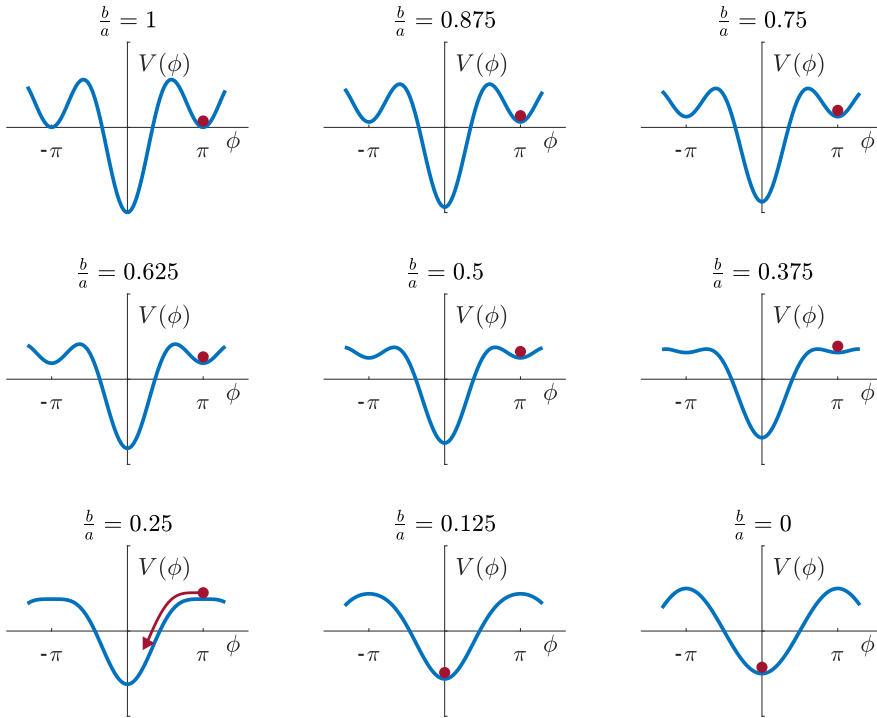

Figure 3.13: Attractor landscapes of Haken et al. (1985) for different values of $\frac{b}{a}$.

is at $\phi = \pi$ which is the phase angle of 180° in radians. When going from left to right and top to bottom in the plot, the tempo of the finger movement increases and the ratio $\frac{b}{a}$ decreases. However, at first, the ball stays in the initial anti-phase attractor. This attractor basin becomes shallower until it is not an attractor basin anymore. At this critical value of $\frac{b}{a}$ (see lower left corner), the ball will be set into motion and roll down to the in-phase attractor which is at $\phi = 0$.

Figure 3.14 shows what happens to the oscillatory finger movements (upper pannel) and their relative phase (middle panel) over time when the ratio $\frac{b}{a}$ drops beyond the critical value. As can be seen in the middle panel, the relative phase changes from anti-phase to in-phase abruptly – with a small portion of instability after which the oscillations of the fingers in the top panel are exactly synchronous (and thus plotted on top of each other).

Hence, this dynamical model that is able to account for two qualitatively distinct coordination modes of finger motion using a potential function that is mod-
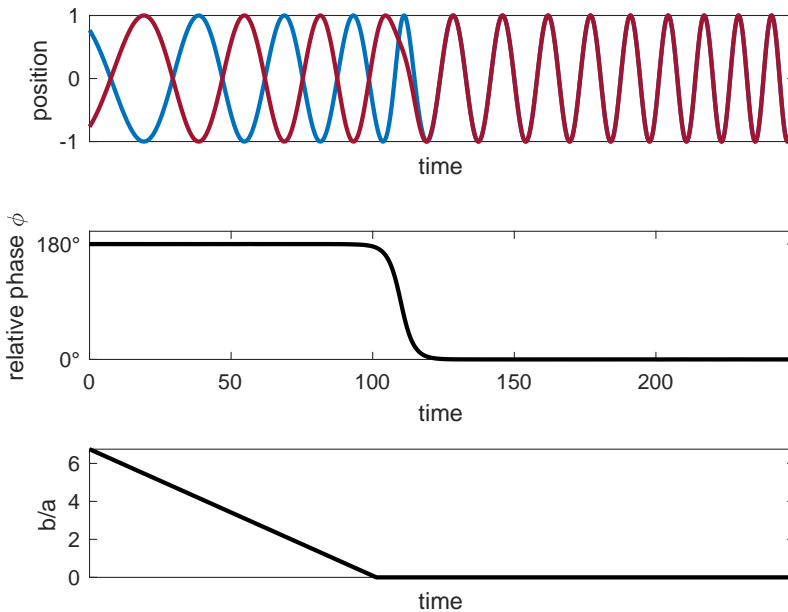

Figure 3.14: Simulation of the model of Haken et al. (1985) starting in anti-phase mode and turning into in-phase mode.

ulated by scaling a continuous parameter (or the ratio of two parameters). The model was in fact seminal and had great impact on models of speech production as will become clear in what follows. First, a closer look at the use of dynamics to model the *gesture* in Articulatory phonology will be taken. Second, the coupled oscillator model for *intergestural coordination* building upon Haken et al. (1985) will be presented.

Code used in this section: hkb_phase_modes.m, hkb_potential.m, hkb_simulation.m

3.2.1.2 Articulatory gestures

The framework of Articulatory phonology, as described in Chapter 2, views gestures as the primitives of phonology. Gestures are orchestrated to build higher forms like syllables and words. A gesture is defined in terms of a dynamical model. Chapter 2 outlined how the model is able to account for variation in the speech signal, as for example the case of assimilation. In the current chapter, the model will be reviewed from the dynamical perspective.

Articulatory phonology builds upon a model known as *task dynamics* to describe the gestures of speech production. Task dynamics (see Saltzman 1986; Saltzman & Kelso 1987; Saltzman 1991; as well as Hawkins 1992 for an introductory overview) has been developed to describe various patterns of movement that include multiple effectors like the shoulder, the upper and lower arm as well as the hand when grasping an object. Similar to other motion systems, speech involves the coordination of multiple effectors, in this case articulators, to form the relevant constrictions (Browman & Goldstein 1989). The task dynamics approach to articulation does not focus on the motion of the individual articulators but on the motion of *tract variables* (Browman & Goldstein 1992). Tract variables describe the coordinative structures that jointly yield the relevant constrictions. For example, the upper lip, the lower lip and the jaw contribute to the tract variable of lip aperture (Browman & Goldstein 1992). Table 3.1 gives an overview of the tract variables and the articulators involved in these tract variables. Most tract variables come in "horizontal-vertical" pairs: Constriction location (horizontal) is combined with constriction degree (vertical). The gestures in Articulatory phonology are formed from the set of tract variables. When constriction degree and location are present, both dimensions are used to build the gesture.

At the heart of the modelling approach of task dynamics is a dynamical system known as the *damped harmonic oscillator* that specifies the control of a tract variable (Hawkins 1992). It is formulated mathematically as the second order differential equation given in Equation 3.6 (Saltzman & Kelso 1987). The systems reviewed in this chapter so far were given by first-order differential equations.

Table 3.1: Tract variables of Articulatory phonology from Browman &
Goldstein (1989: 73).

	tract variable	**articulators involved**
LP	lip protrusion	upper and lower lips, jaw
LA	lip aperture	upper and lower lips, jaw
TTCL	tongue tip constriction location	tongue tip, body, jaw
TTCD	tongue tip constriction degree	tongue tip, body, jaw
TBCL	tongue body constriction location	tongue body, jaw
TBCD	tongue body constriction degree	tongue body, jaw
VEL	velic aperture	velum
GLO	glottal aperture	glottis

In a second-order differential equation the second derivative, denoted by $\ddot{x}$, occurs.

$$m\ddot{x} + b\dot{x} + k(x - x_0) = 0 \qquad (3.6)$$

The damped harmonic oscillator describes the motion of a mass attached to a spring that is stretched vertically (Feynman et al. 1963), and is therefore often referred to as a spring-mass system. The state variable x is the position of the mass attached to the spring, $\dot{x}$ is its velocity and $\ddot{x}$ is its acceleration. Besides x and its two derivatives, a collection of parameters occurs in the equation: m, b, k and x_0. These parameters are discussed shortly in this section; examples are given to illustrate the consequences of manipulating the parameters.

The parameter m refers to the mass attached to the spring. This parameter is usually set to 1 and not changed. For the sake of completeness, however, Figure 3.15 shows what happens when the mass is changed (for this simulation, the parameter b is set to 0, the parameter k is set to 1, the parameter x_0 is set to 0). For higher values of m, the system oscillates at lower frequencies. It is easy to understand intuitively that greater masses need more time to be transported over the same distance compared to smaller masses.

The same effect is achieved by changing the parameter k, the stiffness of the spring, as shown in Figure 3.16 (for this simulation, the parameter b is set to 0, the parameter m is set to 1, the parameter x_0 is set to 0). For higher values of k the system oscillates at higher frequencies. Here as well, an intuitive understanding is facilitated by imagining the consequences of pulling two springs which differ with regard to their stiffness: The stiffer spring snaps back faster.
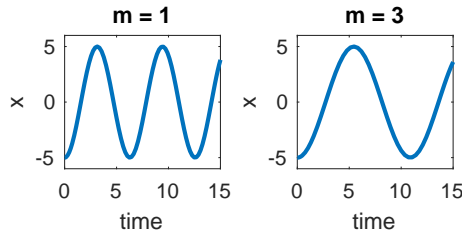
Figure 3.15: Consequences of scaling the parameter m (mass) in the harmonic oscillator.
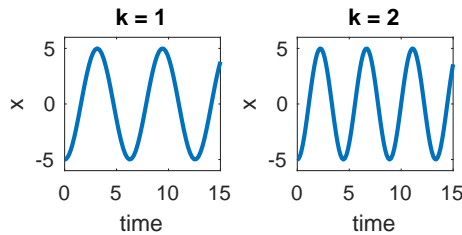

Figure 3.16: Consequences of scaling the parameter k (stiffness) in the harmonic oscillator.

More important for the modelling of articulatory gestures is the parameter b, the damping of the system. Damping leads to dissipation of energy stored in the system due to friction and reduces (or even prevents) the system's oscillation. Figure 3.17 illustrates four interesting cases of damping. In all cases, the mass m and the stiffness k are set to 1, x_0 is set to 0. In the top left plot, the undamped case is shown, $b = 0$. The system is *not* damped, just like in the plots illustrating the changes of mass and stiffness. In this case, the system will oscillate forever. In the top right plot, the case of an *underdamped* system is shown. The system oscillates but the amplitudes shrink and the system converges towards a resting position. In the bottom panels, the cases of a *critically damped* system (left) and an *overdamped* system (right) are illustrated. In these cases, the system does not oscillate and converges towards a resting position, similar to the constant growth model $\dot{x} = kx$ for $k < 0$ discussed above.

Finally, x_0 denotes equilibrium or resting position. Figure 3.18 gives an overview of the consequences of manipulating this parameter for a critically damped system. As illustrated by the plots, for a critically damped system or an overdamped system, the parameter x_0 is the position of the point attractor that the system converges to. In Articulatory phonology, the control of a tract variable is modelled with a critically damped harmonic oscillator. The resting position x_0
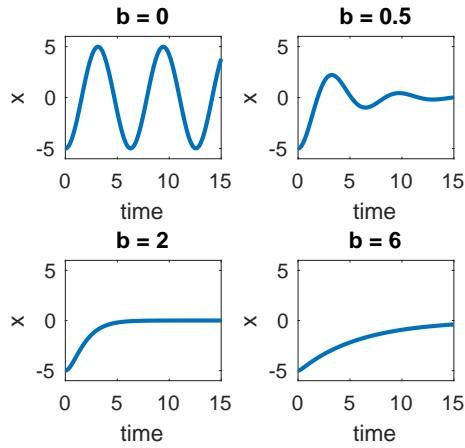
Figure 3.17: Consequences of scaling the parameter *b* (damping) in the harmonic oscillator.

corresponds to the constriction location or degree depending on the tract variable.
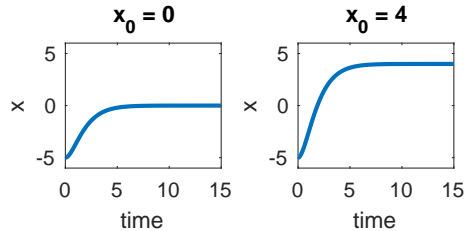

Figure 3.18: Consequences of scaling the parameter x_0 (equilibrium position) in the harmonic oscillator.

In fact, a critically damped harmonic oscillatory never reaches its resting position but only approaches it infinitesimally close. For a concrete implementation scenario of the model, a point has to be defined at which the target is said to be reached. As shown above, an undamped oscillator repeats in even cycles and its oscillation can be described by the phase angle as already shown for the finger movements when discussing the Haken et al. (1985) model. Articulatory phonology defines the target of the critically damped oscillator as the point of 240° of the undamped corresponding oscillator (Browman & Goldstein 1990), as shown in Figure 3.19. In this figure, one full undamped oscillator cycle is plotted as a solid red line, and the corresponding undamped oscillator is plotted as a dotted

blue line. The vertical black line indicates the point where a phase angle of 240°
is reached in the cycle of the undamped oscillator, i.e. the *target* of the gesture
described by the damped harmonic oscillator. The second x-axis on the bottom
relates the time points on the first x-axis to the phase angles in degrees.

Code used in this section:
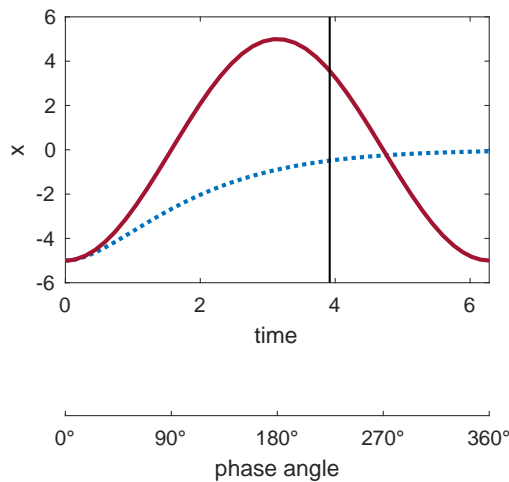*damped_harmonic_oscillator.m, damped_harmonic_oscillator_target.m*

Figure 3.19: The target of the critically damped harmonic oscillator is
reached at 240° of the phase angle of the corresponding undamped os-
cillator.

3.2.1.3 Inter-gestural timing

Articulatory phonology views words and syllables as made up of gestures. Al-
though timely ordered in some sense, these gestures are, crucially, not "beads on
string" (Pouplier 2011), i.e. they are not strictly sequentially ordered. Structures
like words and syllables should rather be seen as molecular structures (Nam et
al. 2009) consisting of gestures and connections between the gestures that deter-
mine their relative timing. This section will shed some light on the modelling
of timing in Articulatory phonology. Here again, a dynamical systems approach
is used which will be reviewed in some detail since it ties together what has
been presented in the last two sections on the Haken et al. (1985) model and the
modelling of gestures in Articulatory phonology.

It has been highlighted in different passages of this work that *timing* is important in the context of modelling articulatory gestures. The gestural score in Figure 2.3 in Chapter 2 emphasised this fact by showing how fine-grained timing differences can account for more subtle processes, like for example assimilation. To illustrate the point in the present context and for the sake of completeness and clarity, another gestural score is presented in Figure 3.20 for the words *ban* and *mad*, adapted from Goldstein et al. (2009). The only difference between the scores is the timing of the *velar wide* gesture. The result are two completely different words. This simple example demonstrates that the relative timing of the gestures (i.e. the timing of a gesture in relation to another gesture) involved in forming the word plays a significant role in determining phonological structures and lexical contrast.
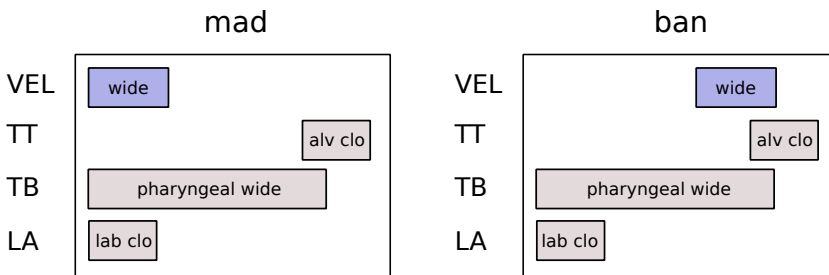

Figure 3.20: Gestural scores for the words *ban* and *mad* adapted from Goldstein et al. (2009).

The timing structure of articulatory gestures has been described in detail in a model using *coupled oscillators* (Saltzman & Byrd 2000; Nam & Saltzman 2003; Goldstein et al. 2009; Tilsen 2017). The central idea of this model is that each gesture is associated with a planning oscillator. This planning oscillator is not to be confused with the oscillator that describes the trajectory of the gesture itself. The planning oscillators of multiple gestures are connected with a coupling relation that determines the relative phases of the gestures. As in the model of Haken et al. (1985), the two stable patterns of the relative phasing are in-phase and anti-phase. During planning, the oscillators adjust their phases either in an in-phase or anti-phase manner. When a stable pattern is achieved, the actual production gestures are activated by the oscillators. The adjustment of the oscillators towards a stable pattern is modelled using a potential function similar to that of Haken et al. (1985). There exist slightly different ways to formulate the coupled oscillator model (Saltzman & Byrd 2000; Tilsen 2017). Tilsen (2017)

presents the potential functions as formulated in 3.7, where V^+ is the potential for in-phase coupling and V^- is the potential for anti-phase coupling (as in the model of Haken, Kelso & Bunz, ϕ is relative phase of the oscillators, given as the difference between the individual phases θ: $\phi_{ij} = \theta_i - \theta_j$). The evolution of the relative phase can be described using the negative derivative of the potential, the force function of the system, as given in Equation 3.8.

$$V^+(\phi) = -\cos\phi, \qquad V^-(\phi) = \cos\phi \tag{3.7}$$

$$F(\phi) = -\frac{dV(\phi)}{d\phi} \tag{3.8}$$

Each planning oscillator i can be expressed in polar coordinates with the phase θ_i such that the evolution of planning oscillator's phase without coupling can be described as in Equation 3.9, where f_i represents the intrinsic frequency of the oscillator (Tilsen 2018).

$$\dot{\theta}_i = 2\pi f_i \tag{3.9}$$

To model the effect of coupling, the expression of Equation 3.9 that models the evolution of the phase angle is extended by the force function of the coupling dynamics $F(\phi)$, see Equation 3.10 (Tilsen 2017). The force that is exerted on the planning oscillator is proportional to the coupling strengths of the planning oscillators. These coupling strengths are given as a matrix in which the coupling strength of each planning oscillator i to another oscillator j is defined. This matrix for three coupled oscillators looks like the one given in Equation 3.11. The diagonal elements c_{ii} of the matrix are 0 because they denote the coupling of the oscillator to itself.

$$\dot{\theta}_i = 2\pi f_i + \sum_j c_{ij} \frac{-dV(\phi_{ij})}{d\phi_{ij}} \tag{3.10}$$

$$C = \begin{pmatrix} 0 & c_{12} & c_{13} \\ c_{21} & 0 & c_{23} \\ c_{31} & c_{32} & 0 \end{pmatrix} \tag{3.11}$$

Figure 3.21 provides an example of two oscillators that are coupled in an in-phase manner and that start with a relative phase of 110°. The three rows of the figure show three different points in time. The first row presents a time point

that is shortly after the beginning of the simulation, the relative phase at this time point is still close to the initial relative phase of 110°. The last row presents a time point at which the two oscillators have almost the same phase, i.e. the relative phase is close to 0°. The middle row presents a time point in between. In the left column, the graph of the potential function is shown with the red dot indicating the current state of the system (it is possible to think of the dot as the ball in the metaphor used above). In the mid column, the phase angles of the oscillators are presented. In the right column, the position of the oscillators are shown in a time window around the time point corresponding to the potential and the phase angles in the same row (the dashed line specifies this time point).
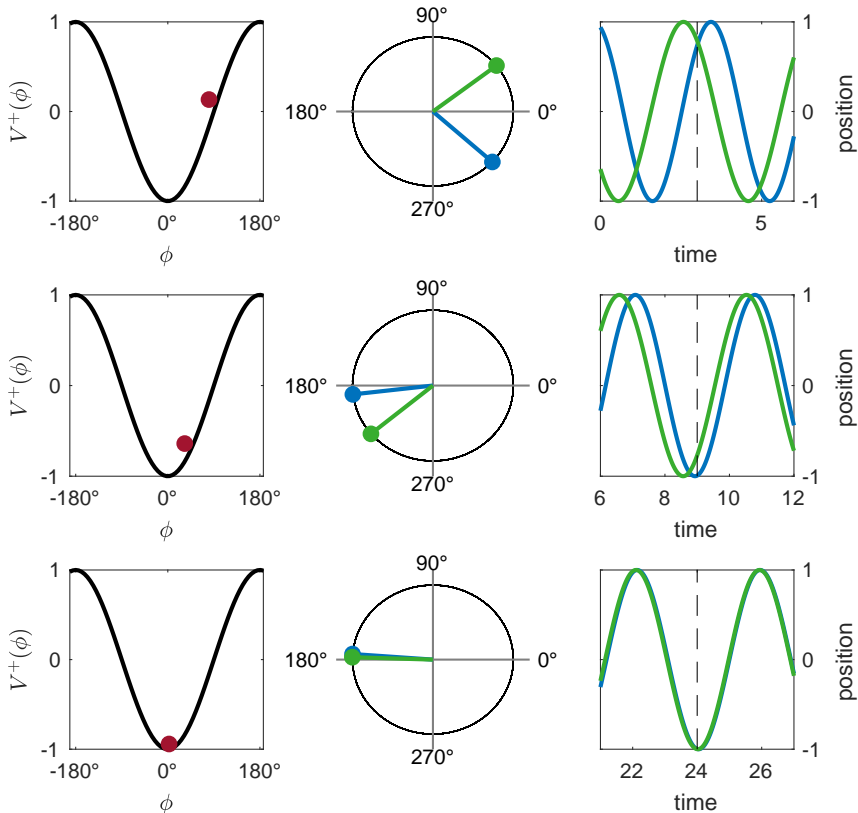

Figure 3.21: Potential, phase angles and position corresponding to two oscillators coupled in-phase at three time points (from top to bottom).

The plot illustrates how the state of the system approaches the attractor at 0°, the minimum of the potential, and how the relative phase of the two oscillators decreases over time.

Usually, syllables involve more than two gestures and, thus, more than two oscillators are coupled in a pair-wise fashion. As a result, a network of coupled oscillators emerges.[3] The target phasing relations of these oscillators are represented in *coupling graphs*. The coupling graphs for the English words *bud* [bʌd] and *dub* [dʌb], adapted from Mücke (2018), are shown in Figure 3.22. The solid lines indicate in-phase coupling, the dashed lines indicate anti-phase coupling. The onset consonant of the syllable in both cases is modelled as being coupled in-phase with the vowel of the syllable while the coda consonant is coupled anti-phase with the vowel. This organisation reflects the fundamental hypothesis that CV structures in a syllable are coupled in-phase while VC structures are coupled anti-phase (Goldstein et al. 2006; 2009).

Figure 3.22: Gestural scores and coupling graphs for *bud* [bʌd] and *dub* [dʌb] adapted from Mücke (2018), solid lines indicate in-phase coupling, dashed lines indicate anti-phase coupling.

If two consonants or more are present in the onset of a syllable, a pair-wise in-phase coupling of all of the consonants to the vowel alone would lead to an overlapping of the consonants. To achieve at least a partially sequential realisation of the onset consonants, the onset gestures of each consonant can be coupled with anti-phase links to the vowel. The result is a *competitive coupling* in which

---

[3]An alternative view is that the oscillator of each gesture is coupled to a master clock. This perspective is discussed in relation to the network of coupled oscillators in Goldstein et al. (2009).

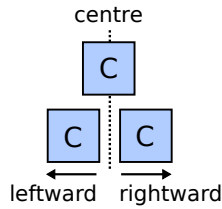Figure 3.23: Schematic illustration of the c-centre effect adapted from Hermes et al. (2013).

the final phases represent a compromise of the competing coupling forces (Nam 2007; Saltzman et al. 2008; Goldstein et al. 2009). Competitive coupling is able to explain an effect known as the *C-centre effect* in branching onsets (Browman & Goldstein 1988; Byrd 1995). The C-centre effect, illustrated in Figure 3.23, describes the following situation: The timing of C1 in relation of the vowel V in a complex onset structure C1C2V is earlier compared to the simple structure C1V. In other words, when a second consonant C2 is added after C1, the earlier consonant C1 shifts leftwards. Likewise, the timing of C2 in C1C2V is later compared to the simpler structure C2V, i.e. when a second consonant C1 is added before C2, the later consonant C2 shifts rightwards. The timing of the centre of the consonant cluster in relation to the vowel, however, stays constant (Browman & Goldstein 1988; Goldstein et al. 2009).

Code used in this section: coupled_osc.m

3.2.2 Modelling dynamics of categoriality and continuity

In the last section, dynamical models of motion and articulation have been presented. Crucially, these models do not only entail mere physical descriptions of the mechanism of speech production but put the dynamical approach they employ in a phonological, cognitive perspective: The gestures of Articulatory phonology are viewed as phonological primitives, the coordination patterns of the coupled oscillators model shape phonological patterns of syllable structure. This perspective maintains that both categorical aspects and continuous aspects of speech can be described jointly by dynamical models. The current section explores this idea more explicitly by presenting some interesting and influential approaches to modelling categoriality and continuity in the sound pattern of language.

3.2.2.1 Perceptual categories

Tuller et al. (1994) present a dynamical model that is able to account for interesting results in the perception of speech sound categories. Their work illustrates how the stability of speech sound categories and their perception can be modelled while including flexibility of the perceptual responses.

The authors exposed participants to ordered continua between English *say* and *stay* and between *stay* and *say*. To be more specific, the gap duration between *s* and *ay* increases and then decreases with each stimulus during one experiment run. The participants were tasked to categorise each stimulus in a forced choice task. Two dominant response patterns are evident in the data. In one response pattern, the category switch in the increasing order (the gap duration becomes larger) is later compared to the decreasing order (the gap duration becomes smaller). In other words, for the change from *say* to *stay* a larger gap is necessary than for the change from *stay* to *say*. This response pattern is called *hysteresis*. In the other response pattern, the category switch is earlier in the increasing order compared to the decreasing order. For the change from *say* to *stay* a smaller gap is necessary than for the change from *stay* to *say*. This response pattern is called *enhanced contrast*. Both patterns are illustrated in the two top panels of Figure 3.24. The figure also shows the response pattern *critical boundary* in which the category switch occurs at the same point in the acoustic continuum in both increasing and decreasing order. In the data of Tuller et al. (1994), the response patterns hysteresis and enhanced contrast dominate while the pattern critical boundary is rarely found.

Tuller et al. (1994) propose a model that is similar to the double well potential introduced in the first part of the chapter in Equation 3.4 and illustrated in Figure 3.10 with a slight difference: The control parameter k occurs with a positive sign in the present model, see Equation 3.12. The effect of this difference is simply that the scaling of the control parameter has the opposite effect. For positive k values, e.g. $k = 1$, the landscape is tilted to the left, for negative k values, e.g. $k = -1$, it is tilted to the right. To connect this attractor landscape to the perceptual data, one attractor is associated with the percept of *say*, the other with *stay*, see Figure 3.25. While listening to the ordered continuum from *say* to *stay*, the control parameter increases and the attractor landscape gradually tilts to the side of *stay*. When the critical boundary k_c is reached, the *say* attractor is destabilised and the percept changes to *stay*. The process takes place analogously from *stay* to *say* when the control parameter decreases (in this case, the critical boundary is $-k_c$).

$$V(x) = kx - \frac{x^2}{2} + \frac{x^4}{4} \qquad (3.12)$$
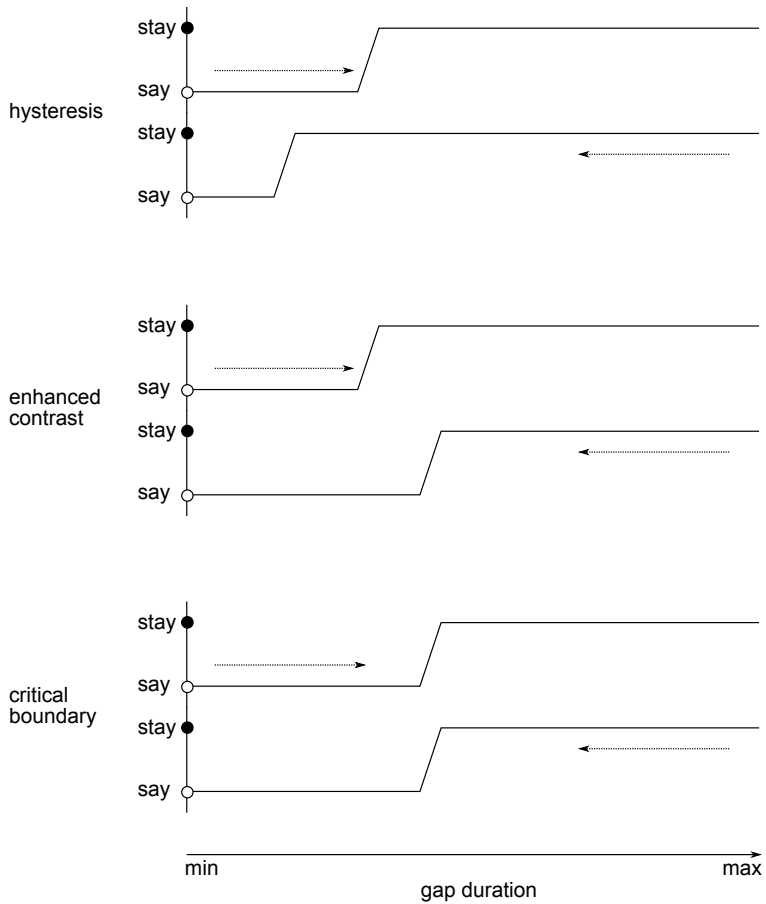
Figure 3.24: Possible response pattern of Tuller et al. (1994): hysteresis (top), enhanced contrast (middle), critical boundary (bottom).
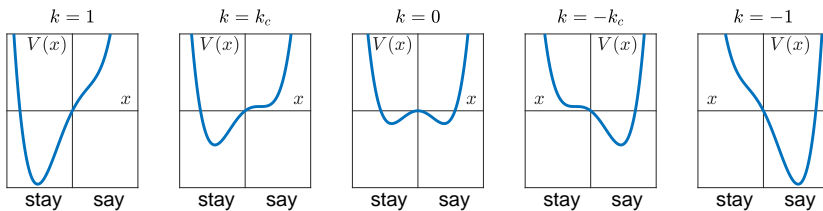

Figure 3.25: Potential function of Tuller et al. (1994) for different values of the control parameter k.

In the response patterns described above, the switch from one category to the other (i.e. *say* to *stay* or *stay* to *say*) is at different points in the continuum. For the model, this means that the critical value of the control parameter k in both sides (towards *stay* or towards *say*) has to occur with different gap durations in the response patterns. To explain this phenomenon, Tuller et al. (1994) hypothesise that k depends on a variety of variables and is determined by the function given in Equation 3.13. In this function, k_0 is the value of k at the beginning of the run, i.e. -1 when the participant starts listening to the continuum with increasing gap durations from *say* to *stay*. λ is a variable that is proportional to the gap duration and thus represents the position on the acoustic continuum. λ_f is the value of λ at the other end of the continuum, i.e. the maximal gap duration when the trial starts with *say* (without a gap). The variable n represents the number of stimuli that the participant already listened to, n_c denotes a critical number of trials defined as 50% of the trials. $\theta(n - n_c)$ is a step function that is defined as 0 when $n < n_c$, i.e. in the first half of the trial, and 1 when $n \geq n_c$, i.e. in the second half of the trial. The variable ϵ "represents the lumped effect of learning, linguistic experience, and attentional factors" (Tuller et al. 1994: 8). This last parameter is a very important parameter for the model because it plays a major role in explaining the response patterns introduced above.

$$k(\lambda) = k_0 + \lambda + \frac{\epsilon}{2} + \epsilon\theta(n - n_c)(\lambda - \lambda_f) \tag{3.13}$$

Figure 3.26 provides an illustration of the relation between ϵ, gap duration represented by λ and the control parameter k as a colour map. The colours represent the values of k determined with the formula of Equation 3.13. The left plot presents the predictions for the first half of the run (increasing gap durations), the right plot presents the predictions for the second half of the run (decreasing gap durations). The colours are only shown for the range of k values in the interval of $[-1, 1]$. Of course k further increases (left plot) or decreases (right plot) through the white area but the restriction to this range makes the colour contrasts stronger and thus visualises the differences better. Both plots show that the colours reflecting the values of the control parameter k are distributed roughly diagonally over the plots. This structure illustrates that for the same gap durations, k values are higher when ϵ is large in the first half of the run. The opposite is true for the second half of the run where k values are lower when ϵ is large.

Recall that the system stays in the attractor as long as the critical value of k that destabilises the attractor is not crossed. When it is crossed, the system moves to the remaining attractor and the percept changes. In a forced choice
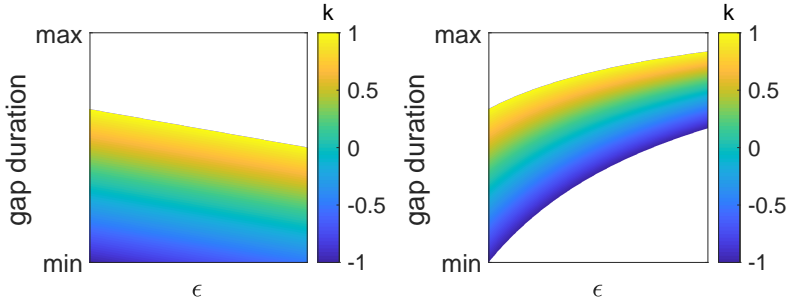
Figure 3.26: Colour maps representing the values for the control parameter k in relation to ϵ and gap duration in the Tuller et al. (1994) model. Left: first half of the run with increasing gap durations (small n). Right: second half of the run with decreasing gap durations (large n).

experiment, the participant changes the response at that point. It is thus sensible to investigate which value of λ, i.e. which gap durations, yield the critical value of k for different values of ϵ. The thick lines in Figure 3.27 show at which gap duration (y-axis) the critical boundary k occurs as a function of ϵ (x-axis). The shaded area in the left and middle panel is the span of gap durations for which the system has two attractors, i.e. the control parameter k is between the critical values on both sides $-k_c$ and k_c.

The left panel of 3.27 presents the predictions for the first half of the run. The plot is to be read from bottom to top as the arrows indicate, i.e. from no gap to the maximal gap duration (*say* to *stay*). The number of perceived stimuli n is small in this first half and under the threshold n_c (and thus $\theta(n - n_c) = 0$). When ϵ is small (left on the x-axis), the categorisation is only dependent on the gap duration. However, for larger ϵ values, the gap duration needed to shift the percept from *say* to *stay* decreases. In other words, the larger ϵ, the earlier the switch from one category to the other when going from no gap to the maximum gap.

The middle panel of 3.27 shows the predictions for the second half of the run. This plot is to be read from top to bottom, i.e. from the maximal gap duration to no gap (*stay* to *say*). The number of perceived stimuli in the second half is large and above the threshold n_c (and thus $\theta(n - n_c) = 1$). In this case, the model predicts that for larger ϵ values, the gap duration needed to switch the percept from *stay* to *say* is larger compared to smaller values of ϵ. The larger ϵ, the earlier the switch from from *stay* to *say* when going from the maximum gap to no gap.
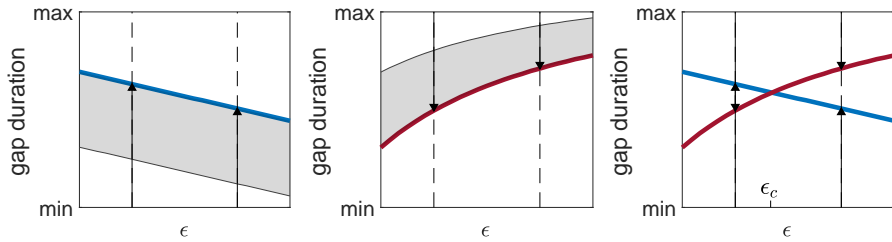
Figure 3.27: Location of critical values of the control parameter k in the acoustic continuum of gap duration as a function of the parameter ϵ in the Tuller et al. (1994) model. Left: first half of the run (increasing gap duration, small n). Middle: second half of the run (decreasing gap duration, large n). Right: superposition of critical boundary in first and second run.

The right panel of 3.27 combines the thick lines of the two neighbouring plots. There is a critical value of ϵ, namely ϵ_c, for which the category switch occurs at the same gap duration in both halves of the run. At this point the two lines intersect. When ϵ is below ϵ_c, the percept changes later and the observed response pattern is *hysteresis*. This is visualised by the line of the first half of the run being positioned above the line of the second half of the run. Going from bottom (no gap) to top (maximum gap) in the first half of the run, the critical boundary of k is reached at a longer gap duration. Going from top (maximum gap) to bottom (no gap) in the second half of the run, the critical boundary of k is reached at a shorter gap duration.

When ϵ is above the critical ϵ_c, the percept changes earlier and the observed response pattern is *enhanced contrast*. This is illustrated by the fact that the line of the first half of the run is positioned under the line of the second half of the run. Going from bottom (no gap) to top (maximum gap) in the first half of the run, the critical boundary of k is reached at a shorter gap duration. Going from top (maximum gap) to bottom (no gap) in the second half of the run, the critical boundary of k is reached at a longer gap duration.

The model of Tuller et al. (1994) shows how the flexibility and context dependency of perceptual categories can be modelled using the double well potential with two attractors presented earlier in this chapter. The next section will present work adapting a similar potential for the modelling approach to capture continuous and categorical variation found in production data.

Code used in this section:
tuller_1994_potential.m, tuller_1994_crit_lambda.m, tuller_1994_map.m

3.2.2.2 Incomplete neutralisation

As introduced in the previous chapter, the phenomenon of incomplete neutralisation of syllable final obstruents in German poses a major problems for purely symbolic approaches to phonology and a modular separation of phonetics from phonology. To remind the reader, a large body of work has centred around the question whether the voicing contrast of obstruents in syllable coda positions in German is complete or not. Numerous studies have shown that there are indeed differences between the final obstruents of words like *Rat* and *Rad* such that the acoustic features of the devoiced final obstruent [t] in *Rad* are modulated subtly in the direction of the voiced [d].

In addition, a study by Port & Crawford (1989) suggests that the communicative context modulates the differences between the obstruents. When the speaker produces the words containing the obstruents in direct contrast ("Ich habe Rad gesagt, nicht Rat") and a listener is tasked to write down the correct word, the supposedly neutralised obstruent shifts more in the direction of the voiced variant compared to a task in which the speaker simply reads the words in a list.

Gafos & Benus (2006) propose a dynamical model that is able to capture the differences between the obstruents in relation of the speaker's intent to maintain the contrast. In this model, the categorical nature of the phonological voicing contrast can be maintained while allowing for fine-grained differences. In the first part of the model, the intention of a speaker to produce a voiced or a voiceless obstruent is described by defining one attractor for each voicing value on a continuum of voicing. The continuum of voicing are all possible states x of the systems. The intention to produce a contrast is formally modelled with the force function $F(x)$ in Equation 3.14, where x_{req} is the *required* (i.e. intended) value on the voicing continuum x. Crucially, x_{req} is the location of the attractor of this system. For the voiceless obstruent, a location in the positive range of x denoted by x_0 is chosen. For the voiced obstruent, a location the negative range of x denoted by $-x_0$ is chosen. The exact values do not play a role in the modelling approach, it is only important that they are distributed on both sides of zero.

The second line of Equation 3.14 presents the potential energy function $V_F(x)$ that is obtained by integration of the negative of the force function. Figure 3.28 displays $F(x)$ and $V_F(x)$ with $x_{req} = -x_0$ on the left and $x_{req} = x_0$ on the right. The control parameter of this system is θ, a quantity representing the intent of the speaker to produce this value of voicing x_{req}. In the further description of the model, the role of the parameter θ will become clearer.

$$F(x) = \theta(x_{req} - x)$$
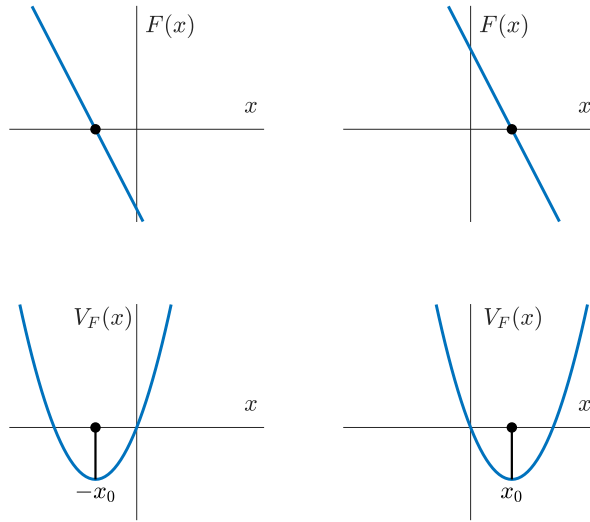$$V_F(x) = \theta\frac{x^2}{2} - \theta x_{req}x$$

(3.14)

Figure 3.28: Force (top) and potentials (bottom) for $x_{req} = -x_0$ (voiced) and $x_{req} = x_0$ (voiceless) in the Gafos & Benus (2006) model. Vertical lines in the potential visualise the location of the minimum at the value of x_{req} ($-x_0$ or x_0).

In the second part of the model, an additional force is introduced to account for the fact that German allows for voiceless obstruents only in syllable coda. Here, the same attractor landscape is used as in Tuller et al. (1994), its force $M(x)$ and potential $V_M(x)$ are given in Equation 3.15. For the control parameter k, a value beyond the critical threshold of $-k_c$ is chosen (-1 in the illustration) such that the landscape is tilted to the voiceless side and the voiceless attractor is the only attractor that remains, see Figure 3.29. The presence of only on attractor reflects the fact that there is one possibility for obstruents in syllable codas: voiceless.

$$M(x) = -k + x - x^3$$

$$V_M(x) = kx - \frac{x^2}{2} + \frac{x^4}{4}$$
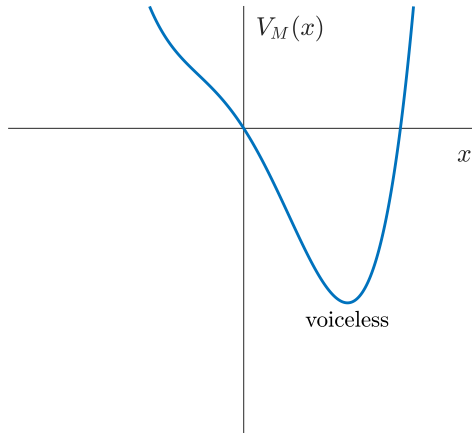
(3.15)

Figure 3.29: Double well potential $V_M(x)$ of the Gafos & Benus (2006) model tilted to the right side to represent the fact that only the voiceless attractor is available in coda position.

Gafos & Benus (2006) draw parallels of their dynamical model to an analysis of the phenomenon in optimality theory (OT), a purely symbolic framework. In such an analysis, the presence of one attractor for voiced and one attractor for voiceless obstruents as in the first part of the model corresponds to a *faithfulness* constraint. This constraint is violated when the output form deviates from the underlying representation, see Chapter 2. In other words, the constraint entails that it is the intention of the speaker to produce outputs as close as possible to the underlying representation. The second part of the model in which only an attractor for voiceless is present corresponds to a *markedness* constraint that requires coda consonants to be voiceless in German.

The interaction of the two parts of the model is achieved by adding up the two forces $F(x)$ and $M(x)$ to obtain the final force function of the system (see Equation 3.16 for the combined force and potential). In an OT analysis, the markedness constraint would be ranked higher than the faithfulness constraint eliminating

any influence of the latter. In the interaction of the present model, however, the force $F(x)$ can influence the outcome of the whole system even if the force $M(x)$ might be stronger. This means that despite the pressure to realise only voiceless obstruents in syllable-final position, the "underlying" voicing can still have an impact. The size of this impact can be scaled by virtue of a scalar value.

$$M(x) + F(x) = -k + x - x^3 + \theta(x_{req} - x)$$

$$V_M(x) + V_F(x) = V(x) = kx - \frac{x^2}{2} + \frac{x^4}{4} + \theta\frac{x^2}{2} - \theta x_{req}x \tag{3.16}$$

The resulting patterns are illustrated in Figure 3.30. The top panel presents the outcomes for the underlying voiced obstruent, i.e. $x_{req} = -x_0$, for three possible values of θ: 0.1, 0.3, and 0.5. With increasing θ the attractor basin drifts subtly towards the negative, voiced part of the continuum x while it stays in the general region of the positive, voiceless part of x. As a result, the voiceless obstruent becomes *slighty more voiced*. The production of words like *Rat* with underlyingly voiceless obstruent, i.e. $x_{req} = x_0$, does not lead to the same conflict. The lower panel of Figure 3.30 shows for the same three values of θ that the location of the attractor does not change.

To summarise, producing the intended obstruent and adhering to voiceless obstruents in codas leads to a conflict in words like *Rad*. While this conflict is resolved by a constrain ranking and a single resulting outcome in OT, the interaction can be modulated continuously in the model of Gafos & Benus (2006). The effect of $F(x)$ on $M(x)$ is a "pull" towards more voiced productions on the voicing continuum x. This "pull" is modulated by the scaling of the control parameter θ.

Code used in this section: gafos_benus_incomplete_neutralisation.m

3.2.2.3 Transparent vowels

In the previous chapter, the phenomenon of transparent vowels in Hungarian has been introduced. It was explained there that the front unrounded vowels function as transparent vowels as they can occur between the vowel triggering the vowel harmony and the target of the vowel harmony but have been described as not affecting the process of vowel harmony. For example, the back vowel /a:/ of the stem *kávé* /ka:ve:/ ('coffee') determines the vowel /ɔ/ of the suffix *nak* /nɔk/ regardless of the intervening unrounded front vowel in the stem.

The supposedly insignificant role of the transparent vowels in the determination of suffixes, however, is questioned by observations of the behaviour of

Figure 3.30: Combined potentials of the model of Gafos & Benus (2006) for three values of the control parameter θ. Vertical lines indicate the location of the attractor. Blue: $\theta = 0.1$, red: $\theta = 0.3$, green: $\theta = 0.5$. k is constant at -1 in all graphs.

transparent vowels. Stems that only have transparent vowels can trigger both front and back suffixes (Vago 1980; Gafos & Benus 2006) although the distribution of the suffixes is not even, as the majority of stems triggers front suffixes (Hayes & Londe 2006; Gafos & Benus 2006). In addition, the probability of se-

lecting a back suffix decreases with increasing numbers of transparent vowels intervening between a back stem vowel and the suffix (Gafos & Benus 2006).

Gafos & Benus (2006) (as well as Benus 2005) hypothesise that systematic articulatory differences in transparent vowels are responsible for the suffix choice. In consequence, transparent vowels may participate in the process of vowel harmony. The authors report data from a study employing electromagnetic articulography and ultrasound to track the position of the tongue and investigate the tongue shape. They show that the tongue is more advanced when articulating transparent vowels in stems that trigger front suffixes compared to transparent vowels in stems that trigger back suffixes. A dynamical model is proposed that resembles the model for incomplete neutralisation outlined in the previous section although the two deal with rather different phonological phenomena. Nevertheless – like the model for incomplete neutralisation – the present model links *continuous* and *categorical* aspects in one formal approach.

The first part of the model is a formalisation of articulatory gestures by point-attractor dynamics which is in line with the description of gestures in Articulatory phonology (see also the damped harmonic oscillator presented above). The dynamics of the spatial dimension of constriction location of the tongue body is modelled with a monostable potential as given in its general form in the first line of Equation 3.17. x_0 represents the target constriction location. The factor γ represents the strength with which this gesture controls the articulator, the consequence of its scaling will become clearer in the further description. The equation also presents two potentials in the second and third line, one for back vowels and one for front vowels. The factor γ is now called α for back vowels and β for front vowels. For the back vowel, x_0 is chosen as -2; for the front vowel, x_0 is chosen as 2. The two potentials $V_B(x)$ (back vowels) and $V_F(x)$ (front vowels) are shown in Figure 3.31 (with $\alpha = \beta = 1$).

$$V(x) = \gamma(x - x_0)^2$$
$$V_B(x) = \alpha(x + 2)^2 \qquad\qquad (3.17)$$
$$V_F(x) = \beta(x - 2)^2$$

The effect of coarticulation of vowels can be modelled as gestural blending by assuming linear combinations of the respective vowel potentials, $V_B(x) + V_F(x)$, and adjusting the values for the factors α and β. Simply speaking: The factors α and β represent weights that specify which gesture will have a dominant effect in the blending. In this way, the influence of a preceding vowel on the target constriction location of a vowel can be captured. Figure 3.32 illustrates two examples
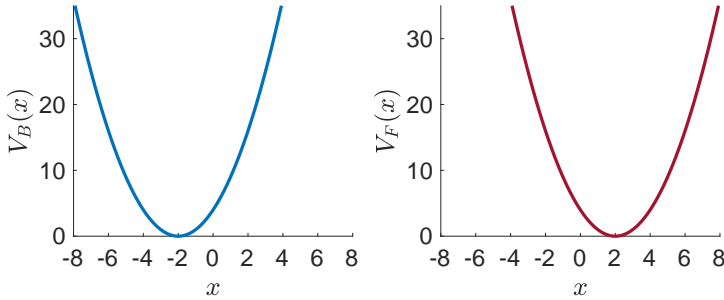
Figure 3.31: Monostable potentials modelling the tongue body gesture with regard to constriction location. Left: $V_B(x)$ for back vowels, right: $V_F(x)$ for front vowels.

for combined potentials produced by linear combinations of the vowel potentials for different values of α and β. In this figure, the solid lines represent the two potentials for back and front vowels $V_B(x)$ and $V_F(x)$ as in the previous figure. The dotted line shows the graph of the combined potential: $V_B(x) + V_F(x)$. The position of the minimum of this graph on the x axis indicates the vowel target that results from the blending. In the left panel, the case $\alpha = \beta = 1$ is shown, the hypothetical vowel has a target midway between the front and the back vowel. The experimental data of Gafos & Benus (2006) showed that the retraction of the front vowel is only subtle. Therefore, the right panel shows an example where the weight of the back vowel, β, is smaller than the weight of the front vowel, α: $\alpha = 1, \beta = 3$. The outcome is a vowel that has a constriction location target slightly lower than the original front vowel modelled by $V_F(x)$.

The degree of retraction of a front vowel is assessed by the quantity R in this model of vowel gesture blending. In the example in the right panel of Figure 3.32, the front vowel described through the potential $V_F(x)$ has a target constriction location of 2, the vowel resulting from blending has a target location of 1. In this case, R, representing the difference between the basic front vowel and the vowel retracted by influence of the back vowel, is 1.

The choice of a front vowel or back vowel suffix can now be modelled employing a dynamical system with a double-well potential that uses the retraction degree R as a control parameter. In the system, one attractor is associated with the back suffix, the other attractor is associated with the front suffix. Hence, the second part of the vowel harmony model of Gafos & Benus (2006) is given by the force $F(x)$ and potential $P(x)$ in Equation 3.18. These equations are similar to those in the previous sections. Note, however, that the exposition deviates
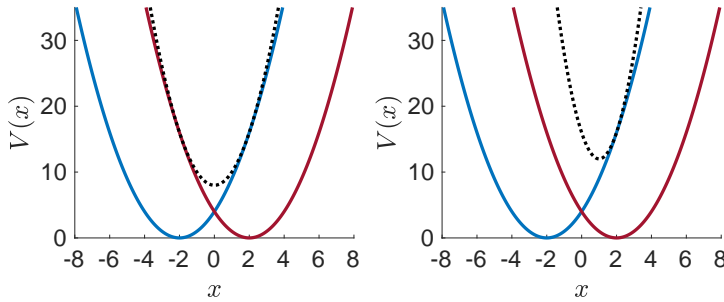
Figure 3.32: Examples for blending of vowel gestures by linear combination of the potentials corresponding to each vowel. Solid lines: individual vowel potentials (blue for $V_B(x)$ and red for $V_F(x)$). Dotted line: combined potential $V_B(x) + V_F(x)$. The two panels show cases for different weightings of the potentials. Left: $\alpha = \beta = 1$, right: $\alpha = 1, \beta = 3$.

slightly here and integrates the coefficients for the quartic term x^4 and the linear term x from Benus (2005: 262, footnote 85).[4] The choice of this coefficient does not change the general quality of the model; it merely locates the attractor in regions around -2 and 2 corresponding to the assumed values for front and back constriction locations.

$$F(x) = (2 - 3R) - 0.4x^3 + x$$
$$P(x) = -(2 - 3R)x + 0.1x^4 - 0.5x^2$$

(3.18)

Figure 3.33 presents the graph of the potential $P(x)$ for different values of the control parameter R. If the degree of retraction of the transparent vowel is high, e.g. $R = 1.2$ or $R = 1.0$, the attractor landscape only predicts one possibility for the choice of the suffix: back. The reverse is true when the retraction is small, e.g. $R = 0.2$ or $R = 0.4$, for which only front suffixes are possible. In between these values, the attractor landscape becomes bistable. In this region of intermediate R values, the suffix choice can vary between front and back due to random fluctuations and the initial state.

The graphs of the potential $P(x)$ shown in Figure 3.33 illustrate how the scaling of a continuous parameter can lead to categorical changes. Depending on the magnitude of the adjustment of the continuous parameter, only one category remains as the possible output of the system or the system is in principle

---

[4]The graphs of Gafos & Benus (2006) are also based on the coefficient 0.1 for x^4, the coefficient of the linear term x is simply put differently.
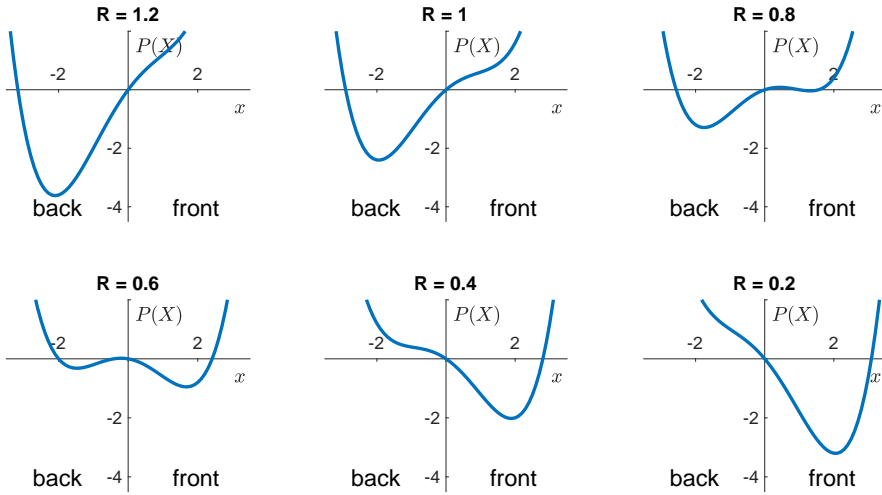
Figure 3.33: Suffix choice potential $P(x)$ for different values of the control parameter R.

able to produce two different outcomes. When the system is conceptualised as a stochastic dynamical system, the outcomes of the system depending on the control parameter can be described as statistical distributions. A stochastic system is obtained by the introduction of random fluctuations or noise (Haken 1977), denoted by ξ_t in Equation 3.19. ξ_t is conceptualised here as Gaussian white noise with a scaling factor q that determines the strength of the noise (Gafos & Benus 2006). These statistical distributions of the system's outcome are modulated as a result of the relative stabilities of the attractors. In other words, in the region of bistability, a deeper attractor in the potential is more resistant to random fluctuations than a shallower attractor, it takes more noise to move the system out of the attractor basin. The system is thus more likely to end up in this deeper, more stable attractor.

$$\dot{x} = F(x) + \text{Noise} = \frac{-dV(x)}{dx} + \xi_t \qquad (3.19)$$

A description of the distributional patterns of a stochastic system can be obtained in two ways: either analytically by finding a stationary solution to the *Fokker-Planck* equation (Haken 1977; Freidlin & Wentzell 1984) or by computa-

tional simulation of the system with noise (Gafos & Benus 2006). The solution to the Fokker-Planck equation is used to compute a probability density distribution. Figure 3.34 presents probability distributions obtained in this way for the potential $P(x)$ and the same R values as in Figure 3.33.
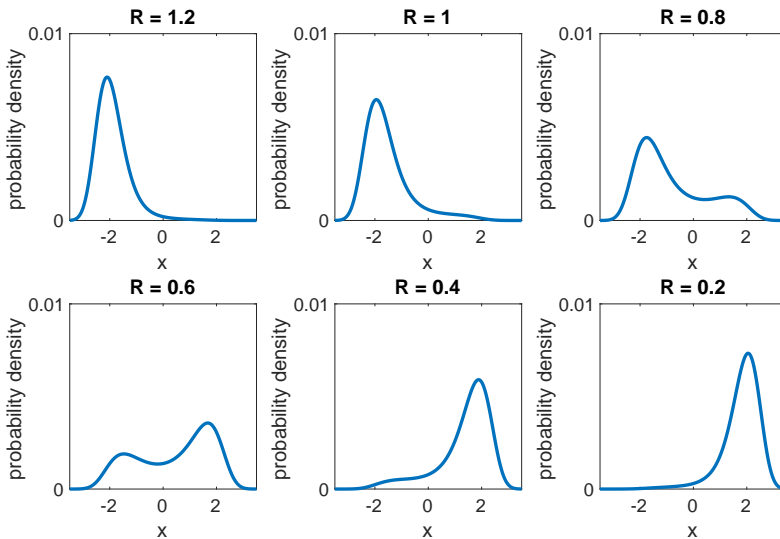

Figure 3.34: Probability density distributions corresponding to the stochastic version of the system described by the potential $P(x)$ for different values of the control parameter R.

The simulation of the stochastic system can be carried out by solving the differential equation as described above in Section 3.1.2. During each time step, noise from a Gaussian distribution is added. After a fixed number of steps, the simulation is finished and the solution is recorded. This procedure is repeated for a number of times in order to obtain a distribution of final solutions. For the solutions shown in the histograms of Figure 3.35, the simulation was run for 5000 time steps and repeated to obtain 5000 solutions.[5] Again, the same values for R as before are chosen in these plots. The outcome of the simulation reproduces the probability density functions shown in the previous figure.

The concept of the stochastic dynamical system plays a major role for the modelling approach presented in the second part of this book. It makes it possible to leverage the differences in stability of the attractors when multiple attractors are present. This section is completed by taking up the metaphor of a ball rolling

---

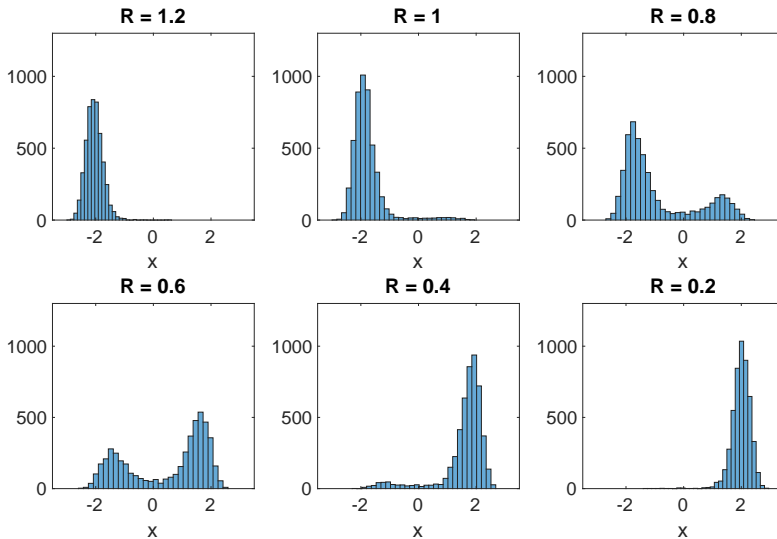[5]The simulation is based on the code accompanying Gafos (2006).

Figure 3.35: Simulation results for different values of the control parameter R in the system given by the potential $P(x)$.

in the attractor landscape and adding the notion of noise or random fluctuations to it. Imagine that the ball's trajectory is perturbed by tiny strokes. Sometimes these strokes are stronger, sometimes they are weaker. It fits our intuition well that it takes a stronger stroke to push the ball out of an attractor basin if it is deeper, i.e. more stable, compared to a shallower attractor basin.

Code used in this section: gafos_benus_vowel_harmony.m

3.3 Summary

This chapter has presented an overview on dynamical systems and their application in the science of language and speech. It has touched upon the fundamental concepts of attractors and multi-stability, and outlined how dynamical systems can be used together with simulation techniques to investigate their predictions in more detail. The most important point to take away from this chapter is that dynamical systems have proven to be suited in reconciling categorical and continuous phenomena in an integrated view and thus provide a toolbox to gain insights into the relation of phonetics and phonology.

4 Prosody, prosodic prominence and focus

Prosody denotes all aspects of the rhythmical and tonal organisation of spoken communication (Calhoun 2007). Every utterance in every language has prosodic properties (Grice 2006). Languages employ prosody to chunk the stream of speech and to single out certain elements in this stream, like words or syllables. Both mechanisms lend rhythmical and tonal structure to speech. The use of prosodic aspects to single out certain elements has been termed *prosodic prominence*. Prosodic prominence is understood as a relative property as one element is made *prominent* in comparison to other elements (Wagner & Watson 2010). Making a word or a syllable in a word more prominent often reflects the speaker's intention to highlight importance or newness of the word or referent denoted by that word in the discourse. Prominence and its relation to discourse play a major role in this work but the demarcating aspects and the marking of edges of prosodic constituents are also discussed to some extent in this chapter.

Among the physical parameters that have attracted most attention in the analysis of prosodic patterns are fundamental frequency (F0), intensity or amplitude, and duration (Ladd 2008; Grice 2006). Their perceptual counterparts are pitch, loudness and length. These parameters form a conglomerate of prosodic cues and neither of them is used exclusively for a certain function of prosody. They all play a role in both chunking as well as prominence. F0 movements on or close to a syllable as well as an increase in amplitude and duration can make this syllable stand out from neighbouring syllables. But an F0 movement and an increase in duration are also often used to mark the end of a phrase. In addition, spectral properties of vowels and consonants have been reported to be affected by prosody (Cho 2011; Gussenhoven 2004). For example, lending prosodic prominence to a syllable can affect the formant values of the vowel in that syllable. Vowels in more prominent positions tend to be less centralised compared to their counterparts in less prominent positions, revealing a more distinct articulation (de Jong 1995; Cho 2005; Mücke 2018).

Decades of prosody research have emphasised that prosody has decomposable structure and that the relations of the elements within that structure are impor-

tant to understand prosody (Ladd 2008; Arvaniti 2011). The next section sheds light on prosodic structure (Section 4.1). Section 4.2 then takes to a first look at one of the most important functions of prosody in West-Germanic languages, namely the marking of *focus*, and uses the knowledge of prosodic structure to explain the relation of focus and prosody. Section 4.3 investigates the categorical and continuous aspects of prosody and argues that prosody is characterised by both. In the light of this argumentation, Section 4.4 takes a second look at focus marking and shows that categorical and continuous aspects of prosody are involved in this function of prosody. Section 4.5 gives an overview on how prosody affects movements of supra-laryngeal articulation and points out a strong connection of prosodic structure and what has been described as segmental. Section 4.6 takes a third look at focus beyond tonal properties presenting evidence that prosodic strengthening plays an important role here as well. Section 4.7 summarises the chapter.

4.1 Prosodic structure

A very influential framework for the analysis of prosody is the auto-segmental model (Pierrehumbert 1980; Ladd 2008), henceforth abbreviated as AM. This framework stresses the idea that prosodic structure is organised hierarchically and that the tune of an utterance, the *intonation contour*, can be decomposed into smaller units (tones) that are to some extent independent of the words they are produced on. Different hierarchies of prosodic structure have been proposed in the literature (Nespor & Vogel 1986; Pierrehumbert & Beckman 1988; Hayes 1989; Selkirk 1996; Shattuck-Hufnagel & Turk 1996). All proposals share the assumption that utterances can be decomposed into hierarchically organised constituents although they disagree as to the existence of some levels. Most researchers in the field agree upon a minimal structure that can be outlined as follows (Grice 2006): An utterance consists of one or more intonational phrases which contain one or more smaller phrases. Here, an intermediate phrase is assumed as the level between the intonational phrase and the word. 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. This prosodic hierarchy is illustrated with an example in Figure 4.1.

An intonation contour in the AM framework is modelled as a string of high and low tones, H and L, that is associated with the constituents in the prosodic hierarchy. However, not every word or syllable in a phrase receives a tone. The string of tones is associated with the "text", i.e. the syllables of a sentence, at
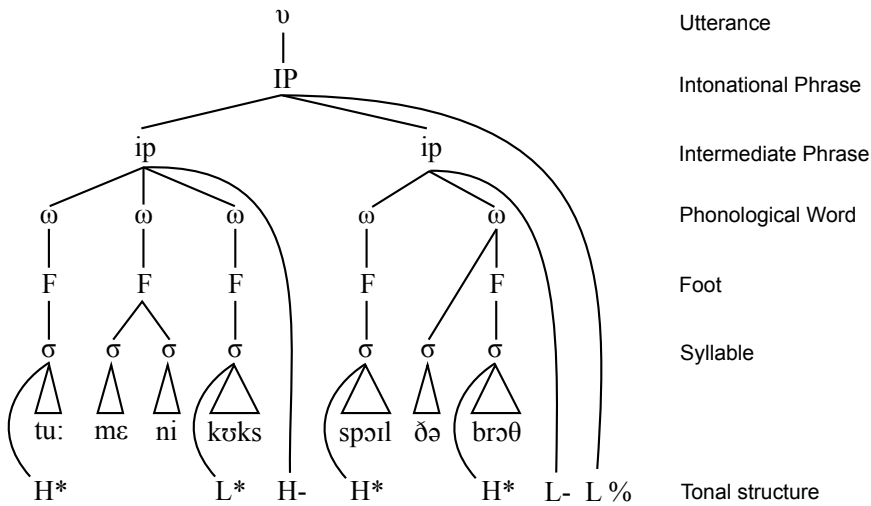
Figure 4.1: Prosodic hierarchy for the sentence "too many cooks spoil the broth" inspired by Grice (2006) and Gussenhoven (2004).

certain points. Between the syllables where tones occur, the intonation contour is interpolated (Ladd 2008).

Tones can occur as *pitch accents* or as *boundary tones*. In many intonational languages – like English, German and Dutch – pitch accents are associated with primary stressed syllables in the prosodic hierarchy. The term pitch accent is used here in the sense of *intonational pitch accent* as opposed to *lexical pitch accent* (Ladd 2008). Intermediate phrases can contain more than one pitch accent. Of all pitch accents, the last (fully fledged) pitch accent is called the *nuclear pitch accent* (Gussenhoven 2004). The placement of the nuclear pitch accent can be seen as a form of head-marking (Beckman & Edwards 1994; Shattuck-Hufnagel & Turk 1996) in which the most prominent syllable of the intermediate phrase is the head of this prosodic constituent and receives the nuclear pitch accent (Grice 2006). A pitch accent is marked by the symbol * and can consist of one or two tones, e.g. H* or L+H*. In the latter case, i.e. in bitonal pitch accents, only one of the two tones receives the * and is called the *starred tone*. The other tone occurs either before the starred tone, as in the case of a *leading tone* (L+H*), or after the starred tone, as in the case of a *trailing tone* (L*+H).

Boundary tones occur on two levels. Both intermediate and intonation phrases have boundary tones. One convention (see for example GToBI: Grice & Baumann 2002 and Grice et al. 2005) is to mark the boundary tone of an intermediate phrase with the symbol -, e.g. L-, and the boundary tone of an intonation phrase with

the symbol %, e.g. L%. Since intermediate phrases are contained in intonation phrases, the end of an intonation phrase co-occurs with the end of the last intermediate phrase of the intonation phrase. The notation of the two boundary tones is T_1-T_2%, where T stands for tone. If T_1 and T_2 are identical, only one of them is written, i.e. L-L% is written as L-%.

The motivation of modelling an intonation contour as a string of tones rather than a holistic melodic unit comes from the observation that the intonation contour is not simply stretched or compressed when the number of syllables present in the sentence increases or decreases. Instead, the shape of the intonation contour is affected by the alignment of the tones relative to the syllables they are associated with (Ladd 2008; Arvaniti 2011). The example in Figure 4.2 illustrates this point: It shows a schematised version of a contour prominently discussed in the literature by Ward & Hirschberg (1985), Hirschberg & Ward (1992), Arvaniti (2011), Ladd (2008) and many others. The contour is used to express incredulity towards a suggestion made by the interlocutor. On the left, the realisation on the mono-syllabic phrase "Me?!" is shown. The intonation contour rises, falls and rises again. This contour could be described as a holistic rise-fall-rise pattern (Ladd 2008). On the right, the figure shows two options of adjusting the contour to be realised on a longer phrase. In the upper plot, the holistic rise-fall-rise shape of the contour is simply stretched to span the longer phrase. However, this is not the way the contour is realised on longer phrases (Arvaniti 2011). The actual realisation of the contour is represented by the schematisation in the lower plot and demonstrates that the contour is best described as a composition of separate tonal events. The phrase starts low, the rise-fall movement is timed roughly with the syllable "ball" and not stretched over several syllables. The final rise of the contour is realised at the very end of the utterance, most probably on the last syllable of "designer". There is a low stretch of F0 in between these two events. An analysis within the AM framework decomposes this contour into a pitch accent and a rising boundary tone at the end of the phrase. The pitch accent is associated with the most prominent syllable in the phrase, "ball", and gets aligned with it in time. The boundary tone is associated with the phrase and marks the edge of this phrase. As a consequence, its rising tonal movement is realised towards the end of the phrase. In between these two tonal events, the low stretch of F0 represents an interpolation that can be extremely short as in the case of "Me?!" or longer depending on the number of syllables between the accent and the edge (Ladd 2008) .

A possible analysis in the AM tradition (as in ToBI: Beckman et al. 2005) would describe the contour as L*+H L-H%[1] with the L of the boundary tone spreading

---

[1]The fact that the phrase starts low is accounted for with a default boundary tone that is described as mid or low in the speaker's range and, being the default, is omitted from the transcription.
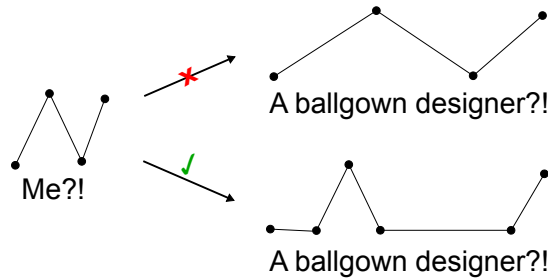
Figure 4.2: Schematic illustration of an intonation contour expressing incredulity (Hirschberg & Ward 1992). The right side shows two hypothetical options how the short version can be enlarged for longer sentences.

"leftwards" to the end of the pitch accent to create a low stretch on the non-prominent syllables after the pitch accent. Depending on the inventory of pitch accents and boundary tones (e.g. presence of leading and/or trailing tones), models in the AM framework can differ in their description of the same contour. The common assumption, however, is that the contour can be decomposed into tonal events and that such a description is better able to capture generalisations on intonational structure and meaning than a holistic approach.

4.2 A first look at prosodic focus marking

In West-Germanic languages, like German, Dutch and English, the placement of a pitch accent is used as a means to package the information conveyed by a sentence. In these languages, prosody plays a role in signalling which parts of the utterance are in *focus*, and which are in the *background* (Lambrecht 1994). This structure, the *focus structure* of the sentence, manifests the speaker's beliefs about how the information of the sentence fits the knowledge of the listener (Vallduví 1992). The focus constituent includes the information that the speaker believes to be significant, the background constituent contains information that is less relevant (Gussenhoven 2004). Often, the focus constituent contains new information while the background contains given information, but this is not always the case. In the above-mentioned languages, the focussed part usually receives the nuclear pitch accent. The background constituent usually does not receive the nuclear pitch accent and is often completely unaccented, in particular when the background part follows the focus part (also known as *post-focal deaccentuation*).

Consider the example in 4.1. The answer contains only new information (Ladd 1980; Gussenhoven 2004), nothing is given by the question, everything is in focus. This condition is often called *broad focus* (Ladd 1980). In this case, the nuclear pitch accent is usually placed on the last noun, "Jana". Even though "Jana" receives the nuclear pitch accent, this does not mean that it is the only word in focus. The mechanism that assigns the nuclear pitch accent within the focussed constituent is called *focus projection* (Büring 2003; Ladd 2008). The word that receives the nuclear pitch accent functions as the *focus exponent* for the larger focus domain (Ladd 2008).

$$
\begin{array}{ll}
\text{Q:} & \text{Was gibt's Neues?} \\
& \text{What's up?} \\
\text{A:} & \text{[Melanie will Jana treffen.]}_F \\
& \text{[Melanie wants to meet Jana.]}_F
\end{array}
\qquad (4.1)
$$

In example 4.2, the question triggers a different focus structure in the answer. The new information is contained in the word "Jana" only. The focus structure of this example is called *narrow focus* (Ladd 1980; Gussenhoven 2004). As in the example before, the pitch accent is placed on the word "Jana".

$$
\begin{array}{ll}
\text{Q:} & \text{Wen will Melanie treffen?} \\
& \text{Who does Melanie want to meet?} \\
\text{A:} & \text{Melanie will [Jana]}_F \text{ treffen.} \\
& \text{Melanie wants to meet [Jana.]}_F
\end{array}
\qquad (4.2)
$$

A special case of narrow focus is illustrated in example 4.3. Again, all information except for "Jana" is given by the question. In addition, "Jana" serves as a direct alternative to the referent "Peter" given in the question. The nuclear pitch accent is again placed on the word "Jana". This focus structure is often labelled *contrastive focus* or *corrective focus* (Gussenhoven 2004).

$$
\begin{array}{ll}
\text{Q:} & \text{Will Melanie Peter treffen?} \\
& \text{Does Melanie want to meet Peter?} \\
\text{A:} & \text{Melanie will [Jana]}_F \text{ treffen.} \\
& \text{Melanie wants to meet [Jana.]}_F
\end{array}
\qquad (4.3)
$$

So far, the pitch accent location did not change in the examples as a function of the focus structure of the sentence. Either the last noun was the only focussed element (narrow and contrastive focus) or it served as the focus exponent for a larger focus domain (broad focus). Example 4.4 demonstrates what happens if the focus is earlier in the sentence. The example in fact shows a case of contrastive focus. This time, however, the focussed element is "Melanie". As a consequence,

the nuclear pitch accent is placed on "Melanie" and the part of the utterance after this word, "will Jana treffen", is in the background. This part is deaccented and characterised by a low stretch of F0, lower amplitudes and shorter durations.

Q: Will Paul Jana treffen?
Does Paul want to meet Jana?
A: [Melanie]$_F$ will Jana treffen.
[Melanie]$_F$ wants to meet Jana.

(4.4)

Moving the nuclear pitch accent to a different word as a means of packaging the information is a very well described phenomenon in intonation languages like English, German and Dutch. Other intonation languages like Spanish and Italian do not use the manipulation of the nuclear pitch accent location to mark focus structure or do so only to a limited extent (Ladd 2008). Languages that use nuclear pitch accent location to mark focus have been called *plastic* languages, those that do not have been called *non-plastic* languages (Vallduví 1991).

Although manipulating the location of the nuclear pitch accent is a very well documented and salient feature of focus marking, this is not to say that focus structures with the nuclear pitch accent in the same place are identical with regard to their prosodic realisation. Later in this chapter, evidence for fine-grained strategies of prosodic focus marking will be discussed. The second part of this book presents a large data pool supporting and extending this evidence. It will become clear that both categorical and continuous modulations are used in the differentiation of focus structures. Therefore, before turning to this central issue, it is useful to take a closer look at what is known about categorical and continuous characteristics of prosody.

4.3 The nature of prosody: categorical and continuous

The outlined view of intonation in the AM framework is characterised by the assumption of discrete events in tonal structure. The last two sections sketched the motivation for this perspective. First, as illustrated by the example of the "incredulity contour" (see Figure 4.2), it is plausible to analyse intonation contours as composites. Some units that form these composites are associated with strong syllables while others are associated with phrase constituents. Hence, the wide-spread approach of AM intonation analysis posits that categoriality is a major feature of the structure of intonation contours: Tones can belong to the categories pitch accent or boundary tone.

Second, the examples of focus realisation in German (4.1, 4.2, 4.3, and 4.4) have added another important aspect of categorical patterning in the analysis of intonation. It is a discrete choice which syllable a pitch accent (in this case the nuclear pitch accent) is associated with. The pitch accent is either associated with syllable A or B (or C or…). Moreover, different types of pitch accents exist in the AM analysis. This means that while there is a possible syntagmatic contrast between a syllable that receives a pitch accent and one that does not, there are also possible paradigmatic contrasts between pitch accent types. A pitch accent can be of type A or B (or C or…), or in AM terms: H* or L* or H+L* (or L+H* or…). Models within the tradition of AM draw pitch accents from a closed inventory and consider these pitch accent types as phonologically distinct. Analogously, boundary tones are also given as a closed set of distinctive types.

The identification of categoriality in the structure of intonation forms the basis of the "intonational phonology" (Ladd 2008) put forward by scholars in the AM framework. However, it is not restricted to this framework and has antecedents in other models. For instance, intonation analysis in the tradition of the *British School* (O'Connor & Arnold 1973) identifies categorical distinctions on multiple levels. Pitch accented syllables are distinguished from other categories of prominence (secondary stress, tertiary stress, unstressed). Moreover, intonation contours are decomposed into parts like head, nucleus and tail. With regard to the *nuclear tone*, consisting of the nucleus and the tail, the British School has attempted to identify distinct tunes that evoke differences in meaning, like "high fall" or "low rise" (Cruttenden 1997). In the *IPO* model of intonation ('t Hart et al. 1990) – named after the Institute of Perception Research, in Dutch: Instituut voor Perceptie Onderzoek – similar to the analysis of the AM framework, a distinction is made between pitch movements associated with boundaries and pitch movements associated with prominences. Tonal patterns are described in terms of a composition of "categorically distinct entities" (Ladd 2008: 14), e.g. "Type A Fall" or "Type 1 Rise".

To conclude, the assumption of the categorical, phonological status of intonation contours or building blocks of contours plays a significant role in various aspects of the description of intonation. Experimental approaches have been developed to assess the notion of categoriality in intonation empirically. For example, Pierrehumbert & Steele (1989) manipulated the timing of the accentual peak of a rise-fall-rise contour on the sentence "Only a millionaire" from relatively early to relatively late with 13 intermediate steps.[2] The resulting 15 stimuli

---

[2]The "relatively early peak" L+H* in this experiment should not be confused with the general notion of an early peak accent that is usually transcribed as H+L* or H+!H*.

represented a continuum between two contours that can be described as L*+H L-H% and L+H* L-H% with the pitch accent associated with the syllable "mil". Apart from the alignment of the accentual peak with the accented syllable, the contour was unaltered. The subjects participating in the study were asked to listen to a stimulus and imitate it. The authors' prediction was the following: If the nature of the difference between the two pitch accent types is categorical, the subjects should not be able to reproduce the continuum exactly. The data should instead cluster around two alignment patterns, one representing L*+H and the other one representing L+H*. If, on the contrary, alignment is used as continuous dimension in the intonation of American English, the subjects should be able to reproduce the continuum faithfully.

The result showed that almost all participants produced patterns that more or less clustered around two points on the continuum of peak alignment, yielding bimodal distributions for almost all speakers. Interestingly, although the results of the study suggest that peak alignment is used in a categorical fashion rather than continuously, two interesting observations can be made. First, the relative sizes of the two modes of the bimodal distributions differ across speakers, indicating that the continuum of stimuli is divided differently by individual speakers. Most speakers categorise a larger portion of the continuum as L+H* but there is also one speaker that divides the continuum into two approximately equal parts. Second, the modes of the alignment distributions of individual speakers are located differently in time. Both observations show that categories like L*+H and L+H* are not invariantly linked to patterns of phonetic implementation.

Results pointing towards a categorical use of tonal alignment were also obtained by D'Imperio & House (1997), investigating the intonation of narrow focus statements and yes/no questions in Neapolitan Italian. In this variety of Italian, both narrow focus statements and yes/no questions are characterised by a rising nuclear pitch accent followed by low boundary tone. The pitch accent of the narrow focus statement is analysed as H* while the pitch accent of the yes/no question is analysed as L+H*. According to D'Imperio & House (1997), this difference is at least partially expressed with respect to timing of tonal events: The accentual peak of the H* is reported to be reached earlier than the peak of L+H*. The stimuli of this study comprise a five step continuum from relatively early (H*) to relatively late peak (L+H*). One set of stimuli was synthetically produced on the basis of the natural production of a statement and another set was produced on the basis of a question.

In contrast to Pierrehumbert & Steele (1989), a different experimental design was employed by D'Imperio & House (1997) to elicit the subject's responses. Participants were asked to listen to the continuum and categorise the stimuli they

just heard as a question or a statement. The results showed that subjects categorise the earlier peak (H*) as a narrow focus statement and the later peak (L+H*) as a yes/no question as expected. Around the mid-point of the continuum, the proportions of ratings for both categories are around 50% and the variability in the responses is high – a response pattern expected for categorical oppositions. It should be noted, however, that these effects are clearest for stimuli that were resynthesised from a natural production of a statement. Additional evidence for the role of temporal alignment in the categorisation of pitch accent types comes from Dilley & Heffner (2013) for American English as well as Kohler (1987) for German.

Gussenhoven & Rietveld (2000) provided support for categoriality of pitch accents in Dutch. The authors investigated the difference between a "high rise" and "low rise" question contour. In the former contour, the nuclear accented syllable is mid-pitched. In the latter contour, the nuclear accented syllable is low. Both contours are characterised by a rise from the accented syllable to the end of the phrase. In terms of an AM analysis, the contours could be described as H* H-H% and L* H-H% (Gussenhoven 2004). The experimental paradigm of Gussenhoven & Rietveld (2000) is based on the observation that speakers expand the pitch range when speaking in an emphatic or surprised fashion. Accordingly, subjects were presented with artificial contours in which the height of both the accentual target (L* or H*) and the boundary tone target (H-H%) were manipulated.

The results are in line with the prediction that what is analysed as a phonological low tone (L) behaves differently from what is analysed as a phonological high tone (H) when the pitch range is expanded. A L* accent is perceived as "more surprised" when realised lower whereas a H* accent is perceived as "more surprised" when realised higher. These results suggest that it is plausible to describe the accent as belonging to two different categories. At the same time, they reveal that intonational meaning, in this case "surprise", is carried by continuous variation as well. In addition, the results of Gussenhoven & Rietveld (2000) show that the implementation of categories is subject to contextual variation and not invariant.

The experiment of Gussenhoven & Rietveld (2000) was based on the assumption that pitch range constitutes a parameter continuously varied by speakers. Ladd & Morton (1997) questioned the purely continuous nature of pitch range as a parameter of intonation. The authors sought empirical evidence for the existence of *normal* and *emphatic* peak pitch accents as two distinct categories in British English. This distinction is assumed to be (at least partially) expressed by pitch range with the emphatic accent having a larger pitch range than the normal accent. Since in this study phrases with only one accent are used, it remains

open whether pitch range is interpreted "globally" on the level of the phrase or "locally" on the level of the accent. Regardless of which interpretation is to be favoured, one can speak of a difference in *pitch excursion*, manifested phonetically in the relation between the F0 target of the peak on the accented syllable and the F0 target of the low boundary following.

The findings of Ladd & Morton (1997) on the categorical use of pitch excursion remain inconclusive and thereby provide an interesting example of the study of categoriality and continuity in intonation. The methodology of the study included both rating on a scale as well as forced choice tasks. In a first perception experiment – which was designed to be a norming study to test the usefulness of the speaker's voice and the manipulation strategies – the authors randomly played stimuli from a continuum of pitch excursions to 35 subjects. This continuum consisted of resynthesised contours based on natural utterances produced in either a "normal" or an "emphatic" fashion. The pitch excursion on the continuum was increased in steps of 6 Hz. Two instructions to elicit the participants' answers were used. In the first instruction, the participants were asked to rate the degree of emphasis on a ten-point scale. In the second instruction, two different interpretations were presented, "a neutral statement" and "a contradiction or correction". The participants were asked to rate on a ten-point scale how likely the interpretations were. The results show that stimuli manipulated from an emphatic natural basis were interpreted as more emphatic, a finding that indicates that emphasis is expressed by more than just pitch excursion. Apart from this effect, both sets of stimuli evoked similar responses: The larger the pitch excursion, the more ratings towards an "emphatic" interpretation. This trend showed no clear discontinuities that would be expected for a categorical difference.

In a second experiment, Ladd & Morton (1997) used a subset of the stimuli from the first experiment and employed a classic categorical perception paradigm consisting of two forced choice tasks, an *identification task* and a *discrimination task*. In the identification task, subjects listened to a stimulus from the continuum and judged whether the sentence described an "everyday occurrence" or an "unusual experience". In the discrimination task, subjects listened to pairs of stimuli in quick succession and were asked to choose whether they were the same or different. The stimuli of a pair were either neighbours on the manipulated continuum and thus differed by 6 Hz or were identical.

The responses of the identification task provided the classic S-shaped curve expected for the presence of two categories. The curve showed that in the region between 138 and 156 Hz, the responses abruptly become more variable and tend towards 50%, i.e. subjects were less certain about the interpretation of the stimuli. This result alone does, however, not suffice to attest the existence of two

distinct categories of normal and emphatic. It has to be kept in mind that the subjects had only two options to choose from. To provide clearer evidence for categoriality, the discriminability of stimuli must increase in the region of the putative category boundary. Remember that the pairs of different stimuli played in the discrimination task consisted of neighbours on the continuum differing by 6 Hz. Subjects should perform better at discriminating stimuli belonging to different categories compared to stimuli belonging to the same category. In fact, in the study of Ladd & Morton (1997), subjects were better at discriminating stimuli pairs at the putative category boundary. However, the discrimination rate was rather low even in this region. In addition, the discrimination rates at the ends of the acoustic continuum, i.e. within the assumed categories, were also larger than zero.

While the findings of Ladd & Morton (1997) do not support categorical perception of normal and emphatic pitch accents in a strict sense, they suggest that the continuum was not completely linearly perceived. These observations reflect the fact that pitch excursion may be used in a categorical and continuous way at the same time. For example, transcription systems in the AM tradition acknowledge that H* and L+H* accents are, among other parameters, distinguished in terms of in the magnitude of the rise towards the starred tone H* (Grice & Baumann 2002). The rise is described as steeper and higher for L+H* compared to H*. Thus, in this case, local pitch excursion is one phonetic parameter to differentiate phonological categories. At the same time, this parameter is used in a continuous fashion. The continuous use can be the result of global or local pitch range expansions to signal emotional affection or surprisal (as in the case of Gussenhoven and Rietveld's experiment) or to make a distinction between different accent types clearer. The latter case might be classified as a kind of intonational hyperarticulation in the sense of Lindblom's (1990) Hyper- and Hypo-Articulation theory (H&H theory).

Striking similarities are found for the phonetic parameter of temporal alignment of an accentual peak. At the beginning of this section, evidence has been presented for the existence of a categorical implementation of this parameter. However, it is also known that later alignment of an accentual peak may be used in addition to increased pitch excursion – or even as a substitute for increased pitch excursion – to increase prominence in intonation languages like German and English (Gussenhoven 2004; Ladd & Morton 1997). A possible explanation might be that higher pitch excursions take longer to be executed. Therefore, higher peaks have a potential physiological correlate of delay, a mechanism that speakers may have "tacit knowledge" of (Gussenhoven 2004: 90). Regardless of the validity of this explanation, the observation is interesting for two reasons.

First, it shows that alignment is used in a continuous manner as well. Second, while the symbiosis of pitch excursion and alignment is conceptualised as a phonetic effect, it is also manifested in the description of intonational categories: In addition to a difference in the magnitude of the rise towards the high accentual target (as already mentioned above), the distinction of H* and L+H* in German ToBI is based on a difference in alignment. A later peak alignment is attributed to the L+H* accent in comparison to the H* accent in this model.

Essentially, these observations suggest a deep intertwining of the use of prosodic parameters to establish phonological categories on the one hand and their direct, continuous (or: phonetic) use on the other. The next section will present evidence on the realisation of focus types that provides support for the idea of an interrelation of phonetic and phonological aspects of intonation developed in the present section.

4.4 A closer look at prosodic focus marking

Earlier in this chapter, the use of nuclear pitch accent placement in West-Germanic languages to mark focus structure was introduced. It was noted that the placement of the nuclear accent might not change as a function of focus structure in some cases. In the examples, the nuclear accent remained on the word "Jana", the last noun, regardless of whether the focus structure was broad focus, narrow focus or contrastive focus. The literature nevertheless discusses prosodic ways of differentiation between these focus types.

For American English, Pierrehumbert & Hirschberg (1990: 296) state that the L+H* pitch accent "evokes a salient scale". It is described to mark contrastive focus as opposed to the H* accent that – combined with a low boundary tone – contributes to a "neutral declarative intonation" (Pierrehumbert & Hirschberg 1990: 290). Watson et al. (2008) add supporting evidence for the interpretation of L+H* as marking contrastive focus in an eye-tracking study. They note, however, that the uses of L+H* and H* seem to overlap.

For German, Baumann et al. (2007) found that the proportion of downstepped H* accents, i.e. !H*, is highest in broad focus. The number of upstepped H* accents, ^H*, is highest in contrastive focus. Plain H* accents are predominantly found in narrow focus. While the distributions are overlapping, in line with the perceptual findings of Watson et al. (2008), these results by Baumann et al. (2007) indicate that different pitch accent types are used to signal different focus types. Crucially, the predominantly used accent types are characterised by increasing acoustic salience from broad focus to narrow focus, and from narrow focus to

contrastive focus (see Baumann & Röhr 2015). The results are supported by data presented in Baumann et al. (2006) that show fewer downstepped accents in narrow compared to broad focus, and fewer downstepped accents in contrastive compared to narrow focus.

Contrary to these findings, Féry (1993: 63) reported no difference between broad and narrow focus. The divergence of the results might be attributed to the fact that the speech material of this study is different from the sentences used in above-mentioned studies by Baumann et al. (2007) and Baumann et al. (2006) (and the examples given earlier in the present chapter). Another explanation might be sought in the fact already mentioned several times in this discussion: The distributions of pitch accent types used for different functions are overlapping. Hence, there does not seem to be a one-to-one mapping of pitch accent type to focus type.

Grice et al. (2017) investigated the categorical and continuous aspects of intonational patterns used to mark focus structures by 5 German speakers (data also reported on in Mücke & Grice 2014). The speech material used was very similar to the examples 4.1, 4.2, and 4.3. Their analysis comprised both symbolic annotations as well as measurements of continuous, phonetic parameters. Figures 4.3 and 4.4 give a first insight into their data by showing the GToBI nuclear pitch accent distributions based on consensus annotations by two annotators. Figure 4.3 gives the distributions for all speakers together and confirms the finding outlined above, namely that there is not a one-to-one mapping of pitch accent type to focus type but that there are tendencies: Most falling accents, H+!H*, can be found in broad focus, most accents with a steep rise, L+H*, can be found in contrastive focus. Narrow focus seems to be positioned in between with a rather high number of H* accents but also a considerable number of L+H* and even H+!H* accents. Figure 4.4 presents the distributions for each speaker separately. While for some speakers (F1 and F2) the overall tendency of Figure 4.3 can be observed at least for the differentiation of broad focus and contrastive focus, for other speakers (F3 and M2), a different picture is obtained. These speakers use one pitch accent category predominantly for all three focus types. In the case of speaker F3, it is H*; in the case of speaker M2, it is L+H*. Speaker M1 seems to use a mixed strategy between the two groups. This speaker differentiates broad and narrow focus by using different pitch accent types, but shows almost exclusively L+H* accents in both narrow and contrastive focus.

The data of Grice et al. (2017) was used in a perception test by Cangemi et al. (2015) and Krüger (2009). Their results showed that all five speakers were "able to convey their pragmatic intentions to listeners in a comparable way" (Grice et al. 2017: 91). This conclusion is rather surprising with regard to the different
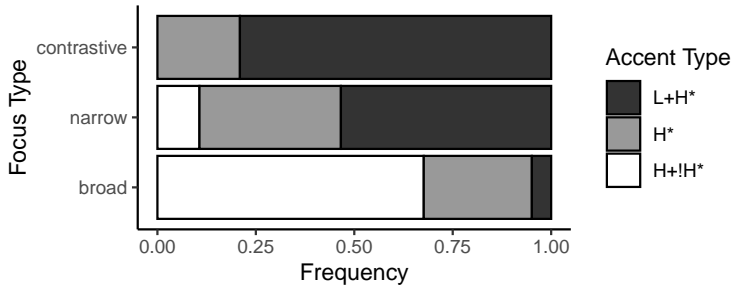
Figure 4.3: Nuclear pitch accent type distributions in the data set of Grice et al. (2017).
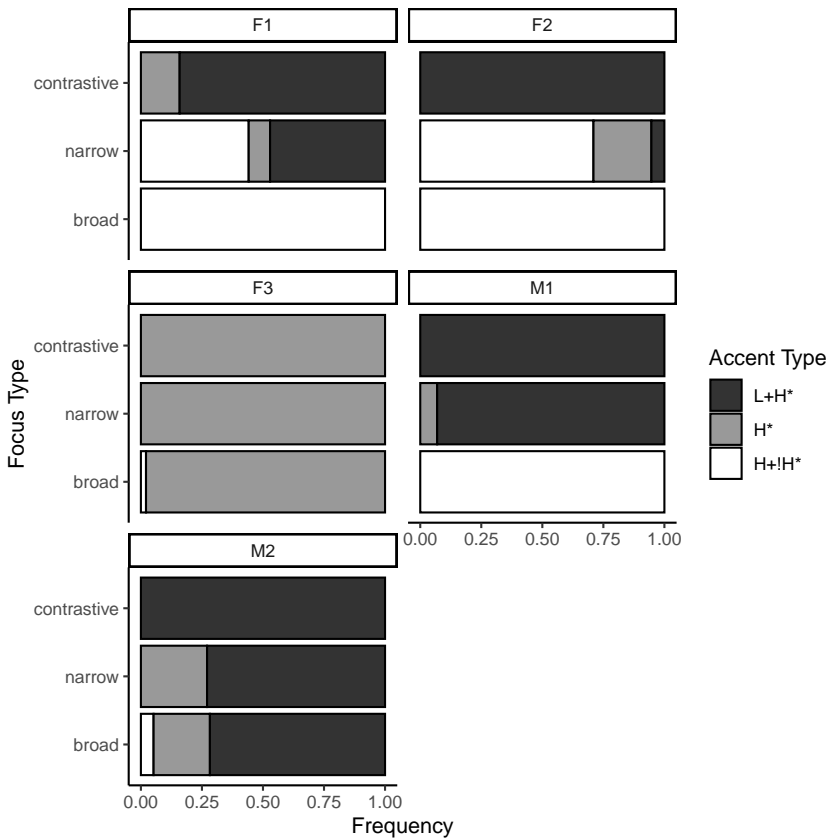

Figure 4.4: Nuclear pitch accent type distributions in the data set of Grice et al. (2017) for each speaker separately.

pitch accent distributions and considering that some speakers do not differenti-ate the focus types clearly in terms of their pitch accent choice. Motivated by the discrepancy of pitch accent choice and scores in perception testing, the analysis by Grice et al. (2017) went beyond the description of intonational patterns using pitch accent types and investigated various continuous parameters. Interestingly, their results reveal that the speakers' strategies are not as disparate as the pitch accent distributions suggest. Figure 4.5 shows the mean alignment values of the peak of the nuclear pitch accent in relation to the onset of the accented syllable.[3] Note that the peak does not have to belong to the starred tone (as in the case of H* and L+H*). In H+!H* accents, the peak is part of the leading tone (H+). The alignment data reveal the same pattern for all five speakers: Broad focus has the earliest peak alignment, contrastive focus has the latest peak alignment, narrow focus is in between the two extremes: broad < narrow < contrastive. For the speakers that showed pitch accent type differentiation (F1 and F2, partly M1), the alignment modification "crosses" category boundaries of pitch accent types. H+!H* accents in broad focus, for example, contribute negative alignment values as opposed to clearly positive values for L+H* accents in contrastive focus. For the other speakers (F3 and M2, partly M1), the alignment is modified within the boundaries of a single category. In other words, the parameter alignment is used in both continuous, "phonetic" and categorical, "phonological" ways but the gen-eral trend of modification, or the direction of modification, is invariant across speakers.

In addition to the alignment of the peak, Grice et al. (2017) measured the *tonal onglide* (Ritter & Grice 2015), a parameter that plays an important role through-out the following parts of this work. The tonal onglide is a means to assess the F0 movement towards the main tonal target of the accent – in AM terms, the starred tone. In the version that Grice et al. (2017) employed, the tonal onglide is quantified as the difference in semitones between the starred tone target and a fixed point 30 ms before the onset of the accented syllable. If the accent is falling, the resulting tonal onglide is negative, if the accent is rising, the tonal onglide is positive. This means that the tonal onglide gives an indication of the movement of the accent (falling vs. rising) but also of the magnitude of this movement. It is comparable to the measure of alignment by answering the question whether the accent is an early peak accent like H+!H* or a mid to late peak accent like

---

[3]The original paper uses the alignment relative to the accented vowel onset which yields com-parable results. The current figure is based on the openly available data set which contains both alignment relative to the onset of the accented syllable (that has a CV structure) as well as alignment relative to the onset of the accented vowel.
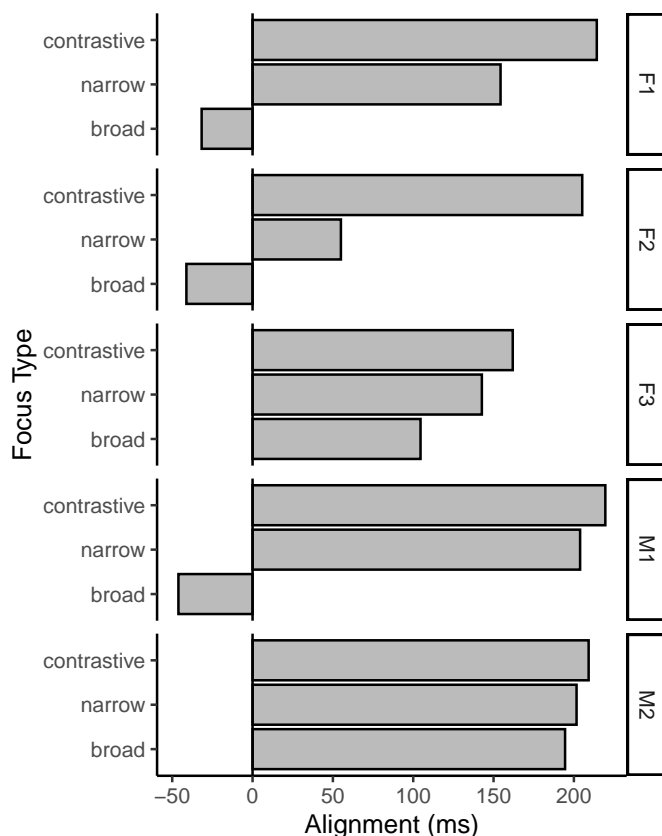
Figure 4.5: Mean alignment of the peak relative to the onset of the accented syllable in the data set of Grice et al. (2017) for each speaker separately.

H* or L+H*. But it also assesses pitch excursion continuously, for example by quantifying the magnitude of the rise in accents like L+H*.

In Figure 4.6, the results of the tonal onglide measure from Grice et al. (2017) are shown. The picture obtained here is comparable to the results for the peak alignment – it is even clearer in some cases. All speakers manipulate the tonal onglide in the same direction. Broad focus has the smallest tonal onglide values, contrastive focus has the largest, narrow falls in between: broad < narrow < contrastive. As in the case of alignment, some speakers cross clear category boundaries. For example, the negative tonal onglide values of F1, F2 and M1 in broad focus are an indication of accents of the falling type H+!H* whereas the positive onglide values in contrastive focus result from the frequent use of the
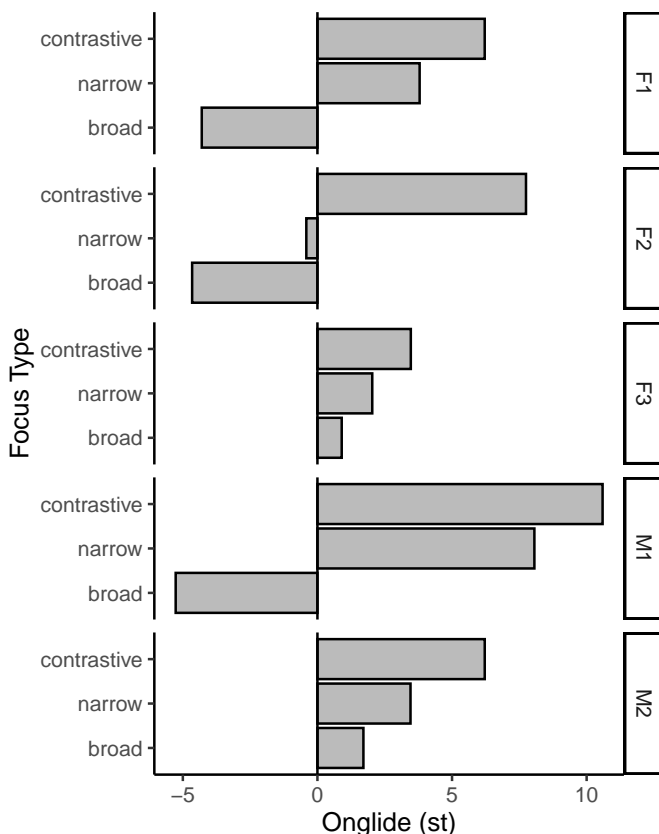
Figure 4.6: Mean tonal onglide values in the data set of Grice et al. (2017) for each speaker separately.

rising accent type L+H*. Crucially however, a speaker like F3 that exhibits only H* accents in the categorical analysis of Figure 4.4 differentiates the focus types in terms of onglide values. The same is true for speaker M2 and the difference between narrow and contrastive focus in speaker M1.

Both alignment and tonal onglide give a deeper insight into strategies of focus marking. What appears to be disparate strategies in a symbolic account is shown to be very similar when looking at the continuous parameters in addition to the symbolic transcription. Moreover, the results suggest that there might be no clear boundary between a categorical and a continuous use of intonation to mark focus types. Rather, speakers seem to integrate the two uses with a clear mutual trend. As the perception results of Cangemi et al. (2015) and Krüger (2009) show,

listeners are able to use the information encoded in the speech signal, regardless of whether the differences are in terms of what is described as pitch accent categories or more subtle and subsymbolic variation. This finding resonates well with the remark of Ladd (2014: 88) discussed in Chapter 2 that categorical and continuous variation commonly go hand in hand and very often "are likely to interact and reinforce one another".

4.5 Prosodic strengthening

So far, this chapter has concentrated on tonal aspects of prosody. A large body of research shows, however, that many other aspects play an important role as well. One important aspect is that the articulation of consonants and vowels – also often called *supra-laryngeal* articulation as opposed to *laryngeal* articulation that produces tone – is affected by prosodic structure (Mücke 2018). More specifically, it has been shown that some positions in the prosodic hierarchy are characterised by "stronger" articulation to increase syntagmatic and paradigmatic contrasts. Often, articulatory gestures are expanded in the spatio-temporal domains (Cho 2011). This phenomenon, termed *prosodic strengthening* (Cho 2006), can be observed as a result of marking the boundaries of prosodic units, like an intonational phrase, but also as a concomitant of accentuation, e.g. in syllables with the nuclear pitch accent. This section attempts to give an overview on parts of this research that are most relevant to the topic of the present work, and will thus focus on prominence-induced strengthening.

Speakers have multiple possibilities to express prominence in the speech signal. As has been discussed extensively in this chapter, the placement of a (nuclear) pitch accent is a very important prosodic feature – perhaps the principal feature in this regard. A pitch accent is characterised by a pitch movement around the syllable that bears it. However, as introduced above, amplitude, and duration (Ladd 2008; Grice 2006) as well as spectral properties of vowels in the accented syllable (Cho 2011; Gussenhoven 2004) have been identified as important prosodic parameters. Under the influence of accent, amplitude and duration are increased, and the spectral properties of vowels and consonant are modified (Cho 2011). The source of all these modulations can be found in various modifications of supra-laryngeal articulatory patterns during the production of syllables under accent (Mücke 2018; Cho 2005; de Jong 1995; Beckman et al. 1992; de Jong et al. 1993; Harrington et al. 2000; Cho & McQueen 2005; Mücke & Grice 2014).

The research literature describes different strategies of supra-laryngeal articulatory marking of accent in the case of vowels. One strategy that has been discussed is *sonority expansion* (Beckman et al. 1992). Sonority expansion enhances

the sonority of the vowel to strengthen the syntagmatic contrasts between accented and unaccented syllables. This strategy reflects the speaker's intention to produce louder and more sonorous syllables by opening the mouth wider as a more open oral cavity allows for a greater radiation of acoustic energy from the mouth.

Another strategy is referred to as *localised hyperarticulation* (de Jong 1995). Based on the H&H theory developed by Lindblom (1990), it follows the observation that prosodic prominence is expressed by a more extreme articulation of the tongue body in vowel productions. This strategy is often connected to the enhancement of paradigmatic features such as the place feature for a specific vowel. The tongue body position becomes lower in hyperarticulated 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 & McQueen 2005).

Sonority expansion and hyperarticulation are highly overlapping strategies in low vowels like /a/. Opening the mouth wider and lowering the jaw and the tongue both achieve a higher sonority as well as a hyperarticulated, more distinct lower vowel. In other vowels, for example high vowels, the strategies may compete. Opening the mouth wider and lowering the jaw counteracts the raising of the tongue that is necessary to hyperarticulate a high vowel, i.e. to make it higher. There is, however, evidence that it is possible to combine the two strategies in the coordination of different articulatory subsystems that are used independently to some degree. Speakers may use the lingual system to hyperarticulate the place feature in high vowels such as /i/ and /ʊ/, while the mandibular and the labial system contribute to sonority expansion by increasing the degree of lip opening (Harrington et al. 2000).

Regardless of the strategy pursued by the speaker, modifications found in the articulatory movements can be described in terms of variation in the parameters of the gestural model of Articulatory phonology that is based on the critically damped harmonic oscillator. Figure 4.7 presents an overview following Cho (2006) on the possible parameter changes. In (a) a change in stiffness is shown. In the exposition of the harmonic oscillator model in Chapter 3, stiffness corresponds to the parameter k. With higher stiffness values, the target is reached earlier, the velocity of the movement is increased.[4] In (b) a change in the intergestural timing is shown. If a gesture is activated earlier, it can truncate a preceding

---

[4]While the parameter stiffness relates to the temporal organisation of the gesture only in this definition, it is defined in different ways in the literature. Munhall et al. (1985: 465) define stiffness as a tempo-spatial parameter such that "higher stiffness corresponds to shorter duration [of] movements and greater peak velocity/movement amplitude ratios".
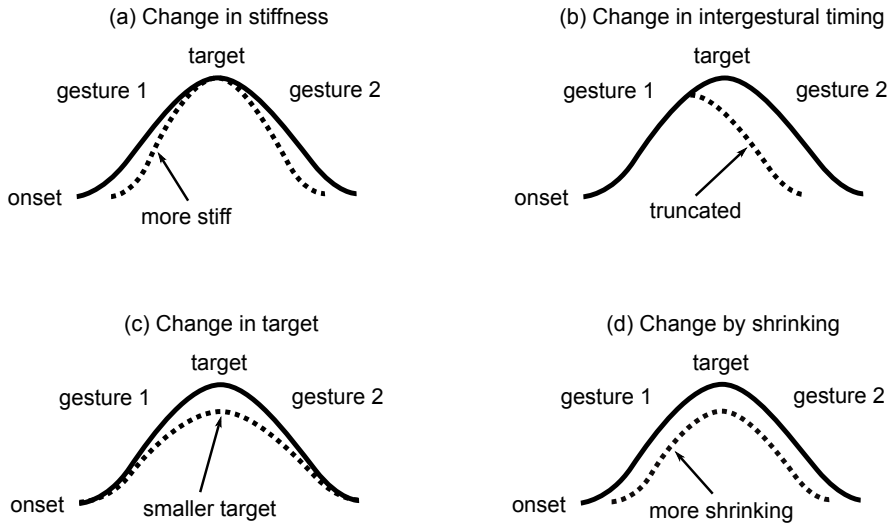
Figure 4.7: Modifications of articulatory gestures induced by prominence following Cho (2006).

gesture before this gesture reaches its target. Remember that temporal overlap was used in modelling assimilation in Chapter 2. There, overlap took place on different tiers, i.e. in different tract variables. The result was that one gesture was hidden. If overlap takes place in the same tract variable, truncation is the outcome. In (c) a change in the target is illustrated. The target corresponds to the parameter x_0 in the harmonic oscillator model. With a larger target to be reached in the same time, the velocity of the movement increases as well. In (d) a simultaneous change in target and stiffness is shown, resulting in a scaling (shrinking or enlargement) of the gesture.

The effect of prosodic structure on articulation, resulting in the modification of gestures as shown above, can be modelled using *prosodic gestures*. Prosodic gestures are conceptualised as special gestures without any constriction. They exert control on the individual gestures responsible for the constrictions to produce vowels and consonants. This approach was developed to account for the fact that gestures are executed slower and overlap less in time at prosodic boundaries (Byrd 2000; Cho & Keating 2001; Fougeron 2001; Tabain 2003; Cho 2006; Krivokapić & Byrd 2012). The proposal of Byrd et al. (2000) and Byrd & Saltzman (2003) is a prosodic gesture, called π-gesture, that modulates the intergestural timing at prosodic boundaries by modifying the stiffness of the constriction gestures co-activated with it. The extent to which the "local speaking rate", i.e. the

stiffness of gestures, is affected depends on the strength of the π-gesture (Byrd 2000: 13). The activation level of the π-gesture is defined by the rank of the constituent in the prosodic strength.

While the π-gesture models what happens at prosodic boundaries, the μ-gesture has been developed to account for the modifications of articulatory gestures under prominence. Initially proposed to account for the effect of stress (Saltzman et al. 2008), it comes in two different versions. The μ_T-gesture is slows down the constriction gestures in the interval of activation like the π-gesture (Krivokapić 2014). The μ_S-gesture modulates the spatial properties of the constriction gestures. Although developed for stress, μ_S-gestures can be applied to the level of accent as well.

4.6 Prosodic focus marking beyond tone

With regard to focus structure, the results of Mücke & Grice (2014) give interesting insights into how speakers use the modification of supra-laryngeal articulation. The study used the same data set as Grice et al. (2017) with the three focus types broad, narrow and contrastive focus. In addition, the data set contains material of a fourth condition in which the target word is in the background, i.e. out-of-focus. This condition corresponds to example 4.4. Remember that in this condition the target word does not receive an accent.

All target words in this data set are disyllabic with stress on the first syllable that is either /biː/, /baː/ or /boː/. The authors measured the duration and the displacement of the opening gesture of the lips from the maximal constriction of the lips for the bilabial stop to the maximal opening for the vowel.

The data of Mücke & Grice (2014) show that there is a gradual increase in the duration and displacement of the opening gesture from background to contrastive focus with intermediate values for broad and narrow focus, see Figure 4.8. Interestingly, there does not seem to be a big step between background and broad focus, i.e. between unaccented and accented. In fact, Mücke & Grice (2014) find a statistically significant effect between the two conditions for only one speaker (F1). In contrast to this finding, the modifications between broad focus and contrastive focus seem to be much larger.

These data suggest that accented words do not necessarily have to be articulated differently from unaccented words (at least in the investigated articulatory dimensions). In addition, modifications of the supra-laryngeal articulation are used to encode focus structures that have the nuclear pitch accent in the same place (broad, narrow and contrastive focus all have the nuclear pitch accent on
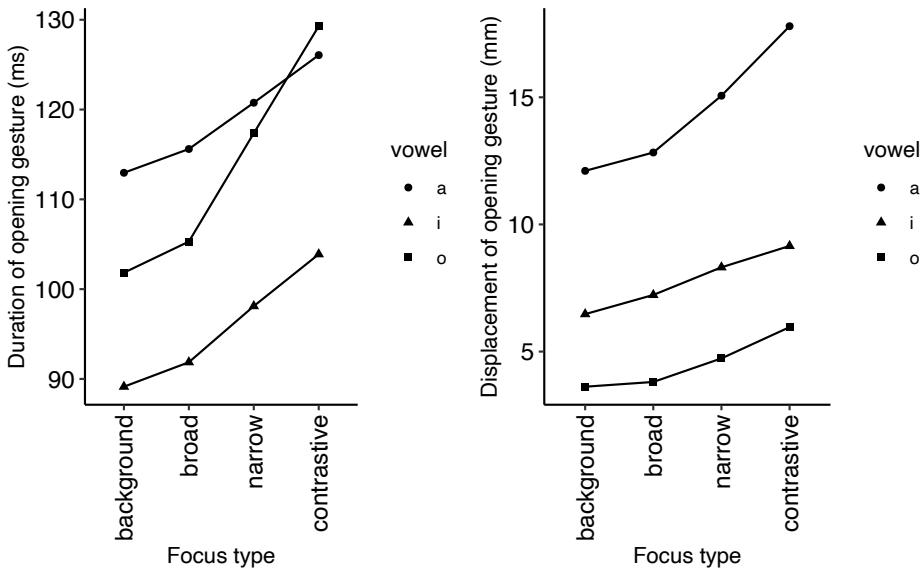
Figure 4.8: Mean duration and displacement of the vocalic opening gesture of the lips in the stressed syllable from the study of Mücke & Grice (2014).

the target word). The latter result aligns well with the findings of Grice et al. (2017) that show tendencies to enhance prosodic prominence by choice of pitch accent type and the phonetic parameters of the pitch accent (alignment, onglide) from broad to narrow focus and from narrow to contrastive focus. It indicates that placing a nuclear pitch accent on a word is not the only important mechanism and that prominence is marked beyond accentuation. In addition to the categorical property of accentuation, continuous modulations play a significant role in the supra-laryngeal as well as the laryngeal system. These observations provide strong support for a view that acknowledges the importance of categoriality and continuity in prosody.

4.7 Summary

This chapter has been a whirlwind exploration of important concepts of prosody, prosodic structure and prosodic prominence. It has reviewed arguments for why it is reasonable to consider prosody as shaped by both categorical and continuous traits. The prosodic expression of focus was considered from different perspectives as focus is one of the most important functions of prosody in West-

Germanic intonation languages. In addition, studies dealing with the prosodic marking of focus have revealed interesting results concerning the categorical and continuous aspects of prosody and their integration. The data presented in the following chapters of this book will also concentrate on focus and take the studies of Mücke & Grice (2014) and Grice et al. (2017) as a basis for the experimental design, attempting to advance their conclusions by providing a formal modelling account.

5 Data collection: a controlled corpus of prosodic focus marking

Forming the second part of the book, the next few chapters are concerned with a study on prosodic prominence in German that is designed to explore integration in a two-fold manner: First, it concentrates on the intertwining of categorical and continuous aspects of prosodic prominence. Second, it takes into account aspects of tonal structure as well as supra-laryngeal articulation, since both have been shown to be relevant for prosodic prominence in Chapter 4. Special attention is going to be paid to the theoretical implications that the data reveal. Therefore, a major part of this second part of the book is going to be concerned with the development of a model that is able to explain categoriality and continuity as the outcome of a single system. This model owes a lot to the dynamical models presented in Chapter 3 and aims to contribute to a theoretical symbiosis of the categorical and the continuous aspects of speech. This symbiosis can be stated more concretely in viewing phonetics and phonology as a unified system in which "a single formal language, nonlinear dynamics, makes it possible to model the relation between the discreteness of phonological form and the continuity of phonetic substance" (Gafos & Benus 2006: 937).

Building on the studies of Mücke & Grice (2014) and Grice et al. (2017), the current work investigates the expression of focus by prosodic means in a controlled laboratory setting. A large data pool was elicited that allows extensive examinations of the interplay of categorical and continuous aspects of prosodic focus marking and prosodic prominence. The following chapters describe the results obtained from the various analyses performed on the corpus and the modelling approaches designed to capture the main generalisations that the data provides for linguistic theory. Both chapters make use of the same data set, or subsets of the same data set, and employ similar measurement methods. These methods are explained in the current chapter. But before turning to the methodology, the research goals of the present study shall be stated more clearly.

The aims pursued by the research presented in this second part of the present work can be formulated as follows: First, it will be investigated whether the relation of nuclear pitch accent type to focus type can be described as one-to-one or

whether the mapping can best be characterised as fuzzy and probabilistic with different proportions of pitch accent types present in each focus type. Second, in addition to tendencies of the use of categories to differentiate focus types – even if their distributions might not be clear-cut or non-overlapping – the continuous aspects of the realisation of the pitch accent types will be investigated. Third, it will be examined whether there is co-variation of categorical and continuous aspects of the pitch accent patterns found in the data. Fourth, as stated above, the study will take a look beyond F0 and investigate the supra-laryngeal modifications (tongue and lip kinematics) used to express prominence and mark focus types. In this context, modifications found between unaccented and accented words as well as modifications within the group of accented words in different focus types will be considered. This last research goal is suited to contribute to our understanding of the interplay between categories and continuous modifications. Since *accentuation* is conceptualised as a categorical quality of prosodic structure, it is interesting to consider whether articulatory modifications mirror this categorical nature or whether articulatory modifications are used in a continuous way to enhance prosodic prominence beyond accentuation. Furthermore, including tonal and articulatory aspects in the analysis has the potential to advance our understanding of the multiple dimensions that are recruited to express prosodic prominence.

5.1 Speakers and recordings

Twenty-seven monolingual German native speakers were recorded acoustically using a head-mounted condenser microphone and with 3D Electromagnetic Articulography (EMA) using a Carstens AG501 articulograph. 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. For the EMA recordings, sensors were placed on the jaw, upper and lower lip, tongue tip, tongue blade, and tongue body to track the movements of the articulators. To compensate for head movements, reference sensors were placed on the bridge of the nose and behind the ears. A bite plate measure was used to rotate the occlusal plane. The articulatory data were recorded at 1,250 Hz, downsampled to 250 Hz and smoothed with a 3-step floating mean. In this work, the data from the lip sensors and the backmost tongue sensor are analysed.

All recordings took place at the I*f*L – Phonetik department of the University of Cologne. The speakers were aged between 19 and 35 at the time of recording. 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 recording session, after the participant had been prepared, lasted about 45 minutes including a training session.

The participants were involved in an interactive animated game sitting in front of a computer screen. They were told that the game revolved around two robots working in a factory. One of the robots likes to move around the tools while the other, 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 and the second robot entered the room, took the tool, and left the scene.

5.2 Speech material

Natural productions by a male, native German speaker were used for the robot's questions that served as triggers for the focus structures of the answers. They 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 5.1 shows examples for such question-answer-pairs. In the examples, square brackets and subscript *F* mark the focus domain. Each question was given auditorily and additionally shown as a combination of pictures in a thought bubble above the robot's head: 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 presented 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 "nein" ('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.

To control for the segmental context and for word frequency, twenty German sounding disyllabic nonce words with a C_1V_1:C_2ə structure were chosen as target words. The words were designed to have word stress on the first syllable. The

Table 5.1: Example question-answer-pairs to elicit the focus structures.

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

stressed vowel V_1 was either /aː/ or /oː/, the second vowel always schwa. The consonants (C_1 and C_2) 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 consonants and vowels were combined such that each first consonant C_1 of /n m b l v/ occurred twice with each first vowel V_1 /aː/ or /oː/. Each second consonant-schwa-combination /nə mə zə lə və/ 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" /diː/. All participants pronounced the words as expected. The target words are listed in Table 5.2.

In the setting of the experiment, each target word was associated with a fictitious visual object. The association remained fixed throughout the whole experiment and across all participants. There was no connection between the appear-

Table 5.2: Target words.

Nohme	Mohme	Bohme	Lohne	Wohme
Nohse	Mohwe	Bohwe	Lohle	Wohse
Nahne	Mahne	Bahle	Lahse	Wahne
Nahle	Mahse	Bahwe	Lahle	Wahwe

Table 5.3: Tools with English translation.

Tool	Translation	Tool	Translation
Amboss	anvil	Pinsel	Paint brush
Besen	broom	Rolle	Paint roller
Bohrer	drill	Säge	Saw
Bürste	brush	Schere	Scissors
Hammer	Hammer	Zange	Pliers

ance of the object and the sound of the word. 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 or had any difficulties pronouncing the words. 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. All tools are given in Table 5.3. Since there were 10 tools and 20 target words in the game, each tool had to occur twice. Furthermore, for the background condition and the contrastive focus condition, a competitor tool or object was needed: 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 all participants, yielding 20 quadruples of target object, tool, competitor object, and competitor tool.[1] The competitor object was chosen randomly with the restriction that the first consonant or the

---

[1] In the case of trials with broad focus or narrow focus, the competitor object and the competitor tool are not needed. Thus, in the quadruples for broad and narrow focus trials, these positions are empty.

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 target words occurred with all four focus conditions resulting in 80 trials per speaker. As a training phase, 16 trials with different objects and tools (everyday objects, e.g. a table) preceded the actual experiment session.

For each of the 27 participants the order of trials was randomised with some restrictions: 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 sat behind the participant and 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 seconds and the screen transitioned through a series of light, muted colours. This short break detached the trials from one another to make sure that the focus structure of the target sentence made reference to the current trial only and not to the previous trial. In order to make the task more game-like, points were counted for each complete trial in the lower right corner of the screen. Figure 5.1 presents an example of the experiment screen where the second robot has just asked his question and is waiting for the answer of the participant. The code of the experiment app is available under open source and creative commons license: http://doi.org/10.5281/zenodo.2611287.

5.3 Measures

This section describes the various measures performed on the data set. All measures are in some form based on the annotation of the stressed syllable and the vowel of the stressed syllable. One trained annotator labelled the beginning and end of the stressed syllable of each target word using the waveform and the spectrogram in the emuR speech database system (Winkelmann et al. 2018). Another trained annotator labelled the beginning of the vowel of this syllable in Praat (Boersma & Weenink 2018).

Figure 5.1: Example screen from the experiment during a trial with contrastive focus condition.

5.3.1 Tonal onglide

To assess the differences in the F0 contours, the tonal onglide of each nuclear pitch accent was measured. Figure 5.2 provides a schematic depiction of the tonal onglide measure. The tonal onglide characterises the portion of the F0 movement towards the main tonal target of the pitch accent (Ritter & Grice 2015). 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?"). Albeit being a continuous variable that represents both the direction of the pitch movement as well as the magnitude of this movement, the tonal onglide does not capture all relevant details of pitch accents (Grice et al. 2017). Nevertheless, it has been shown that the tonal onglide movement is a perceptually relevant parameter of pitch accents in German (Baumann & Röhr 2015; Ritter & Grice 2015).

F0 movements were annotated by two labellers with training in prosody using a simple labelling scheme without having access to the intended focus structures

of the sentences. First, the labellers identified all utterances in which the speaker did not place the nuclear pitch accent on the object. Only cases which exhibited a clear nuclear accent on the target word for both annotators were labelled as accented. Second, the labellers judged perceptually whether the nuclear pitch accent was falling or rising. Third, the labellers identified the beginning and the end of the onglide movement manually within a window of three syllables including the accented syllable in the centre, the syllable before and the syllable after the accented syllable.

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 labelled 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 labelled in the pre-accented syllable or the accented syllable itself as the beginning of the onglide movement. Since F0 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 was 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 characterised by a low stretch of F0. This situation – deaccentuation of the target word – was found in almost all cases of the background condition and in a minority of cases of the other conditions. When deaccentuation of the target word occurred, an "onglide" measure was carried out with fixed time points (5 ms before the start and 50 ms before the end of the stressed syllable). It is not possible to speak of tonal onglide in a strict sense here since there is no tonal movement of a pitch accent. Hence, there is no beginning and end of a tonal movement that can be identified. 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, like gender effects, normalisation is needed to make the speakers more comparable. To do so, each rising onglide value is divided by the mean of the speaker's rising onglides, and each falling onglide value is divided 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 it is not possible to speak of rises and falls, the overall mean of the absolute onglide values for each speaker was used in the normalisation procedure.

5.3.2 Alignment of the peak

The labels described above for the tonal onglide were used to determine the alignment of the peak. In the case of falling accents, the beginning of the fall was used as the peak; in the case of rising accents, the end of the rise was used as the peak. The alignment of the peak is calculated as the difference between the time point of the peak and the time point of the beginning of the vowel in the accented syllable in ms. No equivalent for deaccented target words could be applied for this measure. Figure 5.3 shows a schema of the measure.

5.3.3 Lip aperture

Lip aperture was evaluated as the Euclidean distance between the lips, as given in Equation 5.1 (Byrd 2000), within the boundaries of the syllable. An automatic procedure was used to determine the maximum of the trajectory within the boundaries of the labelled acoustic syllable. The maximal lip aperture represents the widest opening of the lips during the production of the vowel. Figure 5.4 presents a schema of this measure.

$$
\begin{aligned}
\text{lip aperture } x &= \text{upper lip } x - \text{lower lip } x \\
\text{lip aperture } y &= \text{upper lip } y - \text{lower lip } y \\
\text{lip aperture} &= \sqrt{(\text{lip aperture } x)^2 + (\text{lip aperture } y)^2}
\end{aligned}
\tag{5.1}
$$

5.3.4 Position of the tongue body

The vertical (high-low) and the horizontal (front-back) position of the back-most tongue sensor was measured as an indication of the position of the tongue body. Because the stressed syllable is followed by a syllable containing schwa, the exact target of the horizontal tongue body movement could not consistently be determined for the vowel /o/ in the stressed syllable as schwa can be lower than /o/. Therefore, a different method was applied for all tongue trajectory measures: The first 50% of each acoustic vowel was taken as a time window to measure the

Figure 5.2: Schema of tonal onglide measure.
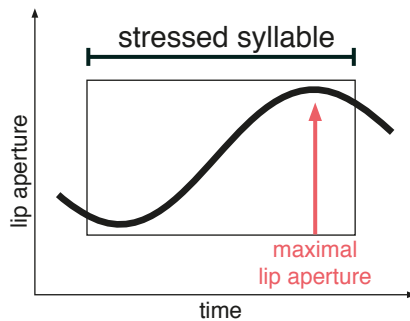
Figure 5.3: Schema of alignment measure.

Figure 5.4: Schema of the lip aperture measure.

tongue body position. The mean of all points on the trajectory falling in this 50% window was calculated for each token. The length of the window (50%) is, of course, arbitrary. However, it seems plausible to assume that the acoustic vowel is a perceptually relevant time interval. It thus makes sense to ask how high or low, back or front the vowel is articulated in this window. Using only the first 50% reduces the influence of the following vowel on the measure. Figure 5.5 shows a schema of the measure for two different trajectories and the potential outcome of the measure (as bar graphs). The measure was used on both the horizontal and the vertical dimension of the movement as well as for both vowels /a/ and /o/ to make all measures comparable.

Figure 5.5: Schema of the tongue body measure.

5.4 Data and availability

Of the 2160 planned utterances (80 utterances per speaker, 27 speakers), 29 (1.3%) productions had to be excluded due to technical problems during the recording or because the participant did not pronounce the words of the sentence correctly (and the trial was not repeated). The total dataset comprised 2131 recordings.

As stated above, the target word is unaccented in the background condition in the majority of cases. Conversely, the target word is accented in the majority of cases in all other conditions (broad focus, narrow focus, and contrastive focus). This situation was expected and planned in the design of the study (see Chapter 4). The opposition background vs. {broad focus, narrow focus, contrastive focus} is used for the comparison of unaccented and accented. Therefore, all tokens with accented target words are excluded from the background condition (0.14% of the productions in the complete corpus of 2131 utterances). Likewise, all tokens with unaccented target words are excluded from the conditions broad focus, narrow focus and contrastive focus (1.88% of the productions in the complete corpus of 2131 utterances). It has to be noted that only those cases were labelled as accented that exhibited a clear nuclear pitch accent on the target word for both annotators. All in all, 2088 utterances enter the analysis, i.e. 96.67% of the 2160 planned utterances. The data set is accessible for download: https://osf.io/4g6s2/.

6 Integrating categorical and continuous aspects of pitch accents

The first analysis of the controlled corpus of prosodic focus marking is concerned with intonation contours, more specifically with nuclear pitch accents. The results outlined in this chapter show that parameters of pitch accents are used in a categorical as well as a continuous manner at the same time. The two kinds of using F0 modulations seem to be applied by speakers to achieve the same – or at least comparable – communicative ends. This striking parallel is implemented in a first sketch of a dynamical model. The model conceptualises pitch accents as attractors in a continuous phase space. In doing so, it is able to capture the categorical nature of pitch accents, continuous variation within one pitch accent category and the generalisation that the categorical and the continuous use of intonation are related. In addition, two subgroups of speakers are identified that exhibit different patterns of intonational focus marking. Despite their differences, the behaviour of both groups can be described by means of the same model.

6.1 Results of F0 measures

As outlined in Chapter 5, the nuclear pitch accents in the corpus are assessed using two continuous parameters: tonal onglide and alignment of the peak. This section outlines the results of the two measures. Since the present chapter deals with nuclear accents only, the background condition is not included in the analyses since the vast majority (99%) of target words in the background condition were unaccented. A treatment of this condition is added to the analysis in the next chapter (Chapter 7).

6.1.1 Tonal onglide

Before turning to the results of the tonal onglide, it is useful to consider some example contours from the corpus. Gaining an impressionistic perspective of the data is particularly helpful in discussing tonal onglide since this measure is far less established in the research literature than the alignment of the peak.

Figure 6.1 presents one instance of broad (top), narrow (middle) and contrastive focus (bottom) each produced by male speaker. The stressed syllable of the target word is marked by the blue box, the arrows illustrate the F0 movement that is captured by the tonal onglide measure. The informative value of these examples is of course limited since they only represent individual utterances. However, accompanying the quantitative results, they help to give a thorough insight into the data. The figure shows that the speaker uses 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. As the following quantitative results show, this first impression is supported by the data set but there is more variation involved.

Figure 6.1: Example intonation contours for broad (top), narrow (middle), and contrastive focus (bottom). The blue box marks the nuclear accented syllable. The red arrows indicate the direction of the tonal onglide.

Figure 6.2 plots the distributions of normalised onglide values of all speakers together for broad focus, narrow focus and contrastive focus. Note that negative onglides indicate a falling tonal movement while positive onglides indicate a rising tonal movement. Broad focus is characterised by an almost symmetrical bimodal distribution, reflecting the fact that both falling and rising accents are equally possible in this focus condition. Moving on to narrow focus, the right mode of the distribution grows, indicating that the majority of accents are rising in this condition. This trend is continued in contrastive focus for which an even higher proportion of rising accents is found. The numbers and proportions of falling and rising onglides are summarised in Table 6.1. In a minority of cases (6 cases, 0.38% of the data), it was not possible to track the F0 at the locations marked by the labellers, rendering the onglide measure impossible to apply (labelled "NA" in the table).

Figure 6.2: Distributions of normalised tonal onglide values for broad, narrow and contrastive focus.

Table 6.1: Proportions of falling and rising onglides.

focus types	falling accents (negative onglides)		rising accents (positive onglides)		NA		all
broad focus	241	(47.07%)	270	(52.73%)	1	(0.2%)	512
narrow focus	115	(21.78%)	411	(77.84%)	2	(0.38%)	528
contrastive focus	47	(9.04%)	470	(90.38%)	3	(0.58%)	520
all	403	(25.83%)	1151	(73.78%)	6	(0.38%)	1560

The distributions reveal that there is no one-to-one mapping of accent type (rising/falling) to focus type, a finding that is in line with the literature. Rather, a

probabilistic mapping of accent type to focus type with a large degree of overlap can be attested. Figure 6.3 provides a closer look at the rising portions of all three distributions. In addition to the increase in the proportion of rising accents, the rises themselves tend to become larger in magnitude.
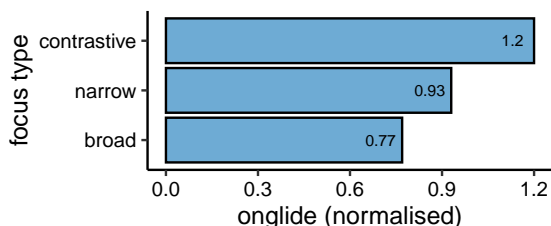

Figure 6.3: Means of rising onglides (normalised) for broad, narrow and contrastive focus.

The results are analysed using a Bayesian linear mixed model in R (R Core Team 2018) with the package brms (Bürkner 2018) which implements an interface to Bayesian inference with MCMC sampling in Stan (Carpenter et al. 2017). The estimated differences between focus conditions in terms of posterior means are reported in addition to 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, the analysis subtracts 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).

Normalised onglide is included as the dependent variable in the model, focus type as a fixed effect, and random intercepts for speakers and target words as well as by-speaker slopes for the effect of focus type. Since the distribution of the dependent variable is bimodal, a prior for the predictor is used that is characterised by a mixture of two Gaussian distributions centred around −0.5 and 0.5 respectively. The model estimates the parameter theta that represents the extent to which the two Gaussian distributions are mixed. This parameter is referred to as the mixing parameter. For this parameter, a prior centred around zero is used. Differences in the mixing parameter indicate differences in the proportions of the two modes in the onglide data. The model runs with four sampling chains of 3,000 iterations each.

The presentation of the results starts with the mixing parameter. Given the model and the data, the analysis yields evidence for differences in the posterior

probabilities for the mixing parameter between broad focus and narrow focus ($\hat{\beta} = 1.32, CI = [0.60, 1.95], Pr(\hat{\beta} > 0) = 1$), narrow focus and contrastive focus ($\hat{\beta} = 1.59, CI = [0.56, 2.65], Pr(\hat{\beta} > 0) = 1$) as well as broad focus and contrastive focus ($\hat{\beta} = 2.91, CI = [1.61, 4.09], Pr(\hat{\beta} > 0) = 1$).

To assess the differences between the focus conditions regarding the rising distributions, the mean estimates of the right Gaussian sub-distribution are investigated. Given the model and the data, the analysis yields evidence for differences in the posterior probabilities for the mixing parameter between broad focus and narrow focus ($\hat{\beta} = 0.14, CI = [0.07, 0.20], Pr(\hat{\beta} > 0) = 1$), narrow focus and contrastive focus ($\hat{\beta} = 0.24, CI = [0.16, 0.32], Pr(\hat{\beta} > 0) = 1$) as well as broad focus and contrastive focus ($\hat{\beta} = 0.37, CI = [0.30, 0.45], Pr(\hat{\beta} > 0) = 1$).

6.1.2 Alignment of the peak

In addition to the scaling of an accent – an aspect that is captured by the tonal onglide – the alignment of the peak has been reported to be an important dimension of pitch accents (Gussenhoven 2004; Ladd 2008; Ladd & Morton 1997: see also Chapter 4). The measure includes not only the peak alignment of rising accents but also the peak alignment of falling accents. This means that the peaks investigated here are of two kinds in terms of an autosegmental-metrical analysis: In rising accents, H* and L+H*, the peak belongs to the starred tone, for falling accents, H+!H* or H+L*, it belongs to the leading tone.

Negative alignment values indicate that the peak is early, i.e. before the onset of the vowel, while positive alignment values characterise a mid or late peak, i.e. after the vowel onset. The latter are referred to as *mid to late peaks* in what follows. Figure 6.4 presents the distributions of alignment values of all speakers together for broad, narrow and contrastive focus. Similar to the distributions of the tonal onglide, the right mode increases in relation to the left mode when going from broad to narrow focus and from narrow to contrastive focus. The numbers and proportions of early and mid to late peaks are summarised in Table 6.2. Note that the numbers of NA data points is different from the tonal onglide as the alignment measure is only based on the time points and not on the calculation of F0 that may fail due to technical issues in some cases.

Figure 6.5[1] presents the means of the positive alignments (mid to late peaks) and shows that in addition to lower proportions of early peaks, the positive alignments increase, i.e. the instances within the group of mid to late peaks become later, although the difference between narrow and contrastive focus appears to be very subtle.

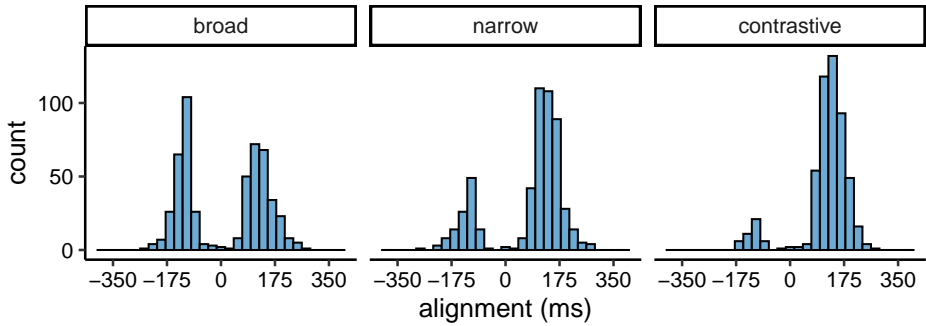[1]Note that the x-axis does not start at zero.

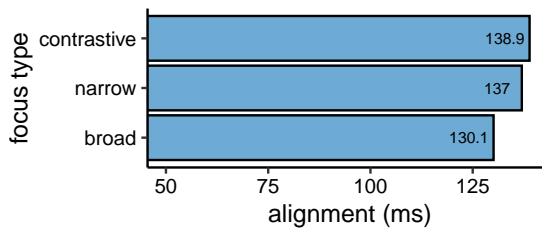Figure 6.4: Distributions of alignment values for broad, narrow and contrastive focus.

Figure 6.5: Means of positive alignment values for broad, narrow and contrastive focus.

Table 6.2: Proportions of early and mid to late alignments

focus types	early peak (negative alignment)		mid to late peak (positive alignment)		NA		all
broad focus	241	(47.07%)	271	(52.93%)	0	(0%)	512
narrow focus	117	(22.16%)	411	(77.84%)	0	(0%)	528
contrastive focus	47	(9.04%)	473	(90.96%)	0	(0%)	520
all	405	(25.96%)	1155	(74.04%)	0	(0%)	1560

A similar Bayesian model to the one employed for the tonal onglide is used. The dependent variable is changed to alignment. For the mixing parameter a bimodal prior centred around −130 and 130, respectively, is used. Everything else is parallel to the statistical modelling of tonal onglide. Again, the presentation of the results starts with the mixing parameter. Given the model and the data, the analysis yields evidence for differences in the posterior probabilities for the mixing parameter between broad focus and narrow focus ($\hat{\beta} = 1.29, CI = [0.56, 1.91]$, $Pr(\hat{\beta} > 0) = 1$), narrow focus and contrastive focus ($\hat{\beta} = 1.56, CI = [0.50, 2.60]$, $Pr(\hat{\beta} > 0) = 0.99$) as well as broad focus and contrastive focus ($\hat{\beta} = 2.85, CI = [1.48, 4.10], Pr(\hat{\beta} > 0) = 1$).

As in the case of tonal onglide, the mean estimates of the right Gaussian subdistribution are investigated to assess the differences between the focus conditions regarding the rising distributions. Given the model and the data, the analysis yields evidence for differences in the posterior probabilities for the mixing parameter between broad focus and contrastive focus ($\hat{\beta} = 7.72, CI = [2.22, 13.50]$, $Pr(\hat{\beta} > 0) = 1$). The evidence is weaker for differences between broad focus and narrow focus ($\hat{\beta} = 5.86, CI = [-0.22, 11.36], Pr(\hat{\beta} > 0) = 0.98$) and between narrow focus and contrastive focus ($\hat{\beta} = 1.86, CI = [-3.56, 6.74], Pr(\hat{\beta} > 0) = 0.75$), confirming the impression given by Figure 6.5 that the peak alignments of narrow and contrastive focus are closer compared to those of broad and contrastive focus.

6.2 Modelling account

The data presented in the last sections reveal two major results: First, distributions of pitch accent types (rising/falling, early/mid to late) are overlapping between focus types but there are clear trends for more rising and more mid to late accents in narrow compared to broad focus, and in contrastive compared to narrow focus. Second, more subtle modifications in the positive portions of both distributions (i.e. rising accents and mid to late accents) take place as well. The magnitudes of the rising onglides increase and the mid to late peaks are aligned later. These findings point towards an intertwining of categorical and continuous aspects of prosodic focus marking. It could be described as a case of a parallel between the phonetics and the phonology of intonation in which one "mimics" the other.

This interpretation motivates a modelling approach that describes both categorical and continuous aspects in the same formal system. Chapter 3 discussed the role of dynamical systems for such a modelling approach. The current section

applies a dynamical model to the description of the pitch accent data outlined above for the dimension of tonal onglide. The alignment of the peak exhibits comparable results to those of the tonal onglide and, hence, the outlined model for tonal onglide will in general be applicable to this parameter as well. Since the distributions of tonal onglide show bimodal patterns, it seems natural to employ a system with two attractors as the one given by the potential $V(x)$ and the force $F(x)$ in Equation 6.1.

This dynamical system is similar to the systems with double-well potentials discussed extensively in Chapter 3. The reader is reminded that the scaling of the control parameter in these systems (e.g. k in Equation 6.1) tilts the potential, the attractor landscape, to one side or the other, lending more stability to one of the attractors. In a stochastic conception of the system, it is more likely that the system settles in the more stable attractor as more noise is needed to push the system out of the "deeper", more stable attractor. When the control parameter is zero, both attractors are equally stable. For all other values, one attractor is more stable than the other. Past critical thresholds of the control parameter on both sides (negative and positive), the system bifurcates and only one attractor remains.

$$V(x) = -kx + \frac{x^4}{4} - x^2$$
$$F(x) = k - x^3 + 2x$$

(6.1)

The emphasis of the approach outlined here is on the qualitative correspondence of the experimental observations and the theoretical model. The coefficients for the model are chosen for presentation purposes, the system does not produce the same scale as the real data. As will become clear in the course of the section, the presented system is able to capture at least the most important aspects of the structure of the data. The choice of two attractors, as stated above, is motivated by the structure of the data. Using a system with two attractors for the model does not imply that there are only two pitch accent categories in German. The GToBI model of German intonation (Grice & Baumann 2002) posits six distinct accent types, although the importance of the differentiation between H+L* and H+!H* has decreased as there is only little evidence for a difference between the two. It is thus not uncommon to treat falling accents as one group. Furthermore, the current data set does not exhibit instances of L* accent types (L*+H and L*).

As to the differentiation between the rising pitch accent types H* and L+H*, it can be stated that neither the positive portion of the alignment data, nor the

positive portion of the tonal onglide data reveal a bimodal shape. Assuming only a single rising accent type is thus data-driven and does not claim that the intonational system of German is structured this way.

The control parameter k which occurs in the functions of Equation 6.1 is used to modulate the attractor landscape such that one attractor becomes more stable and hence more resistant to noise. The attractor landscape is symmetrical with $k = 0$. For broad focus, a value of k slightly above 0 can be used, e.g. $k = 0.1$, tilting the landscape subtly to the right while retaining the characteristics of two nearly symmetrical attractors. For narrow focus, the value of the control parameter k has to increase such that the right attractor, the "rising attractor", becomes more stable. In the present model, k is set to 0.45 in narrow focus. For contrastive focus, the system has to tilt even more to the right side. Hence, a higher number, $k = 0.9$, is chosen. Figure 6.6 plots the graphs of the resulting potential and force functions with this parametrisation of k.

Figure 6.6: Potential and force functions for k values modelling broad, narrow and contrastive focus.

Simulations are run on these attractor landscapes. The methodology of the simulation is as described in Chapter 3: The system is conceptualised as a stochastic system that exhibits random fluctuations. The differential function describing the system (the force function) is solved using small discrete time steps. Noise is added in each of these time steps. The noise in this case is a random number

from a Gaussian distribution with a mean of 0 and standard deviation of 0.03. The simulation is finished after 5,000 time steps and records the solution. This procedure is repeated to obtain 2,500 solutions for each value of k.

Figure 6.7 presents the results of one simulation for each of the k values chosen to model the focus types (broad: 0.1, narrow: 0.45, contrastive: 0.9). The histograms show the characteristic pattern of the real tonal onglide data (see Figure 6.2 for comparison): almost equal numbers of rises and falls in broad focus and increasing numbers of rises in narrow and contrastive focus. In Figure 6.8, the means of the positive sub-distributions of the stimulation results are displayed. This figure also indicates a strong correspondance of the simulated data to the real data (compare to Figure 6.3). It becomes evident that while making the right (rising) attractor more stable, the attractor basin also shifts subtly towards more extreme values in the state space. This accounts for the observed pattern of increasing onglide magnitudes from broad to narrow focus and narrow to contrastive focus in the data. The phenomenon can be observed in Figure 6.9. This graph displays the right attractor basin for broad focus, narrow focus and contrastive focus, i.e. $k = 0.1$, $k = 0.45$ and $k = 0.9$, plotted on top of each other. The vertical lines (solid for $k = 0.1$, dotted for $k = 0.45$, dashed for $k = 0.9$) indicate the location of the attractor that moves to the right, i.e. larger values for rising onglides with the increase in k.

Figure 6.7: Distributions of simulation results for k values modelling broad, narrow and contrastive focus (from left to right).

The modelling account outlined here employs a rather simple dynamical model that posits two attractor in a continuous state space. This state space is made up of only one dimension, namely the tonal onglide. Of course, this is a simplification: Pitch accents are best seen as multi-dimensional, as the alignment measure for instance demonstrated. A full model would include all relevant parameters.
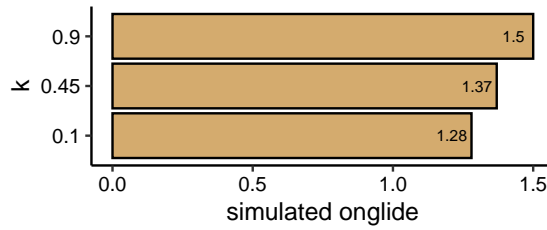
Figure 6.8: Means of positive simulated onglides for *k* values modelling broad (*k* = 0.1), narrow (*k* = 0.45) and contrastive focus (*k* = 0.9).
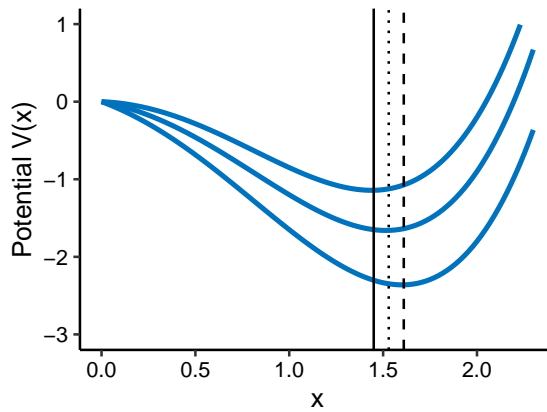

Figure 6.9: Zoom of right (rising) attractor basin. The vertical lines mark the location of the attractor. Solid line: broad focus; dotted line: narrow focus; dashed line: contrastive focus.

However, reducing the model to one dimension makes it more transparent. The present model is best seen as a proof of concept that is able to capture important generalisations in the data about the interplay of categorical and continuous aspects of intonation. The centrepiece of the model is the idea that phonological categories are not separated from their phonetic implementation. The stability relations, i.e. the presence of an attractor, position the categories (falling / rising accents) in the continuum of the state space. The phonetic realisation (magnitude of the movement in terms of onglide) follows from this position on the continuum. The category and the phonetic realisation are thus deeply intertwined and inseparable.

6.3 Speaker groups

With the dynamical model for pitch accents sketched, the question arises as to whether all speakers use the system in the same way. Even with as few subjects as five, Grice et al. (2017) could observe different strategies: One group of speakers used qualitative modifications, i.e. different pitch accent types, to differentiate between focus types, the other group used rising pitch accents only but produced more subtle quantitative variation in the realisation of the pitch accents. To assess these differences in the present data set, speakers were grouped according to their overall pattern of pitch accent productions. Group 1 consists of the 11 speakers who use falling onglides in more than 33% of the cases overall. Group 2 consists of the 16 speakers who use up to 33% falling onglides overall. Group 1 thus pools the speakers that make frequent use of clearly qualitatively different tonal onglide patterns, while the speakers in group 2 only rarely make use of a falling-rising distinction.

The distributions of tonal onglide values for the two groups are given in Figure 6.10. For group 1, the distributions of broad, narrow and contrastive focus are more distinct. In broad focus, falling accents are most frequent; in narrow focus the distribution is dominated by rising accents but there is a considerable number of falling accents; in contrastive focus there is only a small number of falling accents. For group 2, the distributions appear less distinct. Rising onglides are predominantly used in all three focus types, although there is a small number of falling accents in broad focus and an even smaller number of falling accents in narrow focus. Hence, a minimal version of the trend observed for group 1, and the group of all speakers, can be attested here as well, i.e. less falls and more rises from broad to narrow focus and from narrow to contrastive focus.

Figure 6.11 and Figure 6.12 display the means of the rising onglide distributions for the two speaker groups. For both groups, the means increase when going from broad through narrow to contrastive focus, indicating that the magnitudes of rising onglides become larger. In addition to this main trend, the means of group 2 are slightly higher overall than those of group 1.

In terms of the dynamical model outlined above, the behaviour of the speaker groups can be captured by choosing different values for the control parameter k. For group 1, broad focus may be modelled with $k = -0.4$, narrow focus with $k = 0.3$, and contrastive focus with $k = 0.8$. For group 2, broad focus may be modelled with $k = 0.8$, narrow focus with $k = 1.0$, and contrastive focus with $k = 1.2$. Crucially, in all cases, the scaling of the control parameter goes in the same direction: broad focus < narrow focus < contrastive focus. The resulting attractor landscapes are plotted in Figure 6.13 for group 1 (top) and group 2 (bottom), the
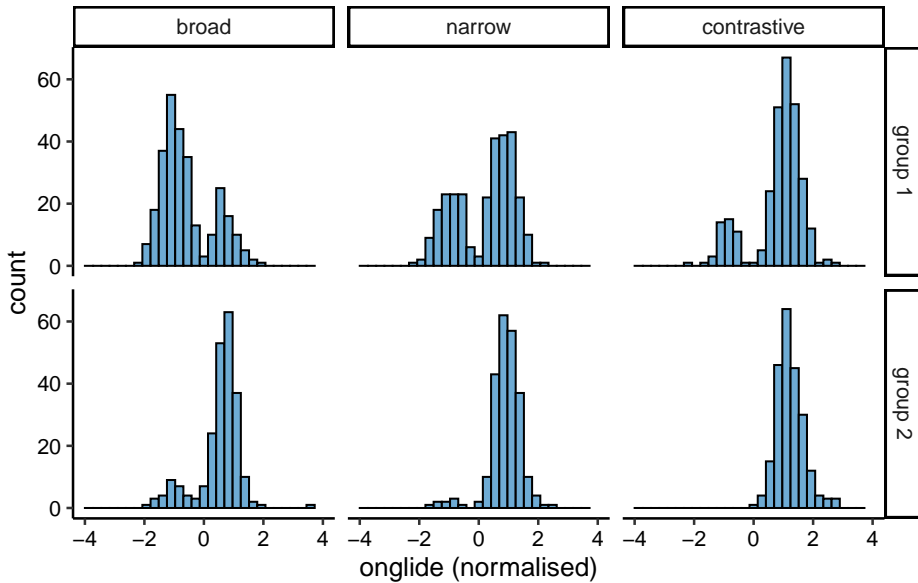
Figure 6.10: Distributions of normalised tonal onglide values for broad, narrow and contrastive focus, separately for the two speaker groups. Top: group 1, bottom: group 2.
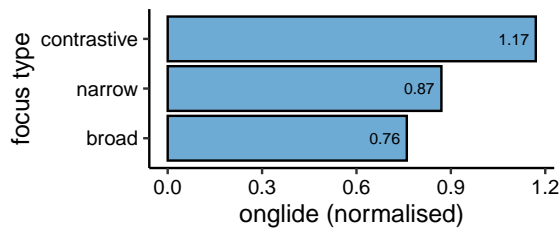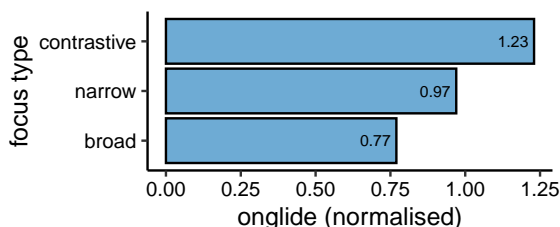

Figure 6.11: Means of rising onglides (normalised) of speaker group 1 for broad, narrow and contrastive focus.

k values chosen for the groups are compared graphically in Figure 6.14. In all cases, the model predicts less falling onglides from broad to narrow focus and from narrow to contrastive focus. The initial proportion of falling and rising accents differs between the group, the direction or mechanism of modification is comparable. In addition, the model predicts increasing magnitudes of the rising onglides and a higher overall level of rises in group 2 compared to group 1. It should be noted that the latter prediction – the overall higher level of rising onglides in group 2 – poses a limitation in the correspondence between the model

Figure 6.12: Means of rising onglides (normalised) of speaker group 2 for broad, narrow and contrastive focus.

Figure 6.13: Potential and force functions for *k* values modelling broad, narrow and contrastive focus, separately for the two speaker groups. Top: group 1, bottom: group 2.

and the real data. As a comparison of Figure 6.11 and Figure 6.12 shows, the means of the rises are higher in group 2, and thus the pattern predicted by the model can be found in the data. The difference between the two groups with regard to the means of the rises is, however, certainly less pronounced than the model predicts it.

The discussion of these differences between speakers raises the question as to how communication is possible if the speakers' behaviour is so different. Since the current analysis includes only production results, no direct evidence for the

Figure 6.14: Control parameter values *k* of the two speaker groups.

effect on perception can be provided here. However, three points appear important to note. First, as pointed out in Chapter 4, the perception results of Cangemi et al. (2015) – in which listeners rated the productions analysed in Grice et al. (2017) and Mücke & Grice (2014) – provide evidence that both speaker types yield comparable scores overall in perception. Some listeners preferred the categorical speakers (comparable to group 1) while other listeners preferred the more subtle speakers (comparable to group 2). Second, it has to be emphasised that tonal onglide only represents one of many dimensions in which pitch accents can vary, or as Cangemi et al. (2015: 143) put it: "Rather than singly necessary and jointly sufficient features for category membership, phonetic cues are better understood as dimensions along which phonological categories cluster, in an individual-specific network of phonological knowledge". Third, in the light of evidence for phonetic accommodation (e.g. Yu et al. 2013; Babel 2009), it may be plausible to assume that listeners are able to "tune in" to a speaker and adjust their perceptual patterns to the production patterns of the speaker. Listeners may be able to quickly learn (or recall) the prosodic patterns of their interlocutor and adjust their mappings of forms and functions.

6.4 Summary

This section provided an analysis of nuclear pitch accents used to mark focus structure in German by 27 native speakers. The parameters investigated were tonal onglide and alignment of the peak. The analysis showed that speakers use categorical as well as continuous modifications to express different focus structures. Crucially, both types of modification appear to go hand in hand. From broad to narrow, and from narrow to contrastive focus, more rising and more

mid to late peak accents (and less falling and early peak accents) are found. In addition, the magnitude of the tonal onglide of the rises and the peak alignment of the mid to late accents increase at the same time. These results support a perspective that emphasises an intertwining of categorical and continuous aspects of prosody and call for an integration of the two in a theoretical treatment. A dynamical model was proposed that does not separate phonological categories from their phonetic implementation and thus provides a first step towards treating both types of modifications as the outcome of the same system. Furthermore, it was shown that speakers differ in how they use the system while maintaining the general direction of scaling of the control parameter.

7 Integrating dimensions of prosody

The analysis of the current corpus of prosodic focus marking so far has concentrated on modulations of parameters related to F0. It was, however, shown in Chapter 4 that modifications of the supra-laryngeal articulatory subsystem play an important role in prosodic prominence as well (Cho 2011; Mücke 2018). This chapter extends the analysis of the corpus to include tongue body and lip movements. In doing so, evidence will be provided that prosodic prominence can be regarded as a multi-dimensional bundle. The chapter also aims to gain a deeper insight into the tonal patterns of prosodic marking of focus by looking at unaccented tokens.

The modelling approach outlined in the previous chapter (Chapter 6) is enriched in a two-fold manner: First, the consequences of accent placement are integrated in the model. This is achieved by proposing a property of the model that results in a bifurcation, i.e. a qualitative change, when the control parameter is changed past a certain threshold. Second, the articulatory dimensions are included in the model, employing the same control parameter for F0 and articulatory patterns.

7.1 Results of F0 measures

The previous chapter (Chapter 6) presented the results of the tonal onglide measure that revealed the categorical and continuous nature of prosodic parameters at the same time. The analysis there was restricted to the nuclear pitch accents. However, focussing on nuclear pitch accents exclusively leaves an important part of the data aside, namely the condition in which the target word is in background. As described in Chapter 5, an equivalent measure of tonal onglide was performed on these unaccented cases using fixed time points (5 ms before the start and 50 ms before the end of the stressed syllable). This measure is employed in what follows to add the missing piece to the picture and to extend the model.

Before turning to the quantitative results, an example of a contour with an early nuclear accent on the direct object is presented to give a better impression of the kind of data that are analysed. As Figure 7.1 illustrates for one production
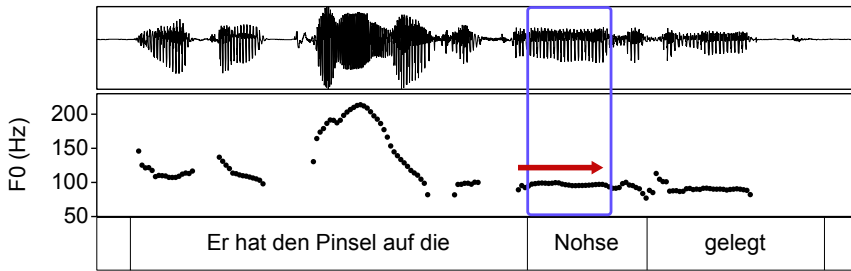
Figure 7.1: Example intonation contour for the background condition. The blue box indicates the lexically stressed (but unaccented) syllable of the target word. The red arrow traces the tonal onglide equivalent.

of a male speaker (the same male speaker as in the contours in 6.1), not placing an accent on the target word yields a flat stretch of F0 on and around the target word.

The fact that the F0 contour is flat on unaccented target words is reflected in the distributions of the tonal onglide displayed in Figure 7.2. The distribution for background is characterised by a single mode slightly below zero. The other three conditions (broad focus, narrow focus, contrastive focus) are as reported in Chapter 6 but are repeated here to be able to compare their distributions to the background condition. Repeated for convenience here are also the means of the rising accents, see Figure 7.3. This figure only includes broad focus, narrow focus and contrastive focus since it is only possible to speak of rises in these cases.

In sum, the full analysis of the tonal onglide reveals the following picture: In the background condition, the data show a single mode located slightly below zero. For broad focus, a bimodal shape of the distribution can be observed, with almost equal numbers of falling and rising onglides. In narrow and contrastive focus, the right mode is more pronounced (more rises in narrow compared to broad focus, and more rises in contrastive compared to narrow focus). In addition to the increase in the number of accents with a rising onglide, the magnitude of the onglides becomes larger, as reflected in the stepwise growth of the mean from broad focus to narrow focus, and from narrow focus to contrastive focus. The next section attempts to capture this pattern with a dynamical description.

7.2 Enriching the tonal onglide model I: accentuation

This section is concerned with the development of a model to account for the results outlined above. In order to be able to exhibit a behaviour that is consistent
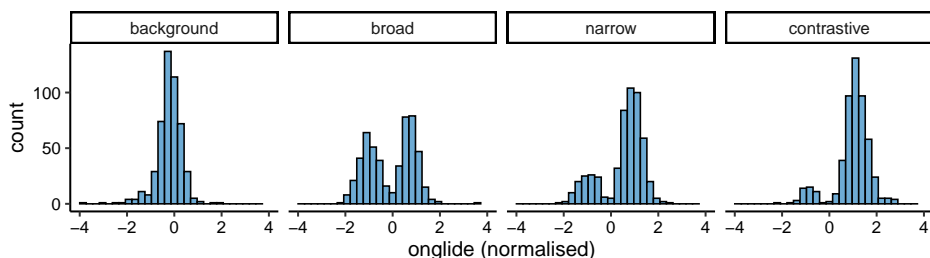
Figure 7.2: Distributions of normalised tonal onglide values for background, broad, narrow and contrastive focus.
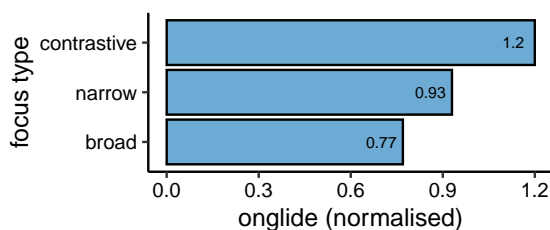

Figure 7.3: Means of rising onglides (normalised) for broad, narrow and contrastive focus.

with the data, the model sought here needs to change from a monostable attractor landscape to a bistable attractor landscape when moving from background to broad focus (first pattern of behaviour). In addition, the attractor landscape needs to be able to tilt to one or the other side giving more stability to one of the attractors (second pattern of behaviour). Both types of changes need to be induced by the scaling of a control parameter.

The first pattern of behaviour can be obtained by a model like the one described by the potential $A(x) = \dfrac{x^4}{4} - kx^2$ as displayed in the upper row of Figure 7.4 for different values of the control parameter k. When k is below zero, the system has only one attractor. When k passes zero, the system's landscape develops into a landscape with two attractors. The second pattern of behaviour is exhibited by a model like the one described by the potential $B(x) = \dfrac{x^4}{4} - x^2 - kx$. This is a version of the typical double-well potential that was discussed several times in the course of this book and is also used in the modelling of the tonal onglide in the previous chapter (Chapter 6). This landscape tilts from left to right when k is scaled from negative to positive values.
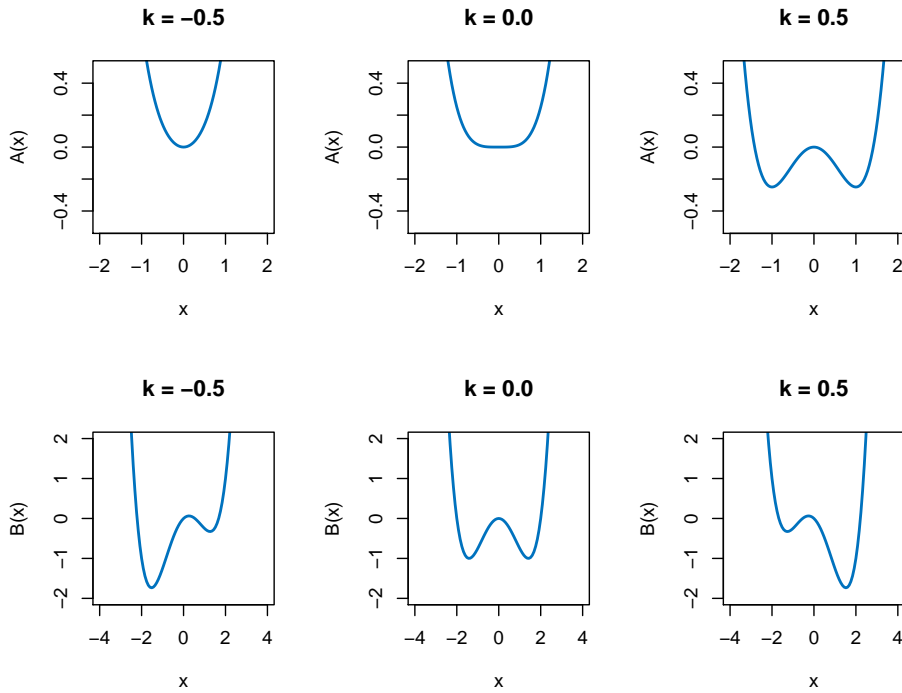
Figure 7.4: Example models $A(x)$ (top) and $B(x)$ (bottom) for different control parameter k values.

A possible model that ties these two patterns together can be formulated as given in Equation 7.1 by the potential $V(x)$ and the corresponding force $F(x)$. Choosing parameter values for background, broad focus, narrow focus and contrastive focus, one can observe the change in the system as shown in Figure 7.5. The attractor landscape is monostable for $k = -0.5$ (background) but is bistable for $k = 2.1$ (broad focus). Scaling the control parameter further tilts the landscapes to the right side (i.e. it stabilises the rising attractor) for $k = 2.5$ (narrow focus) and $k = 3.0$ (contrastive focus). The jump from background to broad focus appears large. It has to be borne in mind, however, that the system has to go through a stage where the falling accents dominate the system's outcome. In this stage, the left attractor has to be more stable, e.g. to model the results of speaker group 1 (see Chapter 6).

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

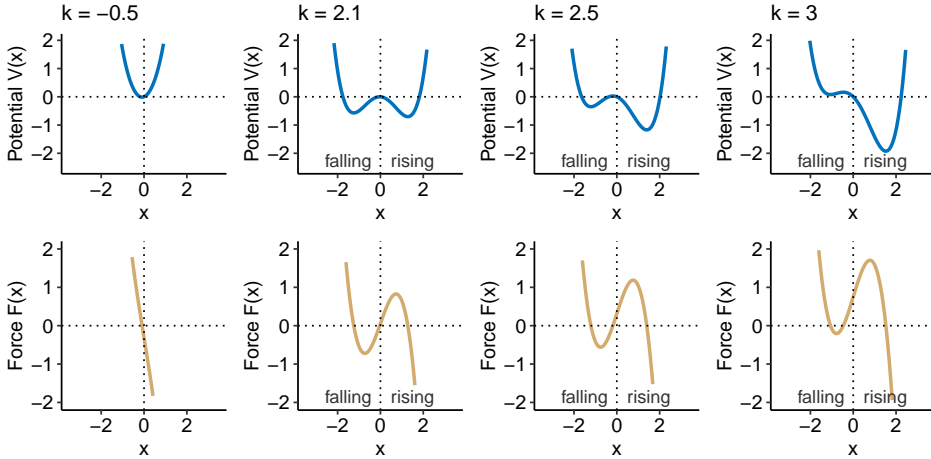$$F(x) = -x^3 + 2(1 - e^{\frac{1}{2}-k})x + \frac{|k|(k-2)}{4}$$

(7.1)

Figure 7.5: Potential and force functions for *k* values modelling background ($k = -0.5$), broad ($k = 2.1$), narrow ($k = 2.5$) and contrastive focus ($k = 3$).

Using a simulation with the addition of noise, the outcome of the model as a stochastic system can be observed; see Figure 7.6 (distributions) and Figure 7.7 (means for accented focus types). The simulation works as described before in Chapter 6. The differential function describing the system (the force function) is solved using small discrete time steps in each of which noise is added in the form of a random number from a Gaussian distribution with a mean of 0 and standard deviation of 0.03. The simulation is finished after 5,000 time steps and records the solution. In total, 2500 solutions are recorded.

The same characteristic pattern as in the results for the tonal onglide can be attested: The system produces a unimodal distribution slightly below zero for background. The bimodal distribution of broad focus is nearly symmetrical. In narrow and contrastive focus, the right mode (rising) becomes increasingly strong. The mean values of the rising (positive) distributions also show essentially the same stepwise increase for the "accented" focus types (broad, narrow and contrastive focus). This result reveals that the rising (right) attractor moves toward more
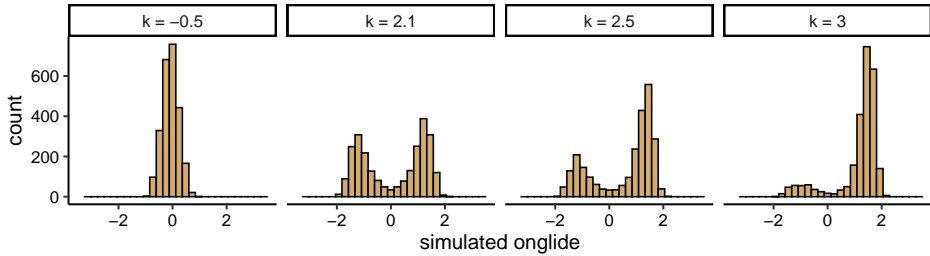
Figure 7.6: Distributions of simulation results for k values modelling background ($k = -0.5$), broad ($k = 2.1$), narrow ($k = 2.5$) and contrastive focus ($k = 3$).
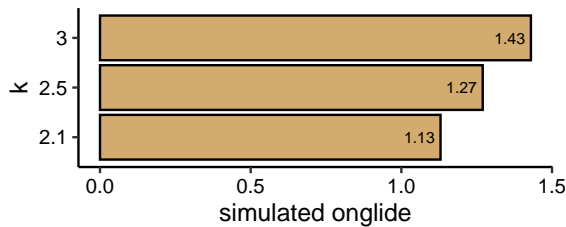

Figure 7.7: Means of positive simulated onglides for k values modelling broad ($k = 2.1$), narrow ($k = 2.5$) and contrastive focus ($k = 3$).

extreme values when the control parameter value is increased and the attractor landscape tilts to the right side.

The proposed model captures the main features of the tonal onglide data and produces a close qualitative correspondence to the empirical data. The change from background to broad focus is modelled with a bifurcation from monostable to bistable. The modifications from broad to narrow focus, and from narrow to contrastive focus are modelled with the tilt of the bistable landscape. Crucially, all this happens as the consequence of scaling the same parameter.

7.3 Results of articulatory measures

F0 has been shown to be a very important parameter in signalling prosodic prominence and modulating the signal in order to encode pragmatic meanings. However, the literature has accumulated numerous examples of modifications of the supra-laryngeal articulation going hand in hand with prosodic structure (for overviews see Cho 2011; Mücke 2018). This section is concerned with the exploration of lip and tongue movements in the data set. The data set does not contain

missing values (NA) for the articulatory parameters. Hence, of the 2088 rows of the complete data set, all data points can be used. One half (1044 data points) contains data for the target words with vowel /a/, the other half contains data for the target words with /o/. All values of the articulatory measures presented here are z-scored for speaker and vowel.

Figure 7.8 presents the distributions and means of the maximal lip aperture during the opening of the vowel in the stressed syllable for both /a/ and /o/ and the four focus conditions. Although – as an inspection of the distributions reveals – the modifications are rather subtle, there is a systematic trend for the lips to be opened wider in broad focus compared to background, in narrow focus compared to broad focus, and in contrastive focus compared to narrow focus, or put more succinctly: background < broad focus < narrow focus < contrastive focus. This trend appears to be more pronounced in the vowel /a/ compared to /o/.

In light of the literature on prosodic strengthening, the opening of the lips to convey higher levels of prominence is interpreted as *sonority expansion.* Employing this strategy the speaker aims to increase the acoustic energy radiating from the mouth (Beckman et al. 1992). In the case of the low vowel /a/, the higher degree of lip opening could also be a concomitant of a lower jaw position. The lower jaw position would allow for a lower tongue position to be employed as *localised hyperarticulation* of the low vowel (de Jong 1995). In either case, it is important to note that the prosody-induced modification of the articulatory gestures that are responsible for lip aperture does not only occur between unaccented and accented (i.e. background vs. broad focus) but also within the group of focus types that all share the same nuclear pitch accent location (i.e. the accent remains on the target word in broad, narrow and contrastive focus).

Figure 7.9 illustrates the results of the vertical tongue body position, extracted using the 50% window method, in distributions and means for both vowels and all focus types. As in the case of lip aperture, the differences between focus types seem to be subtle but systematic: The tongue body position is lowered in both /a/ and /o/ from background to broad focus, from broad focus to narrow focus, and from narrow focus to contrastive focus: background > broad focus > narrow focus > contrastive focus.[1]

For the low vowel /a/, this trend can again be identified as localised hyperarticulation and sonority expansion at the same time. On the one hand, a lower tongue body strengthens the identity of the low vowel. On the other hand, a lower jaw, used to let more acoustic energy radiate from the mouth, may also lead to a lower tongue body position. For the mid vowel /o/, the interpretation

---

[1]The direction is reversed since the tongue body is lowered and the values therefore decrease.
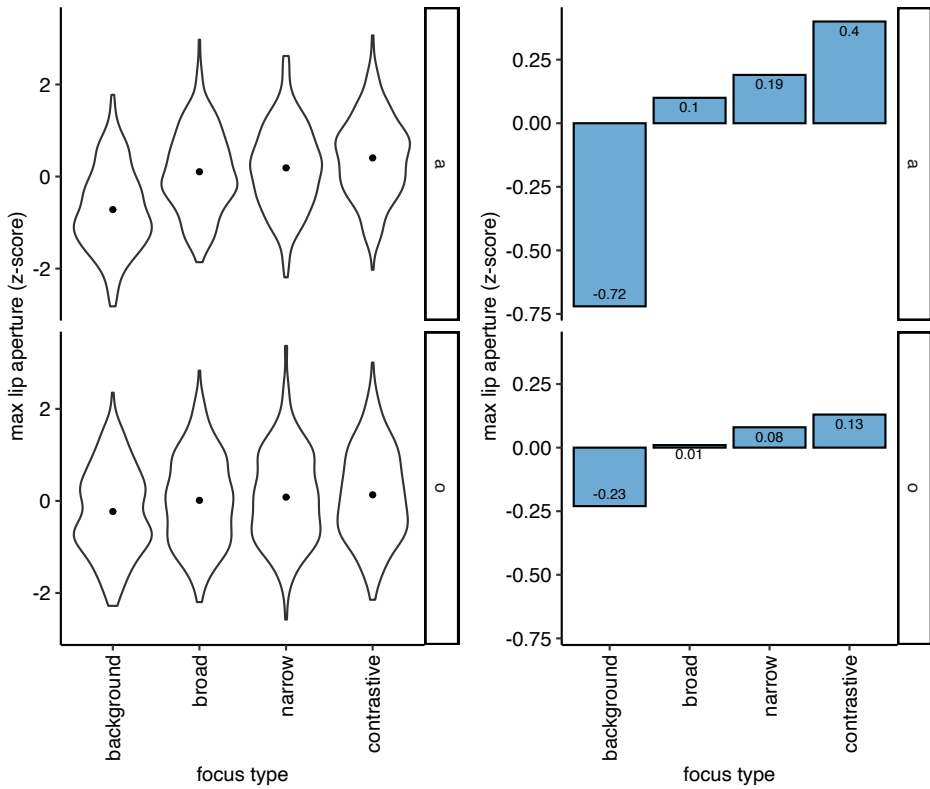
Figure 7.8: Distributions (left) and mean values (right) of the maximal lip aperture for /a/ (top) and /o/ (bottom). The black dots in the distributions on the left indicate the position of the means for comparison with the right panel.

as sonority expansion seems more plausible. Again, the modifications do not exclusively take place under the influence of accentuation (i.e. from background to broad) but also within the group of accented target words to mark different focus types (broad → narrow → contrastive focus).

Finally, Figure 7.10 shows the distributions and means for the horizontal tongue body position for both vowels and all focus types, extracted using the 50% window method. The results for /a/ do not show the clear trend found in the other parameters. In addition, the differences seem to be rather small. For the vowel /a/, hence, no conclusive results can be obtained for the horizontal tongue position. In contrast, for the vowel /o/, the trend observed in the other parameters is present. The tongue body is retracted from background to broad focus, from
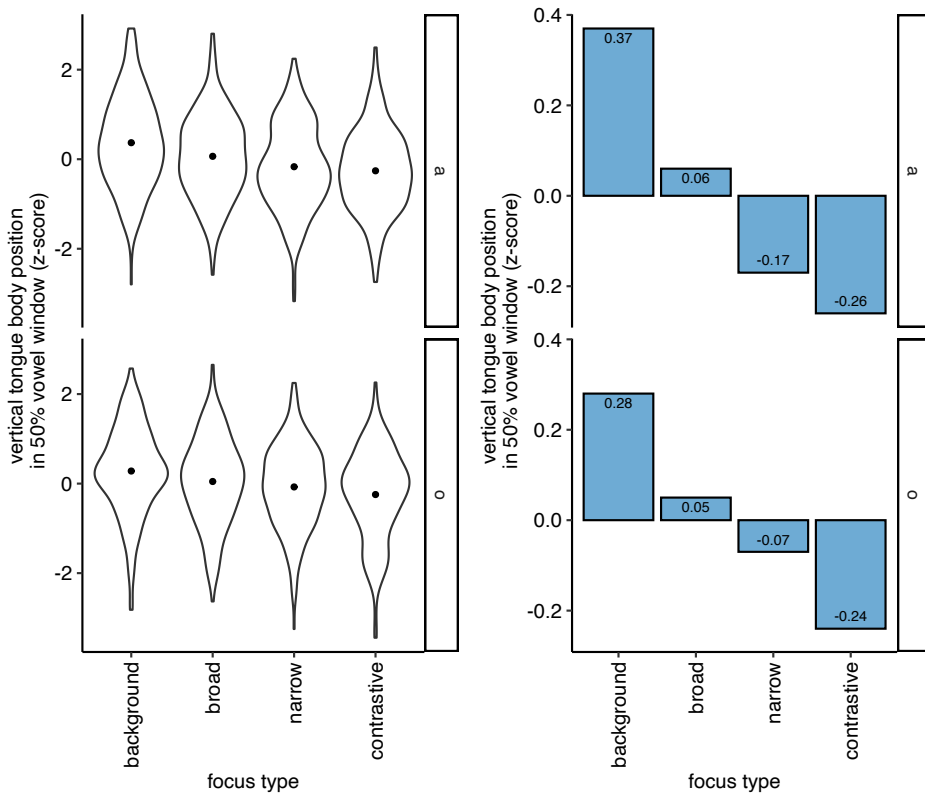
Figure 7.9: Distributions (left) and mean values (right) of the vertical tongue body position for /a/ (top) and /o/ (bottom). The black dots in the distributions on the left indicate the position of the means for comparison with the right panel.

broad focus to narrow focus and from narrow focus to contrastive focus. These results are indicative for localised hyperarticulation of the back vowel /o/.

Analogously to the tonal onglide analysis in the previous chapter (Chapter 6), the results are analysed using Bayesian linear mixed models in R (R Core Team 2018) with the package brms (Bürkner 2018). The estimated differences between focus conditions in terms of posterior means are reported in addition to 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, the analysis subtracts the posterior samples for background from broad focus (broad−background), broad focus from narrow focus (narrow−broad), nar-
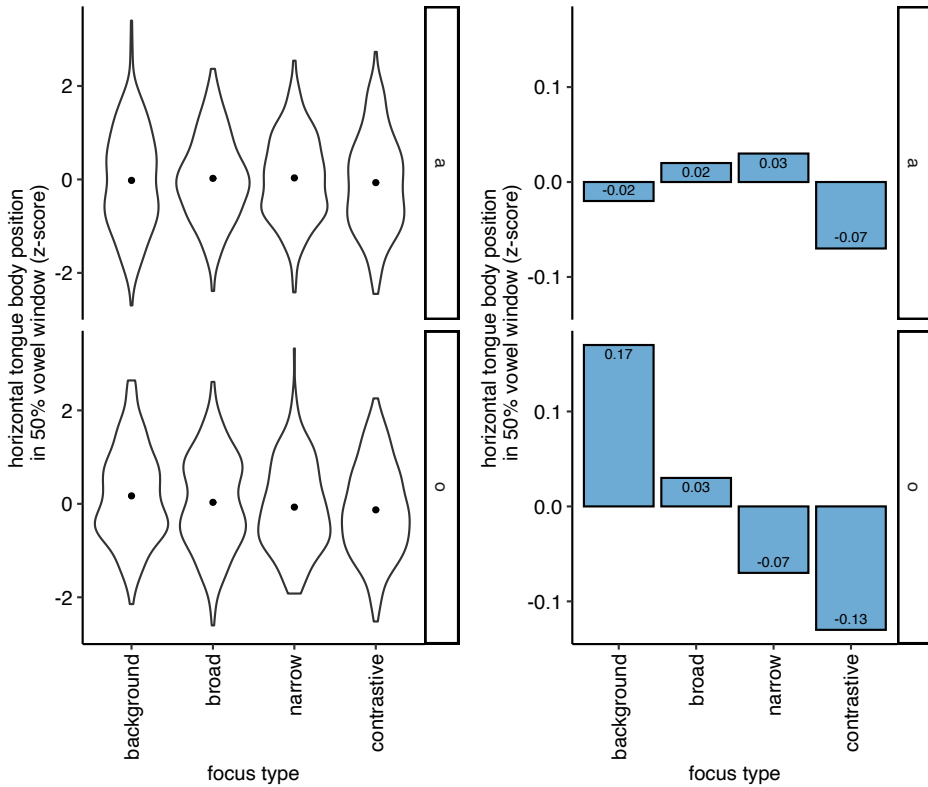
Figure 7.10: Distributions (left) and mean values (right) of the horizontal tongue body position for /a/ (top) and /o/ (bottom). The black dots in the distributions on the left indicate the position of the means for comparison with the right panel.

row focus from contrastive focus (contrastive–narrow), and broad focus from contrastive focus (contrastive–broad).

In the case of the maximal lip aperture, the probability of the estimate being greater than zero is reported because it is interesting whether the lip aperture increases from one focus type to another. In the case of the tongue position measures, the probability of the difference being smaller than zero is reported, because it is interesting whether the tongue position is lower or more retracted, i.e. whether the values decrease, from one focus type to another. All models are run separately for each vowel (/a/ or /o/).

The models include one z-scored articulatory parameter each (maximal lip aperture, vertical tongue position, horizontal tongue position) as the dependent variable. In all 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. The priors are centred around zero, the models run with four sampling chains of 3,000 iterations each.

The presentation starts with the modelling results for the maximal lip aperture for the vowel /a/. Given the model and the data, the analysis yields a clear difference in the posterior probabilities between background and broad focus ($\hat{\beta} = 0.81, \text{CI} = [0.65, 0.97], Pr(\hat{\beta} > 0) = 1$) and broad and contrastive focus ($\hat{\beta} = 0.30, \text{CI} = [0.12, 0.49], Pr(\hat{\beta} > 0) = 1$), i.e. across accentuation (background vs. broad) as well as within accentuation (broad vs. contrastive focus). There is also clear evidence for differences between narrow and contrastive focus ($\hat{\beta} = 0.21, \text{CI} = [0.02, 0.40], Pr(\hat{\beta} > 0) = 0.99$), while the evidence is not as strong for the opposition of broad vs. narrow focus ($\hat{\beta} = 0.09, \text{CI} = [-0.08, 0.26], Pr(\hat{\beta} > 0) = 0.84$).

For the vowel /o/, the model shows evidence for differences between background and broad ($\hat{\beta} = 0.24, \text{CI} = [0.08, 0.39], Pr(\hat{\beta} > 0) = 1$). There is weaker evidence for a difference between broad and contrastive ($\hat{\beta} = 0.12, \text{CI} = [-0.05, 0.30], Pr(\hat{\beta} > 0) = 0.92$). Narrow focus seems to pattern with contrastive focus, but there is no clear evidence for narrow focus to be differentiated from broad ($\hat{\beta} = 0.07, \text{CI} = [-0.11, 0.23], Pr(\hat{\beta} > 0) = 0.79$) or contrastive ($\hat{\beta} = 0.05, \text{CI} = [-0.12, 0.23], Pr(\hat{\beta} > 0) = 0.73$).

To sum up the results of the lip aperture: Modifications of the degree of lip opening are attested from unaccented to accented (background < broad focus) and also within the group of accented focus types (broad < contrastive focus). The modifications always go in the same direction, the lips are opened wider. However, the evidence is stronger for the vowel /a/ than for /o/. In addition, the differentiation of narrow from broad and contrastive focus is not as clear as the differentiation of broad from contrastive focus. Narrow seems to overlap with both "neighbouring" focus types.

For the vertical tongue body position in the vowel /a/, the model provides the following results: There is clear evidence for a differentiation of background from broad focus ($\hat{\beta} = -0.30, \text{CI} = [-0.48, -0.11], Pr(\hat{\beta} < 0) = 1$) and of broad focus from contrastive focus ($\hat{\beta} = -0.32, \text{CI} = [-0.53, -0.11], Pr(\hat{\beta} < 0) = 1$). There is also evidence for a difference between broad and narrow focus ($\hat{\beta} = -0.23, \text{CI} = [-0.46, 0.02], Pr(\hat{\beta} < 0) = 0.98$) while narrow focus seems to overlap more with contrastive focus and there is no clear difference between these two conditions ($\hat{\beta} = -0.09, \text{CI} = [-0.32, 0.14], Pr(\hat{\beta} < 0) = 0.79$).

For the vowel /o/, the analysis of the vertical tongue body position reveals clear evidence for differences between background and broad focus ($\hat{\beta} = -0.23, \text{CI} = [-0.41, -0.04], Pr(\hat{\beta} < 0) = 0.99$) and broad and contrastive focus ($\hat{\beta}=-0.29, \text{CI} =$

$[-0.51, -0.09], Pr(\hat{\beta} < 0) = 0.99)$ – similar to the vowel /a/. The evidence for a difference between narrow focus and contrastive focus is weaker ($\hat{\beta} = -0.17, \text{CI} = [-0.37, 0.04], Pr(\hat{\beta} < 0) = 0.95$). Broad and narrow focus seem to overlap to a higher degree ($\hat{\beta} = -0.12, \text{CI} = [-0.32, 0.09], Pr(\hat{\beta} < 0) = 0.88$).

To sum up the results of the vertical tongue body position: As with lip aperture, there is evidence for modifications of the tongue body position from unaccented to accented (background > broad focus) and within the group of accented focus types (broad > contrastive). Again, all modifications go in the same direction, the tongue body is lowered. Similar to the case of lip aperture, the differentiation of broad from contrastive focus is clear while narrow focus seems to overlap with both.

For the horizontal tongue body position in the vowel /a/, the analysis provides a less clear picture (a result that is congruent with the descriptive statistics). All comparisons yield no or only weak differences and the direction of modification is not systematic: background vs. broad focus $\hat{\beta} = 0.04, \text{CI} = [-0.13, 0.22], Pr(\hat{\beta} < 0) = 0.32$; broad vs. narrow focus $\hat{\beta} = 0.02, \text{CI} = [-0.17, 0.20], Pr(\hat{\beta} < 0) = 0.43$; narrow vs. contrastive focus $\hat{\beta} = -0.10, \text{CI} = [-0.29, 0.08], Pr(\hat{\beta} < 0) = 0.85$; broad vs. contrastive focus $\hat{\beta} = -0.08, \text{CI} = [-0.28, 0.12], Pr(\hat{\beta} < 0) = 0.79$.

The picture is again clearer for the horizontal tongue body position in the vowel /o/, although the evidence for differences is overall not as strong as for lip aperture and vertical tongue body position. There is rather strong evidence that background is differentiated from broad focus ($\hat{\beta} = -0.14, \text{CI} = [-0.30, 0.01], Pr(\hat{\beta} < 0) = 0.96$) and that broad focus is differentiated from contrastive focus ($\hat{\beta} = -0.17, \text{CI} = [-0.33, 0.00], Pr(\hat{\beta} < 0) = 0.98$). The difference between broad and narrow focus ($\hat{\beta} = -0.11, \text{CI} = [-0.28, 0.06], Pr(\hat{\beta} < 0) = 0.90$) and between narrow and contrastive focus ($\hat{\beta} = -0.06, \text{CI} = [-0.22, 0.11], Pr(\hat{\beta} < 0) = 0.75$) is not as strong.

To sum up the results of the horizontal tongue body position: For the vowel /a/, the differences do not appear to be systematic. This comes as no surprise since the vowel /a/ in German is usually associated to a central position in the horizontal (front-back) domain. For the vowel /o/, there is evidence for systematic differences in the form of retraction of the tongue between unaccented and accented (background > broad focus) and within the group of accented focus types (broad > contrastive focus). Overall, the evidence is not as strong as for the other parameters. In addition, the prosodic marking of narrow focus seems to overlap with the marking of both broad focus and contrastive focus.

7.4 Enriching the tonal onglide model II: articulation

The articulatory data reveal the following important results: First, lip aperture is increased continuously from background to contrastive focus with broad focus and narrow focus positioned in between the two. Second, the reverse is true for the vertical tongue body position, indicating a continuous lowering of the tongue in both /a/ and /o/ from background to contrastive focus with intermediate positions for broad and narrow focus. Third, for /o/, also the horizontal tongue body position seems to follow this pattern with a continuous retraction of the tongue from background to contrastive focus (and intermediate positions for broad and narrow focus). Hence, the modifications do not only apply to accented vs. unaccented but can also be observed in the group of focus types with the nuclear pitch accent on the same word (broad focus, narrow focus, contrastive focus). These results are (at least partially) in line with Mücke & Grice (2014) and point towards the importance of subtle continuous modulations of the supra-laryngeal articulation to enhance prominence.

In the light of this finding, it is worthwhile to take a short look at how it relates to the widespread view of prosodic prominence as a characteristic of a hierarchically organised structure. As outlined in Chapter 4, different hierarchies of prosodic structure have been proposed in the literature (Nespor & Vogel 1986; Pierrehumbert & Beckman 1988; Hayes 1989; Selkirk 1996; Shattuck-Hufnagel & Turk 1996). All proposals share the assumption that utterances can be decomposed into hierarchically organised constituents with a minimal structure 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.

Within this framework, one approach is to assume that the levels in the hierarchy are headed by prominences (Beckman & Edwards 1994; Shattuck-Hufnagel & Turk 1996). For example, a nuclear pitch accented syllable is the head of an intermediate phrase. This theory would interpret the modifications of supra-laryngeal articulatory gestures in the target word's stressed vowel as a correlate of the reorganisation in the prosodic prominence structure from background to broad focus. The nucleus is placed on the target word and hence the head status is moved from the stressed syllable of the direct object to the stressed syllable of the target word.

However, the findings of the current analysis go beyond what is conceptualised as a reorganisation 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 looking at the
focus types with the nuclear pitch accent in the same position, i.e. the same as-
signment of the head status, an additional increase in prominence can be ob-
served. Moreover, the results reveal a deep intertwining of the use of tonal and
articulatory cues to prosodic prominence. The modifications in articulatory ef-
fort are correlated with a higher probability of rising accents, and larger tonal
onglides. Therefore, prosodic prominence is best seen as a multi-dimensional
bundle of cues.

The model of tonal onglide as proposed above can be seen as picturing one
dimension of prosodic prominence. It is plausible to assume that there are more
dimensions than onglide in the tonal domain (in fact, the results in Chapter 6
demonstrated this for peak alignment) and more dimensions in supra-laryngeal
articulation than the ones analysed here. The results presented in the current
chapter concentrate on a subset of phonetic dimensions that play a role for pros-
odic prominence. In what follows, an extension of the model is sketched that can
be considered as a proof of concept to demonstrate how we can think of prosody
in a dynamical systems framework.

Equation 7.2 adds a dimension for lip aperture to the model. This dimension
exhibits a different shape and behaviour than tonal onglide. Since the distribu-
tion of lip aperture is uni-modal, only one attractor is assumed. When scaling
the control parameter, the attractor moves towards more extreme values on the
dimension of lip aperture, yielding higher degrees of lip opening. The resulting
attractor landscape is visualised in Figure 7.11 in a two-dimensional phase space
with tonal onglide on one axis and lip aperture on another. In background, when
$k = -0.5$, there is a single attractor basin defined by the two dimensions. When
k is scaled to 2.1 (broad focus), the combined landscape goes through a bifur-
cation and develops into having two basins. On the dimension of tonal onglide,
this means that there can be falling *and* rising accents. On the dimension of lip
aperture, the basins are shifted "forward", i.e. in the direction of higher values.
This change in the parameter k models what happens when the nuclear accent
is placed on the target word.

From broad focus to narrow focus and from narrow focus to contrastive fo-
cus, the parameter k is scaled further. The scaling, however, does not lead to a
qualitative change. Rather, the attractor landscape tilts to produce more rising
accents and also moves towards higher lip aperture values. This change can best
be observed in Figure 7.12. This figure shows the same landscapes as Figure 7.11
for $k = 2.1$, $k = 2.5$ and $k = 3.0$ zoomed in to highlight the differences.

$$V(x, y) = \frac{x^4}{4} - (1 - e^{\frac{1}{2}-k})x^2 - \frac{|k|(k-2)x}{4} + \frac{(y-k)^2}{2} \tag{7.2}$$
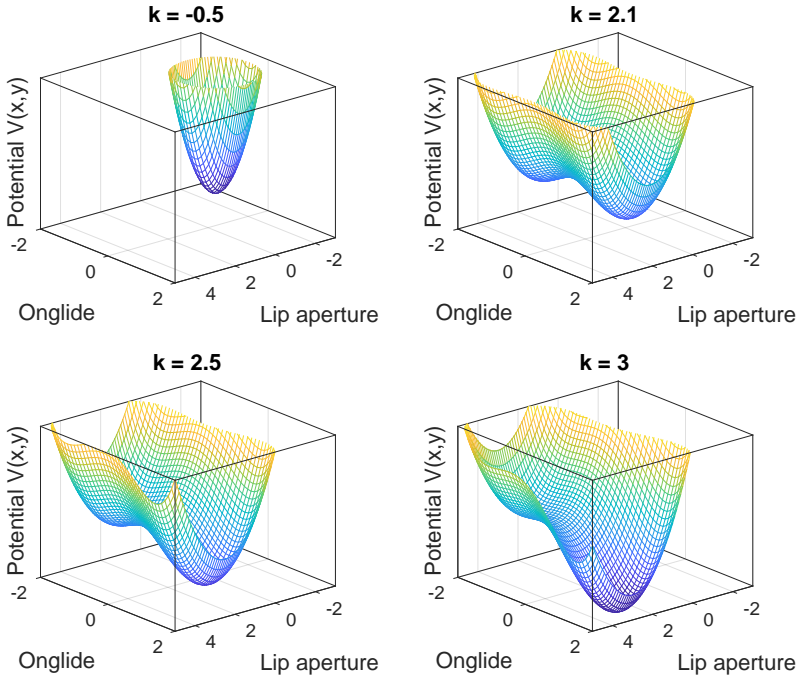
Figure 7.11: Attractor landscapes defined by the dimensions tonal onglide and lip aperture for k values modelling background (top left), broad (top right), narrow (bottom left) and contrastive focus (bottom right).

The probability density function of this non-deterministic dynamical system can be found as a stationary solution to the Fokker-Planck equation for the system (Haken 1977; Gafos & Benus 2006). In Figure 7.13 and Figure 7.14, the graphs of probability functions are given for the system with two dimensions from two perspectives. They reveal the pattern described above for the attractor landscapes with a change from a single peak to two almost equal peaks from $k = -0.5$ to $k = 2.1$ (background to broad focus), and a strengthening of the right peak from $k = 2.1$ to $k = 2.5$ (broad focus to narrow focus) and from $k = 2.5$ to $k = 3.0$ (narrow focus to contrastive focus).

The outlined model can be extended by adding an unrestricted number of dimensions, for example the vertical and horizontal tongue body position. Their shape may resemble the shape of the lip aperture dimension. Since the modifications in these parameters go in the opposite direction, the part added to the potential would be minimally different (i.e. $\frac{(z+k)^2}{2}$ instead of $\frac{(z-k)^2}{2}$, where z de-
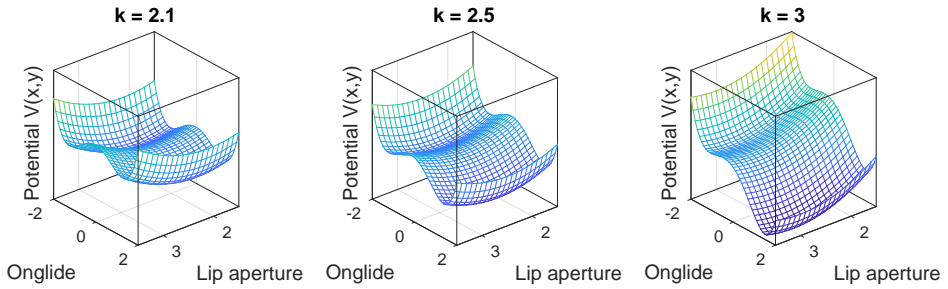
Figure 7.12: Attractor landscapes defined by the dimensions tonal onglide and lip aperture for *k* values modelling broad (left), narrow (centre) and contrastive focus (right).
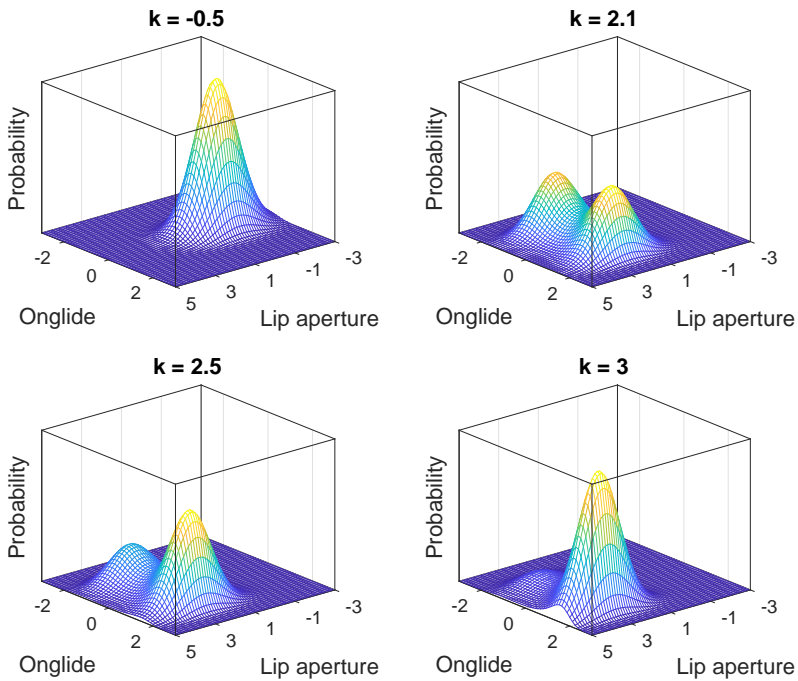

Figure 7.13: Probability density functions for the system with the *k* values modelling background (top left), broad (top right), narrow (bottom left) and contrastive focus (bottom right) – perspective 1.
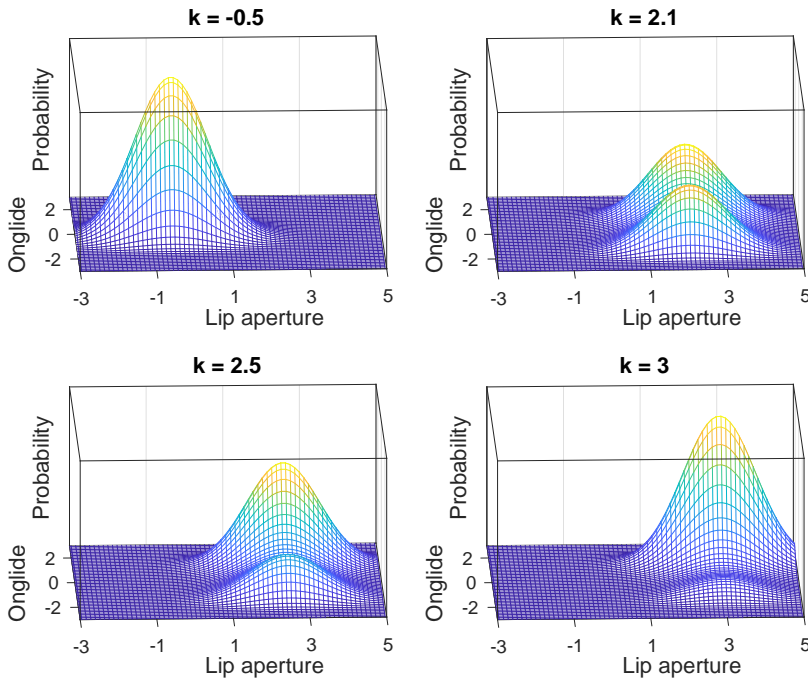
Figure 7.14: Probability density functions for the system with the *k* values modelling background (top left), broad (top right), narrow (bottom left) and contrastive focus (bottom right) – perspective 2.

notes the state of the tongue position). Most of the results presented above make it plausible to think of these dimensions as being modulated by the same control parameter. The horizontal tongue position of the vowel /a/, however, does not fit in this picture.

7.5 Summary

This chapter has dealt with the completion of the picture drawn by Chapter 6 on prosodic marking of focus in German. Tonal and articulatory data of 27 native speakers were presented. The analysis showed that the background condition is characterised by a stretch of flat F0 on and around the target word since no accent is placed on the target word in this condition. The tonal onglide model was extended to be able to capture the change from these flat F0 stretches to falling and rising accents as the result of a bifurcation in the system.

Furthermore, the examination of the articulatory data revealed that lip aperture and tongue body position are modulated as a means of increasing prosodic prominence continuously through the focus types. This implies that the kinematic parameters are modified from unaccented to accented (background to broad focus) but also from broad to narrow focus and from narrow to contrastive focus. The latter finding is important because it shows that prosody-induced modifications go beyond the categorical notion of accentuation and are used to signal prosodic prominence directly. Emphasising the importance of a multidimensional perspective on prosody, a model was sketched that ties the tonal and articulatory dimensions together (and is open to extension to an unrestricted number of dimensions). A key feature of this model is that the dimensions are modulated by the same control parameter.

To conclude, this chapter concentrated on the idea of integration in a twofold manner: First, a full model of tonal onglide modifications was proposed that is able to capture categorical (accentuation *and* accent types) and continuous aspects of intonation. Second, multiple dimensions were tied together with a joint control parameter in a dynamical approach.

8 General discussion

The present work provided a dynamical systems approach to contribute to an integration of categorical and continuous aspects of prosody. In a multi-dimensional account that incorporates various facets of prosodic prominence, this approach fuses intonation and articulatory modulations of prosody within a joint system. The model was developed on the basis of a large EMA corpus of recordings of 27 speakers of German allowing for an extensive analysis of the tonal and articulatory patterns of marking different focus structures. The results showed that prosodic prominence entails a symbiosis of categorical and continuous, as well as tonal and articulatory adjustments and that speakers use the bundle of cues to prosodic prominence flexibly, yet systematically at the same time. The present work thereby participates in a long-standing debate that revolves around categorical and continuous phenomena in speech, and the tension between symbolic and continuous descriptions.

8.1 Summary of the results and modelling approach

The main findings of the study with regard to F0 can be summarised as follows:

1. The data of prosodic focus marking reveal categorical and continuous modulations at the same time.

2. The first, most obvious – and of course unsurprising – categorical modulation is the placement of a nuclear pitch accent on the target word from background to broad focus.

3. In addition, speakers use roughly equal numbers of falling / early and rising / mid to late pitch accents in broad focus but increase the number of rising / mid to late accents in narrow focus. From narrow to contrastive focus, the number of rising / mid to late accents is increased even further.

4. In addition, the magnitude of the rises, assessed here as the quantity in terms of tonal onglide, is increased and the alignment of the peak is delayed from broad to narrow focus and from narrow focus to contrastive focus.

These results are in line with the observation of Bolinger (1961) or Ladd (2014) that categorical and continuous modulations in prosody are often hard to disentangle. The data of the present study show that the two types of modulation are used in symbiosis: A higher frequency of rises goes hand in hand with an increase in the magnitude of the rises. Likewise, a higher frequency of mid to late accents goes hand in hand with a later peak alignment of these accents. The analysis is congruent with the general notion of prosody as a "half-tamed savage" (Bolinger 1978; Gussenhoven 2004). It is, however, questionable that it is possible to draw a "sharp dividing line" (Gussenhoven 2004: 49) between the tamed and the untamed half, as envisioned by many phonological models of prosody. Rather, the symbiosis of the categorical and continuous aspects underline Bolinger's (1978: 475) claim that "to understand the tamed or linguistically harnessed half [...] one has to make friends with the wild half".

The present data reveal that a great deal of fuzziness is involved in the prosodic modulations used by speakers to mark focus. There is no one-to-one mapping between focus types and pitch accent types. Rather, overlapping distributions are found, both in the categorical domain as well as in the continuous domain. For example, broad focus may be expressed by using a falling accent or a "mildly" rising accent. However, narrow focus also exhibits many rather shallow rises and even some falling accents.

This might be due to the fact that meaning differences expressed through prosody are often not as clear-cut as differences in lexical meanings in languages like German and English. However, as the discussions in Chapter 2 and Chapter 3 showed, even when it comes to speech sound phenomena involved in the differentiation of lexical meanings (often termed the "segmental" domain), a lot of fuzziness and variation is found. Purely symbolic approaches often have difficulties in dealing with this fuzziness and variation.

The integration of categorical and continuous, as well as the fuzziness or probabilistic nature can well be captured in the dynamical modelling approach outlined in this book. It resonates with a view "in which the human mind/brain typically construes the world via partially overlapping fuzzy gray areas that are drawn out over time" (Spivey 2007: 3), a perspective referred to as "the continuity of mind".

The models sketched here use the concept of the attractor to induce stability that is comparable to the notion of prosodic categories. The attractors vary in stability when the control parameter is scaled, providing a mechanism to capture the overlapping of the tonal onglide distributions. Since the attractors are stable states on continuous dimensions, the implementation of the categories follows directly from the location of the attractors themselves. In other words, there is no

separation between an abstract category and a concrete implementation. This allows for fuzziness in the physical output of the system. As the control parameter is changed and the attractor landscape tilts towards the "rising" side, the rising attractor stabilises and also shifts subtly to the right. Hence, comparing the rising (positive) parts of outcome distributions of the different control parameter values (e.g. by simulation), there is of course a lot of overlap but also a general trend towards more extreme values in addition to more rises in general – exactly as in the empirical data.

The analysis of speaker groups reveals that speakers use this system in different but comparable ways. For both speaker groups, the differentiation of broad from narrow and of narrow from contrastive focus can be accounted for by a stabilising of the rising attractor, i.e. an increase in the control parameter. However, the speaker groups differ with respect to the range of control parameter values they use. Group 1 starts lower than group 2. As a consequence, the speakers of group 2 have almost exclusively rising accents in their repertoire, the speakers of group 1 have considerable numbers of rising *and* falling accents. The speaker groups mirror the main speaker strategies found in Grice et al. (2017). In this study, it was demonstrated that all speakers manipulate the F0 parameters under scrutiny in the same direction. In doing so, some speakers crossed category boundaries, e.g. negative tonal onglides in broad focus and positive onglides in narrow focus, while other speakers modulated the parameters within the boundaries of the same pitch accent category. The former group of speakers corresponds to group 1 in the current work, the latter corresponds to group 2. The dynamical model provides a formal implementation of the generalisation that the parameters are manipulated *in the same direction*. It even goes a step further in employing the control parameter as a device to scale the prosodic prominence. The common strategy of both speaker groups is that they increase the control parameter from broad to narrow focus and from narrow to contrastive focus, the difference lies in the range of values that are used for the control parameter.

The qualitative change in the data from background to broad – a unimodal distribution of onglide values changes into a bimodal distribution – is accounted for by a bifurcation in the model of Chapter 7. In a classical phonological sense, this change in the system corresponds to the placement of an accent. This part of the modelling reveals one of the major strengths of dynamical systems. In addition to the more gradual modifications (between broad, narrow and contrastive focus), the system is able to exhibit a "dramatic" change (between background and broad focus). Crucially, the cause of the qualitative change is of the same kind as the cause of the gradual changes, both kinds of change are effected by the scaling of the same control parameter.

In addition to an extensive analysis of the tonal pattern in the data, the book provided results in the domain of supra-laryngeal articulation. With 27 speakers and 2088 productions of the target sentence, the current data collection represents a large EMA corpus contributing to our understanding of the role of articulatory modulations in prosody. The results of the supra-laryngeal articulatory parameters can be summarised as follows:

1. The lips are opened wider from background to broad focus (unaccented → accented) and from broad to contrastive focus (within accentuation) with an intermediate step for narrow focus, i.e. within the group of accented target words.

2. The tongue body is lowered from background to broad focus (unaccented → accented) and from broad to contrastive focus (within accentuation) with an intermediate step for narrow focus (within the group of accented target words).

3. In the vowel /o/, the tongue body is retracted from background to broad focus (unaccented → accented) and from broad to contrastive focus (within accentuation) with an intermediate step for narrow focus (within the group of accented target words), although these results are not as clear as those for the vertical tongue body position.

Again, the distributions reveal a great deal of overlap or fuzziness. This is particularly evident when it comes to the differentiation of narrow focus from the two "neighbouring" focus types broad and contrastive focus. Narrow focus overlaps with both of them, the statistical results are not always clear. However, a general trend for a continuous increase in prosodic prominence beyond accentuation expressed by a larger opening of the vocal tract and more peripheral vowel articulation, can be attested. The articulatory results were used to extend the model outlined for tonal onglide to include more dimensions. In this model, the prosodic dimensions contribute differently to the shared attractor landscape. The complexity of the articulatory dimensions is lower compared to the tonal onglide dimensions as they do not exhibit bifurcation or the presence of two attractors. An important point conveyed by the modelling approach is that change on all dimensions is induced by the scaling of the control parameter of the system.

8.2 Limitations and future directions

Like many models, the current approach has some limitations. First, as already discussed in Chapter 6, the results of the speaker groups do not fit perfectly the

predictions of the model. The differences between the rising means of group 1 compared to group 2 are predicted to be larger than they are in the real data. It is plausible to assume that the tonal onglide, albeit being a very important parameter, does not capture all details in the tonal domain. Speakers certainly mark prominences and differentiate pragmatic meanings on many dimensions (Baumann & Winter 2018).

Another regard in which the model is simplifying is the domain of tonal movement. The analyses look at a stretch of the F0 contours that has been shown to be a critical part (Ritter & Grice 2015; Baumann & Röhr 2015). However, when it comes to the flat contours in background, it is not entirely clear how the arbitrary window translates into a production unit. From an AM perspective, where contours are underspecified, it seems unusual to specify tonal onglide values for this low stretch of F0 as it only represents an interpolation between two tonal targets. It is not the intention of the model to state that speakers choose certain F0 values in the arbitrarily chosen time window. However, what is going on in this time window is representative for the intonation contour as such. It is thus better to take the characterisation that the measure is able to provide as a representation of a larger unit (for example, in the case of background this could be something like "everything after the direct object should sound flat or low"). Likewise, with regard to the nuclear pitch accents, it should be noted that – despite choosing the tonal onglide – the study intends to leave open as to whether portions of F0 after the tonal target on the accented syllable ("offglide" portions) might play an important role like the data of Kügler & Gollrad (2015) suggest.

Using a single control parameter for multiple dimensions, in the way the model does in the current version, leads to the simplifying assumption that the relation of scaling on all dimensions is proportional for all speakers. It is, however, not implausible to assume that speakers might weight one dimensions more in relation to another. For example, a speaker could concentrate on sonority expansion and show large differences between the focus types with regard to lip aperture, but use only slight modulations in terms of F0. Another speaker might exhibit large changes in F0 but only subtle changes of lip aperture. These fine-grained differences are not captured by the model in the current form. One the one hand, the analysis needs to be extended on speaker-specific strategies in all prosodic domains, including articulation. On the other hand, the model might need additional weights or control parameters. In extending the model, the interconnection of dimensions should be maintained. How this is to be achieved has to be left for future research.

Another task of future research will be to consolidate the general idea that a dynamical systems approach is the right avenue to pursue in modelling prosody and

that a plausible type of dynamical model is assumed. This work already showed that scaling the control parameter of the system in the current form may lead to a bifurcation. From a dynamical modelling point of view, this property is important, as Kelso (2013: 1538) puts it: "you only know for certain you have identified a control parameter if, when varied, it causes the system's behaviour to change qualitatively or discontinuously". One important problem of the current approach is, however, that it assumed a dynamical formulation by investigating only the equilibrium positions of the putative system, i.e. the attractors where the system comes to rest. The trajectories, i.e. the paths towards the attractors, were not included in the description and in fact they may be impossible to observe. It has also to be borne in mind that the trajectories in the proposed model are shaped by noise to a large degree. Further work on this topic will have to deal with this limitation. In addition, it might be promising to test other important traits of dynamical systems such as sensitivity to initial conditions or hysteresis, although here as well the implications of noise have to be taken into account.

Finally, the work presented here was based on production data alone. In future works on the topic, the relevance of the prosodic modulations for the decoding of pragmatic meanings should be investigated in more detail. Data of the kind presented here can be used in a perception task. It may also be feasible to model the patterns of perception of the prosodic data in a dynamical framework. In particular, the perception of speakers employing different strategies (like group 1 vs. group 2 in the current data set) may be an interesting research topic. From the modelling perspective in this work, it could be assumed that listeners adjust their attractor landscapes or scalings of attractor landscapes according to a given speaker they listen to. In addition, interlocutors may also be able to align their uses of the system in interaction as an effect of accommodation or entrainment.

8.3 Conclusion

The current book provided an analysis of tonal and articulatory strategies of prosodic focus marking in German. It investigated the implications of the obtained results for the relation between the categorical and the continuous aspects of prosodic prominence – thereby contributing to a long-standing debate on the relation between phonology and phonetics. While categorical and continuous modulations seem to go hand in hand and often form a kind of symbiosis, it is desirable to reconcile the two in a joint theoretical treatment. The work presented here followed work on models of cognition in the framework of dynamical systems and sketched out an attractor modelling approach for the patterns of pros-

odic prominence in the data. This modelling account is able to unify the categorical as well as the continuous aspects of prosodic prominence on one level, and therefore provides a promising tool to integrate the phonology and phonetics of prosody. In addition, it offers a first outline of an integration of different tiers or dimensions that contribute to prosodic prominence, tonal and articulatory, that have often been treated separately. Future research will investigate the patterns of prosodic prominence in both production and perception in more detail and consolidate the idea of prosody as a multi-dimensional dynamical system.

References

't Hart, Johan, René Collier & Abraham Cohen. 1990. *A perceptual study of intonation: An experimental-phonetic approach to speech melody.* Cambridge: Cambridge University Press. DOI: 10.1017/CBO9780511627743.

Anttila, Arto. 2007. Variation and optionality. In Paul de Lacy (ed.), *The Cambridge handbook of phonology*, 519–536. Cambridge: Cambridge University Press. DOI: 10.1017/CBO9780511486371.023.

Arvaniti, Amalia. 2011. The representation of intonation. In Marc van Oostendorp, Colin J. Ewen, Elizabeth V. Hume & Keren Rice (eds.), *The Blackwell companion to phonology*. Oxford, UK: John Wiley & Sons, Ltd. DOI: 10.1002/9781444335262.wbctp0032.

Aylett, Matthew & Alice E. Turk. 2004. The smooth signal redundancy hypothesis: A functional explanation for relationships between redundancy, prosodic prominence, and duration in spontaneous speech. *Language and Speech* 47. 31–56. DOI: 10.1177/00238309040470010201.

Babel, Molly. 2009. *Phonetic and social selectivity in phonetic accommodation.* University of California at Berkeley. (Doctoral dissertation). https://escholarship.org/uc/item/1mb4n1mv.

Baumann, Stefan, Johannes Becker, Martine Grice & Doris Mücke. 2007. Tonal and articulatory marking of focus in German. In Jürgen Trouvain & William J. Barry (eds.), *Proceedings of the 16th International Congress of Phonetic Sciences*, 1029–1032.

Baumann, Stefan, Martine Grice & Susanne Steindamm. 2006. Prosodic marking of focus domains – categorical or gradient? In Rüdiger Hoffmann & Hansjörg Mixdorff (eds.), *Proceedings of Speech Prosody, Dresden*, 301–304.

Baumann, Stefan & Christine T. Röhr. 2015. The perceptual prominence of pitch accent types in German. In The Scottish Consortium for ICPhS 2015 (ed.), *Proceedings of the 18th International Congress of Phonetics Sciences, Glasgow.*

Baumann, Stefan & Bodo Winter. 2018. What makes a word prominent? Predicting untrained german listeners' perceptual judgments. *Journal of Phonetics* 70. 20–38. DOI: 10.1016/j.wocn.2018.05.004.

References

Beckman, Mary E. & Jan Edwards. 1994. Articulatory evidence for differentiating stress categories. In Patricia A. Keating (ed.), *Papers in Laboratory Phonology III: Phonological structure and phonetic form*, 7–33. Cambridge: Cambridge University Press. DOI: 10.1017/CBO9780511659461.002.

Beckman, Mary E., Jan Edwards & Janet Fletcher. 1992. Prosodic structure and tempo in a sonority model of articulatory dynamics. In Gerard J. Docherty & D. Robert Ladd (eds.), *Papers in Laboratory Phonology II: Segment, gesture, prosody*, 68–86. Cambridge: Cambridge University Press. DOI: 10.1017/CBO9780511519918.004.

Beckman, Mary E., Julia Hirschberg & Stefanie Shattuck-Hufnagel. 2005. The original ToBI system and the evolution of the ToBI framework. In Sun Ah Jun (ed.), *Prosodic typology: The phonology of intonation and phrasing*, 9–54. Oxford: Oxford University Press. DOI: 10.1093/acprof:oso/9780199249633.003.0002.

Benus, Stefan. 2005. *Dynamics and transparency in vowel harmony*. New York University. (Doctoral dissertation).

Bloomfield, Leonard. 1933. *Language*. New York: Henry Holt & Co.

Boersma, Paul. 1997. How we learn variation, optionality, and probability. In *Proceedings of the Institute of Phonetic Sciences of the University of Amsterdam 21*, 43–58.

Boersma, Paul & Bruce Hayes. 2001. Empirical test of the gradual learning algorithm. *Linguistic Inquiry* 32(1). 45–86. DOI: 10.1162/002438901554586.

Boersma, Paul & David Weenink. 2018. *Praat: Doing phonetics by computer*. http://www.praat.org/.

Bolinger, Dwight Le Merton. 1961. *Generality, gradience, and the all-or-none*. The Hague: Mouton.

Bolinger, Dwight Le Merton. 1978. Intonation across languages. In Joseph H. Greenberg (ed.), *Universals of human language, vol. 2*, 471–524. Stanford: Stanford University Press.

Browman, Catherine P. & Louis Goldstein. 1986. Towards an articulatory phonology. *Phonology Yearbook* 3. 219–252. DOI: 10.1017/S0952675700000658.

Browman, Catherine P. & Louis Goldstein. 1988. Some notes on syllable structure in Articulatory phonology. *Phonetica* 45(2-4). 140–155. DOI: 10.1159/000261823.

Browman, Catherine P. & Louis Goldstein. 1989. Articulatory gestures as phonological units. *Phonology* 6(2). 201–251. DOI: 10.1017/S0952675700001019.

Browman, Catherine P. & Louis Goldstein. 1990. Tiers in articulatory phonology, with some implications for casual speech. In John Kingston & Mary E. Beckman (eds.), *Papers in laboratory phonology I: Between the grammar and physics of speech*, 341–376. Cambridge University Press. DOI: 10.1017/CBO9780511627736.019.

Browman, Catherine P. & Louis Goldstein. 1992. Articulatory phonology: An overview. *Phonetica* 49(3-4). 155–180. DOI: 10.1159/000261913.

Büring, Daniel. 2003. On d-trees, beans, and b-accents. *Linguistics & Philosophy* 26(5). 511–545. DOI: 10.1023/A:1025887707652.

Bürkner, Paul-Christian. 2018. Advanced Bayesian multilevel modeling with the R package brms. *The R Journal* 10(1). DOI: 10.32614/RJ-2018-017.

Bybee, Joan. 2001. *Phonology and language use.* Cambridge: Cambridge University Press. DOI: 10.1017/CBO9780511612886.

Byrd, Dani. 1995. C-centers revisited. *Phonetica* 52(4). 285–306. DOI: 10.1159/000262183.

Byrd, Dani. 2000. Articulatory vowel lengthening and coordination at phrasal junctures. *Phonetica* 57(1). 3–16. DOI: 10.1159/000028456.

Byrd, Dani, Abigail Kaun, Shrikanth Narayanan & Elliot Saltzman. 2000. Phrasal signatures in articulation. In Michael B. Broe & Janet B. Pierrehumbert (eds.), *Papers in Laboratory Phonology V: Acquisition and the lexicon*, 70–87. Cambridge: Cambridge University Press.

Byrd, Dani & Elliot Saltzman. 2003. The elastic phrase: Modeling the dynamics of boundary-adjacent lengthening. *Journal of Phonetics* 31. 149–180. DOI: 10.1016/S0095-4470(02)00085-2.

Calhoun, Sasha. 2007. *Information structure and the prosodic structure of English: A probabilistic relationship.* University of Edinburgh. (Doctoral dissertation). http://hdl.handle.net/1842/8120.

Cangemi, Francesco, Martina Krüger & Martine Grice. 2015. Listener-specific perception of speaker-specific productions in intonation. In Susanne Fuchs, Daniel Pape, Caterina Petrone & Pascal Perrier (eds.), *Individual differences in speech production and perception*, 123–145. Frankfurt a. M.: Peter Lang. DOI: 10.3726/978-3-653-05777-5.

Carpenter, Bob, Andrew Gelman, Matthew D. Hoffman, Daniel Lee, Ben Goodrich, Michael Betancourt, Marcus Brubaker, Jiqiang Guo, Peter Li & Allen Riddell. 2017. Stan: A probabilistic programming language. *Journal of Statistical Software* 76(1). 1–32. DOI: 10.18637/jss.v076.i01.

Cedergren, Henrietta J. & David Sankoff. 1974. Variable rules: Performance as a statistical reflection of competence. *Language* 50(2). 333–355.

Charles-Luce, Jan. 1985. Word-final devoicing in German: effects of phonetic and sentential contexts. *Journal of Phonetics* 13. 309–324. DOI: 10.1016/S0095-4470(19)30762-4.

Cho, Taehong. 2005. Prosodic strengthening and featural enhancement: Evidence from acoustic and articulatory realizations of /ɑ,i/ in English. *The Journal of the Acoustical Society of America* 117(6). 3867–78. DOI: 10.1121/1.1861893.

References

Cho, Taehong. 2006. Manifestation of prosodic structure in articulatory variation: Evidence from lip kinematics in English. In Louis Goldstein, Douglas H. Whalen & Catherine T. Best (eds.), *Laboratory phonology 8*, 519–548. Berlin/New York: Mouton de Gruyter. DOI: 10.1515/9783110197211.3.519.

Cho, Taehong. 2011. Laboratory phonology. In Nancy C. Kula, Bert Botma & Kuniya Nasukawa (eds.), *The Continuum companion to phonology*, 343–368. London/New York: Continuum.

Cho, Taehong & Patricia A. Keating. 2001. Articulatory and acoustic studies on domain-initial strengthening in Korean. *Journal of Phonetics* 29(2). 155–190. DOI: 10.1006/jpho.2001.0131.

Cho, Taehong & Peter Ladefoged. 1999. Variation and universals in VOT: Evidence from 18 languages. *Journal of Phonetics* 27. 207–229. DOI: 10.1006/jpho.1999.0094.

Cho, Taehong & James M. McQueen. 2005. Prosodic influences on consonant production in Dutch: Effects of prosodic boundaries, phrasal accent and lexical stress. *Journal of Phonetics* 33(2). 121–157. DOI: 10.1016/j.wocn.2005.01.001.

Chomsky, Noam. 1964. *Current issues in linguistic theory*. The Hague: Mouton.

Chomsky, Noam & Morris Halle. 1968. *The sound pattern of English*. New York: Harper & Row.

Cohn, Abigail C. 2006. Is there gradient phonology? In Gisbert Fanselow, Caroline Féry, Ralf Vogel & Matthias Schlesewsky (eds.), *Gradience in grammar: Generative perspectives*, 25–44. Oxford: Oxford University Press. DOI: 10.1093/acprof:oso/9780199274796.003.0002.

Cruttenden, Alan. 1997. *Intonation*. Cambridge: Cambridge University Press. DOI: 10.1017/CBO9781139166973.

Culler, Jonathan. 1986. *Ferdinand de Saussure*. Ithaca: Cornell University Press.

D'Imperio, Mariapaola & David House. 1997. Perception of questions and statements in Neapolitan Italian. In George Kokkinakis, Nikos Fakotakis & Evangelos Dermatas (eds.), *Proceedings of the 5th European Conference of Speech Communication and Technology, Rhodes, Greece*, 251–254.

de Boer, Bart. 2001. *The origins of vowel systems*. Oxford: Oxford University Press.

de Courtenay, Baudoin. 1963. *Izbrannye trudy po obščemu jazykoznaniju*. Moskva: Izdat. Akad. Nauk SSSR.

de Jong, Kenneth J. 1995. The supraglottal articulation of prominence in English: Linguistic stress as localized hyperarticulation. *The Journal of the Acoustical Society of America* 97(1). 491–504. DOI: https://doi.org/10.1121/1.412275.

de Jong, Kenneth J., Mary E. Beckman & Jan Edwards. 1993. The interplay between prosodic structure and coarticulation. *Language and Speech* 36(2-3). 197–212. DOI: 10.1177/002383099303600305.

Dilley, Laura C. & Christopher C. Heffner. 2013. The role of f0 alignment in distinguishing intonation categories: Evidence from American English. *Journal of Speech Sciences* 3(1). 3–67.

Dinnsen, Daniel & M. Garcia-Zamor. 1971. The three degrees of vowel length in German. *Papers in Linguistics* 4. 111–126. DOI: 10.1080/08351817109370250.

Durand, Jacques & Bernard Laks. 2002. Phonology, phonetics, and cognition. In Jacques Durand & Bernard Laks (eds.), *Phonetics, phonology, and cognition*, 10–50. New York & Oxford: Oxford University Press.

Ernestus, Mirjam. 2011. Gradience and categoricality in phonological theory. In Marc van Oostendorp, Colin J. Ewen, Elizabeth V. Hume & Keren Rice (eds.), *The Blackwell companion to phonology*, 1–22. Oxford: John Wiley & Sons, Ltd. DOI: 10.1002/9781444335262.wbctp0089.

Ernestus, Mirjam & Harald Baayen. 2006. The functionality of incomplete neutralization in Dutch: The case of past-tense formation. In Louis Goldstein, Douglas H. Whalen & Catherine T. Best (eds.), *Laboratory Phonology 8*, 27–49. Berlin, New York: Mouton de Gruyter. DOI: 10.1515/9783110197211.1.27.

Farkas, Illés, Dirk Helbing & Tamás Vicsek. 2002. Mexican waves in an excitable medium. *Nature* 419(6903). 131–132. DOI: 10.1038/419131a.

Feigenbaum, Mitchell J. 1978. Quantitative universality for a class of nonlinear transformations. *Journal of Statistical Physics* 19(1). 25–52. DOI: 10.1007/BF01020332.

Féry, Caroline. 1993. *German intonational patterns*. Tübingen: Max Niemeyer Verlag. DOI: 10.1515/9783111677606.

Feynman, Richard, Robert B. Leighton & Matthew Sands. 1963. *The Feynman lectures on physics*. Reading, MA: Addison-Wesley.

Flemming, Edward. 2001. Scalar and categorical phenomena in a unified model of phonetics and phonology. *Phonology* 18(1). 7–44. DOI: 10.1017/S0952675701004006.

Fodor, Jerry A. 1975. *The language of thought*. Cambridge: Harvard University Press.

Fodor, Jerry A. & Zenon W. Pylyshyn. 1981. How direct is visual perception? Some reflections on Gibson's 'ecological approach'. *Cognition* 9(2). 139–196. DOI: 10.1016/0010-0277(81)90009-3.

Fougeron, Cécile. 2001. Articulatory properties of initial segments in several prosodic constituents in French. *Journal of Phonetics* 29(2). 109–135. DOI: 10.1006/jpho.2000.0114.

Freidlin, Mark I. & Alexander D. Wentzell. 1984. *Random perturbations of dynamical systems*. Vol. 260 (Grundlehren der mathematischen Wissenschaften). New York: Springer. DOI: 10.1007/978-1-4684-0176-9.

References

Fuchs, Armin. 2013. *Nonlinear dynamics in complex systems*. Berlin, Heidelberg: Springer. DOI: 10.1007/978-3-642-33552-5.

Gafos, Adamantios I. 2006. Dynamics in grammar. In Louis Goldstein, Douglas H. Whalen & Catherine T. Best (eds.), *Laboratory Phonology 8*, 51–80. Berlin, New York: Mouton de Gruyter. DOI: 10.1515/9783110197211.1.51.

Gafos, Adamantios I. & Stefan Benus. 2006. Dynamics of phonological cognition. *Cognitive Science* 30(5). 905–943. DOI: 10.1207/s15516709cog0000_80.

Gahl, Susanne & Alan C. L. Yu. 2006. Introduction to the special issue on exemplar-based models in linguistics. *The Linguistic Review* 23(3). 213–216. DOI: 10.1515/TLR.2006.007.

German, James S., Katy Carlson & Janet B. Pierrehumbert. 2013. Reassignment of consonant allophones in rapid dialect acquisition. *Journal of Phonetics* 41(3-4). 228–248. DOI: 10.1016/j.wocn.2013.03.001.

Goldinger, Stephen D. 1996. Words and voices: Episodic traces in spoken word identification and recognition memory. *Journal of Experimental Psychology: Learning, Memory and Cognition* 22. 1166–1182. DOI: 10.1037/0278-7393.22.5.1166.

Goldrick, Matthew, Jocelyn R. Folk & Brenda Rapp. 2010. Mrs. Malaprop's neighborhood: Using word errors to reveal neighborhood structure. *Journal of Memory and Language* 62(2). 113–134. DOI: 10.1016/j.jml.2009.11.008.

Goldstein, Louis, Dani Byrd & Elliot Saltzman. 2006. The role of vocal tract gestural action units in understanding the evolution of phonology. In Michael A. Arbib (ed.), *Action to language via the mirror neuron system*, 215–249. Cambridge: Cambridge University Press. DOI: 10.1017/CBO9780511541599.008.

Goldstein, Louis, Hosung Nam, Elliot Saltzman & Ioana Chitoran. 2009. Coupled oscillator planning model of speech timing and syllable structure. In C. Gunnar M. Fant, Hiroya Fujisaki & Jiaxuan Shen (eds.), *Frontiers in phonetics and speech science*, 239–250. Bejing: The Commercial Press.

Grice, Martine. 2006. Intonation. In *Encyclopedia of language & linguistics, second edition*, vol. 5, 778–788. Oxford: Elsevier. DOI: 10.1016/B0-08-044854-2/00045-6.

Grice, Martine & Stefan Baumann. 2002. Deutsche Intonation und GToBI. *Linguistische Berichte* 191. 267–298.

Grice, Martine, Stefan Baumann & Ralf Benzmüller. 2005. German intonation in Autosegmental-metrical phonology. In *Prosodic typology: The phonology of intonation and phrasing*, Jun, Sun-Ah, 55–83. Oxford: Oxford University Press. DOI: 10.1093/acprof:oso/9780199249633.003.0003.

Grice, Martine, Simon Ritter, Henrik Niemann & Timo B. Roettger. 2017. Integrating the discreteness and continuity of intonational categories. *Journal of Phonetics* 64. 90–107. DOI: 10.1016/j.wocn.2017.03.003.

Gussenhoven, Carlos. 2004. *The phonology of tone and intonation.* Cambridge: Cambridge University Press. DOI: 10.1017/CBO9780511616983.

Gussenhoven, Carlos & Haike Jacobs. 2011. *Understanding phonology.* London: Routledge. DOI: 10.4324/978020377708.

Gussenhoven, Carlos & Toni Rietveld. 2000. The behavior of H* and L* under variations in pitch range in Dutch rising contours. *Language and Speech* 43(2). 183–203. DOI: 10.1177/00238309000430020301.

Guy, Gregory R. 2014. Linking usage and grammar: Generative phonology, exemplar theory, and variable rules. *Lingua* 142. 57–65. DOI: 10.1016/j.lingua.2012.07.007.

Haken, Hermann. 1977. *Synergetics.* Berlin, Heidelberg: Springer. DOI: 10.1007/978-3-642-96363-6.

Haken, Hermann, J. A. Scott Kelso & Herbert Bunz. 1985. A theoretical model of phase transitions in human hand movements. *Biological Cybernetics* 51(5). 347–356. DOI: 10.1007/BF00336922.

Haken, Hermann & Paul Levi. 2012. *Synergetic agents: From multi-robot systems to molecular robotics.* Weinheim: Wiley. DOI: 10.1002/9783527659524.

Hall, Kathleen Currie, Elizabeth V. Hume, T. Florian Jaeger & Andrew Wedel. 2018. The role of predictability in shaping phonological patterns. *Linguistics Vanguard* 4(s2). DOI: 10.1515/lingvan-2017-0027.

Harnad, Stevan. 1990. The symbol grounding problem. *Physica D: Nonlinear Phenomena* 42(1). 335–346. DOI: 10.1016/0167-2789(90)90087-6.

Harrington, Jonathan, Janet Fletcher & Mary E. Beckman. 2000. Manner and place conflicts in the articulation of accent in Australian English. In Michael B. Broe & Janet Pierrehumbert B. (eds.), *Papers in Laboratory Phonology V: Acquisition and the lexicon*, 40–51. Cambridge: Cambridge University Press.

Hawkins, Sarah. 1992. An introduction to task dynamics. In Gerard J. Docherty & D. Robert Ladd (eds.), *Papers in Laboratory Phonology II: Segment, gesture, prosody*, 9–25. Cambridge: Cambridge University Press. DOI: 10.1017/CBO9780511519918.002.

Hawkins, Sarah. 2003. Roles and representations of systematic fine phonetic detail in speech understanding. *Journal of Phonetics* 31(3-4). 373–405. DOI: 10.1016/j.wocn.2003.09.006.

Hawkins, Sarah & Rachel Smith. 2001. Polysp: A polysystemic, phonetically-rich approach to speech understanding. *Italian Journal of Linguistics – Rivista di Linguistica* 13(1). 99–188.

References

Hay, Jen, Janet B. Pierrehumbert & Mary Beckman. 2004. Speech perception, well-formedness and the statistics of the lexicon. In John Local, Richard Ogden & Rosalind Temple (eds.), *Papers in Laboratory Phonology VI: Phonetic interpretation*, 58–74. Cambridge: Cambridge University Press. DOI: 10.1017/CBO9780511486425.004.

Hayes, Bruce. 1989. The prosodic hierarchy in meter. In Paul Kiparsky & Gilbert Youmans (eds.), *Rhythm and meter*, 201–260. San Diego: Academic Press. DOI: 10.1016/B978-0-12-409340-9.50013-9.

Hayes, Bruce & Zsuzsa C. Londe. 2006. Stochastic phonological knowledge: The case of Hungarian vowel harmony. *Phonology* 23(1). 59–104. DOI: 10.1017/S0952675706000765.

Hermes, Anne, Doris Mücke & Martine Grice. 2013. Gestural coordination of Italian word-initial clusters: The case of 'impure s'. *Phonology* 30(1). 1–25. DOI: 10.1017/S095267571300002X.

Hintzman, Douglas L. 1986. "Schema abstraction" in a multiple-trace memory model. *Psychological Review* 93. 411–428. DOI: 10.1037/0033-295X.93.4.411.

Hirschberg, Julia & Gregory Ward. 1992. The influence of pitch range, duration, amplitude and spectral features on the interpretation of the rise-fall-rise intonation contour in English. *Journal of Phonetics* 20(2). 241–251. DOI: 10.1016/S0095-4470(19)30625-4.

IPA. 1999. *Handbook of the International Phonetic Association*. Cambridge: Cambridge University Press.

Iskarous, Khalil. 2017. The relation between the continuous and the discrete: A note on the first principles of speech dynamics. *Journal of Phonetics* 64. 8–20. DOI: 10.1016/j.wocn.2017.05.003.

Jacobson, Roman, C. Gunnar M. Fant & Morris Halle. 1952. *Preliminaries to speech analysis*. The Hague: Mouton.

Johnson, Keith. 1997. Speech perception without speaker normalization: An exemplar model. In John Mullennix & Keith Johnson (eds.), *Talker variability in speech processing*, 145–165. San Diego: Academic Press.

Jurafsky, Daniel, Alan Bell & Cynthia Girard. 2002. The role of the lemma in form variation. In Carlos Gussenhoven & Natasha Warner (eds.), *Laboratory Phonology 7*, 3–34. Berlin: De Gruyter. DOI: 10.1515/9783110197105.1.3.

Kaplan, Daniel & Leon Glass. 1995. *Understanding nonlinear dynamics*. New York: Springer. DOI: 10.1007/978-1-4612-0823-5.

Keating, Patricia A. 1988. The phonology-phonetics interface. In Frederick J. Newmeyer (ed.), *Linguistics: The Cambridge survey, volume i: Grammatical theory*, 281–302. Cambridge: Cambridge University Press.

Kelso, J. A. Scott. 1981. On the oscillatory basis of movement. *Bulletin of the psychonomic society* 18. 63.

Kelso, J. A. Scott. 1995. *Dynamic patterns: The self-organization of brain and behavior.* Cambridge: MIT Press.

Kelso, J. A. Scott. 2013. Coordination dynamics. In Robert A. Meyers (ed.), *Encyclopedia of complexity and systems science*, 1537–1564. Heidelberg: Springer. DOI: 10.1007/978-3-642-27737-5_101-3.

Kleber, Felicitas, Tina John & Jonathan Harrington. 2010. The implications for speech perception of incomplete neutralization of final devoicing in German. *Journal of Phonetics* 38. 185–196. DOI: 10.1016/j.wocn.2009.10.001.

Kohler, Klaus J. 1987. Categorical pitch perception. In *Proceedings of the 11th International Congress of Phonetic Sciences, Tallinn*, 331–333. Academy of Sciences of the Estonian S.S.R.

Krivokapić, Jelena. 2014. Gestural coordination at prosodic boundaries and its role for prosodic structure and speech planning processes. *Philosophical Transactions of the Royal Society B: Biological Sciences* 369(1658). 20130397. DOI: 10.1098/rstb.2013.0397.

Krivokapić, Jelena & Dani Byrd. 2012. Prosodic boundary strength: An articulatory and perceptual study. *Journal of Phonetics* 40(3). 430–442. DOI: 10.1016/j.wocn.2012.02.011.

Krüger, M. 2009. *Produktion und Perzeption von Fokus im Deutschen.* University of Cologne. (Master's thesis).

Kügler, Frank & Anja Gollrad. 2015. Production and perception of contrast: The case of the rise-fall contour in German. *Frontiers in Psychology* 6. 1254. DOI: 10.3389/fpsyg.2015.01254.

Labov, William. 1969. *The study of nonstandard English.* Urbana: National Council of Teachers of English.

Ladd, D. Robert. 1980. *The structure of intonational meaning: Evidence from English.* Bloomington: Indiana University Press.

Ladd, D. Robert. 2006. "Distinctive phones" in surface representation. In Louis Goldstein, Douglas H. Whalen & Catherine T. Best (eds.), *Laboratory Phonology 8*, 3–26. Berlin: de Gruyter. DOI: 10.1515/9783110197211.1.3.

Ladd, D. Robert. 2008. *Intonational phonology.* Cambridge: Cambridge University Press. DOI: 10.1017/CBO9780511808814.

Ladd, D. Robert. 2014. *Simultaneous structure in phonology.* Oxford: Oxford University Press. DOI: 10.1093/acprof:oso/9780199670970.001.0001.

Ladd, D. Robert & Rachel Morton. 1997. The perception of intonational emphasis: Continuous or categorical? *Journal of Phonetics* 25. 313–342.

References

Lambrecht, Knud. 1994. *Information structure and sentence form.* Cambridge: Cambridge University Press. DOI: 10.1017/CBO9780511620607.

Lindblom, Björn. 1990. Explaining phonetic variation: A sketch of the H&H theory. In William J. Hardcastle & Alain Marchal (eds.), *Speech production and speech modeling*, 403–439. Dordrecht: Springer. DOI: 10.1007/978-94-009-2037-8_16.

Lisker, Leigh. 1986. "Voicing" in English: A catalogue of acoustic features signaling /b/ versus /p/ in trochees. *Language and Speech* 29. 3–11. DOI: 10.1177/002383098602900102.

May, Robert M. 1976. Simple mathematical models with very complicated dynamics. *Nature* 261(5560). 459–467. DOI: 10.1038/261459a0.

Mücke, Doris. 2018. *Dynamische Modellierung von Artikulation und prosodischer Struktur: Eine Einführung in die Artikulatorische Phonologie.* Berlin: Language Science Press. DOI: 10.5281/zenodo.1188764.

Mücke, Doris & Martine Grice. 2014. The effect of focus marking on supralaryngeal articulation – is it mediated by accentuation? *Journal of Phonetics* 44. 47–61. DOI: 10.1016/j.wocn.2014.02.003.

Munhall, Kevin G., David J. Ostry & Avraham Parush. 1985. Characteristics of velocity profiles of speech movements. *Journal of Experimental Psychology: Human Perception and Performance* 11(4). 457–474. DOI: 10.1037/0096-1523.11.4.457.

Nam, Hosung. 2007. Syllable-level intergestural timing model: Split-gesture dynamics focusing on positional asymmetry and moraic structure. In Jennifer Cole & José Ignacio Hualde (eds.), *Laboratory Phonology 9*, 483–506. Berlin, New York: Walter de Gruyter.

Nam, Hosung, Louis Goldstein & Elliot Saltzman. 2009. Self-organization of syllable structure: A coupled oscillator model. In François Pellegrino, Egidio Marsico, Ioana Chitoran & Christophe Coupé (eds.), *Approaches to phonological complexity*, 299–328. Berlin, New York: Walter de Gruyter. DOI: 10.1515/9783110223958.297.

Nam, Hosung & Elliot Saltzman. 2003. A competitive, coupled oscillator model of syllable structure. In Maria-Josep Solé, Daniel Recassens & Joaquín Romero (eds.), *Proceedings of the 15th International Congress of Phonetic Sciences, Barcelona.*

Nespor, Marina & Irene Vogel. 1986. *Prosodic phonology.* Dordrecht: Foris Publications.

Newell, Allen & Herbert A. Simon. 1976. Computer science as empirical inquiry: Symbols and search. *Communications of the Association for Computing Machinery* 19(3). 113–126. DOI: 10.1145/360018.360022.

Nolan, Francis. 1992. The descriptive role of segments: Evidence from assimilation. In Gerard J. Docherty & D. Robert Ladd (eds.), *Papers in Laboratory Phonology II: Segment, gesture, prosody*, 261–280. Cambridge: Cambridge University Press. DOI: 10.1017/CBO9780511519918.011.

Nosofsky, Robert M. 1986. Attention, similarity and the identification-categorization relationship. *Journal of Experimental Psychology: General* 115. 39–57. DOI: 10.1037/0096-3445.115.1.39.

O'Connor, Joseph D. & Gordon F. Arnold. 1973. *The intonation of colloquial English.* London: Longman.

Ohala, John J. 1990. There is no interface between phonology and phonetics: A personal view. *Journal of Phonetics* 18. 153–171. DOI: 10.1016/S0095-4470(19)30399-7.

Ohala, John J. 1997. The relation between phonetics and phonology. In William J. Hardcastle (ed.), *The handbook of phonetic sciences*, 674–694. Chichester: Wiley-Blackwell.

Pierrehumbert, Janet B. 1980. *The phonology and phonetics of English intonation.* MIT. (Doctoral dissertation).

Pierrehumbert, Janet B. 1990. Phonological and phonetic representation. *Journal of Phonetics* 18. 375–394. DOI: 10.1016/S0095-4470(19)30380-8.

Pierrehumbert, Janet B. 2001. Exemplar dynamics: Word frequency, lenition, and contrast. In Joan Bybee & Paul Hopper (eds.), *Frequency and the emergence of linguistic structure*, 137–157. Amsterdam: John Benjamins. DOI: 10.1075/tsl.45.08pie.

Pierrehumbert, Janet B. 2002. Word-specific phonetics. In Carlos Gussenhoven & Natasha Warner (eds.), *Laboratory Phonology 7*, 101–139. Berlin: Mouton de Gruyter. DOI: 10.1515/9783110197105.1.101.

Pierrehumbert, Janet B. 2016. Phonological representation: Beyond abstract versus episodic. *Annual Review of Linguistics* 2. 33–52. DOI: 10.1146/annurev-linguistics-030514-125050.

Pierrehumbert, Janet B., Mary Beckman & D. Robert Ladd. 2000. Conceptual foundations of phonology as a laboratory science. In Noel Burton-Roberts, Philip Carr & Gerard J. Docherty (eds.), *Phonological knowledge: Conceptual and empirical issues*, 273–304. Oxford: Oxford University Press.

Pierrehumbert, Janet B. & Mary E. Beckman. 1988. *Japanese tone structure.* Cambridge: MIT Press.

Pierrehumbert, Janet B. & Julia Hirschberg. 1990. The meaning of intonational contours in interpretation of discourse. In Philip R. Cohen, Jerry Morgan & Martha E. Pollack (eds.), *Intentions in communication*, 271–311. Cambridge, MA: MIT Press. DOI: 10.7916/D8KD24FP.

References

Pierrehumbert, Janet B. & Shirley A. Steele. 1989. Categories of tonal alignment in English. *Phonetica* 46(4). 181–196. DOI: 10.1159/000261842.

Port, Robert F. 2002. Dynamical systems hypothesis in cognitive science. In Lynn Nadel (ed.), *Encyclopedia of cognitive science*, 1027–1032. London: Nature Publishing Group. DOI: 10.1002/0470018860.s00020.

Port, Robert F. & Penny Crawford. 1989. Incomplete neutralization and pragmatics in German. *Journal of Phonetics* 17(4). 257–282. DOI: 10.1016/S0095-4470(19)30444-9.

Port, Robert F. & Michael L. O'Dell. 1985. Neutralization of syllable-final voicing in German. *Journal of Phonetics* 13(4). 455–471. DOI: 10.1016/S0095-4470(19)30797-1.

Pouplier, Marianne. 2011. The atoms of phonological representations. In Marc van Oostendorp, Colin J. Ewen, Elizabeth V. Hume & Keren Rice (eds.), *The Blackwell companion to phonology*. Oxford: John Wiley & Sons, Ltd. DOI: 10.1002/9781444335262.wbctp0005.

Prince, Alan & Paul Smolensky. 2004. Optimality Theory: Constraint interaction in generative grammar. In John J. McCarthy (ed.), *Optimality Theory in phonology: A reader*. Oxford: Blackwell Publishing Ltd. DOI: 10.1002/9780470756171.ch1.

R Core Team. 2018. *R: A language and environment for statistical computing*. Vienna, Austria. http://www.r-project.org/.

Ritter, Simon & Martine Grice. 2015. The role of tonal onglides in German nuclear pitch accents. *Language and Speech* 58(1). 114–128. DOI: 10.1177/0023830914565688.

Roettger, Timo B. & Dinah Baer-Henney. 2019. Towards a replication culture in phonetic research: Speech production research in the classroom. *Phonological Data and Analysis* 1(4). 1–23. DOI: 10.3765/pda.v1art4.13.

Roettger, Timo B., Bodo Winter, Sven Grawunder, James Kirby & Martine Grice. 2014. Assessing incomplete neutralization of final devoicing in German. *Journal of Phonetics* 43. 11–25. DOI: 10.1016/j.wocn.2014.01.002.

Saltzman, Elliot. 1986. Task dynamic coordination of the speech articulators: A preliminary model. *Experimental brain research series* 15. 129–144.

Saltzman, Elliot. 1991. The task dynamic model in speech production. In Herman F. M. Peters, Wouter Hulstijn & C. Woodruff Starkweather (eds.), *Speech motor control and stuttering: Proceedings of the 2nd international conference on speech motor control and stuttering, Nijmegen*, 37–52. Amsterdam: Excerpta Medica.

Saltzman, Elliot & Dani Byrd. 2000. Task-dynamics of gestural timing: Phase windows and multifrequency rhythms. *Human Movement Science* 19(4). 499–526. DOI: 10.1016/S0167-9457(00)00030-0.

Saltzman, Elliot & J. A. Scott Kelso. 1987. Skilled actions: A task-dynamic approach. *Psychological Review* 94(1). 84–106. DOI: 10.1037/0033-295X.94.1.84.

Saltzman, Elliot, Hosung Nam, Jelena Krivokapic & Louis Goldstein. 2008. A task-dynamic toolkit for modeling the effects of prosodic structure on articulation. In Plinio Barbosa, Sandra Madureira & Cesar Reis (eds.), *Proceedings of the 4th International Conference on Speech Prosody, Campinas, Brazil*, 175–184.

Schweitzer, Katrin. 2012. *Frequency effects on pitch accents: Towards an exemplar-theoretic approach to intonation.* Universität Stuttgart. (Doctoral dissertation).

Selkirk, Elisabeth. 1996. The prosodic structure of function words. In James L. Morgan & Katherine Demuth (eds.), *Signal to syntax: Bootstrapping from speech to grammar in early acquisition*, 187–214. Mahwah, NJ: Lawrence Erlbaum.

Seyfarth, Scott. 2014. Word informativity influences acoustic duration: Effects of contextual predictability on lexical representation. *Cognition* 133(1). 140–155. DOI: 10.1016/j.cognition.2014.06.013.

Seyfarth, Scott, Jozina Vander Klok & Marc Garellek. 2019. Evidence against interactive effects on articulation in Javanese verb paradigms. *Psychonomic Bulletin & Review* 26(5). 1690–1696. DOI: 10.3758/s13423-019-01637-2.

Shattuck-Hufnagel, Stefanie & Alice E. Turk. 1996. A prosody tutorial for investigators of auditory sentence processing. *Journal of Psycholinguistic Research* 25(2). 193–247. DOI: 10.1007/BF01708572.

Smolensky, Paul. 1988. On the proper treatment of connectionism. *Behavioral and Brain Sciences* 11(1). 1–23. DOI: 10.1017/S0140525X00052432.

Spivey, Michael. 2007. *The continuity of mind.* Oxford: Oxford University Press. DOI: 10.1093/acprof:oso/9780195170788.001.0001.

Sumner, Meghan, Seung Kyung Kim, Ed King & Kevin B. McGowan. 2014. The socially weighted encoding of spoken words: A dual-route approach to speech perception. *Frontiers in Psychology* 4. 1015. DOI: 10.3389/fpsyg.2013.01015.

Tabain, Marija. 2003. Effects of prosodic boundary on /aC/ sequences: Articulatory results. *The Journal of the Acoustical Society of America* 113(5). 2834–2849. DOI: 10.1121/1.1564013.

Thelen, Esther & Linda B. Smith. 1994. *A dynamic systems approach to the development of cognition and action.* Cambridge: MIT Press. DOI: 10.7551/mitpress/2524.001.0001.

Tilsen, Sam. 2017. Exertive modulation of speech and articulatory phasing. *Journal of Phonetics* 64. 34–50. DOI: 10.1016/j.wocn.2017.03.001.

Tilsen, Sam. 2018. Three mechanisms for modeling articulation: Selection, coordination, and intention. *Cornell Working Papers in Phonetics and Phonology.* https://ling.auf.net/lingbuzz/004167.

References

Trubetzkoy, Nikolai S. 1958. *Grundzüge der Phonologie.* Göttingen: Vandenhoeck & Ruprecht.

Tuller, Betty, Pamela Case, Mingzhou Ding & J. A. Scott Kelso. 1994. The nonlinear dynamics of speech categorization. *Journal of Experimental Psychology: Human Perception and Performance* 20(1). 3–16. DOI: 10.1037/0096-1523.20.1.3.

Vago, Robert M. 1980. *The sound pattern of Hungarian.* Washington D.C.: Georgetown University Press.

Vallduví, Enric. 1991. The role of plasticity in the association of focus and prominence. In *Proceedings of the Eastern States Conference on Linguistics 7*, 295–306.

Vallduví, Enric. 1992. *The informational component.* New York: Garland.

van Gelder, Timothy & Robert F. Port. 1995. *Mind as motion.* Cambridge: MIT Press. DOI: 10.7551/mitpress/4622.001.0001.

Wagner, Michael & Duane G. Watson. 2010. Experimental and theoretical advances in prosody: A review. *Language and Cognitive Processes* 25(7-9). 905–945. DOI: 10.1080/01690961003589492.

Walker, Abby & Jen Hay. 2011. Congruence between 'word age' and 'voice age' facilitates lexical access. *Laboratory Phonology* 2(1). 219–237. DOI: 10.1515/labphon.2011.007.

Ward, Gregory & Julia Hirschberg. 1985. Implicating uncertainty: The pragmatics of fall-rise intonation. *Language* 61(4). 747. DOI: 10.2307/414489.

Watson, Duane G., Michael Tanenhaus & Christine A. Gunlogson. 2008. Interpreting pitch accents in online comprehension: H* vs. L+H*. *Cognitive Science: A Multidisciplinary Journal* 32(7). 1232–1244. DOI: 10.1080/03640210802138755.

Winkelmann, Raphael, Klaus Jaensch, Steve Cassidy & Jonathan Harrington. 2018. *emuR: Main package of the EMU speech database management system.*

Winter, Bodo & Timo B. Roettger. 2011. The nature of incomplete neutralization: Implications for Laboratory Phonology. *Grazer Linguistische Studien* 76. 55–74.

Wright, Charles E. 1979. Duration differences between rare and common words and their implications for the interpretation of word frequency effects. *Memory & Cognition* 7. 411–419. DOI: 10.3758/BF03198257.

Yu, Alan C. L., Carissa Abrego-Collier & Morgan Sonderegger. 2013. Phonetic imitation from an individual-difference perspective: Subjective attitude, personality and "autistic" traits. *PLoS ONE* 8(9). e74746. DOI: 10.1371/journal.pone.0074746.

Name index

Categoriality and continuity in prosodic prominence

Prosody has been characterised as a 'half-tamed savage' being shaped by both discrete, categorical aspects as well as gradient, continuous phenomena. This book is concerned with the relation of the 'wild' and the 'tamed' sides of prosodic prominence. It reviews problems that arise from a strict separation of categorical and continuous representations in models of phonetics and phonology, and it explores the potential role of descriptions aimed at reconciling the two domains. In doing so, the book offers an introduction to dynamical systems, a framework that has been studied extensively in the last decades to model speech production and perception. The reported acoustic and articulatory data presented in this book show that categorical and continuous modulations used to enhance prosodic prominence are deeply intertwined and even exhibit a kind of symbiosis. A multi-dimensional dynamical model of prosodic prominence is sketched, based on the empirical data, combining tonal and articulatory aspects of prosodic focus marking. The model demonstrates how categorical and continuous aspects can be integrated in a joint theoretical treatment that overcomes a strict separation of phonetics and phonology.