



THE UNIVERSITY *of* EDINBURGH

This thesis has been submitted in fulfilment of the requirements for a postgraduate degree (e.g. PhD, MPhil, DClinPsychol) at the University of Edinburgh. Please note the following terms and conditions of use:

- This work is protected by copyright and other intellectual property rights, which are retained by the thesis author, unless otherwise stated.
- A copy can be downloaded for personal non-commercial research or study, without prior permission or charge.
- This thesis cannot be reproduced or quoted extensively from without first obtaining permission in writing from the author.
- The content must not be changed in any way or sold commercially in any format or medium without the formal permission of the author.
- When referring to this work, full bibliographic details including the author, title, awarding institution and date of the thesis must be given.

PHONETIC BIASES AND SYSTEMIC EFFECTS IN
THE ACTUATION OF SOUND CHANGE

Márton Sóskuthy



A thesis submitted in fulfilment of requirements for the degree of

DOCTOR OF PHILOSOPHY

to Linguistics and English Language

School of Philosophy, Psychology and Language Sciences

UNIVERSITY *of* EDINBURGH

March 2013

DECLARATION

I hereby declare that this thesis is of my own composition, and that it contains no material previously submitted for the award of any other degree. The work reported in this thesis has been executed by myself, except where due acknowledgement is made in the text.

Márton Sóskuthy

CONTENTS

DECLARATION •	i
ABSTRACT •	vii
ACKNOWLEDGEMENTS •	xi
1 INTRODUCTION •	1
2 BACKGROUND •	10
2.1 Defining sound change	11
2.2 The actuation problem	13
2.3 Bias-based models	17
2.3.1 The leap model	17
2.3.2 The nudge model	24
2.4 Functionalist models	26
2.5 Sociolinguistic models	32
2.6 The suggested solution	34
3 THEORETICAL ASSUMPTIONS •	37
3.1 Probabilistic category representations	39
3.2 Category-based perception and production	45
3.2.1 Alternatives to the category-based view	45
3.2.2 Evidence for sound categories	49
3.3 The continuous update of categories	60
3.4 Phonetic biases	68
3.5 The filtering listener	74
3.6 Summary	80

4	MODELLING SOUND CHANGE •	82
4.1	The structure of the simulations	85
4.1.1	Category representations and sampling	86
4.1.2	Modelling the effects of phonetic biases	93
4.1.3	Category update and ambiguity	97
4.1.4	Summary of simulation architecture	101
4.2	Comparing prototype and exemplar models	101
4.2.1	Convergence towards the bias	102
4.2.2	Ambiguity-driven dispersion	108
4.2.3	Summary	114
4.3	Single and multi-agent simulations	115
4.4	Conclusions	120
5	A SYSTEMIC VIEW OF SOUND CHANGE •	123
5.1	Adaptive landscapes in sound change	124
5.2	Simulating complex sound systems	130
5.2.1	An example simulation	131
5.2.2	An alternative simulation framework	137
5.2.3	Exploring the adaptive landscape	141
5.3	Why sounds don't change	153
5.4	Why sounds change	158
5.5	The social aspects of sound change	169
5.6	Summary	174
6	LEXICAL FACTORS IN CONTEXTUAL EFFECTS •	176
6.1	Sub-distributions within categories	181
6.2	The voicing effect	191
6.3	Simulating the behaviour of sub-distributions	199
6.3.1	Modelling sub-distributions	200
6.3.2	The separation of sub-distributions	205
6.3.3	Predicting the effect of bias proportion	211
6.4	An empirical study	217
6.4.1	Estimating bias proportion from corpus data	218
6.4.2	The production experiment	220
6.5	Conclusions	228

7 CONCLUSIONS • 231

a APPENDIX: VARIANCE INFLATION • 234

b APPENDIX: THE OBSERVED DISTRIBUTION • 238

c APPENDIX: TARGET LOCATIONS • 241

BIBLIOGRAPHY • 245

ABSTRACT

This thesis investigates the role of phonetic biases and systemic effects in the actuation of sound change through computer simulations and experimental methods. Phonetic biases are physiological and psychoacoustic constraints on speech. One example is vowel undershoot: vowels sometimes fail to reach their phonetic targets due to limitations on the speed of the articulators. Phonetic biases are often paralleled by phonological patterns. For instance, many languages exhibit vowel reduction, a phonologised version of undershoot. To account for these parallels, a number of researchers have proposed that phonetic biases are the causal drive behind sound change. Although this proposal seems to solve the problem of actuation, its success is only apparent: while it might be able to explain situations where sound change occurs, it cannot easily explain the lack of sound change, that is, stasis. Since stability in sound systems seems to be the rule rather than the exception, the bias-based approach cannot provide an adequate account of their diachronic development on its own.

The problem of bias-based accounts stems from their focus on changes affecting individual sound categories, and their neglect of system-wide interactions. The factors that affect speech production and perception define an adaptive landscape. The development of sound systems follows the topology of this landscape. When only a single category is investigated, it is easy to take an overly simplistic view of this landscape, and assume that phonetic biases are the only relevant factor. It is natural that the predicted outcomes will be simple and deterministic if such an approach is adopted. However, when we look at an entire sound system, other pressures such as contrast maintenance also become relevant, and the range of possible outcomes is much more diverse. Phonetic biases can still skew the adaptive landscape towards themselves, making phonetically natural outcomes more likely. However, their effects will often be countered by other pressures, which means that they will not be satisfied in every case.

Sound systems move towards peaks in the adaptive landscape, or local optima, where the different pressures balance each other out. As a result, the system-based approach predicts stability. This stability can be broken by changes in the pressures that define the adaptive landscape. For instance, an increase or a decrease in functional load or a change in lexical distributions can create a situation where the sound system is knocked out of an equilibrium and starts evolving towards a new stable state. In essence, the adaptive landscape can create a moving target for the sound system. This ensures that both stability and change are observed. Therefore, this account makes realistic predictions with respect to the actuation problem.

This argument is developed through a series of computer simulations that follow changes in artificial sound systems. All of these simulations are based on four theoretical assumptions: (i) speech production and perception are based on probabilistic category representations; (ii) these category representations are subject to continuous update throughout the lifetime of an individual; (iii) speech production and perception are affected by low-level universal phonetic biases; and (iv) category update is inhibited in cases where too many ambiguous tokens are produced due to category overlap. Special care is taken to anchor each of these assumptions in empirical results from a variety of fields including phonetics, sociolinguistics and psycholinguistics. Moreover, in order to show that the results described above follow directly from these theoretical assumptions and not other aspects of these models, the thesis demonstrates that exemplar and prototype models produce the same dynamics with respect to the observations above, and that the number of speakers in the model also does not have a significant influence on the outcomes.

Much of the thesis focuses on rather abstract properties of simulated systems, which are difficult to test in a systematic way. The last chapter complements this by presenting a concrete example, which shows how the simulations can be linked to empirical data. Specifically, I look at the effect of lexical factors on the strength of contextual effects in sound categories, using the example of the voicing effect, whereby vowels are longer before voiced obstruents than they are before voiceless ones. The simulations implemented in this chapter predict a larger effect in cases where a given vowel category occurs equally frequently in voiced and voiceless environments, and a smaller difference where one of the

environments dominates the lexical distribution of the vowel. This prediction is borne out in a small cross-linguistic production experiment looking at voicing-conditioned vowel length patterns in French, Hungarian and English. Although this is only one of many predictions that fall out of the theory of sound change developed in this thesis, the success of this experiment is a strong indication that the research questions it brings into focus are worth investigating.

ACKNOWLEDGEMENTS

There are many people who helped me in writing this thesis either directly or indirectly. First of all, I would like to thank Patrick Honeybone, my main supervisor. From the moment I decided to apply to the University of Edinburgh to the very end of the writing process, Patrick was an invaluable source of help and support. He is a brilliant supervisor who helped me turn my ideas into a thesis without ever making me feel pressured or uncomfortable. It was a real pleasure to work with him. James Kirby and Simon Kirby, my other two supervisors, have also contributed greatly to this project. James' ideas on modelling sound change have had an important influence on my own work, and his suggestions inspired me to articulate my arguments as clearly as possible. Simon also helped a lot, especially by involving me in the Language Evolution and Computation Research Unit, which had a crucial role in shaping the way I see language and linguistics.

I am grateful to all the linguists who have commented on various aspects of this work. Andrew Wedel's approach to phonology and sound change has been a major source of inspiration for this thesis, and he has always been more than willing to discuss my ideas with me. Roger Lass was kind enough to listen to a preliminary and somewhat incoherent version of the main argument in this thesis. His encouragement was an important factor in my decision to build my thesis around that argument. Keelin Murray, Péter Rácz and Patrycja Strycharczuk have proof-read various chapters on short notice and provided useful comments. I am also indebted to the following people for their suggestions and ideas: Ricardo Bermúdez-Otero, Richard Ogden, Mits Ota, Jaye Padgett, Sean Roberts, Koen Sebregts, Daniel Silverman, Kenny Smith, Dániel Szeredi, Graeme Trousdale, Marilyn Vihman and Dominic Watt.

I would like to thank the University of Edinburgh for the generous grant that made this work possible. Special thanks are due to Paul Foulkes and the Department of Language and Linguistic Science at the University of York for their support during the last months of my PhD. I am also grateful to the Theoretical Linguistics Department and the Department of English Linguistics at Eötvös Loránd University in Budapest. Without their excellent work, I would likely not be where I am now.

My family and my friends in Edinburgh, Budapest and elsewhere have made the last few years enjoyable despite all the stress and uncertainty that is part of writing a PhD thesis. I am immensely grateful for that (*köszönöm!*). Finally, I would like to thank Esme for not jumping on my face every morning and Amanda Cardoso for her help with the thesis and all the support.

NOW I SHALL tell of the city of Zenobia, which is wonderful in this fashion: though set on dry terrain it stands on high pilings, and the houses are of bamboo and zinc, with many platforms and balconies placed on stilts at various heights, crossing one another, linked by ladders and hanging sidewalks, surmounted by cone-roofed belvederes, barrels storing water, weather vanes, jutting pulleys, and fish poles, and cranes.

No one remembers what need or command or desire drove Zenobia's founders to give their city this form, and so there is no telling whether it was satisfied by the city as we see it today, which has perhaps grown through successive superimpositions from the first, now undecipherable plan. But what is certain is that if you ask an inhabitant of Zenobia to describe their vision of a happy life, it is always a city like Zenobia that he imagines...

Italo Calvino, *Le città invisibili*
(transl. by William Weaver)

Consider the brief excerpt from Italo Calvino's *Le città invisibili* that serves as the epigraph for this thesis. Zenobia is mysterious, because the forces that have shaped it over time are invisible. The city has taken a complex and peculiar form, in many ways reminiscent of a port city, but there is no water, no ships, nothing that would explain its present form. Zenobia has emerged organically through 'successive superimpositions', perhaps best seen as chance responses to the city's transient, ever changing needs; there is no obvious master plan behind it. However, despite all the changes, Zenobia has never ceased to function as a city. Indeed, the inhabitants themselves find no reason to complain about its peculiarities, and see it as an ideal place. That is to say, the Zenobia that has emerged through chance exemplifies a state of orderly chaos – orderly to the extent it needs to be, and chaotic to the extent it is allowed to be.

The development of the city of Zenobia is closely analogous to the way the sound systems of natural languages evolve. Any sound system will present the linguist with a treasure trove of peculiarities, which may appear just as mysterious as Zenobia when taken out of their historical context. Take English as an example. The vowel systems of essentially all varieties of English are extremely complex, especially when compared to other languages. They contain tense and lax vowels, which tend to differ both in quantity and in quality. The tense vowels themselves usually divide into diphthongs (both rising and falling) and monophthongs. Most varieties show at least four different vowel heights, and complex phenomena that result in a number of different values for frontness. On the other hand, English is quite unremarkable in terms of consonantal place of articulation contrasts, with only three contrastive values for oral stops (as opposed to, say, Malayalam with seven different values; Mohanan & Mohanan 1984).

We cannot hope to explain why the sound system of English looks the way it does simply by looking at its present context. Although it is clear that synchronic facts about human physiology and psychology constrain sound systems in many ways, they do not fully determine them: there are many other languages spoken by very similar humans that have completely different sound systems. The sources of such differences lie not in the present context in which a sound system is spoken, but in its history. This is similar to the case of Zenobia: certain features of the city certainly follow from synchronic laws (e.g. the laws of physics), but there are many other facts that we can only hope to understand by looking at its development (e.g. the fact that it is essentially a port city that stands on dry land). Therefore, looking at the present is not enough: we need to enquire into the history of a sound system if we want to understand its present shape. This provides us with a specific type of explanation as to why the system appears as it does by revealing the path through which it has developed. Granted, this is not the only way we can conceive of explanations, but in the context of sound systems it has proven immensely useful.

But is there any way we can describe the historical paths that sound systems walk in a meaningful way? After all, the changes that take place in sound systems are much like the successive superimpositions that Zenobia has undergone: seemingly random and haphazard. There is no master plan. We cannot assume that the people who spoke English five hundred years ago were changing their sound system in an effort to create the sound system that we see today. The solution to this issue has been to take a step back, and try to get a sense of the likelihood of different types of change instead of looking at isolated changes in specific languages. Indeed, when investigated this way, sound systems reveal a much more orderly picture.

One particularly successful historical approach to sound systems makes predictions about the likelihood of changes by focusing on the ‘invisible forces’ behind them: phonetic biases (Ohala 1981, Blevins 2004). Phonetic biases are those aspects of speech that follow directly from the limitations of the physical apparatus that we use to produce and perceive speech sounds. For instance, in some cases vowels may fail to reach their phonetic target due to the sluggishness of the articulators, resulting in vowel centralisation. The main contribution of the bias-based approach is the proposal that frequently observed

changes tend to be exaggerated versions of phonetic biases. To give an example, vowel centralisation can give rise to patterns of vowel reduction, which apply much more categorically and have a more visible influence.

This, however, leads to a paradox. Since phonetic biases are seen as general properties of the physical apparatus used for speech, they are necessarily universal to the extent that humans can be said to use the same ‘hardware’ for production and perception. If this was not the case, the bias-based model could not hope to make predictions about the likelihood of a given type of change in unrelated languages. However, if the same phonetic pressures are there in every language, how is it possible that some languages develop them into robust patterns, while others do not? Even if the predicted likelihood of a phonetically-based change is low, given enough time, every language should eventually yield to the phonetic pressures behind it. This is a point that has been made before in the literature (see e.g. Baker et al. 2011), and it will be explored in much more detail in later chapters. For now, it will be sufficient to note that the bias-based approach seems to ‘overpredict’ sound change. Specifically, it predicts that phonetic biases should result in change in every language where their conditions are met, even though this is clearly not what we observe in natural languages.

The paradox described above is part of a larger issue often referred to as the *actuation problem*. The following passage from Weinreich et al. (1968) provides a clear summary of this problem:

[...] What factors can account for the actuation of changes? Why do changes in a structural feature take place in a particular language at a given time, but not in other languages with the same feature or in the same language at other times? This *actuation problem* can be regarded as the very heart of the matter.

(Weinreich et al. 1968: p. 102)

One might, of course, argue that the actuation problem becomes irrelevant when the focus is not on individual changes but on the relative probabilities of different types of change. In other words, the actuation problem might seem less worrying if we are not trying to predict exactly when and where a given change will take place, but comparing different types of change in terms of their likelihood. This, however, is not a valid argument: even if the goal is not to

make predictions about specific changes, a theory that implies that the effects of a given phonetic bias should be visible in every language cannot be right. Note that there is also another way in which a model of sound change can fail with respect to the actuation problem: by predicting that sound change will not take place in cases where it does. A successful model of sound change will avoid both of these issues by relying on ‘a mechanism of change that does not underapply (failing to predict cases where change occurs) or overapply (failing to predict cases where change does not occur)’ (Baker et al. 2011: p. 349).

The main goal of this thesis is to provide a solution to the actuation problem in the context of bias-based models of sound change. Specifically, the task is to find a mechanism of change that allows us to capture the cross-linguistic tendencies predicted by bias-based models, but does not overapply or under-apply. The solution presented in this thesis is necessarily limited in some ways. Thus, while the mechanism proposed in the following chapters is shown to work well for certain types of sound change, it is not possible to fully explore its implications for all types of change. The main focus is on category-wide shifts that do not lead either to mergers or splits. This is not to say that this account makes no predictions with respect to mergers and splits. Indeed, Chapters 5 and 6 will review many such predictions. However, the arguments presented in the thesis are most directly applicable to changes where the number of categories remains constant.

The task undertaken here is not to predict when and where sound change will take place: given the large number of factors that could influence the development of a sound system, such a prediction may well be impossible. Instead, the account presented in this thesis simply aims to increase the predictive power of bias-based models. This will be achieved by showing how phonetic biases interact with other – perhaps less obvious – pressures within a sound system (e.g. the implicit tendency towards contrast maintenance, described in much more detail in later chapters). One of the most important outcomes of the investigation presented in the next chapters is that an exclusive focus on biases is unlikely to yield a truly explanatory account of sound change.

The central argument of the thesis can be summarised as follows. The notion of phonetic bias is not the main reason why bias-based mechanisms fail to account for cases where no change occurs. The source of the overapplication

problem lies in the way bias-based models are used to approach specific phenomena. Typically, sound changes are viewed as alterations that affect a given sound category, or perhaps a set of categories that share a common feature. For instance, the phenomenon of [u]-fronting could be described as a change where a back vowel shifts its realisation and becomes a front vowel [ɥ]. By ignoring all the other categories in the sound system, this approach treats vowel categories as if they existed in a vacuum – and this is why it runs into problems when trying to explain how they can resist change.

An analogy will make this argument clearer. Consider the effects of gravity on physical bodies in a vacuum. If a feather and a hammer are dropped from the same height in a vacuum, they will accelerate at exactly the same rate, and hit the ground at the same time. Indeed, this experiment has been successfully performed on the surface of the moon by astronaut David Scott during the Apollo 15 mission (which is not to say that the validity of the principle behind this experiment had not been confirmed before). Nevertheless, anyone who has ever dropped a feather on Earth will know that this description of the movement of physical bodies in a vacuum does not carry over to their movement in air. In a vacuum, both of the objects fall at the same speed because gravity is the only force affecting them, and there is nothing that would counter their momentum. However, they fall at different speeds in air because of additional forces like aerodynamic drag.

Similarly, an approach that investigates sound categories in a vacuum will overestimate the likelihood of sound change by disregarding the forces that could hinder it. This thesis places sound change in a more realistic context by considering it in relation to sound systems rather than isolated categories. I will show that sound systems are more resistant to changes because their evolution is determined by a number of different factors, the most important of which are phonetic biases and a tendency towards dispersion. These factors conspire to create a complex adaptive landscape, which guides the evolution of the system. One crucial feature of this landscape is the existence of multiple stable states: a sound system may come to be arranged in a way that the different pressures balance each other out, creating an equilibrium. Although a phonetic bias might make certain stable states statistically more likely, sound systems will often settle into equilibria that do not satisfy a given phonetic bias. Note that the fact

that sound systems inevitably move towards such equilibria exposes this account to criticism based on the underapplication aspect of the actuation problem: it seems that this model predicts stability even where changes occur. As it will be seen, this problem is only apparent. The pressures defining the adaptive landscape (e.g. the functional loads of different oppositions) are themselves subject to change, which can alter the stable states. When such alterations occur, the sound system might be knocked out of an equilibrium and move towards a new state.

In the rest of this thesis, this argument will be developed in a much more rigorous way by deriving testable hypotheses from first principles. In order to do this, I will use computer simulations. The evolution of sound systems under multiple pressures is a complex phenomenon which cannot simply be investigated through thought experiments. Even simple models looking at categories in a vacuum might produce unexpected results, and surprising outcomes tend to be the rule rather than the exception when the systems under investigation are complex. For this reason, every effort is taken to ensure that the underlying theory is linked to the argumentation in a systematic fashion. This will be achieved by exploring the predictions of the bias-based model through a large set of computer simulations. The arguments about how the system-based approach to sound change presents a solution to the actuation problem are based directly on the predictions that emerge from the simulations.

Before presenting an outline of the thesis, I will briefly discuss one further point related to it: its focus on predictions. In the discussion above, I noted that the problem with bias-based approaches lies not in their main assumptions, but in the way these assumptions are employed. Specifically, I argued that the solution to the actuation problem does not require us to abandon the bias-based approach altogether, only to shift our attention from sound categories to sound systems. In a sense, then, the problem is not with the theory, but with its apparent predictions. The main contribution of this thesis is to amend and consolidate these predictions by using a systematic method in deriving them, namely, computer simulations. This, however, creates a paradoxical situation with respect to the audience that this thesis is aimed at. Since the focus is on predictions, much of the discussion is fairly abstract and theoretical, and will likely appeal to theorists who are concerned with similarly abstract

issues. However, the researchers who will benefit the most from the framework presented here are those who wish to apply it in the investigation of concrete phenomena. The predictions discussed in the following chapters may help to see old phenomena in a new light, and suggest lines of research that have not been explored before. In an attempt to bring the results of this thesis closer to researchers focusing on empirical problems, I have included a chapter that shows how the insights gained through this framework can be put to use in a cross-linguistic investigation (Chapter 6). Of course, this does not make the rest of the discussion any less abstract, but it will be sufficient to show that this approach holds significant promise in the realm of empirical research.

In what follows, I present an overview of the structure of this thesis. The outline below will also clarify some of the points above by showing how the thesis will approach them.

CHAPTER 2 presents the problems that form the starting point for the main argument of the thesis. The first half of the chapter defines two key concepts: sound change and the actuation problem. The rest of the chapter discusses a number of theoretical approaches that have important implications for the latter of these. This includes two slightly different bias-based models, one of which will serve as the basis of the simulations presented later in the thesis. There are two further approaches that are considered: functionalist accounts that attribute a special role to functional load in the actuation of sound change; and sociolinguistic approaches that see the problem of actuation as inseparable from the social aspects of language. It will be shown that none of these approaches can provide a satisfactory answer to the actuation problem, mainly because they focus on particular aspects of sound change at the expense of others. I will argue that the system-based view of sound change holds more promise with respect to the actuation problem, and briefly explain how computational simulations can help us approach these issues in a more rigorous way.

CHAPTER 3 elaborates on the main theoretical assumptions of the model of speech production and perception that will serve as the basis of the simulations. This is a crucial step in the main argument of the thesis: the underlying theory has to be specified explicitly if we want to explore its predictions. The following

theoretical assumptions are discussed: (i) speech production and perception are based on probabilistic category representations; (ii) these category representations are subject to continuous update throughout the lifetime of an individual; (iii) speech production and perception are affected by low-level universal phonetic biases; and (iv) category update is inhibited in cases where too many ambiguous tokens are produced due to category overlap. Although these assumptions are shared with a number of existing bias-based approaches, special care is taken to provide independent motivation for each them. This is done by reviewing results from a variety of fields including phonetics, sociolinguistics and psycholinguistics. The detailed arguments presented in this chapter establish the generality of the results discussed in later chapters. While this thesis can be viewed as a response to certain issues that arise specifically in the context of bias-based models, the plausibility of its fundamental assumptions implies that its findings are relevant to other approaches as well.

CHAPTER 4 outlines the technical aspects of the model that serves as the basis of the simulations, and clarifies a number of controversial points related to them. The chapter also provides novel answers to two broader research questions that are significant in their own right. First, I show that exemplar and prototype-based models of speech production and perception produce the same general dynamics with respect to sound change. Second, I present a comparison of multi-agent simulations and simulations relying on an abstract version of the production-perception feedback loop. The conclusion of this comparison is that the dynamics of these models are essentially the same with respect to sound change. These results provide additional support for previous simulation-based investigations of sound change, and serve as the basis of a number of decisions relating to the implementation of the simulations in the following two chapters.

CHAPTER 5 presents the main argument of the thesis based on a large-scale simulation-based investigation. I introduce the notion of adaptive landscapes in sound change, and use it to analyse the results of a large set of simulations. The simulations themselves look at the evolution of artificial sound systems under a number of different pressures, some of which are varied in order to

get a better sense of their effects. The conclusion of this investigation is that the bias-based model can make valid predictions about the actuation of sound change when applied to sound systems. This chapter also helps to highlight the role of non-phonetic factors in sound change, such as lexical distributions, functional load, and individual differences in production and perception. In fact, one of the most significant predictions that emerge from the simulations is that shifts in such non-phonetic factors can play a major role in initiating sound change. The results of the simulations are also linked to sociolinguistic approaches to sound change. It is shown that the account developed in this thesis can explain certain aspects of the distinction between changes from below and changes from above.

CHAPTER 6 illustrates how the predictions of the bias-based model can be tested by investigating the relationship between lexical distributions and gradient contextual effects. The specific prediction examined in this chapter can best be explained through a concrete phenomenon: the voicing effect. The voicing effect is a nearly universally observed interaction between vowel length and the voicing of a following obstruent, whereby vowels are longer before voiced obstruents than they are before voiceless ones. The prediction of the model is that the voicing effect will be stronger in cases where a given vowel category occurs equally frequently in voiced and voiceless environments, and weaker when one of the environments dominates the lexical distribution of the vowel. In effect, this is a prediction about the likelihood of allophonic splits on the basis of lexical information. The first half of the chapter shows how this prediction can be derived from the bias-based model through a set of simulations and mathematical calculations. The second half of the chapter then tests the prediction on vowel length data from a small cross-linguistic production experiment involving English, French and Hungarian. The data will be shown to provide strong support for the prediction, which suggests that the implications of the systemic approach to sound change are well worth exploring.

CHAPTER 7 concludes the thesis with a brief summary of its main points.

As it has already been explained in the introduction, this thesis provides fresh insight into the actuation problem, that is, the question of why sounds change in certain situations and not in others. This chapter sets the scene for the discussion in the rest of the thesis by clarifying some of the key concepts and giving an overview of previous approaches to the actuation problem. The structure of the chapter is as follows. Section 2.1 is a brief discussion of the notion of sound change, which serves to delimit the range of phenomena investigated in the thesis and to illustrate a number of different ways in which sound change can be conceptualised. Then, Section 2.2 provides an explicit statement of the actuation problem. Sections 2.3–2.5 outline three possible solutions to this problem: bias-based, functionalist and sociolinguistic explanations. Each of these solutions is examined in detail, but they are all found lacking in certain respects. Finally, Section 2.6 provides a short summary of the system-based approach taken in the rest of this thesis, suggesting that it can avoid the problems associated with previous approaches by combining some of their key features.

2.1 DEFINING SOUND CHANGE

On an intuitive level, the notion of sound change seems so straightforward that it might not be clear why it should be clarified at all. A sound change is no more and no less than what its name suggests: a change affecting speech sounds that takes place over a given period of time. However, the task of defining sound change turns out much more complicated on closer inspection. There are many questions that a rigorous definition has to answer. First of all, where do speech sounds exist – in the mental grammars of individuals, in the shared language of a speech community, or perhaps only in concrete

utterances? There is no universally accepted answer to this question, and all of these alternatives have been explored by researchers (see Hale 2003 for an example of the first approach, Weinreich et al. 1968 for the second and Croft 2000 for the third). Second, should we include in the definition smaller gradient changes in the realisation of a given category, or should it be restricted to changes that affect the phonemic structure of the language? Again, both approaches have been taken in the past (cf. Bybee 2001 vs. Hoenigswald 1960, respectively). If gradient changes are included as well, where do we draw the boundary between sound change and small chance modifications? Finally, is the locus of sound change necessarily a given sound category, or could sound change be meaningfully discussed at higher and lower levels as well? Does it make sense to talk about changes to an entire sound system?

While all of these questions have important theoretical ramifications, only the last one will be dealt with in any detail. Contrarily to almost all other concepts in this thesis, the use of the term sound change will be guided by convenience rather than theoretical considerations. Thus, I will not take a position on the issue of whether the changing sound categories belong to mental grammars, communities of speakers or concrete utterances. In a sense, all of these play an important role in the account developed in the following chapters, but singling out one of them as fundamental would not help in moving the discussion ahead. Similarly, the issue of gradience will not be central to the argumentation of the thesis. In the present account, both phonemic and non-phonemic alterations are regarded as sound change. As to the point beyond which a gradient shift should be regarded as sound change, no such cut-off will be defined. In a strict sense, even very small shifts should be considered sound change as long as their effects are consistently observed. An informal distinction will be made between sound changes guided by phonetic biases that result in small shifts (such as vowel centralisation), and those that yield robust patterns reflecting the same bias (such as vowel reduction). The term sound change will be used mainly to refer to the latter of these in the context of the actuation problem. However, it is important to note that this distinction will not be afforded any theoretical status (this point is discussed in more detail in Sections 3.4 and 6.1).

The one aspect of the definition that will be given more attention is the locus of sound change. As I explained in the previous chapter, one of the main

innovations of the line of enquiry pursued in this thesis is the shift in focus from individual categories to sound systems. This does not affect the theoretical underpinnings of the account. Categories still play an important role in the model of speech production and perception that underlies the main argument of the thesis. I also do not wish to claim that sound change cannot be analysed at the level of sound categories. However, I believe that the exclusive focus on individual categories that prevails in studies of sound change is not warranted. Sound systems have much richer evolutionary dynamics than categories in a vacuum, and this will prove crucial to the present account. Since there is no *a priori* reason why sound change should only be investigated with respect to individual categories, the present account is fully justified in its focus on systemic changes. Note that the system-based view will only be fully explored in Chapter 5, where the main argument of the thesis is developed. Chapters 3 and 4 will express their main points in the more familiar language of category-based models.

It is worth noting that while mergers and splits are considered sound change in the present account, the simulated changes in Chapter 5 find their closest parallel in chain shifts. This is due to a simplification in the model adopted in this thesis, namely that it cannot increase or decrease the number of categories in a given system. This problem is partly dealt with in Chapter 6, which explores the predictions of the model with respect to splits, but mergers will not be investigated in any detail. Although the inability to simulate mergers is a serious limitation, the model does make some predictions about the likelihood of situations which might serve as the precursors to mergers. Section 5.3 uses the example of [u]-fronting to explain how phonetic biases and lexical factors can determine the probability that a large amount of overlap will arise between two sound categories. In this sense, the account does capture some properties of mergers, even if further modifications are needed to fully account for them.

2.2 THE ACTUATION PROBLEM

The previous chapter has already introduced the actuation problem briefly. However, since this concept plays a central role in this thesis, a more careful discussion will be necessary. One way to formulate the actuation problem is to ask why a specific sound change takes place in a given language at a given

time, but not in a different language or in the same language at a different time (cf. Weinreich et al. 1968). It is unlikely that we will ever be able to answer this question with any certainty, given the large number of factors that might interact in the actuation of sound change. Therefore, if the actuation puzzle is formulated this way, there is little hope that any model of sound change will provide a satisfactory solution. There is, however, a somewhat more productive way of approaching the same issue. The actuation problem can be viewed as a question about the likelihood of a given change under specific circumstances. We might not be able to predict exactly when and where a change will occur, but it should still be possible to investigate the factors that facilitate or hinder it. This significantly widens the range of models that can be said to be successful with respect to the actuation problem. Given the focus on the likelihood of change, such a model will satisfy the following set of criteria.

- (2.1) A model of sound change can be said to contribute to the solution of the actuation problem if
- a. it does not overestimate the probability that change should occur;
 - b. it does not underestimate the probability that change should occur;
 - c. it contributes to our understanding of the factors that make sound change more or less likely.

For the purposes of this thesis, ‘overestimating’ or ‘underestimating’ the probability of change simply means predicting that a given type of change will always occur or that it will never occur. As it has already been noted, calculating the precise probability of a sound change in a given language may not be possible. However, models implying that cross-linguistically variable changes should always (or never) take place are clearly wrong. Sections 2.3–2.5 demonstrate how such ‘catastrophic’ instances of overprediction and underprediction can emerge in models of sound change.

Approaching the actuation problem this way might be seen as a compromise. If all the factors behind sound change are understood, we should be able to make accurate predictions about when and where it will occur. While this is certainly true, it is questionable whether any scientific endeavour can realistically aim to uncover all the predictors relating to a given phenomenon. Moreover, even if

the set of relevant factors is finite and possible to explore, simply including all of them in a model does not guarantee a better understanding of the phenomenon. This follows from the observation that correlation does not imply causation. We can construct an excellent predictive model based on correlations among various factors, but our understanding of the phenomenon will only be improved if we can show how these factors facilitate or inhibit it. Thus, creating accurate predictions cannot be our only aim in approaching the actuation problem. In this sense, focusing on the likelihood of changes is not only a more realistic, but also a scientifically more justifiable stance.

Note that the issues outlined above are closely related to the question of stasis *versus* change in the history of a language. While the foregoing discussion has focused mainly on change, stability is an equally important property of sound systems (cf. Lass 1997: p. 303–304). It is not clear whether there can be stages in the evolution of a language where all aspects of the sound system remain stable. However, sound change tends to be relatively slow in the sense that language states several generations apart are usually mutually intelligible and easily recognised as representing the same language. Moreover, particular features of a given sound system can resist changes for long periods of time. For instance, the word initial [θ] in English words like *three* has remained unchanged since Proto-Germanic times.¹ This is especially interesting given that many other Germanic languages have lost this sound in their history. While it might be difficult to estimate how stable a given sound system is over a certain period of time, it is clear that a successful theory of sound change will be able to account both for stasis and change (see Milroy 1992: p. 10 for a similar conclusion).

A further important requirement with respect to the solution to the actuation problem is that it cannot explain sound change by attributing goal-directed behaviour to speakers. For example, it has been argued that sound change can be inhibited in cases where it would lead to the loss of a contrast. This is a legitimate approach to the actuation problem inasmuch as it focuses on a factor that can influence the likelihood of sound change. However, Lass (1997: p. 361)

1. There are many dialects of English where this statement would not hold (e.g. Cockney). Note that it is only the phonetic makeup of this sound that has remained the same – the appearance of [ð] in word-initial position has likely changed its specific role in the system of contrasts in English.

argues that such explanations cannot be taken seriously if they assume that the behaviour of speakers during sound change is guided by ‘intuitions about the “efficiency” or “optimality”, etc. of their language for its communicational tasks’. It is not clear where these intuitions would come from, and how they would be put to use in the course of sound change. We cannot simply propose a causal mechanism for sound change that is based on vague notions of optimality and has no independent support. Therefore, in this thesis, extra care will be taken to ground the proposed model of sound change in well-supported observations about speech production and perception. Sound change will be seen to emerge from simple properties of language use, not from goal-directed behaviour.

There are a number of different models that make relatively clear predictions about the actuation of sound change, and can therefore be evaluated with respect to the criteria in (2.1). Since these approaches will be the focus of Sections 2.3–2.5, I only provide a brief outline here. Bias-based models have already been introduced in the previous chapter. These models view sound change as the exaggeration of low-level phonetic biases, and can successfully account for a number of interesting cross-linguistic parallels. While they satisfy the criteria in (2.1b) and (2.1c), they run into problems when it comes to (2.1a) – that is, they seem to overpredict sound change. A second class of models – often labelled *functionalist* – approach the notion of sound change from a slightly different angle, by focusing on the interplay between ease of articulation and communicative efficiency. The main advantage of such approaches is their ability to account for cases where sound change seems to be inhibited to avoid mergers. These models appear to do slightly better with respect to the overapplication problem, given that they can deal with certain cases where sound change fails to take place. However, we will see that functionalist models cannot provide a satisfactory solution to the actuation problem on their own, and that they run into problems regarding the goal-directedness of sound change. The last type of approach to be discussed in this chapter comprises sociolinguistic models which view the actuation problem as a question about the social environment in which the sound system is embedded. As we will see, such models satisfy the criteria in (2.1a) and (2.1b), but they are somewhat problematic with respect to (2.1c). This is because sociolinguistic models tend to focus mainly on the social aspects of sound change, and they often make no predictions about the effects of intra-linguistic factors on the likelihood of change.

2.3 BIAS-BASED MODELS

In this section, two different bias-based models are discussed, one of which is labelled the *leap model* (2.3.1) and the other one the *nudge model* (2.3.2).² The models are built around the same general idea, namely that gradient phonetic biases can lead to more robust patterns through sound change. However, they implement this idea in rather different ways. Despite their differences, both of the models will be shown to run into trouble with regard to the actuation problem, by overestimating the likelihood of sound change.

2.3.1 The leap model

The term ‘leap model’ will be used to refer to the account of sound change originally proposed by Ohala (1981, 1989, 1993) and further developed by Blevins (2004, 2006). The reason for choosing this name lies in the nature of the mechanism of change adopted in these accounts: sound change is seen as an abrupt shift – that is, a leap – from one phonetic target to a different one. The description here focuses on Ohala’s original approach, mainly for reasons of brevity. It should be noted that the arguments in this section carry over to Blevins’ model relatively straightforwardly, given the general similarity between her and Ohala’s approach.

The leap model is centred around one specific class of diachronic phenomena: frequently observed sound changes. Ohala (1989) is quite explicit about this point:

In my own work I impose a (for me) useful restriction: I study those sound changes attested in similar form in diverse languages. This helps to guarantee that they will owe something to universal and timeless physical or physiological factors [...] and not to language- or culture-specific factors.

(Ohala 1989: p. 174)

When looking at such changes, an important observation emerges: they nearly always have parallels among patterns of synchronic variation (see Blevins 2004

2. The latter term is not my own invention: it was coined by Andrew Wedel in a conversation about different models of sound change.

for a more detailed discussion of this point). One of the examples provided by Ohala (1989: p. 177) concerns aerodynamic constraints on voicing in stops. In order to produce voicing, supraglottal pressure needs to be lower than subglottal pressure – otherwise, the air will simply not flow through the glottis. During the closure phase of voiced stops, supraglottal pressure increases continuously, since the air cannot escape the oral cavity due to the obstruction. At a certain point, this increase results in a situation where voicing simply cannot be sustained any longer. The exact point at which this happens is affected by the place of articulation of the stop: if the obstruction is in the front (as opposed to the back) region of the oral cavity, more air can accumulate and therefore voicing can be sustained for longer. These aerodynamic constraints are reflected in the sound patterns of numerous languages. Thus, languages with a single laryngeal category tend to have only voiceless obstruents, obstruent devoicing is an extremely widespread phenomenon (especially in final position), and back consonants are more likely to be devoiced than front ones. Although these are all synchronic observations, there is no doubt that such patterns emerge through sound change.

Of course, these parallels have been noted by other researchers as well. For instance, Hyman (1975, 1976) presents an account in which low-level phonetic forces are seen as the sources of robust language-specific patterns. He uses the term ‘phonologisation’ to refer to this process. However, such accounts typically leave the details and the causes of phonologisation unspecified. In this respect, the leap model improves significantly on previous approaches: it describes an explicit mechanism through which phonologisation can take place.

Ohala (1989) argues that the variation caused by phonetic biases does not normally lead to change given that experienced listeners will have learnt to filter it out. To give an example, when a velar stop is partially devoiced, the listener will be able to restore the original voiced stop by relying on their own phonetic experience. They will be aware that velar stops are often only partly voiced, and will adjust their perceptual expectations accordingly. Ohala (1989) suggests two ways in which this type of perceptual compensation may fail, leading to sound change. First, the listener may not apply the appropriate corrective rule to the token produced by the speaker, and assume that the distorted variant is the intended target (this is referred to as ‘hypocorrection’

by Ohala). In the case of partial devoicing in velar stops, this would mean that the listener interprets the devoiced token as the speaker's target. This leads to a mini-sound change. There is also another way for perceptual compensation to fail: it may overapply, mistakenly undoing phonetic effects that correspond to the speaker's intentions. This mechanism is labelled 'hypercorrection', and it provides a straightforward explanation for many cases of dissimilation. For instance, in Shona, [w] changed to [ɣ] in contexts where it was preceded by another labial sound: [-bwa] > [-bɣa] 'dog', [kumwa] > [kumɣa] 'to drink' (Ohala 1989: p. 188). Ohala suggests that this is due to hypercorrection: Shona listeners misinterpreted the labiality of the original glide as a coarticulatory effect due to the preceding labial consonant, and erroneously factored it out.

A few remarks are in order about the general properties of this model. First of all, it treats sound change as a categorical process: while the synchronic variation that gives rise to different phonetic variants may be gradient, the results of misperception are necessarily discrete. That is to say, in Ohala's account a phonetically devoiced stop can only be interpreted as voiced or voiceless, but not as partially devoiced. Similarly, Shona listeners will either correctly perceive post-labial tokens of [w] or misinterpret them as [ɣ], but they will not adopt an intermediate form. The changes proposed by Ohala necessarily proceed in categorical leaps rather than smaller gradient steps. Another important property of this approach is its avoidance of goal-oriented explanations. Hypercorrection and hypocorrection are both innocent mistakes on the part of the listener, who seems to be the victim of sound change, not the perpetrator. Since perceptual compensation is a well-documented phenomenon (cf. Section 6.1), and it is plausible to assume that it might fail under certain circumstances, Ohala's approach can account for sound change without invoking *ad hoc* explanations based on optimality.

The leap model is certainly impressive in its ability to capture the phonetic grounding of sound change, and it might well be indispensable in accounting for phenomena like dissimilation and metathesis (see Blevins & Garrett 2004 for a more detailed discussion of the latter). Nevertheless, the model has a number of serious flaws as well. First of all, it is not clear to what extent hypocorrection and hypercorrection can lead to observable instances of sound change. Let us first discuss hypocorrection. Consider one of the examples provided by Ohala

(1989: p. 185). A sound like [t] is likely to undergo some degree of affrication when followed by a high vowel such as [i], yielding a realisation that could be transcribed as [tʃ]. If the listener fails to undo the effect of the context, they might erroneously assume that the speaker's original target was [tʃ]. According to Ohala, this misinterpretation constitutes a mini-sound change in itself. The question is whether this mini-sound change has any observable effects. Ohala seems to suggest that the change is in the underlying representation of the sound, although he is careful not to use this term. However, not much seems to change in terms of surface realisations: the speaker originally pronounced the sound as [tʃ], and the listener continues to use the same realisation, even if their underlying representations are now different. Thus, it appears that hypocorrection may not have any observable effects, which makes it somewhat questionable as a mechanism of sound change.

We can now turn to hypercorrection. For instance, imagine that the speaker produces an affricate [tʃ] followed by an [i], where the affricate is the intended target. The listener may attribute the affrication to the following vowel, and reconstruct the target as [t]. Again, this will result in a change in target productions, but it is not clear how it will affect surface realisations. Ohala 1989 suggests that the affrication of coronal stops before high vowels is the result of a universal phonetic bias. Therefore, the listener's future productions of the reconstructed consonant should still show affrication in this position. As a result, hypercorrection may be similar to hypocorrection in that it may not have observable effects.³

The arguments presented above clearly go against Ohala's (1989) own account, who seems to suggest that such reanalyses do have observable effects. However, Ohala does not seem eager to spell out how exactly such effects arise. The only relevant mention of changes in phonetic realisation is in a passage about the emergence of contrastive vowel nasalisation: '[i]t is undoubtedly the case that when the vowel nasalization was taken to be distinctive by the listener

3. Note that the effects of hypercorrection would likely be visible if speakers could differ in whether they show affrication or not. A listener who does not actively affricate consonants in their own speech may still be aware that other speakers do, and erroneously interpret the sequence [tʃi] as [ti]. In this case, their own future productions would differ from the speaker's original production. However, Ohala (1989) is quite explicit about the universality of such phonetic tendencies, which speaks strongly against such a scenario.

it would have an exaggerated quality in his speech vis-à-vis the speech of the original speaker' (Ohala 1989: p. 187). The idea of such a phonetic exaggeration might seem appealing on an intuitive level, but Ohala does not explain why it would take place. In fact, this proposal seems to go against the idea of innocent misinterpretation as the source of change. If speakers are aware of fine details of phonetic realisation, how can they reinterpret a contextual variant as a target production when the latter are 'undoubtedly' more exaggerated in their realisation than contextual variants? Should they not perceive such differences in production and use them to distinguish between phonetic targets and contextual variants? Paradoxically, then, the leap model predicts either that no observable change will occur, or that hypoarticulation/hyperarticulation should be blocked due to the listener's awareness of differences between target productions and phonetic effects. Even if this criticism is a little too harsh, the leap model would certainly benefit from a more explicit statement of how reanalysis leads to changes in surface phonetic realisations.

Another weakness of the model concerns its inability to account for the spread of a change in a speech community. It could be argued that this is not a problem: the phonetic initiation of sound change constitutes a legitimate research area on its own, and does not necessarily have to be studied in conjunction with the social aspects of sound change. In fact, the leap model could be viewed as the source of the 'inherent variation in speech' that can '[assume] direction and [take] on the character of orderly differentiation' within the speech community, leading to community-wide changes (Weinreich et al. 1968: p. 187). However, on closer inspection, the mechanism of change proposed by Ohala turns out to be incompatible with such accounts. The reason for this incompatibility should be obvious if hypercorrection and hypocorrection have no observable effects (cf. the discussion above): even the most observant speech community will have trouble picking up on invisible patterns of variation. Yet this is exactly what Blevins (2004) seems to propose: she claims that certain changes might 'slip into a child's grammar almost unnoticed, providing the seeds for one source of phonetically based sound change within the wider speech community' (Blevins 2004: p. 35; note that Blevins refers to the child's grammar since in her account the transmission errors leading to change take place during language acquisition). I believe it is relatively uncontroversial that children aim

to approximate the language varieties spoken in their environment as closely as possible during language acquisition. Therefore, it is highly implausible to suggest that a child will simply ignore a change that the speech community can easily perceive and propagate.

Since unobservable changes cannot be propagated, let us assume that hypocorrection and hypercorrection can lead to exaggerated variants as suggested by Ohala (1989). The question, then, is whether such variants can make their way into the speech community at large. In principle, this should be possible. Note, however, that both Ohala and Blevins suggest that the reanalyses leading to sound change are most likely to occur in ‘inexperienced listener[s]’ (Ohala 1989: p. 186), that is, children acquiring language. If this is the case, it is very unlikely that a misperception-driven change will take off in the community, given that young children have virtually no social status, which is essential for spreading an innovative pronunciation (see e.g. Labov 2002). Therefore, the leap model cannot easily account for the propagation of sound change in a speech community.

As it has been suggested earlier in this chapter, the leap model also makes implausible predictions with respect to the actuation problem. Ohala (1989) argues that the leap model is about predicting the likelihood of change, but not when and where a given change will occur – and in this respect it is highly compatible with the approach to the actuation problem outlined in the previous section. However, a closer look at the predictions of the model reveals that it does not satisfy all the criteria listed in (2.1): it greatly exaggerates the likelihood of change, and therefore it seems unable to account for stasis. The arguments presented below are based mainly on Baker et al. (2011). The inevitability of sound change in the leap model follows from the simple mathematical observation that even low-probability events become highly likely given a sufficiently long period of time or a sufficiently large sample. Thus, it might be the case that the probability that hypocorrection or hypercorrection will take place in a single individual during language acquisition is extremely low – say, $p = 0.01$ (i.e. one per cent). But the fact that misapprehension-based sound change is unlikely to take place in a single individual does not mean that it will be entirely absent from the speech community. Consider a community with 10,000 children: in this case, there is a nearly one hundred per cent chance

that there will be at least 50 children for whom a mini-sound change takes place (assuming that the probability of change within one child is $p = 0.01$; this can be calculated exactly using the binomial distribution, as Baker et al. 2011 do). The probability of change will be even higher if we look at several generations of speakers. In fact, no matter how small the probability of a mini-sound change within a single speaker, the overall probability that at least some speakers will exhibit such a change in the community always converges to one hundred per cent as the number of generations or the size of the community is increased (cf. Baker et al. 2011: pp. 361–364).

It is not clear whether 50 out of 10,000 children are sufficient to initiate a sound change. However, proponents of the leap model appear to suggest that these sporadic misperceptions are all it takes for sound change to take off. None of the works cited above mention a critical threshold for a mini-sound change to start spreading within a speech community, and these two kinds of change are often implicitly equated in the actual analyses. As long as the relationship between mini-sound changes and community-wide changes is viewed in such simplistic terms, Baker et al.'s (2011) criticism remains valid. If the primary mechanism behind sound change is misapprehension, it is not at all clear how any language could resist the effects of a given phonetic bias.

To sum up, the leap model provides an interesting way of capturing parallels between phonetic biases and the evolution of sound systems, but it appears unable to account for certain finer details of sound change. In the discussion above, three particularly problematic areas were identified. First, it is not clear how hypocorrection and hypercorrection result in changes that are observable in the surface realisation of a given sound category. Second, the leap model makes the propagation of sound change appear somewhat of a mystery, partly because some of the changes it describes have no observable effects, and partly because the proposed agents of sound change – that is, children – are very unlikely to have an effect on the speech patterns of a community. Finally, the leap model makes problematic predictions with respect to the actuation of sound change: it overestimates the likelihood of change, which makes it unable to account for stasis, thereby violating the criterion in (2.1a).

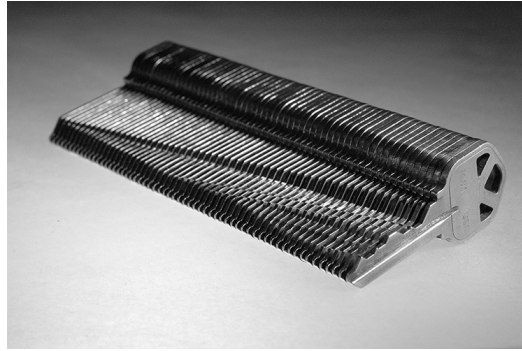


Figure 2.1: A unique object created by artist Daniel Bejar, which consists of a series of keys obtained through iterative copying (image by courtesy of the artist).

2.3.2 The nudge model

In this section, I discuss another bias-based approach to sound change, sometimes referred to as ‘usage-based’ modelling. The account presented here is based mostly on Pierrehumbert (2001, 2002) and Wedel (2006), but there is a wide range of other works that rely on very similar assumptions, including Bybee (2001, 2007), Hay & Sudbury (2005), Phillips (2006) and Silverman (2006, 2012), among others. In this thesis, I relabel this approach as the *nudge model*, as it views the effects of phonetic biases on sound categories as small gradient nudges in a given direction in phonetic space. Since this model will serve as the basis of much of the discussion in the following chapters, only a brief outline is given here. Chapters 3 and 4 present the theoretical assumptions and technical aspects of the nudge model in far greater detail.

Pierrehumbert (2001) argues that many instances of phonetically driven sound change can be explained if we assume that the effects of weak phonetic biases can accumulate in category representations through the so-called production-perception feedback loop. The following analogy will make this point clearer. Consider the picture in Figure 2.1, which shows a unique object created by artist Daniel Bejar.⁴ This object consists of a series of keys that are glued together. What is important for our present purposes is the way the keys were created. The process was initiated by taking the key on the right-hand side, and making a copy of it – this is the key that can be seen immediately after the

4. Also available on the artist’s website: <http://www.danielbejar.com>.

original one. This copy was then copied again, giving the third key. The whole series was obtained by iterating this procedure over and over again. Importantly, each copy is nearly identical to the key it is based on, but it is not entirely the same. The machine that performs the copying introduces small changes at each step. These changes are relatively consistent: the original outline of the key seems to be shifted leftward, and some details are lost. By repeating this procedure many times, the small bias inherent in the machine is amplified into a robust change. Indeed, it is very unlikely that the key at the far end would fit into the lock that the original key was intended for.

The production-perception loop in Pierrehumbert's (2001) account is based on a very similar mechanism. When an example of a sound category is produced, it is influenced by phonetic biases, which can cause small but consistent distortions. For instance, any production of a given vowel has a chance of becoming slightly centralised through vowel undershoot. If this production is then fed back into the category representation from which it originated, the category will be nudged towards the centre of phonetic space. Of course, this nudge will be nearly imperceptible – a single deviant token is unlikely to produce a large restructuring of category representations. If, however, the phonetic bias applies consistently, the effects of these nudges can accumulate in category representations, resulting in robust patterns of change. The mechanism that gives rise to such changes is essentially identical to the one that led to the large-scale changes in Bejar's keys.

This model can answer some of the criticism levelled against the leap model in the previous section. First, I argued that it is not clear how the leap model generates observable changes, given that it seems to apply at the level of underlying representations. This is not an issue for the nudge model: in this case, it is the phonetic realisation of the category that changes, and not the underlying label associated with it. Therefore, the shifts emerging through the production-perception feedback loop are always visible. The second argument I made in the previous section concerns the propagation of innovative variants in the speech community. Since the nudge model produces changes in the surface realisations of sound categories, it avoids the 'invisible changes' problem (namely that shifts which are only apparent at the level of underlying representations cannot be propagated). Moreover, the feedback mechanism underlying the nudge model

applies both in children and adults, which means that this approach does not require the implausible assumption that children are the leaders of sound change (see Section 3.3 for an overview of the evidence that adult speakers can change their phonetic representations).⁵

However, the nudge model does not perform any better than the leap model when it comes to the actuation problem: it also appears to overpredict sound change. The reasons for this are fairly straightforward: since the category is consistently nudged in a given direction by the phonetic bias, it is difficult to see what could inhibit sound change (see Baker 2008 for a similar point). The bias-driven simulation in Pierrehumbert (2001) is an excellent illustration of this point: the simulated category is inexorably nudged forward by the phonetic bias, and it does not seem clear if its progress could be hindered in any way. This is because the model in Pierrehumbert (2001) includes no mechanism that would act against sound change. Thus, this simple version of the nudge model appears to predict that sound change will take place in every case where a sound category is affected by a phonetic bias. In this sense, it appears to do even worse than the leap model, where sound change could at least be delayed by perceptual compensation. In conclusion, although the nudge model avoids some of the issues that arise in connection with the leap model, it does not present an improvement in terms of the actuation problem, as it is also in violation of the criterion in (2.1a). It should be noted that the deterministic nature of the nudge model will be illustrated in much more detail in Chapter 4, where the claims above will be shown to be borne out by the results of computer simulations.

2.4 FUNCTIONALIST MODELS

Functionalist models of sound change are based on the assumption that some sound systems may function better than others with respect to certain criteria. Although the set of such criteria does not need to be predetermined, approaches labelled as functionalist almost always focus on the communicative efficiency

5. The version of the nudge model adopted in this thesis may be insufficient to account for changes that proceed in large shifts rather than small continuous steps. For instance, changes like $[k^w] > [p]$ are ‘articulatorily discontinuous’ (Hansson 2008), which means that gradual change is unlikely. Ultimately, a fully explanatory account of sound change may have to include elements both from the leap model and the nudge model. However, this possibility is not explored in this thesis.

of sound systems. The term communicative efficiency is usually used more or less synonymously with discriminability: a sound system is efficient if most pairs of lexical items are sufficiently distinct from each other. A dysfunctional system will exhibit unusually high rates of homonymy, and will therefore be suboptimal from the point of view of communication. The main idea in functionalist approaches is that sound change may be inhibited in cases where it would result in dysfunctional sound systems (see e.g. Blevins & Wedel 2009). Importantly, this seems to provide a solution to the overapplication problem that plagues bias-based models by suggesting ways in which the likelihood of a change may be reduced. While there is a large number of models that share this insight, the exact way in which it is implemented differs substantially across specific approaches. In what follows, I provide a critical discussion of some of the different functionalist models exemplified in the literature, with special emphasis on the actuation problem.

Before, however, we turn to the models themselves, it will be useful to examine a concrete example of functionalist argumentation. Specifically, I discuss the case of the inhibited Banoni vowel length merger, based on Blevins & Wedel (2009: pp. 151–154). Banoni (a Western Oceanic language) has a traditional vowel length contrast, but this contrast is disappearing as a result of an ongoing merger between the short and the long series of vowels. According to Blevins & Wedel (2009), this merger applies more or less indiscriminately within the language, with one important exception: it seems to have spared certain sets of forms where it would lead to the loss of morphological contrasts. Thus, bare nouns and first singular possessed noun forms are distinguished solely by the length of the final vowel. Some examples are [tama] ‘father’ *versus* [tamaa] ‘my father’ and [kasi] ‘brother’ *versus* [kasii] ‘my brother’. A similar contrast is seen between the first and the third person singular transitive verbal suffixes ([-aa] and [-a], respectively). Crucially, the vowel length merger is inhibited precisely in these two sets of forms, leaving the length contrast unaffected in a small subset of the language (see Blevins & Wedel 2009 for further details, especially with respect to the transitive forms). This is a clear example of a phonetically-driven change that appears to be set back by functional factors.

A number of authors present functionalist analyses where the inhibition of sound change is portrayed as an act of ‘self-defense’ by the sound system (see e.g. Martinet 1952, Campbell 1998). Croft (2000) has criticised such approaches

for reifying languages, arguing that ‘languages don’t change; people change language through their actions’ (Croft 2000: p. 4). This, of course, is a trivial point: sound systems and languages cannot exhibit agency in the same way as speakers do. Precisely because of the triviality of this argument, it is hard to believe that otherwise careful scholars like Martinet and Campbell would make such a mistake. I believe that the real problem lies in the fact that they remain neutral with respect to the low-level mechanisms that lead to the inhibition of sound change. They are content to observe that the effects of inhibited changes are often beneficial with respect to the communicative function of language, but they do not make any assumptions about how this inhibition is implemented at the level of individuals (although Martinet 1952 certainly acknowledges the role of individuals in contrast maintenance; cf. Section 3.5). In this sense, their approach is not explanatory – it observes a parallel between communicative efficiency and certain types of change, but does not elucidate the causal link between these two domains.

The issue of linking communicative efficiency and contrast maintenance through individual-level behaviour has been addressed in a number of different ways in the literature. One influential approach is based on the idea that speakers have an active role in avoiding dysfunctional language states. For instance, Flemming (2001, 2004) and Padgett (2003) propose that optimality theoretic grammars can contain explicit statements to the effect that contrasts should be maintained (*MERGE; Padgett 2003) and that the distance between two categories should not fall below a certain threshold (the MINDIST constraint family; Flemming 2001, 2004). Importantly, such constraints can interact with markedness constraints that penalise articulatory effort (not unlike the phonetic biases proposed in the present account), thereby giving rise to a rich and in many ways realistic set of predictions. Since constraints like MINDIST and *MERGE are supposedly universal, it is not surprising that contrast maintenance effects should be observed in a wide range of languages. Furthermore, optimality theoretic constraints are violable, which means that mergers can and should take place in certain cases under the effects of markedness constraints. Therefore, it appears that this type of model does not overpredict or underpredict phonetically-driven sound changes.

The main problem with such accounts lies in their goal-oriented nature. If constraints like *MERGE and MINDIST are to be included in individuals’ gram-

mars, we have to assume that speakers can ‘groom’ their sound systems through their application. Moreover, speakers have to be able to calculate various measures of communicative efficiency such as the distance between two categories or whether a merger has taken place or not. This clearly goes against the principle established in Section 2.2, according to which accounts of sound change cannot be based on vague notions of optimality without independent support (see also Lass 1997). Constraints such as *MERGE and MINDIST are arbitrary. Although they are successful in the sense that they can account for cross-linguistic differences in sound inventories, neither Flemming (2001, 2004) nor Padgett (2003) present any evidence that speakers actively employ such constraints in their production or perception. It appears that these authors see such constraints as part of universal grammar, and therefore do not require any evidence for them at all apart from the data that they are meant to explain. In the present case, this is almost like saying that sound systems tend to maintain contrasts because it is in the nature of speakers to do so, which – as far as I can see – is no explanation at all.

It should also be noted that while accounts relying on *MERGE and MINDIST might be able to account for synchronic facts about contrast maintenance, it is not clear how such constraints could influence the evolution of sound systems. If sound change simply consists in reranking optimality theoretic constraints, systems where *MERGE and MINDIST are outranked by markedness constraints should be just as likely as systems with the opposite ranking. It seems unlikely that individual constraints could act against constraint reranking. As a result such accounts are unlikely to contribute significantly to the solution of the actuation problem.

de Boer (2001) presents an agent-based computational model of the emergence of vowel systems, which also incorporates an account of contrast maintenance. Although the scope of this model is very different from the optimality theoretic approach described above, the mechanisms used to explain contrast maintenance are rooted in similar principles: speakers actively promote functional systems. This model simulates changes within a population of agents, where each agent has their own vowel system. At each iteration of the simulation, a randomly chosen pair plays the so-called ‘imitation game’, where one of the agents has to imitate a given vowel production by the other agent. If the imitation game is successful, the imitator shifts their category represen-

tation closer to the original production by the other agent. If the imitation game fails, the imitator can either delete one of their categories, or add a new category, depending on a number of factors. This simple framework leads to the emergence of surprisingly realistic vowel systems that match many existing cross-linguistic observations.

Crucially, the imitation game will be more likely to fail when the vowel system is dysfunctional. Therefore, the fact that convergence between the agents' systems is inhibited in such situations (and various repair strategies are applied) means that the agents are actively avoiding dysfunctional systems. While this account does not require speakers to evaluate the communicative efficiency of phonological contrasts, it still attributes a number of goal-oriented strategies to them. Thus, de Boer (2001) assumes that the agents can keep track of the 'success rate' of a given sound category, add and delete sound categories and decide whether they should adjust their category representations based on how successful the imitation game is. Even though de Boer (2001) claims that 'no completely unrealistic hat tricks were used to make the sound system emerge [...]' (de Boer 2001: p. 54), as far as I am aware there is no evidence that speakers perform any of the actions listed above during speech production and perception.

There is one final approach that should be mentioned here, which successfully avoids most of the problems associated with the accounts described so far. Labov (1994), Silverman (2006), Wedel (2006) and Blevins & Wedel (2009) argue that a pressure towards functional sound systems can arise through non-goal-oriented means as well. This point is described in greater detail in Section 3.5. The discussion here serves only to indicate that functionalist arguments do not crucially depend on explicit optimisation in speech production and perception. The approach outlined in the works cited above relies strongly on the notion of the production-perception feedback loop (cf. Section 2.3.2). The main idea is that misperceived tokens are less likely to be fed back into category representations than correctly perceived ones. As a result, areas of phonetic space where misperception is frequent will be underrepresented in the production-perception feedback loop. Importantly, the probability of misperception is particularly high when a given token is ambiguous with respect to its category membership (e.g. a vowel token intermediate between [i] and

[u] in English). This means that unambiguous tokens will have higher ‘rates of survival’ in the production-perception feedback loop, and will therefore have greater influence on category representations. This creates an implicit pressure for categories to remain well-separated and efficient from a communicative perspective.

Note that the exclusion of tokens from the production-perception feedback loop is not a conscious strategy on the part of the listeners. If a token is not identified correctly, the listener might not even be aware of the category label that should be assigned to it. When the category label is not known, the token simply cannot be fed back into the appropriate category representation. The failure of such tokens to participate in the production-perception feedback loop is a direct consequence of misperception. The listener has no influence over this process, and therefore cannot be accused of goal-directed behaviour. In this sense, the ambiguity-driven model provides a much more satisfactory explanation for the avoidance of dysfunctional systems.

Functionalist models provide a plausible explanation for why sound changes fail to take place under certain circumstances. However, functionalism cannot be regarded as a solution to the actuation problem in itself, as it can only account for the absence of sound change. It makes no predictions about the likelihood of change in situations where contrast maintenance is not relevant. Moreover, it is also not clear how functionalist models can account for cases where a contrast is neutralised in one dialect but not another. For instance, Weinreich et al. (1968) describe the case of the neutralisation of a four-way vocalic contrast in Yiddish. Proto-Yiddish had the following high front and high back vowels: short *ɪ*, long *ī*, short *ʊ* and long *ū*. As Proto-Yiddish broke up into regional dialects, this contrast came to be neutralised in many of them. However, the neutralisation seems to have applied differently across the dialects. Southern Yiddish neutralised the contrast between back and front vowels leaving the length contrast intact. North-Eastern Yiddish, on the other hand, neutralised the length contrast but preserved the back-front distinction. To make matters even more complicated, some dialects appear to have levelled all of these contrasts, leaving a single *i* vowel. In its current form, the functionalist approach cannot say much about such differences.

2.5 SOCIOLINGUISTIC MODELS

It appears that neither bias-based models nor functionalist models are capable of providing a satisfactory solution to the actuation problem. In bias-based models, the suggested mechanism of sound change overapplies, violating the criterion in (2.1a). Functionalism seems to suffer from a different problem: while the pressure towards functional systems is relevant to certain types of sound change, it is clear that the idea of functionalism on its own cannot fully account for the actuation of all changes. These issues have long been recognised by researchers studying language variation and the social aspects of sound change. Partly in response to the shortcomings of the approaches outlined in the previous sections, they have proposed a somewhat different solution to the actuation problem. Specifically, researchers such as Croft (2000), Labov (1994, 2002), Milroy & Milroy (1985), Milroy (1992) have argued that the actuation of sound change is strongly dependent on social factors, which can counteract the influence of phonetic and functional pressures.

All of these approaches assume that sound change is a two-stage process. First, intra-linguistic variation may sometimes result in innovative forms in certain speakers. Milroy & Milroy (1985) term this ‘speaker innovation’. It is often suggested that the predictions of bias-based and functionalist models apply at this level, since the range of forms that can come by through speaker innovation is determined mainly by intra-linguistic factors. Speaker innovation in itself does not constitute sound change: the innovative variants also need to diffuse through the speech community. This is the second stage of sound change, which is often referred to as ‘propagation’ (see e.g. Croft 2000). It is at this level that social considerations become relevant: the spread of a change within the speech community is guided by social factors, which likely include prestige and social network structure.⁶

Both stages have an important role in determining the likelihood of changes. Even phonetically well-motivated innovations may fail to take off within a speech community if the social conditions do not favour change. Milroy (1992)

6. The exact nature of such factors is the matter of debate within sociolinguistics, and will not be discussed in any detail in the present thesis. The interested reader is referred to Labov (2001, 2002) and Milroy (1992), who devote considerable attention to the social motivation of sound change.

argues that this provides an answer to the overapplication problem in bias-based models: the effects of phonetic biases are checked by social factors.

Sociolinguistic approaches typically focus on the propagation of sound change and say very little about speaker innovation. This is not a problem *per se*: if speaker innovation and propagation are indeed separate processes, it should be possible to investigate them more or less independently. The other option is exemplified by bias-based and functionalist approaches, which focus on speaker innovation at the expense of propagation. However, Section 2.3.1 showed that the question of propagation cannot be ignored completely even if the main topic is speaker innovation. As I argued there, one of the main problems with the leap model is that it makes unrealistic predictions with respect to the propagation of sound change. This is a serious problem regardless of whether the focus is on propagation or not.

The same point also holds for sociolinguistic approaches. It might be possible to treat speaker innovation and propagation separately, but a successful model of propagation will also have to be compatible with existing models of speaker innovation. As it turns out, this is not the case for sociolinguistic models which assume that sociolinguistic variation simply ‘piggy-backs’ on intra-linguistic variation. Let us assume that speaker innovations do indeed take place, and are governed by linguistic factors such as phonetic biases and contrast maintenance. If this is the case, how is it possible that only a small subset of speakers produce innovative forms? To make this point clearer, consider the nudge model introduced in Section 2.3.2. This model could arguably serve as the origin of speaker innovations.⁷ However, the nudge model predicts that every speaker should exhibit the same innovations, given the universal nature of phonetic biases and the inevitability of change in such models (cf. Section 2.3.2). This prediction is highly problematic: if all speakers show the innovation, why would there be any need for propagation? The sound change should emerge automatically without any social conditioning. It appears that combining models like the nudge model with sociolinguistic models of propagation does not solve the overapplication problem. The only possible way out of this issue is to assume that speaker innovations arise through different mechanisms.

7. The leap model is not discussed here given that it does not seem compatible with sociolinguistic models of propagation (see Section 2.3.1).

Baker et al. (2011) propose a solution to the problem outlined above. They argue that while many phonetic biases are universal, there may be differences in the extent to which a bias affects the production or the perception of individuals. For instance, they show that the extent of *s*-retraction in clusters like [str] in English is highly variable even across individuals who arguably do not have a conventionalised pattern of retraction. According to their argument, sound change may occur when an individual who is particularly strongly affected by a bias also happens to be in a position where the social prerequisites of propagation are met. Since such accidental correlations between individual phonetic factors and social conditions will rarely arise, this mechanism of sound change does not overapply. Note that Baker et al.'s (2011) model makes an important prediction: it implies that sound change will only take place in cases where the strength of a universal phonetic bias differs across individuals. Currently, this prediction is only supported by a single case (*s*-retraction in American English). Therefore, more research is needed to determine whether this approach is viable.

Even if it is possible to find a model of speaker innovation that is compatible with the idea of propagation, sociolinguistic models only present a partial solution to the actuation problem. The social conditioning of propagation might indeed explain why sound change only takes place in certain cases, which means that the proposed mechanism does not underapply or overapply. However, the criterion in (2.1c) is only partially satisfied: while sociolinguistic models shed light on many social factors that may affect the likelihood of change, they do not say much about the linguistic factors involved in sound change.

2.6 THE SUGGESTED SOLUTION

None of the models discussed above provide a satisfactory answer to the actuation riddle. Bias-based models can capture parallels between phonetic biases and robust language-specific patterns, but they overpredict sound change. Functionalist models can account for certain cases where sound change fails to take place, but they have little to say about cases where functional considerations are less relevant. Finally, sociolinguistic models may be capable of explaining

both stability and change, but they make no predictions about the effects of linguistic factors on sound change.

The account presented in this thesis takes the bias-based model as its starting point. The reasons for this are as follows. Bias-based models have a clear explanation for the pervasive parallels between phonetics and sound change. It is difficult to see how such parallels would emerge if not through phonetically-driven change. In fact, the success of bias-based models is usually acknowledged in functionalist and sociolinguistic approaches as well, which tend to include phonetic biases in some form. Since the actuation problem – as formulated in Section 2.2 – is about predicting the likelihood of change, it would be somewhat ill-advised to discard one of the most successful predictors by deeming bias-based models implausible.

This thesis proposes that the bias-based model is correct in its theoretical assumptions, and that the main problem lies in the way these assumptions are typically put to use. The bias-based approaches reviewed in Section 2.3 usually focus on sound categories in a vacuum, and disregard the potential influence of interactions among categories. It is not surprising that the predictions of such a simplified approach fail to capture the subtleties of sound change: sound categories ‘in the wild’ hardly ever occur on their own, and this has important consequences with respect to sound change. This thesis breaks with the tradition of looking at individual categories and proposes to investigate the influence of phonetic biases on sound systems incorporating multiple categories.

As it will be shown in Chapter 5, the systemic view provides a straightforward solution to the overapplication problem (why sounds don’t change; Section 5.3) and the underapplication problem (why sounds change; Section 5.4). Phonetic biases are only one of the many pressures that affect sound systems. Together these pressures define a complex adaptive landscape. In the course of its evolution, the sound system will likely end up in a stable state where the pressures balance each other out. This stable state need not necessarily satisfy a given phonetic bias, although in many cases it will. Importantly, the pressures that determine the adaptive landscape can themselves undergo changes, which can knock the sound system out of stable states. I will also briefly show that the system-based model is compatible with sociolinguistic

approaches, and can make interesting predictions about the propagation of sound change. Of course, these arguments will be presented in a much more principled way in later chapters.

Note that the individual elements of this account are not new. The interaction between phonetic biases (often embodied in the notion of ‘ease of articulation’) and systemic effects (such as contrast maintenance) have been investigated in a number of accounts, including Martinet (1952) and Labov (1994). The novelty of the present approach lies in the methods it uses to look at these interactions and the conclusions it draws with respect to the actuation problem. As opposed to previous accounts, the predictions with respect to the behaviour of sound systems are derived directly from a model of speech production and perception (closely related to the nudge model described in Section 2.3.2). This model rests on assumptions which are supported by independent evidence from a variety of fields. Since the predictions of the model with respect to sound systems are by no means trivial, I will use computer simulations to explore them systematically.

Computer simulations are a useful tool in situations where the object of enquiry is a complex system, but they have to be used with caution. It is easy to build a simulation which is engineered specifically to produce the expected results, but the scientific validity of such an approach is highly questionable. The real challenge lies in showing that the simulation is based on independently plausible principles, and that its predictions derive directly from the underlying theory and not from specific details of implementation. Therefore, the argument presented in the rest of this thesis follows a strict logic. Chapter 3 outlines the main theoretical assumptions that serve as the basis of the present approach. Each of these assumptions is motivated in detail in order to show that the underlying theory is well-grounded in existing research. Then, Chapter 4 demonstrates how the principles in Chapter 3 can be implemented in a formal framework. It is only in Chapter 5 that I turn to the actuation problem itself, and present simulation results showing that the model advocated in this thesis provides a coherent and plausible solution.

This thesis uses computer simulations to investigate changes in sound systems under the influence of phonetic biases and other pressures. This is not an uncontroversial undertaking. Computer simulations are undoubtedly useful in that they make it possible to explore the behaviour of complex systems in a way that might not be feasible through simple thought experiments. For instance, the simulations in Chapter 5 allow us to look at the predictions of a complex model that involves phonetic biases and interactions among multiple categories. The simulations allow us to derive the predictions of the model from its underlying assumptions in a rigorous and controlled way. However, in order to build a simulation, it is crucial that the underlying theoretical assumptions are stated explicitly and justified in the context of previous research. Otherwise it would be unclear what the simulations investigate exactly, even if they are intuitively easy to interpret. Therefore, the present chapter provides a detailed outline of the theoretical assumptions that serve as the basis of the simulations described in the following chapters.

The discussion of the theoretical background of the simulations draws strongly on a distinction between *theories* and *models* explicated in Norris (2005). Norris argues that a computational model in itself does not necessarily constitute a theory, and it may have little explanatory power on its own (a point that is also made in Forster 1994 with regard to connectionist models). The following thought experiment should make this point clearer. Let us assume that we could construct a machine that is capable of creating a map of all the synaptic connections in the human brain and transforming this map into a neural network model that responds to external stimuli in exactly the same way as human subjects do. This would certainly be a majestic feat of engineering.

However, there is no way in which the model created by the machine could be called a theory: simply reproducing the complex interactions that take place within the brain does not take us any closer to understanding them.

The reason for this is that the assumptions underlying this hypothetical model cannot be interpreted as useful scientific hypotheses. For instance, let us take a look at the decision to use a neural network architecture to implement the transmission of information through synaptic connections. Translating this decision into a testable hypothesis yields a general statement of the following form: ‘a neural network model can replicate the responses of human brains.’¹ Under a very generous interpretation of this statement, the success of the model can be taken as an indication that the brain itself utilises a neural network architecture for processing information. Unfortunately, this conclusion is of little value given that it contributes nothing to our understanding of the brain beyond what is already known. Thus, while the decision to use a neural network is important from an engineering point of view, it does not endow the model with any explanatory power.

The hypothetical scenario presented above is an example of a model without a theory in Norris’s (2005) terminology. To avoid such dead ends, Norris urges researchers to take a look at what goes into their models and carefully distinguish between theoretical assumptions (i.e. those that constitute useful and testable hypotheses about the area under investigation) and modelling assumptions (i.e. those that are necessary to make the model work). If both of these sets of assumptions are made explicit, the model can serve as a way of linking theory and data by clarifying the empirical predictions of the theory. However, even when all the theoretical and modelling assumptions are clear, it may turn out that the results of the simulations follow from the specific way they are implemented rather than the theory behind them. To avoid such situations, it is useful to try several different ways of implementing the same theoretical principles, that is, to vary the set of modelling assumptions while keeping the theoretical assumptions the same.

1. There is every reason to be skeptical about the validity of this statement: while the underlying principles of neural network models are inspired by the physiology of the human brain, a single processing unit in a neural network may often stand for a large assembly of individual neurons and its behaviour does not necessarily mimic the behaviour of actual neurons very closely (Forster 1994). As a result, it is not clear whether such a model could be used to represent neuron-level interactions accurately.

The distinction between theoretical and modelling assumptions is reflected in the division of labour between the present chapter and the following one: the present chapter focuses exclusively on the theoretical underpinnings of the simulations and the discussion of the details of the models is relegated to the next chapter. In the present chapter, each of the theoretical assumptions is discussed at length and justified in the context of empirical research. In the next chapter, it will be shown that there are a number of different ways of implementing these assumptions, which all yield very similar results. This is a strong confirmation of the generality of these results, suggesting that they follow directly from the underlying theory, rather than the specific modelling assumptions.

Here is a brief outline of the present chapter. In Section 3.1, I discuss a general view of speech production and perception as processes based on multivariate probability distributions, drawing on ideas from Ashby & Alfonso-Reese (1995) and Kirby (2010). The discussion in Section 3.1 also serves to identify a number of points within this probabilistic theory of speech production and perception where further theoretical elaboration is needed. Sections 3.2-3.5 take up these issues and provide justification for each of the following four assumptions: (i) speech production and perception are based on more abstract sound categories such as segments (3.2), (ii) sound categories are subject to continuous update throughout the life of an individual (3.3), (iii) speech production is affected by weak but universal phonetic biases (3.4), and (iv) ambiguous productions resulting from category overlap participate less in category update (3.5).

3.1 PROBABILISTIC CATEGORY REPRESENTATIONS

All the simulations presented in the next chapter are based on the following simple assumption: sound categories can be represented as probability distributions over a multidimensional phonetic space.² The main advantage of this view of category representations is that it allows us to abstract away from the details of particular models of production and perception, and lay out a

2. The exact nature of such sound categories will be made clearer later in this chapter. The specific units that are the focus of this thesis map relatively closely onto the notion of segments, although they could be interpreted as standing for larger or smaller units as well.

general conceptual framework for the discussion of the simulations in the next chapter. Moreover, it also allows us to describe production and perception as operations over the same stored probability distributions, thereby establishing an explicit link between these two separate domains (cf. Kirby 2010: 49). Note that the idea of probability distribution-based production and perception relates to phonetic realisation and not to phonological patterns. That is to say, the framework described here and in the following chapter links sound categories to concrete physical realisations, but it does not say anything about how these categories interact at a more abstract level. Pierrehumbert (2003) argues that a probabilistic model of speech production and perception may well be compatible with abstract models of phonological competence. Although this thesis does not explore such abstract models in any detail, Sections 3.4 and 6.1 do make the relationship between the current framework and the phonology/phonetics divide clearer.

To get a better idea of how probabilistic category representations should be conceived of, consider the two high vowels [i] and [u] in American English. Figure 3.1a shows the distribution of these vowels along the dimensions of F1 and F2, based on 139 samples for [i] and 138 for [u] taken from Hillenbrand et al. (1995)'s study of American vowels.³ The main idea is that a learner exposed to these realisations can set up probability distributions representing each category, as is shown in Figure 3.1b (which only shows the distributions along F2 for simplicity's sake). This process will be referred to as the 'estimation of the underlying probability distribution of a category' (Ashby & Alfonso-Reese 1995). Estimation in this context simply means taking all the available observations (i.e. tokens of [i] and [u]) and constructing probability distributions that could plausibly serve as the sources of these observations. This is the equivalent of learning the phonetic realisation of a sound category. There are many different ways of performing this estimation, although not all of these are equally accurate. The probability density functions in Figure 3.1b were obtained through a method called 'kernel density estimation' (Silverman 1986; see Section 4.1.1 for a more detailed discussion), but this choice was arbitrary. The exact method of estimation is not crucial for our present purposes – indeed, the next chapter will demonstrate that the results of the simulations follow directly from the

3. Available online at <http://homepages.wmich.edu/~hillenbr/voweldata.html>.

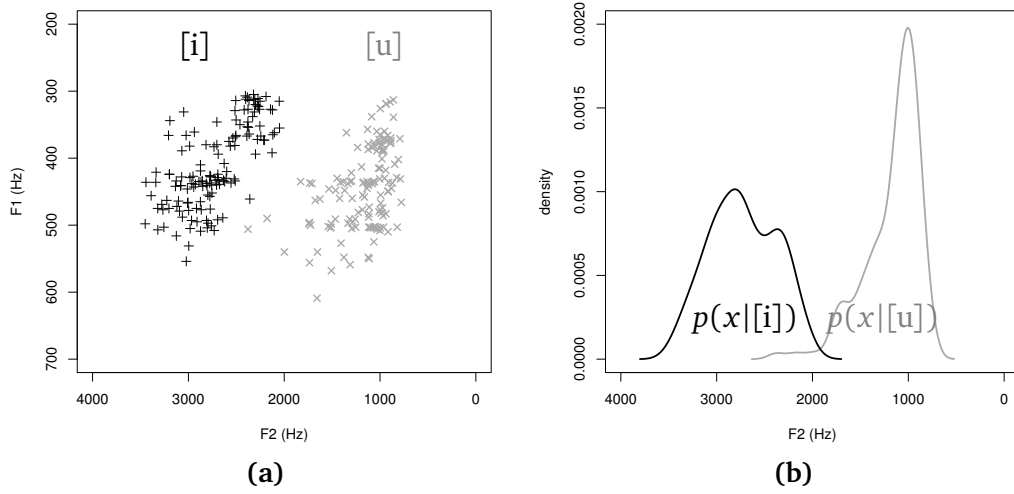


Figure 3.1: (a): the distribution of American English [i] and [u] along the dimensions of F1 and F2; (b): two probability density functions on the dimension of F2 based on the realisations of [i] and [u] on the left (the distributions were determined through kernel density estimation).

general assumption of probability distribution-based representations and not from any particular method of estimation.

The estimated distributions can then be used both for the generation of new stimuli (i.e. production) and establishing the category of a given stimulus (i.e. perception). Production can be modelled as the random sampling of the distribution representing a given category (Kirby 2010).⁴ Of course, randomly drawn samples will be more likely to come from higher density areas of the probability distributions, which means that larger samples of productions will mirror the underlying probability distribution. Perception, on the other hand, is modelled by calculating for each category the probability that a given stimulus comes from that category and then making either a probabilistic or a deterministic choice based on these probabilities. To illustrate, consider a vowel with an F2 value of 2000. The distributions in Figure 3.1b can be used to calculate the values of the probability density functions associated with [i]

4. As Kirby (2010) notes, this is not a trivial problem in the case of continuous probability distributions, since the probability associated with any given value of a continuous random variable is 0. There are, however, a number of ways of obtaining random samples for continuous random variables as well, some of which are described in Devroye (1986).

and [u] at $F2 = 2000$: $p(2000 | [i]) = 0.000152$ and $p(2000 | [u]) = 0.000044$. If we assume that listeners are not biased towards either category *a priori*, the confidence that they can have in a given category decision will be proportionate to these values. The exact probabilities can be derived using Bayes' formula (where $p(c_i|x)$ is the probability that x belongs to category c_i , $p(c_i)$ is the prior probability of category c_i and n is the overall number of categories):

$$p(c_i|x) = \frac{p(x|c_i)p(c_i)}{\sum_{j=1}^n p(x|c_j)p(c_j)}. \quad (3.1)$$

In essence, what this formula says is that an appropriate category label will be chosen based on (i) which categories are the most likely to be produced at a given location in phonetic space and (ii) which categories are the most frequent in general (this is represented by the prior probabilities). In cases where the categories are equally frequent, $p(c)$ will be equal for all the categories, which means that the formula in (3.1) can be simplified as follows:

$$p(c_i|x) = \frac{p(x|c_i)p(c_i)}{\sum_{j=1}^n p(x|c_j)p(c_j)} = \frac{p(x|c_i)p(c_i)}{p(c_i) \sum_{j=1}^n p(x|c_j)} = \frac{p(x|c_i)}{\sum_{j=1}^n p(x|c_j)} \quad (3.2)$$

That is, the probabilities of different category decisions can be obtained simply by comparing the probability density functions representing the different categories. In the present case, such a comparison can be done visually as well: the black line in Figure 3.1b is higher than the grey line at $F2 = 2000$, and therefore [i] is a more likely candidate. This is confirmed by applying the equation in (3.2) to the present case, which yields the following results: $p([i]|2000) = 0.78$ and $p([u]|2000) = 0.22$. Figure 3.2 shows the categorisation probabilities for each category plotted against $F2$ (the dotted line marks the location along the x -axis where $F2 = 2000$). These functions are very similar to the identification functions based on the results of experiments where subjects are instructed to perform the same task, that is, assign phoneme labels to different stimuli varied along a phonetic continuum (e.g. Liberman et al. 1957).

The task of calculating the categorisation probabilities for each category is not identical to that of making a particular categorisation decision for a given stimulus (Ashby & Maddox 1993, Kirby 2010). Ashby & Maddox (1993) argue that the latter process should be treated separately under the label of 'response

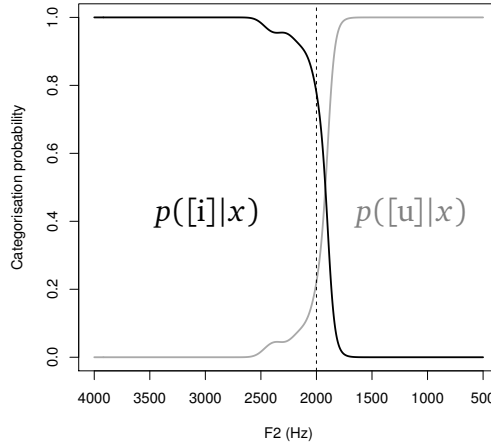


Figure 3.2: Categorisation probabilities for [i] and [u] calculated on the basis of the formula in (3.2). The dotted line indicates an F2 value of 2000.

selection’ and distinguish between two response selection rules: *deterministic* and *probabilistic*. A subject with a deterministic response selection rule will always make the same decision when faced with a given set of categorisation probabilities: using the example from above, the category label assigned to a stimulus with $F2 = 2000$ will always be [i], since [i] has the higher categorisation probability. This rule can be described as follows:

- (3.3) Given stimulus x and a set of categories $C = \{c_i | i = 1, \dots, n\}$, respond c_k , where $p(c_k|x) = \max(p(c_1|x), \dots, p(c_n|x))$.

On the other hand, a probabilistic response rule can make different decisions even when the categorisation probabilities are the same. In this case, the categorisation probabilities determine the relative frequencies of different responses. This can be expressed as follows:

- (3.4) Given stimulus x and a set of categories $C = \{c_i | i = 1, \dots, n\}$, respond c_k with probability $p(c_k|x)$.

Thus, a subject with a probabilistic response rule will respond [i] about 78 percent of the time when $F2 = 2000$.

It should be noted that most existing theories of categorisation, including exemplar theory (e.g. Nosofsky 1986, 1988) and prototype theory (e.g. Posner & Keele 1968) can be reformulated in terms of probability density estimation (Ashby & Alfonso-Reese 1995). In fact, Ashby & Alfonso-Reese (1995) show that the main difference between such models lies in the way they estimate the underlying probability distributions for the individual categories, and is therefore not strictly relevant to the argument developed in the next chapter, which only requires the assumption of some type of probability density estimation. Since this point is not a trivial one, the next chapter will present both exemplar and prototype-based simulations to show that they produce essentially the same results.

Having discussed the role of probability distributions in models of production and perception, I now present the four main assumptions of the theoretical framework adopted in this thesis. The first assumption is a rather general one, namely that speakers establish sound categories (represented by probability distributions) based on their experience with speech, and use these in production and perception. Although this may seem to be a trivial claim, the usefulness of sound categories has been questioned by some phonologists, especially within exemplar theory (Bybee 2001, Kirchner et al. 2010), where it has been claimed that ‘segmentation into a priori phonological units seems contrary to the spirit of Exemplar Theory’ (Kirchner et al. 2010: 541). I show that such claims are not justified and that there is good evidence that speech production and perception are based on sound categories. The second assumption relates to the way the probability distributions representing a given category are updated: I argue that sound categories are subject to continuous update throughout the life of an individual. Again, this claim is not entirely uncontroversial, but there is clear evidence that category representations can be modified even after a speaker has learnt the phonology of a given language. The third assumption concerns a discrepancy between the stored distributions and the produced/perceived stimuli. The claim is that phonetic biases can distort stimuli either during perception or during production, and that many of these pressures are universal (although their effects may be relatively weak in individual languages). Finally, the fourth assumption deals with cases where a stimulus is ambiguous with respect to its category-membership: such ambiguous instances may fail to contribute to the

probability distributions representing the categories, resulting in a situation where non-ambiguous tokens play a more important role in category update, creating a pressure towards contrast maintenance.

3.2 CATEGORY-BASED PERCEPTION AND PRODUCTION

In this section, I argue that speakers' production and perception relies on abstract sound categories established on the basis of their experience with speech. While such sound categories are widely assumed both in phonetics and phonology, little has been said about whether this assumption is justified. In what follows, I provide some evidence for the existence of one specific type of abstract category, namely phonological segments. This choice is admittedly somewhat arbitrary, given the wide range of underlying categories that have been proposed in phonological analyses, including features (Jakobson et al. 1952, Chomsky & Halle 1968), syllables (Kahn 1976), autosegments (Goldsmith 1979) and phonological feet (Selkirk 1980). However, while the notion of language-specific segment inventories appears to be almost universally accepted within phonology, most other categories are the subject of debate and are often challenged on theoretical grounds. Furthermore, it is possible to find relatively direct evidence for the existence of segmental categories in the psycholinguistic literature. Such evidence is reviewed less often for features, syllables and other phonological categories, as these are usually proposed on the basis of more abstract theoretical considerations.

3.2.1 *Alternatives to the category-based view*

It might not seem obvious why a concept as generally accepted as that of segmental categories needs to be justified at all. However, not all researchers agree on the validity of segments as the basic units of speech production and perception. Port (2010b) presents evidence from a number of sources supporting the idea of rich memory representations in speech. Specifically, he argues that larger chunks of speech such as words and phrases are the only psychologically relevant units of representation, while units such as phones or phonemes are only useful when the goal is to describe sound patterns in a community of

speakers. Most of his arguments are grounded in the extreme variability of the speech signal. First, he claims that discrete units like features and segments cannot capture the variation at the level of fine phonetic detail observed in all natural languages. Moreover, he argues that representations employing abstract linguistic units go against the observation that speaker-specific details in memory can aid performance on recognition tasks. Finally, he suggests that the ubiquitous variability of the speech signal makes any attempt to identify invariant phonetic features hopeless.

Similarly, Silverman (2006) rejects the idea of segment-based representations on the grounds that the speech signal does not easily lend itself to segmentation. He points out that articulatory gestures exhibit a large amount of overlap, which results in overlapping acoustic cues as well. Furthermore, he argues that temporally discrete units such as consonants and vowels do not correspond to any coherent acoustic events, given that most of the information is concentrated in the transitions between such proposed units (p. 55). Both Silverman (2006) and Port (2010b) suggest that the popularity of the notion of segments derives from linguists' experience in using alphabetic writing systems, and conclude that segment-like units are merely illusory.

I agree that the notion of segments has to be approached with the same scientific rigour as less well-established units within phonetics and phonology. However, I do not think that the arguments summarised above rule out sequentially arranged units of organisation in speech. In fact, it seems that both authors argue specifically against an extreme version of 'alphabetism' (the term has been borrowed from Silverman 2006), and largely disregard arguments from more moderate proponents of segment-based approaches. It is true that early generative models of phonology treated speech as consisting of strings of invariant feature-bundles (e.g. Chomsky & Halle 1968), but this view has been challenged in numerous ways over the last few decades. Few phonologists would seriously consider a model that (i) does not allow variation in the realisation of speech units and (ii) cannot represent any degree of gestural overlap (and even Chomsky & Halle's model allows for variation and overlap in performance). Pierrehumbert's (2002, 2003) hybrid model is an excellent example of how variation can be tied to abstract units through probabilistic modelling (this model is very similar to the one presented in the previous section); and

approaches such as articulatory phonology (e.g. Browman & Goldstein 1992) have demonstrated that overlap among phonetic events can be represented in models with some degree of abstractness as well. Therefore, while Port and Silverman make a strong case for detailed storage of whole-word units, I do not think that the evidence they present is incompatible with the notion of segments. It is likely that such segments are not invariant or temporally discrete, but that does not mean they do not exist at all.⁵

The next section will present some evidence that speech perception and production are at least partly based on sound categories like segments. However, there is a more practical reason as well for adopting a category-based model: the alternatives are not normally presented in sufficient detail for systematic testing. A brief review of the competing models will help to elucidate this point.

Bybee (2001) argues that sound categories emerge from ‘network connections built up among stored units’, where the stored units are ‘pronounceable linguistic forms – words or phrases stored as clusters of surface variants organized into clusters of related words’ (Bybee 2001: p. 85). Similarly, Lindblom (2000) suggests that category structure emerges in a network of exemplars as a side-effect of ‘systematic covariations among stimulus dimensions’ (Lindblom 2000: p. 304) in a multidimensional psychological space (which probably includes both meaning and phonetic shape). According to Bybee and Lindblom sound categories do not exist as underlying units of storage: categorical effects emerge from complex interactions among phonetically detailed memory representations of speech events. While I do not want to argue that this view is necessarily wrong, there is some reason to treat such claims with skepticism. These assumptions have a certain amount of intuitive appeal, but the lack of explicit models based on the holistic storage of words makes it difficult to see what their implications exactly are and how they relate to phonology or sound change. Bybee (2001) discusses the general outlines of a word-based model of phonology at length, relying on concepts familiar from experimental psychology (e.g. ‘connectionist networks’, ‘exemplar theory’, etc.). However, her

5. Port (2010b) himself acknowledges this in the following quote: ‘[o]f course, speakers might store abstract representations as well, but evidently they are not limited to these. At the very least, any arguments claimed to support a realtime role for abstract segments will need to be much more critically evaluated in future than they have been in the past’ (Port 2010b: p. 48).

discussion lacks the formal rigour that typically accompanies these concepts (for examples of such formal rigour with regard to connectionist networks and exemplar theory, see Rumelhart & McClelland 1986, Nosofsky 1988), which makes it difficult to see whether the theory she advocates does indeed make the predictions she attributes to it.

There are, however, two recent papers that present computationally explicit models that do not assume discrete segment-like units. The first of these is Kirchner et al. (2010), which describes an algorithm that is capable of producing novel utterances on the basis of a pool of stored word-length exemplars, which are not segmented into discrete units. The algorithm can successfully generalise certain phonological patterns to novel stimuli in ways that are similar to phonological repair mechanisms found in natural languages. For example, when most of the stored exemplars conform to a pattern where [x] only appears intervocalically and [k] is found in every other position, a pattern-violating input with intervocalic [k] will usually surface with a [x] (e.g. [ækæ] is realised as [æxæ]). Note that the segment-like units in the description above are not part of the model – they are used purely for descriptive convenience. While these results are promising, the authors themselves admit that their algorithm is rather tentative and that it is not strongly grounded in psycholinguistic research on the mechanisms underlying speech production. Even more problematically, the fact that the algorithm was specifically engineered to perform phonological extensions without much regard to previous theoretical results in the field leaves it open to criticism on the grounds that it is merely a model without a theory in Norris's (2005) terms. Therefore, Kirchner et al.'s (2010) algorithm cannot be taken as a convincing alternative to models based on sound categories until further links are established with theoretical and empirical work on speech production.

Another paper that looks into how a model operating without explicit sound categories can account for certain aspects of phonology is Zuidema & de Boer (2009). Zuidema and de Boer's approach is very different from that of Kirchner and his colleagues. They do not make any claims about what they term 'productively combinatorial phonology' (Zuidema & de Boer 2009: p. 126), that is, the cognitive mechanisms behind the patterns of speech production and perception that segment-like units are meant to capture. Instead, they

look at the evolution of a system under pressure for distinctiveness where the initial signals are random holistic trajectories in phonetic space. One way to imagine such a system is to think of a language where every morpheme consists of completely random movements of the articulators. The paper presents an algorithm that gradually optimises these trajectories for distinctiveness (without any built-in tendency towards segment-based patterns). As the random trajectories become perceptually more distinct, they are also reorganised in a way that they all pass through a small number of ‘way stations’ in phonetic space. The appearance of a limited set of points connected by the trajectories can be interpreted as the emergence of (superficially) combinatorial phonology, where the points themselves are the basic units of combination, that is, sound categories. However, the fact that such structures emerge in these simulations without category-based storage is not an argument against sound categories in itself. Indeed, the authors themselves argue that further mechanisms are needed in order to use these emergent categories in a productive way.⁶ Thus, Zuidema & de Boer’s (2009) results are not incompatible with a category-based view of speech production and perception.

3.2.2 *Evidence for sound categories*

The previous section has demonstrated that the evidence against segment-like categories is rather weak and that there are no convincing alternatives to category-based models at present. However, this in itself does not justify adopting sound categories as the basic underlying units of speech production and perception – it remains to be shown that there is substantive evidence for categories. This type of justification is more often than not completely absent from discussions of category-based models both in phonology and experimental psychology. The convenience and the intuitive appeal of the notion seems to preclude discussions of why underlying categories should be assumed in the first place. This is nicely illustrated by the following quote from Kornai (1996: 400):

6. In their account, the automatic emergence of superficial combinatorial structure under pressure for discriminability is simply the first step in an evolutionary scenario, where the existence of such structures creates an advantage for agents in the population who can better exploit these structures in their language use. This advantage can then drive the biological evolution of the speech system towards phonemic coding.

‘[s]ome [...] units, most notably the phoneme, are instrumental in describing such a broad range of phenomena that their psychological reality can hardly be disputed.’ As Port (2010b) and Silverman (2006)’s work shows, such arguments do not stand up to scrutiny: descriptive usefulness and psychological reality are simply not the same thing. In this section, I review two types of evidence for segments: statistical and psycholinguistic.

STATISTICAL EVIDENCE FOR SEGMENTS The first set of evidence for sound categories comes from certain statistical properties of the speech signal. Kornai (1996) argues that traditional claims about *recurrence* and *discriminability* as the defining properties of segmental categories can be rephrased as statements about the statistical distribution of sounds in phonetic space. Specifically, he proposes that recurrence can be redefined as the presence of distinct density peaks in the overall distribution of speech sounds in phonetic space, and discriminability as the separation of these density peaks. To put it slightly differently, sound categories can be proposed if we find that phonetic tokens are organised into well-defined clumps (recurrence) that do not overlap too much (discriminability). Kornai (1996) provides some evidence that this might be the case for at least some of the sound categories proposed for natural languages by investigating a database of steady-state formant representations of American English vowels (Peterson & Barney 1952). Since I believe that Kornai’s arguments are worth discussing in more detail, I will attempt to replicate and develop some of his results based on a similar but more recent (and phonetically more detailed) database from Hillenbrand et al. (1995). The discussion of recurrence below follows Kornai (1996) quite closely, but the argument about the discriminability of vowel categories in American English is based on novel work carried out specifically for this thesis.

Kornai uses an unsupervised clustering algorithm to test whether the criterion of recurrence holds for vowels in American English. Unsupervised clustering techniques can be used to partition a set of unlabelled data points into smaller clusters by exploiting density peaks in the data set. To put it more simply, unsupervised clustering looks for clumps of tokens in a given data set. Unsupervised clustering is particularly suitable as a test for recurrence given its reliance on density peaks. If vowel phonemes satisfy the criterion of recurrence (i.e. each

vowel phoneme corresponds to a density peak in the overall distribution of vowel realisations), an unsupervised clustering algorithm should be able to find the clusters corresponding to the original vowel phonemes in a set of vowel tokens where the phoneme labels have been removed.

Kornai (1996) looks at F1 and F2 values in a large set of measurements from Peterson & Barney (1952). He uses a *k*-means clustering algorithm to search for density peaks among the F1 and F2 values. This algorithm looks for a pre-specified number of clusters, where the number is *k* (hence the name *k*-means). The technical details of the algorithm are not important for our present purposes – it is sufficient to point out that the behaviour of the algorithm is heavily dependent on density peaks in the input data set. The output consists of a classification of the vowel measurements into clusters along with mean F1 and F2 values for the clusters. Kornai (1996) demonstrates that the *k*-means algorithm is highly successful at finding the clusters corresponding to the original vowel categories in Peterson & Barney's (1952) data set.

I have replicated Kornai's results using a similar set of vowel realisations from Hillenbrand et al. (1995). The unlabelled F1 and F2 measurements are shown in Figure 3.3 (only productions from men are included in this sample to avoid issues related to normalisation). The *k*-means clustering algorithm does remarkably well at finding the original phoneme clusters, as is shown in Figure 3.4, where the orthographic transcriptions⁷ indicate the mean values for the original clusters and the grey dots the mean values for the clusters calculated by the algorithm. The estimations are extremely accurate for most vowels, even though the only information the algorithm has about the phoneme labels is their overall number. This is all the more impressive given the considerable amount of overlap between some pairs of categories, especially in the front region of the vowel space, as can be seen in Figure 3.4 (cf. the pairs 'ih' vs. 'ei' and 'ae' vs. 'eh'). The success of the clustering algorithm implies that the vowel tokens are indeed organised into well-defined density peaks around the individual phoneme centres. Therefore, the suggested units, that is, segments satisfy the criterion of recurrence.

7. The transcriptions correspond to the following lexical sets (cf. Wells 1982): 'iy' = FLEECE, 'ih' = KIT, 'ei' = FACE, 'ae' = TRAP, 'eh' = DRESS, 'ah' = PALM, 'aw' = THOUGHT, 'uh' = STRUT, 'oa' = GOAT, 'oo' = FOOT, 'er' = NURSE, 'uw' = GOOSE.

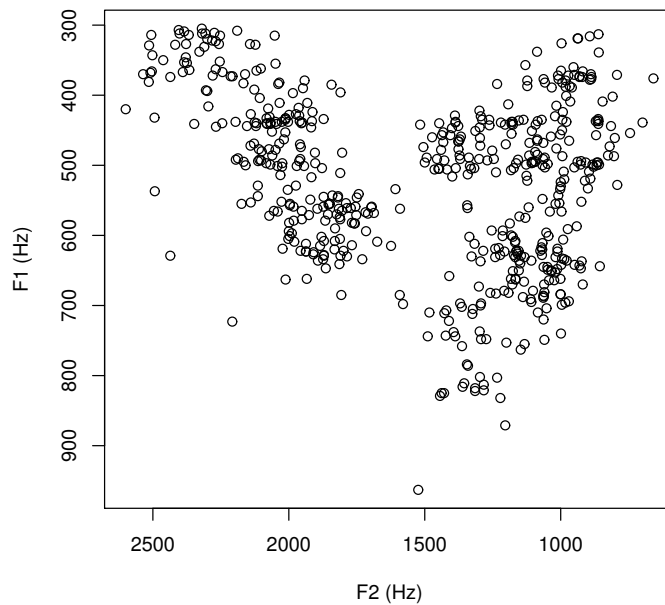


Figure 3.3: Male productions of American English vowels from Hillenbrand et al. (1995) without phonemic labels.

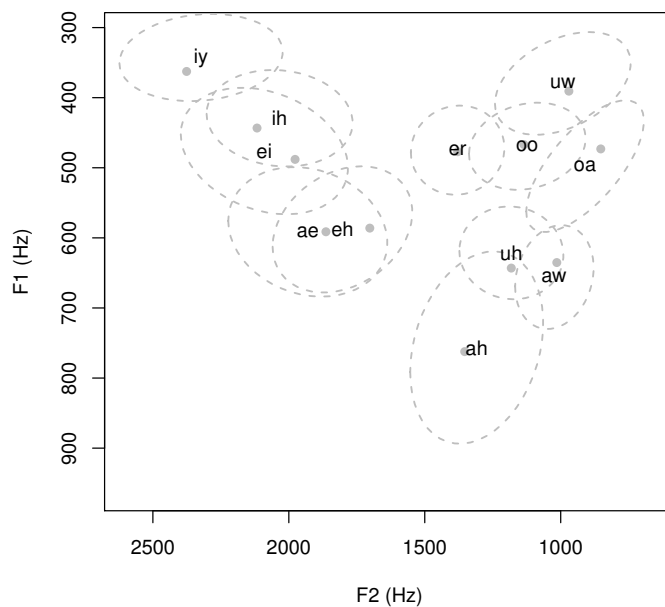


Figure 3.4: A comparison of the actual means (orthographic transcriptions) and the estimated means (grey dots) for the phoneme clusters in Hillenbrand et al.'s (1995) data set; 90% confidence ellipses are also indicated for the original clusters.

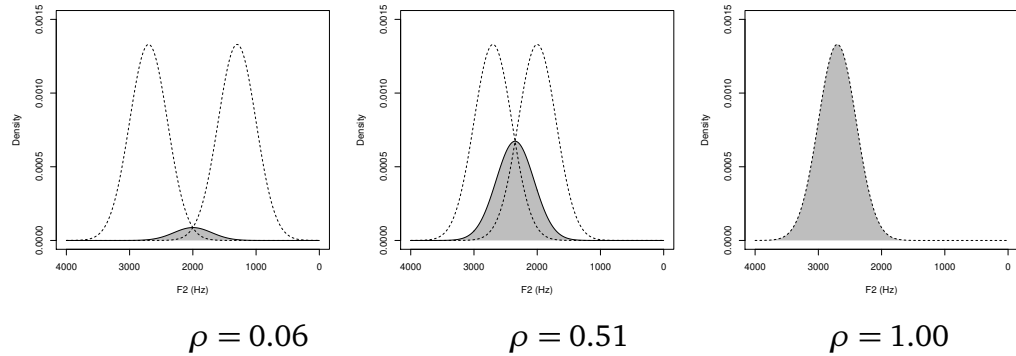


Figure 3.5: Three examples for the calculation of ρ ; the dashed lines indicate the probability density functions, the continuous lines the function from which ρ is obtained through integration (with the area under the function indicated by shading); all of these functions overlap in the panel on the right-hand side.

Another criterion proposed by Kornai is that of discriminability, which is defined as the extent to which the density peaks corresponding to the proposed phonetic units can be separated in phonetic space. Unfortunately, Kornai's arguments for the separability of the proposed phoneme clusters in his data set are somewhat informal. Therefore, a more rigorous test for separability has to be devised, especially because of the substantial amount of overlap between certain phoneme pairs mentioned above. The test I propose below is based on the idea that the separability of two categories is a decreasing function of the amount of overlap between their underlying probability distributions. There are several different methods for estimating the amount of overlap between two probability distributions. I will use Matusita's measure (Matusita 1966), which is defined for the probability density functions $f_i(x)$ and $f_j(x)$ as follows:

$$\rho = \int_{-\infty}^{\infty} \sqrt{f_i(x)f_j(x)} \quad (3.5)$$

This measure produces a value between 0 and 1, which increases with the amount of overlap between the two distributions, reaching 1 only when the distributions are identical. Figure 3.5 shows three pairs of overlapping probability density functions along the dimension of F2 and the function from which ρ is obtained through integration (that is, by calculating the area under the function indicated by shading in the graphs).

Since Matusita's measure is defined on continuous probability density functions, we first have to estimate the probability distributions underlying the vowel clusters to be able to calculate the extent to which they are separable. In this particular case, multivariate normal distributions are used to represent the individual phonemes. This choice was motivated by the existence of a closed-form expression for the calculation of ρ for normal distributions, which makes the present task computationally simple (the interested reader is referred to Lu et al. 1989 for a presentation of this closed-form expression). The normal distributions themselves were obtained through maximum likelihood estimation; the confidence ellipses in Figure 3.4 provide a rough idea of how they should be conceived of. Although such a representation possibly discards some information about the shape of the underlying distributions (e.g. it does not allow any irregularity or 'bumpiness' in the probability density functions), multivariate normal distributions are standardly used in the statistical literature for similar tasks.

Two multivariate normal distributions were obtained for each of the phoneme clusters in Hillenbrand et al.'s (1995) data set (using male tokens only): a bivariate normal distribution based on steady-state F1 and F2 measurements and a multivariate normal distribution based on fundamental frequency, F1, F2, F3 and vowel duration. The value of ρ for each pair of phonemes is shown in the top two matrices in Figure 3.6. The results based on bivariate distributions using only F1 and F2 measurements (shown in Figure 3.6a) correspond very closely to what is seen in Figure 3.4. Most clusters are well-separable (i.e. they have low ρ values), but there is usually a certain degree of overlap between neighbouring clusters, especially in the case of 'ih' *versus* 'ei' and 'eh' *versus* 'ae', which are nearly indistinguishable (i.e. they have ρ values close to 1).

These results change rather dramatically when fundamental frequency, F3 and vowel duration are also included in the calculations: suddenly even the closest neighbours appear as relatively distinct as all the ρ values drop by about 15–75 percent (Figure 3.6b). Thus, while some categories appear inseparable when only a limited number of phonetic dimensions are considered, this inseparability turns out to be an artefact of the choice of phonetic dimensions when other measurements are included as well. Note that this could be a simple result of including extra dimensions, as the separability of clusters of data

	iy	oo	ei	eh	ah	uw	oa	ih	uh	aw	ae	er
iy												
oo	.00											
ei	.12	.00										
eh	.00	.00	.20									
ah	.00	.01	.00	.01								
uw	.00	.33	.00	.00	.00							
oa	.00	.29	.00	.00	.03	.10						
ih	.17	.00	.74	.04	.00	.00	.00					
uh	.00	.03	.00	.01	.30	.00	.15	.00				
aw	.00	.00	.00	.00	.20	.00	.10	.00	.47			
ae	.00	.00	.36	.85	.01	.00	.00	.08	.00	.00		
er	.00	.33	.00	.04	.01	.07	.01	.00	.02	.00	.01	

(a)

	iy	oo	ei	eh	ah	uw	oa	ih	uh	aw	ae	er
iy												
oo	.00											
ei	.08	.00										
eh	.00	.00	.08									
ah	.00	.01	.00	.00								
uw	.00	.20	.00	.00	.00							
oa	.00	.18	.00	.00	.02	.05						
ih	.07	.00	.19	.03	.00	.00	.00					
uh	.00	.02	.00	.00	.23	.00	.11	.00				
aw	.00	.00	.00	.00	.17	.00	.08	.00	.28			
ae	.00	.00	.22	.36	.01	.00	.00	.03	.00	.00		
er	.00	.10	.00	.01	.00	.03	.00	.00	.01	.00	.00	

(b)

	iy	oo	ei	eh	ah	uw	oa	ih	uh	aw	ae	er
iy												
oo	.00											
ei	.11	.00										
eh	.00	.00	.18									
ah	.00	.01	.00	.01								
uw	.00	.26	.00	.00	.00							
oa	.00	.28	.00	.00	.02	.08						
ih	.14	.00	.68	.03	.00	.00	.00					
uh	.00	.03	.00	.00	.24	.00	.12	.00				
aw	.00	.00	.00	.00	.17	.00	.09	.00	.43			
ae	.00	.00	.28	.71	.01	.00	.00	.03	.00	.00		
er	.00	.28	.00	.04	.01	.05	.00	.00	.02	.00	.00	

(c)

Figure 3.6: Matrices showing Matusita's ρ for each pair of vowels in the data set. The shading is a function of ρ : the higher the value, the darker the cell. (a) is based on F1 and F2; (b) on fundamental frequency, F1, F2, F3 and vowel duration; (c) on F1, F2 and three dummy dimensions.

points always increases as further dimensions are added (this is analogous to the phenomenon of overfitting in statistics). To control for the increased number of dimensions, I included a third matrix (shown in Figure 3.6c) based on F1 and F2 measurements from the data set and three ‘dummy dimensions’ consisting of random numbers between 0 and 1 (these random numbers were not significantly different across the vowel categories). While the vowel categories appear more separable in the dummy five dimensional condition than in the original two dimensional case, the separability is much lower than in the real five dimensional condition (Figure 3.6b). This means that duration, fundamental frequency and F3 contribute significantly to the separability of the vowel categories.

In sum, all of the phoneme clusters are relatively well-separable in the higher dimensional case, which suggests that the criterion of discriminability is also satisfied by segments (at least in the case of American English vowels). The replication of Kornai’s results along with the refinements presented above lend statistical support to the plausibility of the segment as a unit of representation.

PSYCHOLINGUISTIC EVIDENCE FOR SEGMENTS In the rest of this section I review two further important sources of evidence for segments: speech errors and categorical perception. In a reply to Port’s (2010b) criticism of segment-based accounts of speech production and perception, Fowler (2010) cites evidence from studies of speech errors as support for the notion of the segment. Segment substitution errors have long been used to make the point that segments have a distinguished role in speech production. One of the earliest studies exemplifying this line of reasoning is Fromkin (1971). After reviewing a large set of naturally occurring speech errors, Fromkin concludes that ‘[b]y far the largest percentage of speech errors of all kinds show substitution, transposition (metathesis), omission, or addition of segments of the size of a phone’ (Fromkin 1971: p. 30). Some of her examples are presented below:

- (3.6) a. *also share* → *alsho share* [ɔlʃo ʃer]
 (anticipation; Fromkin 1971: p. 30)
 b. *week long race* → *reek long race*
 (anticipation; Fromkin 1971: p. 30)

- c. *Spanish speaking people* → ... *speaping people*
(perseverance; Fromkin 1971: p. 30)
- d. *Chomsky and Halle* → *Chomsky and Challe*
(perseverance; Fromkin 1971: p. 30)
- e. *keep a tape* → *teep a cape*
(spoonerism; Fromkin 1971: p. 31)
- f. *ad hoc* [æd hak] → *odd hack* [ad hæk]
(spoonerism; Fromkin 1971: p. 31)

Similarly, Shattuck-Hufnagel (1983) claims that the majority of the errors in her data set involved single segment substitutions. Speech errors are arguably constrained by the units that mental representation are based on. Thus, the observation that segment substitutions are the most common type of speech error strongly suggests that segments play an important role in the psychological representation of speech.

One potential issue with traditional studies of speech errors (including the ones referred to above) is that they are based on simple phonetic transcriptions of naturally occurring speech (Mowrey & MacKay 1990, Port 2010a). This is problematic inasmuch as phonetic transcription is necessarily constrained by the perceptual system of the experimenter, and might not provide an accurate representation of actual speech events. Speech errors that seem to involve discrete segments might, in fact, be based on more gradient processes that do not constitute evidence for segmental categories. This is indeed what Mowrey & MacKay (1990) find in an electromyographic study of muscular activity in tongue twisters.⁸ Many speech errors seem to involve intermediate levels of muscle activity, and Mowrey & MacKay (1990) argue that most of them take place at a 'subphonemic' and 'subfeatural' level (Mowrey & MacKay 1990: p. 1310). Frisch & Wright (2002) present another phonetically detailed study of speech errors based on acoustic analysis. Their findings suggest that while gradient errors of the kind found in Mowrey & MacKay (1990) are frequent, categorical errors also occur. In sum, while the evidence from speech errors is not entirely uncontroversial, it seems that even phonetically detailed data

8. It should be noted that Mowrey & MacKay (1990) used themselves as subjects in their experiment, which, in my view, is no less problematic than evaluating speech errors based on potentially biased transcriptions.

sets provide some support for the idea of the segment as a unit of mental representation.

Another source of evidence for segmental categories comes from experiments investigating a range of phenomena that can be subsumed under the broader label of ‘categorical perception’. Categorical perception ‘[...] refers to the experience of discontinuity as a continuously changing series of stimuli crosses a category boundary, together with the absence of clearly perceived changes within a category’ (Repp 1984: p. 251–252). There are two important observations encapsulated within this definition:

- (3.7) While phonetic events vary in a continuous phonetic space, speakers divide this space up in a more or less discrete fashion.
- (3.8) Gradient differences within a given area covered by one phonetic label are less perceptible to listeners than similar differences across differently labelled areas.

To make these points clearer, let us briefly review the main findings of Liberman et al. (1957), one of the earliest papers looking at categorical perception. Liberman et al. (1957) performed two experiments using an 11-step synthetic continuum of syllables varying only in the place of articulation of the initial sound, yielding stimuli like *be*, *de* and *ge* (the specific acoustic parameter they manipulated is the starting point of the F2 formant transition into the vowel). In the first experiment, they simply asked subjects to label the stimuli as *b*, *d* or *g*. Although the stimuli were varied continuously, the categorisation responses were stable for most of the continuum. The subjects’ responses only showed variation near category boundaries. In the second experiment, subjects were given ABX stimuli triads, where A and B were always different (by one, two or three steps along the continuum), and X was identical to one of them. They were asked to determine whether X was identical to A or to B. The results of this experiment clearly demonstrate an increased sensitivity to differences between stimuli close to category boundaries (e.g. stimuli intermediate between *be* and *de*), and decreased performance within categories. Both of the experiments yielded results that are consistent with the two main observations associated with categorical perception. The fact that speakers show categorical perception related to different segments is strong evidence for the segment as a unit of

representation. It should be noted that such effects have been found for a wide range of other segmental contrasts as well, including manner contrasts (e.g. fricative vs. affricate), contrasts between nasal *versus* oral consonants and vowel contrasts (see Repp 1984 for an extensive summary).

A more specific manifestation of categorical perception is the so-called perceptual magnet effect (Kuhl 1991, Iverson & Kuhl 1995), whereby within-category discrimination performance decreases near the category prototype. Note that this effect is related to the one described in (3.8). However, it is clearly distinct from it, as it looks only at discrimination between pairs of stimuli within the same category (but not pairs from different categories). The perceptual magnet effect has also been demonstrated for a number of segmental categories, although almost all of these categories are vowels (see Feldman et al. 2009 for a review). This provides further support for the notion of segments.

To sum up, there is substantial evidence both from the statistical distribution of phonetic events and psycholinguistic experiments that segmental categories are an important part of the mental representation of speech. Note that I do not wish to argue that mental representations consist solely of a series of abstract segments – merely that segments have a certain amount of psychological and statistical reality. It is perfectly possible that mental representations also involve phonetically detailed memory traces (as Silverman 2006 and Port 2010b suggest), and that information about the segmental make-up of a given utterance exists alongside such episodic representations. The evidence presented above also does not crucially bear on the debate about whether segments are an integral part of mental representations or if they emerge dynamically from detailed memories during language processing (as suggested by Bybee 2001). What is clear, though, is that segments play an important role at some level of language processing. Finally, it should be noted that the units referred to as segments in the discussion above are not identical to the types of segments proposed in early generative approaches to phonology such as Chomsky & Halle (1968). The probabilistic category representations proposed in Section 3.1 allow for a large amount of variation in the realisation of segments, and they do not make any *a priori* assumptions about the sequential arrangement of segments (i.e. they do not rule out overlapping segments). They are also

flexible in terms of the phonetic dimensions that they are based on, and it is, in principle, possible to link them to social, lexical and other types of non-phonetic information (although this thesis does not explore this possibility).

3.3 THE CONTINUOUS UPDATE OF CATEGORIES

In this section, I review evidence for the claims that (i) the category representations belonging to a given sound are subject to continuous update, and that (ii) such update can occur within the lifetime of a single individual. These statements are clearly not uncontroversial: throughout the 20th century, many popular approaches to sound change have assumed that category representations can only change across generations, and, by extension, that the agents of sound change are language acquirers. In the first half of this section, I take a brief look at some representative examples of the latter point of view, labelled ‘acquisitionism’ by Honeybone (2006), Honeybone & Salmons (to appear). This is followed by an overview of the research on changes to category representations within individuals, which constitute strong evidence against the strict acquisitionist approach. Finally, I briefly discuss some findings suggesting that category update is subject to certain conditioning factors that can aid or inhibit it.

Before introducing the debate on acquisitionism, it will be useful to clarify the notion of changes to category representations, and how this might be conceived of in acquisitionist and anti-acquisitionist approaches. In the previous sections, it was suggested that knowledge of categories includes probabilistic details. If categories are represented as probability distributions defined over phonetic space, these details can easily be captured. A change to a category representation will then consist in altering the parameters that define the corresponding probability distribution. For instance, many of the simulations described in the next two chapters use normal distributions to represent sound categories. A normal distribution is defined by two parameters: its mean and its standard deviation. Thus, the category representation will change when one or both of these parameters are modified. The anti-acquisitionist approach holds that such changes can affect the category representations of a single individual, causing an observable (or at least detectable) change in their speech

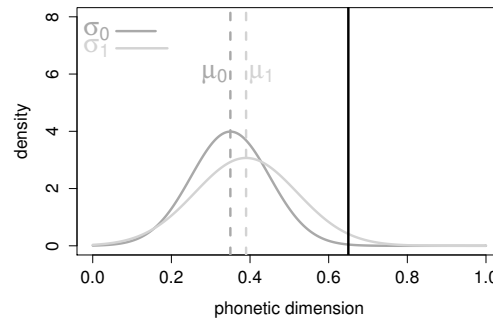


Figure 3.7: Changes to the representation of a category after the presentation of a deviant example (black vertical line). The dark grey line corresponds to the category before the presentation of the example, and the light grey line after. The means and the standard deviations of the original vs. changed distributions are also indicated.

over time even after the critical period.⁹ Typically, this type of category update is envisioned as a reaction to examples of a given category that differ significantly from those predicted by the speaker's own representations. Figure 3.7 provides an illustration, where both the mean (μ) and the standard deviation (σ) of a given category change slightly after the presentation of a deviant example (represented by the black line). Of course, this example is somewhat exaggerated, as a single token is unlikely to cause a clearly observable shift in category representations. Once the anti-acquisitionist perspective is understood, acquisitionism needs little explanation. Acquisitionists claim that changes to category representations are limited to the critical period, and cannot take place in adult speakers. Since adult speakers are not capable of altering their speech even if they are exposed to patterns different from their own, sound change will only be observable across generations.

ACQUISITIONISM In the following few paragraphs, I briefly review some works that adhere to the idea of acquisitionism, starting with Jakobson's famous

9. This is not equivalent to saying that adults can change their representations just as much as language acquirers, only that a certain amount of flexibility remains even after the critical period. Note also that the exact length and nature of the critical period is not particularly relevant to the present account.

Kindersprache (Jakobson 1942/1968). This work is not fully committed to acquisitionism, as it suggests that adults can actually adapt to changes in the ambient language (see e.g. Jakobson 1942/1968: p. 18). Nevertheless, this book already contains some of the characteristic arguments of acquisitionism, and acquisitionist thinking still seems to be greatly influenced by its claims. Early on in the book, Jakobson argues that children can initiate linguistic changes by rejecting ‘a certain component of their linguistic inheritance’ (Jakobson 1942/1968: p. 18). This claim forms the basis of a larger investigation, in which Jakobson finds striking parallels between language acquisition and implicational universals in the sound systems of the world. To give an example, Jakobson claims that children acquire coronal and labial consonants before velar ones, which corresponds well with the cross-linguistic observation that velar consonants imply the presence of coronal/labial ones (i.e. there are no languages with only velar consonants). Although almost all the parallels suggested by Jakobson have been found to be problematic (see e.g. Menn & Vihman 2011 for counterexamples), they continue to be seen as compelling evidence for acquisitionism.

A stronger version of acquisitionism is often presented in generative accounts of sound change, where children are seen as the primary or even the only source of sound change. A particularly clear statement of this position is found in Hale (2003: p. 345):

[...W]e believe that “change” is to be conceived of as the set of differences between the grammar generating the primary linguistic data (PLD) used by an acquirer and the grammar ultimately constructed by that acquirer.

In this model, sound change is cross-generational by definition: changes only occur when the learner mistakenly constructs a grammar that is different from the adult grammar. Such mistakes on the part of the language acquirer are assumed to result either from distributional imbalances in the limited corpus of data they have to rely on, or the misinterpretation of performance errors as grammar-internal effects (Hale 2003: p. 349). Moreover, Hale (2003) clearly states that the grammars he refers to are ‘unique entit[ies] established at the end-point of the acquisition process and not subsequently modified during the lifetime of the speaker’ (Hale 2003: p. 365, fn. 9). Although he acknowledges that speakers are capable of modifying their production by constructing

additional grammars, he claims that such changes are qualitatively different from changes introduced during language acquisition. Thus, this model is truly acquisitionist in that it dismisses changes in the speech of adults as irrelevant and restricts the locus of sound change to language acquisition.

Although generative treatments of sound change are not always so clear about the role of language acquisition, this might simply be due to the fact that acquisitionism is so widely assumed that it is not even necessary to make an explicit statement about it. Indeed, if first language acquisition is viewed as a domain-specific process guided by a language acquisition device that ‘turns off’ at the end of the critical period (a standard assumption in generative linguistics; see e.g. Chomsky 1980), it is difficult to see how any linguistically significant changes could take place outside the critical period. In other words, it is not surprising to see that acquisitionism is so widely accepted within generative linguistics given the close affinity between its main ideas and the standard assumptions of generative linguistics.

While acquisitionism is particularly popular within generative linguistics, it is by no means restricted to it. Acquisitionist ideas are also present in approaches that are more or less agnostic with respect to strong generative assumptions. In fact, one of the central tenets of variationist sociolinguistics, the ‘apparent time hypothesis’ is strongly tied to the notion of acquisitionism. Consider the following quote from Chambers & Trudgill (1980: p. 165–166; as cited in Harrington et al. 2000):

[...] the validity of [studies based on apparent time] hinges crucially upon the hypothesis that the speech of, say, 40 year olds today directly reflects the speech of 20 year olds twenty years ago, and is thus comparable for diffusion research to the speech of 20 year olds today [...]

That is, the idea that sound change can be investigated by looking at different age groups within the same community assumes that individuals’ speech patterns become fixed at some point during their lifetime – otherwise all age groups could potentially end up with the same pattern, even if there was an ongoing sound change.

One final approach that should be mentioned here is the leap model described in Section 2.3.1 (see e.g. Ohala 1981, 1993, Blevins 2004, 2006). As the

reader might recall, this approach locates the source of sound change in errors and reanalyses during the transmission of language (Blevins 2006: pp. 126–129). Although Greenlee & Ohala (1980) argue that transmission errors can occur both in children and adults, it is not clear how the blatant reanalyses that form the basis of this approach could take place in a listener whose representations are already fully established. Accordingly, Blevins' work focuses exclusively on cross-generational transmission. This means that the leap model is clearly acquisitionist: the learner is seen as the primary agent of linguistic change.

ANTI-ACQUISITIONISM As the preceding paragraphs have demonstrated, there is a wide range of different approaches within the study of sound change which argue for some form of acquisitionism. Consequently, the notion of acquisitionism is intimately tied up with a number of further issues including universal grammar, apparent time and transmission errors. A convincing argument against acquisitionism would have to engage with each of these, which is well beyond the scope of the present chapter. Therefore, my goal is not to argue against the existence of sound changes rooted in language acquisition, but to show that the strong version of acquisitionism is not tenable: the speech patterns of individuals can and do change after the critical period.¹⁰ Indeed, this was all that I claimed at the beginning of this section. In what follows, I describe three different types of changes that can take place in the speech of an individual (this classification and part of the discussion below has been adapted from Bane et al. 2010): short-term changes, long-term changes and 'medium term' changes.

There are numerous studies investigating short-term changes in the production patterns of individuals (e.g. Natale 1975, Gregory et al. 1993, Goldinger 1998, Pardo 2006, Nielsen 2008, Babel 2009; such changes are also sometimes referred to as phonetic or sociolinguistic accommodation). The general research themes in these studies are similar: they all investigate rapid changes in the speech of their subjects that occur as a reaction to some type of input that differs

10. It should be noted that the burden of proof rests on acquisitionists: I do not know of any convincing empirical investigation that demonstrates a clear change between two generations that cannot be the result of anything other than some kind of reanalysis or transmission error during language acquisition. As the rest of this section shows, such investigations do exist for the anti-acquisitionist position.

from their own speech. However, the research methods used to investigate this question differ greatly across the studies. Some of them are conducted in a controlled laboratory setting, where the input consists of a pre-recorded (and potentially artificially manipulated) list of stimuli (Goldinger 1998, Nielsen 2008, Babel 2009), while others use more natural data from dyadic conversations (Natale 1975, Gregory et al. 1993, Pardo 2006). The measures used to determine whether a change has taken place also differ from study to study. Thus, Natale (1975) and Gregory et al. (1993) look at intensity and pitch, which are both relatively broad properties of speech; Goldinger (1998) and Pardo (2006) use perceptual judgments from a separate set of subjects to see if the speech of the speakers becomes more similar to the input over time; and Nielsen (2008) and Babel (2009) rely on instrumental measurements of specific acoustic features such as voice onset time and formant values. Regardless of the techniques employed, all studies come to very similar conclusions: there are significant changes in the subjects' speech, which are clearly related to the input they receive. In the case of conversational studies such as Natale (1975), Gregory et al. (1993) and Pardo (2006), the pairs converge to each other over the conversation, whereas in a laboratory setting, subjects shift their pronunciations towards the stimuli they are exposed to. This provides a certain degree of support for anti-acquisitionism, as it shows that individuals' pronunciation patterns are not entirely fixed, and can change substantially under the right circumstances. However, it is not clear whether this type of short-term phonetic convergence can have lasting effects on the speech of individuals. In order to determine this, we have to look at long-term studies of changes in individuals.

Although real-time studies of changes in individuals are notoriously difficult to implement, there is now a solid body of evidence suggesting that such changes do occur. One particular line of research has been focusing on a phenomenon whose existence has probably always been acknowledged, but was previously only supported by anecdotal evidence: the late adoption of non-native dialects. This typically happens when speakers relocate to a different dialect area, and receive considerable exposure to the new dialect. To cite an example, Sankoff (2004) presents a longitudinal study of two Northern English speakers spanning 35 years. While the lexical sets STRUT-FOOT and TRAP-BATH have the same realisations in Northern English dialects, most other dialects have a distinction

between STRUT and FOOT, and some of them also between TRAP and BATH. After being exposed to dialects where all of these lexical sets are realised differently, both speakers successfully unmerged STRUT and FOOT, and one of them also TRAP and BATH. Evans & Iverson (2007) also look at long-term changes in Northern English speakers exposed to Southern English speech (as university students). They demonstrate gradient shifts in the realisations of the STRUT and BATH lexical sets over time. A related study by Sancier & Fowler (1997) shows that a native speaker of Brazilian Portuguese studying in the United States exhibits gradient shifts in the voice onset time (VOT) of fortis consonants: her VOT is shorter after spending several months in Brazil, and longer after spending time in the US. All of these studies suggest that adults are capable of substantially modifying their speech after being exposed to speech patterns different from their own.

Perhaps even more relevant is the series of studies conducted by Harrington and colleagues (Harrington et al. 2000, Harrington 2006, 2007), which investigates long-term gradient changes in an individual that parallel changes in the speech community they belong to. These studies focus on Queen Elizabeth II's annual Christmas Broadcasts, looking for changes in her vowel realisations over a period of 50 years. Harrington et al. (2000) and Harrington (2007) find several changes which seem to mirror shifts that took place in Received Pronunciation: the Queen's vowel space undergoes a significant expansion along the dimension of F1, mainly due to the lowering of the TRAP vowel, and a compression along the dimension of F2, at least partly due to the fronting of the GOOSE vowel. Similarly, Harrington (2006) demonstrates that the Queen's realisation of the KIT vowel in word-final position has become more tense (a process called 'happy-tensing'), following a similar shift among speakers of Standard Southern English. These observations are particularly relevant to this thesis inasmuch as they show that individuals well past the critical period can participate in ongoing sound changes through gradient shifts in their category realisations. The results relating to long-term changes presented in the preceding two paragraphs provide strong support for the assumption that individuals are capable of updating their category representations.

One final study that should be mentioned here is Bane et al. (2010), which investigates lasting changes in the speech of individuals at a much finer time-

scale. Following Bane et al. (2010), I will refer to these as ‘medium term’ changes. The data that serves as the basis of this paper comes from a reality television show produced in the United Kingdom, and consists of VOT measurements of fortis plosives from four speakers taken regularly over a period of three months. They find significant changes in VOT as a function of time, which are related to perturbations in the social dynamics of the group. Interestingly, the patterns of change observed in this study are non-monotonic in that some speakers’ VOT shows both increases and decreases within the same three month period. While the changes reported in Bane et al. (2010) are not simple patterns of convergence (as opposed to the changes reported in the short-term and long-term studies discussed above), they clearly demonstrate that adjustments to category realisations can take place in a continuous fashion, providing further support for the main claim of this section.

Although the results presented above clearly support the anti-acquisitionist position (even if they do not rule out the possibility of changes that occur in acquisition), a few minor qualifications are in order before we move on to the next section. Most of the studies discussed above find that the extent of the changes is affected by a number of social and linguistic variables. It has already been noted that the shifts in VOT reported in Bane et al. (2010) are conditioned by social perturbations. Pardo (2006) and Babel (2009) also find that social factors have an important role in predicting whether a subject will shift towards a new pattern. Pardo (2006) reports that the conversational role of the subject and their sex affect the extent to which they change their speech. Moreover, Babel (2009)’s study reveals that convergence towards the input pattern can be inhibited when the participant has a relatively negative social attitude towards the speaker who produces the input stimuli. Linguistic factors also play an important role in the phonetic shifts reported above. Babel (2009) reports that certain vowels are more likely to exhibit shifts than others, and that this might be related to the large amount of variability shown by these vowels. Nielsen (2008) finds that there are no phonetic shifts along the dimension of VOT when the shifted pronunciation would endanger the contrast between fortis and lenis obstruents. Word frequency can also affect short-term changes: according to Goldinger (1998), low-frequency words are more likely to shift towards the input stimuli than high-frequency words. I believe that these

conditioning factors will prove crucial in understanding the dynamics of sound change at a fine-grained level, although many of these effects will need to be replicated in further experiments to ensure their validity. However, this thesis looks at the behaviour of sound systems at a much broader level. Therefore, the conditioning factors on phonetic shifts in individuals will not form part of the simulations presented in the next chapters.

3.4 PHONETIC BIASES

This section presents an argument to the effect that speech production and perception are affected by weak phonetic biases that apply universally and automatically. One example has already been briefly discussed in Chapter 1: it is widely reported that vowels tend to undergo centralisation in prosodically weak positions (see e.g. Gendrot & Adda-Decker 2007 for an investigation that demonstrates this effect for Arabic, English, French, German, Italian, Mandarin Chinese, Portuguese and Spanish). This tendency is usually attributed to a phonetic bias that causes vowels to ‘undershoot’ their target articulation when the vowel is not sufficiently long, or produced with less articulatory energy (Lindblom 1963; Szeredi 2010). Some type of relationship between purely phonetic factors and properties of sound systems is almost universally assumed. A number of hypotheses about the nature of this relationship have already been discussed in Chapter 2. This might make it somewhat unclear why phonetic biases need to be argued for in a separate section. The motivation for the foregoing discussion lies not in the absence of phonetically oriented approaches to sound change and phonology, but in the lack of rigour with which the notion of ‘phonetic bias’ is treated. It might not be necessary to argue for the general role of phonetics in shaping sound systems, but any specific operationalisation of phonetic biases needs to be presented explicitly and compared to the existing alternatives. Accordingly, this section starts with an attempt to define phonetic biases in a clear and concise way, while the rest of the section ties this definition to existing accounts and provides arguments that support it.

Before I present the definition of phonetic bias adopted in this thesis, it will be useful to provide a schematic description of the processes of speech production and perception. The stages identified in this process will be used to

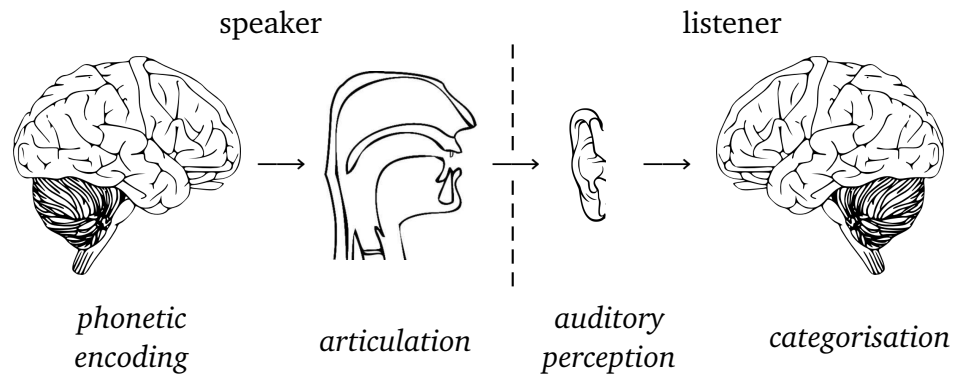


Figure 3.8: *The late stages of speech production and the early stages of speech perception.*

delimit the range of effects that are called phonetic biases. Since our focus is on sound categories, only the late stages of speech production and the early stages of speech perception will be discussed. Processes related to the sequential order of sound categories will be ignored. The diagram in Figure 3.8 presents a summary of the relevant processes in production and perception (based on Kess 1992 and Levelt et al. 1999). The first stage is *phonetic encoding*, where the speaker selects a phonetic target for a given category. In the next stage (*articulation*), this target is transformed into a set of articulatory instructions, which are carried out by the vocal tract. When the acoustic signal reaches the listener, it is first transformed into a raw psychophysical representation by the auditory apparatus in the course of *auditory perception*. This representation serves as the input to the last stage, where a number of higher-level processes apply to find a category label for the stimulus (*categorisation*). Crucially, while phonetic encoding and categorisation rely strongly on the mental representation of sound categories (and will be referred to as *higher-level* processes), articulation and auditory perception are both performed automatically, with little intentional control (and will be referred to as *low-level* processes).

Since this thesis looks specifically at the role of universal phonetic factors in sound change, phonetic biases have to be defined without reference to language-specific category representations. Therefore, the definition is restricted to low-level processes (that is, processes that apply after phonetic encoding and before categorisation). Moreover, it is important to differentiate between *consistent*

and *accidental* effects: production and perception are subject to a certain degree of random noise as a result of the large number of variables affecting them, but some effects are found consistently in certain environments. For example, there is always a certain amount of accidental variation in the quality of a given vowel category, but it is only in non-prominent positions that centralisation takes place consistently. In this thesis, only the latter effect is termed phonetic bias. In the following definition, *target* refers to the output of phonetic encoding, and *percept* to the input to categorisation:

- (3.9) A *phonetic bias* is a consistent mismatch between a class of targets and the corresponding percepts, which results from low-level properties of articulation or perception.

It is important to note that this definition draws a sharp distinction between learnt patterns and biases by asserting that the latter come from low-level processes. This means that phonetic effects can apply at two distinct levels both in production and perception. In other words, there are two different kinds of phonetic patterns: high-level and low-level (or learnt and universal, respectively). The ubiquity of fine-grained cross-linguistic variation in phonetic realisation along with an ability to produce and perceive fine phonetic details argues strongly for the existence of high-level patterns (see Pierrehumbert 1999, Hawkins 2003 and Section 6.1 of this thesis for further discussion). As for low-level phonetic patterns (i.e. phonetic biases), the recurrence of certain types of phonetic effects in genetically unrelated languages suggests that they must share some universal core, even if they differ in their degree and particular details.

The example of vowel centralisation will help to make this distinction clearer. According to the view presented above, vowel centralisation manifests itself in the phonetic make-up of different languages in two forms: as a low-level phonetic bias and a high-level learnt phonetic pattern. Note that I am not suggesting that some languages exhibit only the low-level pattern and some others only the high-level one, but that the two patterns exist alongside each other within the same language. At this point, it will be useful to remind the reader of the account of sound change that serves as one of the main inspirations for the arguments developed in this thesis: the nudge model. As it has already been explained in Section 2.3.2, the nudge model views learnt patterns as

resulting from the accumulation of the effects of phonetic biases in category representations. This accumulation is mediated by a feedback loop based on the process of category update described in the previous section. For instance, each production of a vowel in a non-prominent environment will be slightly displaced towards the center of the vowel space through vowel undershoot. These consistent displacements will be fed back into the representation of the category, nudging the category representation ever closer to the center of the vowel space. It is in this sense that a given language can show the effects of both high-level and low-level manifestations of a given phonetic pattern. The phonetic target will be chosen in accordance with a learnt pattern of vowel centralisation (which can be traced to the low-level effect in a historical sense); and the target will be further displaced towards the centre of the vowel space through the automatic application of a bias towards undershoot.

The effect of the phonetic bias will never be seen on its own, as all observable productions of a given vowel category will already carry the influence of both high and low-level phonetic effects. This is not equivalent to saying that every language will have phonologised the phonetic bias (at least not if the term phonologisation is used in the conventional way). A learnt phonetic effect does not have to be particularly robust and certainly does not have to resemble categorical phonological patterns. What I am proposing here is simply that the low-level phonetic bias will inevitably give rise to higher-level patterns through the production-perception feedback loop. Such high-level phonetic effects may be relatively weak in many cases, but they will still be learnt. The simulations presented in the next chapters will illustrate this point in more detail.

The definition of phonetic biases given in (3.9) is compatible with most existing approaches to the relationship between sound change and phonetics, even if many of these do not explicitly distinguish between low-level and high-level effects. Although a comprehensive review of these accounts is outside the scope of this thesis, I will briefly discuss a few representative examples below (note that most of these approaches have already been discussed in detail in Chapter 2). First, Hyman (1976) suggests that phonological patterns emerge from ‘intrinsic’ phonetic variations, which are ‘the universal, innocent by-products of [...] extrinsic gestures’ (Hyman 1976: p. 407). This is a clear reference to phonetic biases in the same sense as in (3.9). Ohala (1981, 1993)

and Blevins (2004, 2006) also trace sound change to ‘language universal factors, i.e. physiological and psychological factors common to all human speakers at any time’ (Ohala 1993: p. 238) in their transmission-based account of phonetically natural changes. Boersma & Hamann (2008) attempt to account for auditory dispersion in an optimality theoretic framework by invoking a combination of auditory cue constraints and articulatory constraints. The latter of these are very similar to phonetic biases in that they embody a tendency towards minimal effort in articulation, which is clearly a low-level property of production.

Garrett & Johnson (2013) is perhaps even more relevant to the present discussion, since it not only acknowledges the importance of phonetic biases in sound change, but also presents a typology of different bias factors. Although some of their categories are not entirely compatible with the present account in that they cannot easily be classified as high-level or low-level effects (e.g. perceptual confusion), two of them correspond very closely to phonetic biases: aerodynamic constraints and gestural mechanics. It will be useful to briefly discuss some of the bias factors they list under these headings, as this might help to make the notion of phonetic bias a little more concrete. Two different aerodynamic constraints are mentioned in the paper. The first of these has already been discussed in Chapter 2: stops tend to be devoiced as a result of the conflict between the low supraglottal pressure required for voicing and the high supraglottal pressure involved in the production of stops. The second constraint shifts voiced fricatives towards glide-like articulations as a result of a similar conflict between the low supraglottal pressure needed for voicing and the high supraglottal pressure associated with frication. It should be clear that both of these constraints result from universal articulatory factors that have nothing to do with high-level aspects of production. Gestural mechanics – the second type of phonetic bias factor highlighted above – relate to inertial properties of articulators. The most relevant example given by Garrett & Johnson (2013) is ‘gestural blend,’ whereby ‘the phonetic plan for an utterance places competing demands upon a single articulator’ (Garrett & Johnson 2013: p. 63). Gestural blends comprise a wide range of phonetic patterns, most of which are examples of coarticulation, to use more traditional terminology. Garrett & Johnson (2013) discuss the example of the fronting of [k] in the sequence [ki] as opposed

to [ku], which results from the simple fact that the tongue dorsum already anticipates the front articulation of the vowel in [ki] during the initial stop.

Although the discussion above clearly demonstrates that phonetic biases play an important role in other accounts as well, the widespread acceptance of the notion does not justify its use in itself. There are two independent arguments that can be advanced for phonetic biases. The first of these has already been alluded to above. The fact that a wide range of phonetic and phonological patterns are shared among numerous unrelated languages remains somewhat of a mystery unless we assume that they all follow from the same universal phonetic sources. This is not to say that the patterns themselves are simple manifestations of universal biases. As it has been noted above, each language likely shows the effects of high-level and low-level incarnations of the same patterns at the same time, which means that probably all phonetic effects have an important language-specific component as well. Nonetheless, the high-level pattern and the low-level one are linked to each other in a historical sense, which explains how a universal phonetic pressure can lead to a wide range of very similar but still learnt phonetic processes in a variety of different languages.

The second argument is based on the fact that every speech scientist has to assume certain strong universal phonetic biases regardless of their take on the role of learning in phonetics. For instance, the oral cavity provides a limited amount of space in which articulatory manoeuvres can be performed: the tongue cannot move beyond the palate, nor can it fall through the jaw, which limits the range of possible vocalic and consonantal articulations considerably. Moreover, the articulators cannot teleport from one position to another, which means that there will always be a certain amount of coarticulation between neighbouring sounds relying on the same articulator. Since these restrictions result from elementary properties of physics, the assumption of at least some phonetic biases is inevitable regardless of one's view of phonetics.

This takes us to the last topic discussed in this section: the strength of phonetic biases. While there is no doubt that there exist a set of strong biases resulting from the physiology of the vocal tract, the claim at the very beginning of this section explicitly refers to *weak* universal biases. In this thesis, weak phonetic biases are defined as low-level consistent pressures with clear targets

(e.g. the centre of the vowel space in the case of vowel undershoot), whose effects consist in small movements towards these targets without necessarily reaching them. For example, vowel centralisation creates a clear difference between vowels in prominent *versus* non-prominent positions, but it does not necessarily turn all unstressed vowels into schwas. Contrarily, the physical boundaries of the vocal tract constrain articulatory movements in an inviolable way: there is no way in which the articulators can move beyond certain limits. The strength of biases likely varies on a continuous scale from insignificant to inviolable. This thesis focuses on weak biases only, given that these are the phonetic pressures that can lead to sound change, but do not necessarily do so. The potential effects of bias strength will be investigated in more detail in the following chapters.¹¹

3.5 THE FILTERING LISTENER

In Section 2.4, I discussed functionalist models of sound change, which focus on the tendency to maintain contrasts. I noted that many such models are problematic, since they assume that speakers actively optimise their sound systems. Near the end of the section, an alternative model was introduced, which locates the source of contrast maintenance in language use. The main idea was that productions of a given category that are easily perceivable as belonging to a different category will play a diminished role in category update, leading to a pressure for categories to stay well-separated. This hypothesis about the role of the listener in contrast maintenance will be referred to as the ‘filtering listener’ approach after Padgett (2011). In what follows, I give a detailed description of this approach, and cite a number of phenomena where this type of explanation has proven useful. Although this thesis will not model any other factors relating to contrast maintenance, I provide a brief overview of these at the end of the section.

One of the earliest attestations of the filtering listener hypothesis is in Martinet (1952):

11. The notion of bias strength has been examined in two recent papers by Moreton (2008) and Yu (2011), but their measures of bias strength are not applicable to the present problem, since they make no distinction between low-level and high-level phonetic effects.

For each [phoneme], in a given context at least, there must be an optimum which we might call the center of gravity of every range of dispersion, but actual performances will normally fall somewhat off the mark. In the normal practice of speech, some of them are even likely to fall very far off. If too dangerously near the center of gravity of some other phoneme, they may be corrected, and, in any case, *will not be imitated*. If not unusually aberrant, slightly beyond the normal range of dispersion, but not in a direction where misunderstanding might arise, they would in no way threaten to impair mutual understanding.

(Martinet 1952: p. 5; emphasis mine)

What Martinet seems to be suggesting is that listeners are capable of somehow filtering out ambiguous tokens. As it was noted in Section 2.4, Martinet does not explain how this filtering actually happens, and is therefore often accused of attributing implicit knowledge of optimal sound systems to speakers (see e.g. Wedel 2004: p. 129). While I believe that this criticism might be a little too harsh, I agree that a more rigorous description of the filtering mechanism is necessary.

Labov (1994) provides a particularly illuminating discussion of how filtering can take place in a listener. His starting assumption is that '[...] it is not the *desire* to be understood, but rather the *consequence* of misunderstanding that influences sound change' (Labov 1994: p. 586). In other words, he suggests that the source of contrast maintenance lies in the misperception of ambiguous productions. Labov provides the following example for the mechanism he proposes. Consider a partial vowel system like the one illustrated in Figure 3.9. The three categories are well-separated in phonetic space, but a single production of /o/ (as in *block*) happens to fall within the range of /æ/ (as in *cat*). It is possible that the lexical item containing this token will still be recognised, but there is a small chance that recognition will fail. If such a failure takes place, the aberrant example of /o/ will not contribute to the 'pool of tokens' that determine the distribution of /o/. This can be rephrased as follows in the terminology of the present paper: ambiguous tokens might fail to take part in category update. Labov provides another, slightly modified example as well, which is illustrated in Figure 3.10. In this case, the categories /æ/ and /o/ are separated by a wider

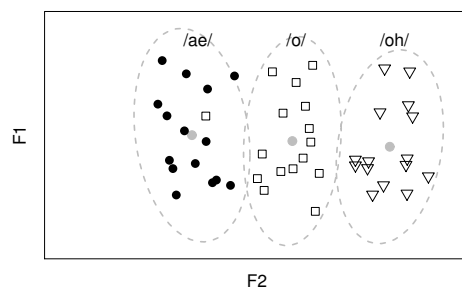


Figure 3.9: Three non-overlapping low vowel categories. A single production of /o/ falls within the range of /æ/, but does not get fed back into the representation of the category /o/ due to misperception.

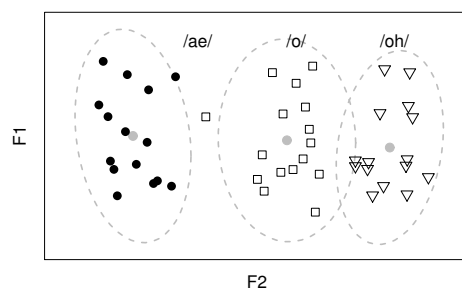


Figure 3.10: Three non-overlapping low vowel categories. Even extreme productions of /o/ will not fall in the range of /æ/, which means that they will be fed back into the representation of the category /o/.

margin, which means that the peripheral token of /o/ will likely be perceived correctly. As a result, it will contribute to category update. These tendencies have important consequences in each case. In the scenario in Figure 3.9, the relative position of the three categories is likely to remain the same, given that /æ/ and /o/ cannot shift any closer to each other without causing a large number of misperceptions. However, the scenario in Figure 3.10 can easily lead to change, as both /o/ and /æ/ are free to shift towards each other due to the large gap between them. It should be noted that the mechanism proposed by Labov does not assume any goal-directed behaviour promoting contrast maintenance on the part of the listener: the observed patterns emerge as a by-product of simple misperception.

Guy (2003) presents a similar account of the interaction of morphological factors with coronal stop deletion. While a large proportion of final coronal stops are deleted in monomorphemic words like *mist*, this proportion is decreased significantly in words where the final coronal stop marks a past tense form or a past participle (e.g. *missed*). Guy locates the source of this difference in misperception. Past tense forms without a final coronal stop can easily be mistaken for present tense forms. If they are, they will be added to the listener's distribution of present tense forms and will not count towards the past tense distribution. If this happens regularly, it might create the illusion that the rate of coronal stop deletion is lower in past tense forms than in monomorphemic ones, simply because many relevant examples are removed from the corpus of past tense forms. This leads to the desired result: coronal stop deletion will occur less in morphologically complex forms than in monomorphemic ones. This mechanism is not entirely identical to the one described above, even though its effects are essentially the same: in the former example, the misperceived token was not added to the representation of either vowel category, while in the latter case the misperceived token is added to the wrong representation. Blevins & Wedel (2009) term the former process 'variant pruning', and the latter one 'variant trading'. The simulations in the next chapters only implement variant pruning, which might be seen as an unwarranted omission, but this is unlikely to make a significant difference given that the two processes have very similar effects.

Wedel (2006) relates the filtering listener hypothesis to an important observation about lexical access in psycholinguistics: lexical items with a large

number of close lexical neighbours are harder to retrieve than lexical items in sparser areas of phonetic space (Luce & Pisoni 1998). Luce & Pisoni (1998) argue that lexical access involves a competition between lexical items similar to the stimulus perceived by the listener. If the stimulus activates a densely populated area of phonetic space, heavy competition among similar items might '[delay processing] to the point that the percept will fail to be assigned to any lexical category at all' (Wedel 2006: p. 264). As Wedel points it out, overlapping category representations will often lead to more competition in lexical access, which will result in higher rates of failure (especially if the contrast between the categories carries a high functional load; cf. Blevins & Wedel 2009). This account also points out an important feature of many approaches relying on the filtering listener hypothesis: their lexical orientation (see e.g. Silverman 2012 for a clear statement of this position). That is to say, most authors suggest that this type of misperception occurs mainly when the neutralisation of a contrast would create homonyms or near-homonyms. This might well be the case, but I believe there is no reason to assume that productions leading to non-neutralising cross-category confusion (e.g. *drop* pronounced as *drap*) cannot be misperceived in a similar way. Unfortunately, there is currently no conclusive empirical evidence for either a strictly lexical approach or one that also allows non-neutralising forms to be filtered out.

Although the filtering listener hypothesis is based mostly on indirect evidence, there is a particular finding in the literature that argues strongly for at least a functionally equivalent mechanism. As it has already been noted in Section 3.3, Nielsen (2008)'s study of phonetic imitation reports a curious asymmetry: fortis stops with artificially lengthened VOT are readily imitated by the subjects, as opposed to fortis stops with shortened VOT, which are not. This result can be interpreted as a manifestation of the tendency towards contrast maintenance. Note that imitation in this study was not immediate: the subjects first listened to all the artificially modified stimuli, and then had to produce the same (and related) words in a separate block. Thus, Nielsen's finding is compatible with the filtering listener account: it is possible that some ambiguous tokens in the listening block did not take part in category update, or were at least downgraded in some way by the participants, and therefore did not significantly influence their production.

Let us now turn to alternative accounts of how contrast maintenance is achieved in speech. The mechanisms described below will not be incorporated into the simulations presented in the next chapter, but it will be instructive to compare them to the filtering listener hypothesis. Perhaps the best-known alternative hypothesis is Lindblom's (1990) H&H theory, which sees hyperarticulation and hypoarticulation (the Hs in H&H) as the speakers' adaptive responses to various constraints on speech. Hypoarticulation reduces the amount of effort needed to produce speech, whereas hyperarticulation increases the clarity of the message being transmitted. It is the latter effect that can result in contrast maintenance when the production of ambiguous tokens would lead to a significant loss of information. Crucially, this approach assumes that the source of contrast maintenance is the speaker, who implicitly monitors the communicative efficiency of the signal, and makes adjustments where necessary (and is therefore termed the 'considerate speaker' hypothesis by Padgett 2011). In this sense, the H&H approach is much like the goal-oriented accounts described in Section 2.4. It is arguably less plausible than the filtering listener hypothesis, which does not assume that speakers actively optimise their speech. However, it will only be possible to decide conclusively between these two hypotheses once they have been compared rigorously through empirical methods.

At this point, it is worth mentioning that there is another account that identifies the speaker as the source of contrast maintenance, but which does not assume the same amount of communicative awareness. Tilsen (in press) suggests that while neighbouring sounds can exert coarticulatory pressures on each other, low-level 'inhibitory interactions between contemporaneously planned articulatory targets result in dissimilatory effects, and over time these effects can prevent speech targets from becoming perceptually indistinct' (Tilsen in press). These inhibitory interactions can keep the assimilatory effects of coarticulation in check. Although Tilsen presents detailed experimental evidence for this view, further investigation is needed to see whether this approach is to be preferred to the ones discussed above.

In conclusion, the hypothesis that category update can be inhibited in situations where the produced variants create ambiguity is widely assumed in the literature, and is capable of accounting for contrast maintenance in an elegant manner without requiring explicit optimisation on the part of the

speakers. There is also a limited amount of empirical evidence for this view from experiments such as Nielsen (2008), but this evidence in itself is not sufficient to conclusively determine which approach is best suited to explaining contrast maintenance. It is, however, important to note that the existence of alternative explanations is not a major issue for this thesis. While the mechanisms discussed above have very different sources, their effects are the same: they create a situation where ambiguous parts of the speaker's category distributions are underrepresented in category update. Whether this occurs because the speaker intentionally avoids such ambiguous productions, or because the listener filters them out does not crucially alter the dynamics of the systems examined in the next chapters.

3.6 SUMMARY

The main goal of this chapter was to explain the theoretical assumptions that form the basis of the simulations presented in the rest of this thesis, and to provide detailed arguments for each of them. This is necessary inasmuch as the usefulness of a model without a well-motivated theory behind it is highly questionable. For the sake of clarity, the theoretical assumptions discussed in this chapter are listed again below:

1. speech production and perception rely on abstract category representations implemented in a probabilistic way (3.1, 3.2);
2. the category representations of speakers are subject to continuous update throughout their lifetimes (3.3);
3. speech production and perception are affected by weak phonetic biases that apply universally and automatically (3.4);
4. productions whose category membership is ambiguous play a diminished role in category update (3.5).

These assumptions will be the main design features of the artificial sound systems examined in the remaining chapters. What I intend to show is that the dynamics of these systems follow directly and inevitably from these principles. That is to say, the behaviour of the simulations is not simply a result of specific

tweaks in their implementation, but a substantive prediction of the theory of category production, perception and update described in this chapter. The task of presenting the details of the simulations will be taken up in the next chapter. In order to consolidate the link between the behaviour of the simulations and the underlying principles, this chapter will also test a number of different ways of implementing these systems, showing that implementational details do not crucially impact their dynamics.

This chapter describes the implementation details of the simulations that serve as the core of the argument presented in this thesis. These simulations all look at artificial sound systems whose behaviours are driven by the theoretical assumptions fleshed out in the previous chapter. The general structure of the simulations is essentially the same throughout this thesis, although some smaller details vary across different runs. Those parts of the simulations that can vary will be referred to as their ‘parameters’. The main focus of the simulations is on the so-called production-perception feedback loop (Pierrehumbert 2001): they attempt to capture the evolutionary dynamics of a system where each production has a chance to influence future productions by being fed back into category representations. In other words, the simulations look at changes in category representations over many thousands of iterations of simple speech interactions consisting of the production of a token and its subsequent perception. Although these types of simulations have already been examined in some detail by Pierrehumbert (2001, 2002) and Wedel (2006), the present chapter expands on these works in two key respects. First, it looks at the ways in which different types of category representations influence the dynamics of these systems by comparing implementations based on exemplar and prototype theory. Second, it provides justification for the idea of simulating an abstract version of the production-perception loop by comparing single-agent simulations to ones with larger communities of agents.

The main motivation for the comparisons mentioned above is to show that the simulation results in Chapters 5 and 6 are direct consequences of the general principles described in the previous chapter. However, their results are also important in their own right. They contribute to ongoing debates about usage-based models of sound change and clarify the fundamental mechanisms underlying the production-perception feedback loop.

The usefulness of comparing exemplar-based and prototype-based implementations can be better appreciated in the light of current trends in computational approaches to sound change. A significant portion of computational research within historical phonology has emerged out of the usage-based models advocated in Bybee (2001) and Pierrehumbert (2001), and therefore takes many of their assumptions for granted. Perhaps the most important of these is the idea of exemplar-based storage, according to which phonetic categories are represented by detailed memories of concrete utterances. For instance, an exemplar theoretic model would represent the category [u] by recording the phonetic details of all heard instances of this vowel. Although there is a considerable amount of evidence that speakers do rely on exemplars in their production and perception (see e.g. Goldinger 1996, 1998, Johnson 1997, Bybee 2001), the notion of exemplar-based representations has been met with a certain amount of skepticism among historical phonologists (see e.g. Bermúdez-Otero 2007, to appear for some criticism). The resistance to the idea of exemplars has the unfortunate consequence that the important results in works such as Pierrehumbert (2001) and Wedel (2006) have also been neglected. While I happen to agree with the hypothesis of exemplar-based storage, I would like to argue that the results in these papers (and this thesis) do not crucially depend on this assumption, insofar as they also emerge in more conservative prototype-based models. This, however, is a non-trivial claim that needs to be investigated systematically. Through this investigation, I hope to bring the arguments of this thesis and other computational work closer to researchers who might otherwise be opposed to some aspects of usage-based models.

The comparison of single-agent and multi-agent simulations is motivated by the fact that many existing simulations (including the ones referred to above) are based on a simplified model of speech interactions that might seem counterintuitive in certain ways. Specifically, they model a situation with a single ‘soliloquising’ agent, who serves both as the source and the target of the productions. Although this model has proven useful and instructive in many ways, it is not clear whether it applies to more natural situations with several speakers.¹ The question is whether a simulation that models speech interactions

1. Wedel (2004) reports on a simulation with two speakers which suggests that at least some aspects of such models do apply to multi-agent simulations as well.

among multiple agents produces similar behaviour with respect to the evolution of sound systems. This is an important issue, since some of the design features of single agent simulations are somewhat unrealistic. For instance, it is not clear in what sense a misperception can take place when there is only a single speaker. The present chapter provides a clear solution to this problem by demonstrating that the general dynamics of the simulated systems are independent of the number of agents, and that the behaviour of abstract single-agent models is identical to that of multi-agent implementations in key respects. The most important consequence of the close parallels between single and multi-agent simulations is that the former can be substituted for the latter in most situations. This is a significant finding to the extent that single-agent simulations are much more tractable, and in some cases their behaviour can be accurately predicted through mathematical methods without having to run the simulations themselves. The next two chapters rely heavily on these observations.

Thus, while the goal of the following sections in the narrow context of this thesis is simply to set the scene for the simulations in the next two chapters, some of their findings have important implications outside this thesis as well. The structure of this chapter follows straightforwardly from the summary presented above. Section 4.1 lays out the formal details of the simulations, discussing a number of different alternatives where necessary. Then, Section 4.2 compares the behaviour of exemplar and prototype models. Finally, Section 4.3 investigates the relationship between single-agent and multi-agent implementations.

4.1 THE STRUCTURE OF THE SIMULATIONS

The simulations reported in this thesis are all based on the same implementation of the production-perception feedback loop, illustrated in Figure 4.1 (note that this implementation is almost identical to the one described in Wedel 2006). Here is a step by step description of this procedure. First, the simulated agent selects a production target by sampling one of their category representations (Section 4.1.1). This roughly corresponds to producing a given segmental category in the context of a specific utterance. In the next step, this production target is slightly displaced by one or more low-level phonetic biases (Section

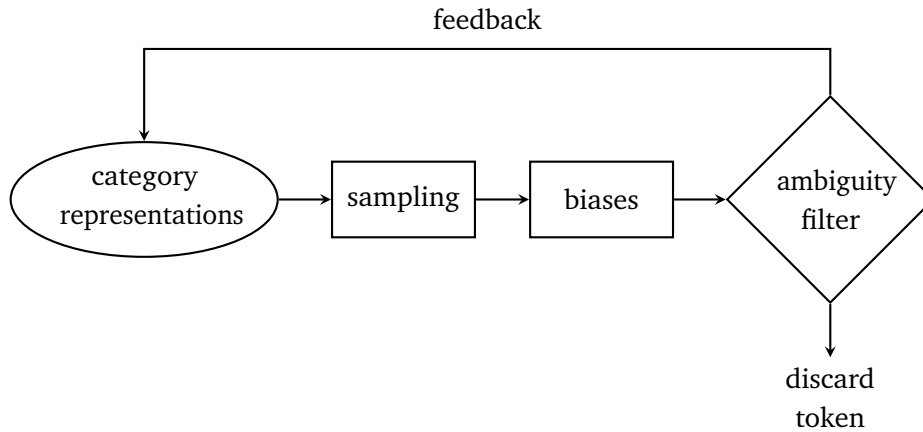


Figure 4.1: A schematic outline of way the production-perception feedback loop is implemented in this thesis.

4.1.2). Finally, the resulting token is fed back into the agent’s category representations or discarded depending on whether it is successfully perceived (Section 4.1.3). Note that the processes referred to above all correspond closely to the aspects of speech production and perception embodied in the theoretical assumptions presented in the previous chapter. The following sections describe each of these steps in detail.

4.1.1 Category representations and sampling

Echoing arguments from Ashby & Alfonso-Reese (1995) and Kirby (2010: pp. 41–45), Section 3.1 suggested that category representations can be modelled as probability distributions created through a process of density estimation. As will be shown in the following sections, the notion of density estimation is particularly useful in the present context insofar as it allows us to treat exemplar and prototype models in the same conceptual framework. This section focuses on the ways in which probability distribution-based representations can be used in production.

Let us first discuss prototype models of category representation. Prototype models have been developed to account for subjects’ behaviour in categorisation experiments, where they are asked to assign category labels to stimuli located in a continuous psychophysical space (see e.g. Posner & Keele 1968). For instance, subjects might be asked to assign different colours to two or more broad colour

categories after a training phase with a limited set of stimuli (e.g. Nosofsky 1988). One pattern that emerges almost universally from such experiments is that subjects can generalise learnt patterns to new stimuli, and that their category decisions are guided by similarities between the training stimuli used to establish the categories in the experiment, and the test stimulus they are asked to label. Thus, if the two categories are BLUE and RED, subjects will likely label a previously unseen violet² stimulus as BLUE, and a pink stimulus as RED. Prototype models explain this behaviour by suggesting that the probability of assigning a category label to a given stimulus is proportional to the amount of similarity between the stimulus and a so-called ‘category prototype’. Category prototypes typically represent the central tendency of all the stimuli belonging to a given category; in the case of BLUE, this would be the specific shade of blue that best represents the broad range of colours normally referred to as blue.

Ashby & Alfonso-Reese (1995) develop a mathematical argument showing that the behaviour of prototype models is identical to that of a model representing categories through parametric probability distributions. Although their original reasoning is based on perception, it readily extends to production as well. Therefore, the illustration below is based on speech production.

Figure 4.2 shows 25 tokens of the vowel [u] from American English (based on Hillenbrand et al. 1995) and a parametric probability distribution estimated from these tokens. Parametric in this context means that the estimation method makes an *a priori* assumption about the distribution of the data, and attempts to find the set of parameter values that create the closest possible fit between the data and the chosen type of distribution. In the case illustrated in Figure 4.2, the assumption is that the data are normally distributed, which means that density estimation consists in finding the appropriate parameter values for a function whose shape is Gaussian (i.e. a normal distribution). Gaussians can be defined in terms of two parameters: mean and standard deviation. Therefore, the problem of density estimation reduces to finding the mean and the standard deviation of the sample (Section 4.1.3 provides more details on how this estimation is performed for Gaussians). If we assume that productions of [u] are based on the normal distribution shown in Figure 4.2, we get the type

2. For those, like the author of this thesis, who are a little shaky on colour names: violet is essentially bluish purple.

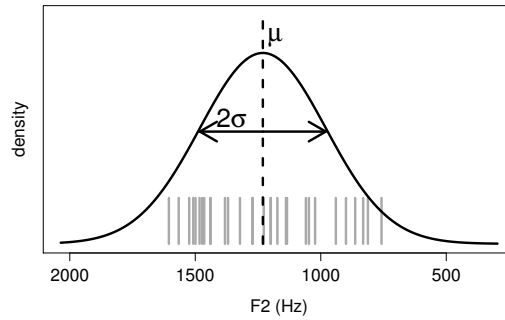


Figure 4.2: 25 tokens of the vowel [u] from American English (grey vertical lines), the corresponding parametric probability density function (the black curve) and its two parameters.

of behaviour characteristic of prototype models: most production targets come from the vicinity of the peak of the distribution (corresponding to the prototype), and the likelihood of producing a token with a given F2 value falls off as we move away from this peak.³ Given this parallelism, prototype-based category representations in this thesis take the form of univariate and multivariate normal distributions, depending on the number of phonetic dimensions in the simulation. The probability density function of a univariate normal distribution is as follows:⁴

$$p(x|c_i) = \frac{1}{\sigma_i \sqrt{2\pi}} \exp\left(-\frac{(x - \mu_i)^2}{2\sigma_i^2}\right), \quad (4.1)$$

where $p(x|c_i)$ is the probability of producing a form with value x along a specific phonetic dimension given a category label c_i , σ_i is the standard deviation and μ_i the mean of category c_i . The multivariate case is more complicated:

3. Of course, this is merely a demonstration – Ashby & Alfonso-Reese (1995) present a much more structured argument showing that the parallel between parametric probability distributions and prototype theory follows from basic mathematical principles of density estimation. It should also be noted that the representations proposed here (which are based on normal distributions with varying means and standard deviations) are more closely paralleled by general quadratic classifiers (Maddox & Ashby 1993) than by traditional prototype models.

4. While these formulae should be straightforward for those familiar with probability theory, some readers might find them a little too arcane. However, it should be noted that understanding the mathematics behind normal distributions is in no way essential for following the arguments developed in the rest of this thesis, and the formulae have only been included for the sake of completeness.

$$p(\mathbf{x}|c_i) = \frac{1}{|\Sigma_i|^{1/2}(2\pi)^{k/2}} \exp\left(-\frac{1}{2}(\mathbf{x} - \boldsymbol{\mu}_i)^T \Sigma_i^{-1}(\mathbf{x} - \boldsymbol{\mu}_i)\right), \quad (4.2)$$

where $\mathbf{x} = (x_1, x_2, \dots, x_n)$ is a vector with a separate value for each of n phonetic dimensions (e.g. F1, F2, F3, etc.), $\boldsymbol{\mu}_i = (\mu_1, \mu_2, \dots, \mu_n)$ is a vector containing the means of category c_i along each of these dimensions, and Σ_i is the covariance matrix of c_i (which specifies its standard deviation along each of the phonetic dimensions and correlations between all pairs of dimensions). Unless otherwise stated, all prototype-based simulations in the rest of this thesis model the selection of production targets by sampling the distributions defined in equations (4.1) and (4.2).

Exemplar models differ from prototype models in that they do not attribute a distinguished role to the category prototype. Instead, category representations are seen as sets of detailed memories, where each of these memories can influence production and perception (Nosofsky 1986, 1988). Much of the criticism levelled against exemplar theory concerns a corollary of this claim, namely that speakers are capable of storing a large amount of redundant information about speech. This is understandable in light of one of the main guiding principles of early generative phonology, namely that phonological representations should be as parsimonious as possible given that ‘[...] storage space is at a premium’ (Kenstowicz 1994: p. 60). Proponents of exemplar theory and related schools of thinking have countered this argument by suggesting that there is no empirical evidence for such ‘aesthetic principles’ (Coleman 2000: p. 111). There is, however, solid evidence for the idea of exemplar storage from experimental psychology, psycholinguistics and related fields (see e.g. Goldinger 1998).

As I have noted above, it is not my goal here to adjudicate between different approaches to category storage. Instead, I am aiming to demonstrate that the question of category storage is irrelevant to the topic at hand (within certain limits, of course). This requires a comparison between prototype and exemplar models, which, however, cannot easily be performed unless both of them are formalised in the same mathematically explicit way. Therefore, let us turn to the formal details of exemplar models.

Figure 4.3a shows a small subset of the [u] tokens from Hillenbrand et al. (1995) along with an exemplar-based density function representing them.

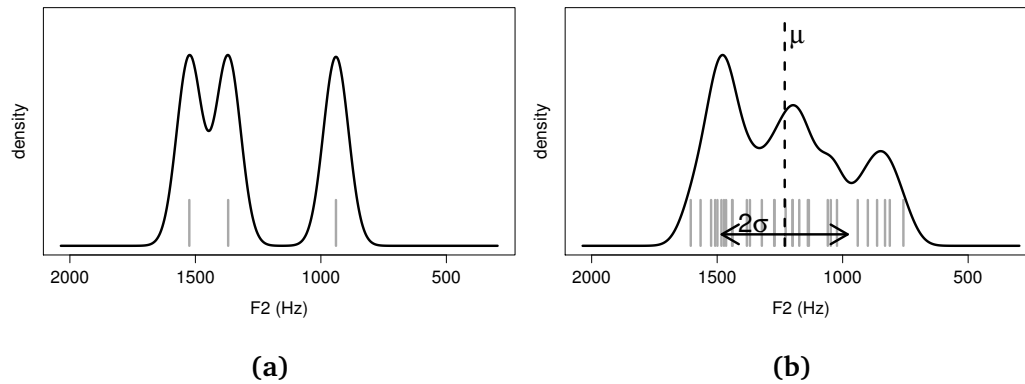


Figure 4.3: (a): 3 tokens of the vowel [u] from American English (grey vertical lines) and the corresponding non-parametric probability density function (the black curve); (b): the same for all 25 tokens.

Before explaining how the probability distribution is obtained from the data points, it will be useful to point out that most approaches to exemplar theory see the memories representing the tokens themselves as the primary units of storage (and not the derived probability distribution). This could be visualised by showing only the grey vertical lines without the curve corresponding to the density function. However, if speech was based entirely on individual exemplars, it is not clear how any intermediate productions could occur (or how previously unheard stimuli could be categorised). Exemplar models overcome this difficulty by suggesting that there is a certain amount of noise in production: while the production targets are concrete exemplars, there is always some inaccuracy in implementing these targets (see e.g. Pierrehumbert 2001). One way to model this is to add a small amount of Gaussian noise to the target values. This is where the curve in Figure 4.3a becomes relevant: the addition of Gaussian noise creates a continuous function, which can be viewed as a probability distribution defined over the set of all possible outcomes.

As it happens, the method of turning discrete distributions (i.e. individual exemplars) into continuous ones (i.e. a probability density function) in this way is a widely used device in statistics, referred to as ‘kernel density estimation’ (Silverman 1986). The basic principle behind this technique should be clear from the example above: the continuous distribution is obtained by summing over a set of functions centred around the data points, each of them having

the same bandwidth or dispersion. Kernel density estimation is not particularly meaningful in examples like Figure 4.3a with only three data points. Figure 4.3b (based on all 25 tokens) is a much more useful illustration which reveals the main advantage of this technique, namely that kernel density estimates can capture irregularities in the distribution of the data that are concealed by parametric estimates like a normal distribution. This ability derives from the fact that kernel density estimates do not have a predetermined shape and cannot be defined through a fixed set of parameters. Because of this, such distributions are commonly referred to as ‘non-parametric’.

Since exemplar models are essentially identical to kernel density estimates (see Ashby & Alfonso-Reese 1995 for a more detailed argument), the properties that make them so popular in experimental approaches can also be traced to their non-parametric nature: they can successfully model patterns of behaviour that cannot easily be accounted for if one assumes parametric representations (e.g. categorisation performance in cases where there are discontinuities in the structure of a category). Below are the formulae that I will use in the exemplar-based simulations in this thesis:⁵

$$p(x|c_i) = \frac{1}{\sum_{j=1}^n w_j} \sum_{k=1}^n w_k K_h(x - x_k) \quad (4.3)$$

$$K_h(x) = \frac{1}{h} \phi\left(\frac{x}{h}\right) \quad (4.4)$$

Equation (4.3) simply says that the representation of category c_i in an exemplar model is the weighted average of n kernels (where n is the total number of exemplars, and w_j is the weight assigned to the j th exemplar), each of which is centred around a given exemplar. The role of the weights is related to the notion of category update and will therefore be discussed in Section 4.1.3. The kernels themselves are defined in equation (4.4), where $\phi(x)$ is the standard normal distribution (i.e. a normal distribution with $\mu = 0$ and $\sigma = 1$), and h is the bandwidth of the kernels. This means that the kernels are Gaussians with variance h^2 . Recall that the original reason for adding the kernels was to

5. Only a univariate version is provided, since all the exemplar-based simulations in this thesis model cases with a single phonetic dimension. It should be noted, however, that this method extends straightforwardly to multivariate cases as well.

account for random noise in production. Therefore, parameter h can be seen as a measure of the amount of production noise in a given simulation.

Modelling production noise through Gaussian kernels has an important additional consequence. A production model based on sampling a kernel density estimate will always produce a set of tokens whose expected sample standard deviation is greater than that of the sample on which the estimate is based. Appendix A presents a mathematical derivation of this claim. The overestimation of the variance of a category will necessarily result in a situation where category variance increases as the production-perception loop is iterated over and over again. Indeed, this is what Pierrehumbert (2001) finds in her own simulations. Pierrehumbert considers this a serious issue and proposes a more complicated sampling mechanism to tackle it. However, variance inflation does not cause any problems in the simulations presented in this thesis, as the phonetic biases, the boundedness of phonetic space and ambiguity-related selection pressures all act against it.

One might also ask whether variance inflation affects prototype models as well. In a nutshell, the answer is no, at least not in their current form. This is because the standard deviation of a normal distribution estimated on the basis of a given sample is equal to the sample standard deviation. This means that the prototype model of category production described in equations (4.1) and (4.2) is not noisy in the same way as the exemplar model. Since there is no reason to assume that a speaker with prototype-based representations should realise articulatory targets more accurately than a speaker with exemplar-based representations, a small modification has to be made to the prototype model. This modification consists in displacing target productions slightly through the addition of Gaussian noise, just as in the exemplar model. Although this step might seem somewhat *ad hoc*, it is not without precedent: Feldman et al. (2009) use the same strategy in their psycholinguistically grounded model of speech production. Appendix A shows that adding Gaussian noise to target productions has exactly the same variance inflating effect as the use of kernel density estimates in exemplar models.

Below is a summary of the speech production mechanisms proposed for exemplar and prototype models.

(4.5) SPEECH PRODUCTION IN PROTOTYPE MODELS:

Production targets are chosen by randomly sampling a normal distribution (equations (4.1) and (4.2)). A small amount of Gaussian noise is added to these targets.

(4.6) SPEECH PRODUCTION IN EXEMPLAR MODELS:

Production targets are chosen by randomly sampling a continuous kernel density estimate (equations (4.3) and (4.4)).

We can now move on to the next stage of the production-perception loop illustrated in Figure 4.1, that is, phonetic biases.

4.1.2 *Modelling the effects of phonetic biases*

This section provides a formalisation of phonetic biases that is compatible with the simulation architecture in this thesis. I will also briefly discuss a related notion, namely ‘bias proportion’, the relative frequency of a given biasing environment within a category. Note that the difference between exemplar and prototype models is not relevant to phonetic biases, which means that a single equation will be sufficient.

The implementation described below is a somewhat modified version of the notion of ‘systematic bias’ proposed in Pierrehumbert (2001). Pierrehumbert models low-level phonetic effects as small displacements of target productions. Since phonetic targets are conceptualised as coordinates in a multidimensional space, this means that a constant is added to each target. The main problem with Pierrehumbert’s approach is that the same constant is added to all production targets, regardless of where they are in phonetic space. In other words, the phonetic bias does not specify a target, only a direction. Since there is no target, the phonetic category will keep moving in the direction specified by the bias until it reaches the boundaries of phonetic space.

A more realistic approach to phonetic biases is to model them as point-like attractors in phonetic space. To make this clearer, consider the case of [u]-fronting in the context of coronal consonants (see e.g. Harrington et al. 2008, 2011). Harrington et al. (2011) describe the source of this effect as follows:

‘[...] a more advanced constriction for /u/ is likely in a coronal context such as /tut/ because the tongue body is dragged forward resulting acoustically in a raised second formant frequency’ (Harrington et al. 2011: p. 122). There is a natural limit on the extent to which a high back vowel can undergo fronting as a result of coarticulation with coronals: it cannot become more front than the coronal consonant itself. This limit can be modelled as a target location in an articulatory (e.g. tongue height and frontness) or an acoustic space (e.g. F1 and F2). Realisations of [u] in a coronal context will be consistently displaced towards this front target location, but the extent of this displacement will be very small. Moreover, this movement is target-oriented, which means that the size of the displacement will be smaller when a given production is already close to the attractor (otherwise the biased production might end up ‘overshooting’ the bias attractor).

There are several different ways of formalising the above model of phonetic biases. In the simulations presented in this chapter and the next one, I use a logistic function to implement the displacements towards the target location:

$$\text{bias}_i(\mathbf{x}) = \mathbf{x} + s_i \left(\frac{1}{1 + \exp\left(\frac{\mathbf{x} - \mathbf{b}_i}{d}\right)} - \frac{1}{2} \right), \quad (4.7)$$

where \mathbf{x} is a vector representing the production target, s_i is a parameter that determines the strength of bias i , \mathbf{b}_i is the location of the bias attractor and d is a scaling factor (set to 1 in all the simulations in this thesis). Figure 4.4 illustrates the size and the direction of the displacement caused by a phonetic bias as a function of where the original production target is in phonetic space ($\mathbf{b}_i = 0$ and $s_i = 0.01$). This function produces the expected results: when a production target has a value that is lower than the bias attractor, the function increments it by a small amount. When the production target has a value that is higher than the bias attractor, the sign of the function changes, which means that it now decrements the original value. Note also that the size of the displacement increases as we move away from the bias attractor, but the logistic function imposes an upper limit on this increase.⁶

While phonetic biases apply to specific target productions, their effects can also be viewed as a transformation of the distribution representing a given cate-

6. This upper limit is 0.005 in the present case, and is given by $\frac{1}{2}s_i$.

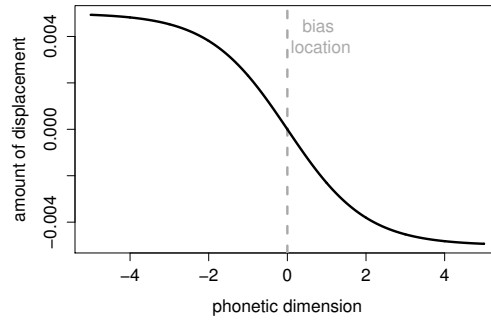


Figure 4.4: *The size and the direction of the displacement caused by a phonetic bias as a function of position along a given phonetic dimension. Parameter settings: $b_i = 0$; $s_i = 0.01$; $d = 1$.*

gory. Consider the sample of productions that form the basis of the probability density estimation, and the sample of surface values after the application of the bias. Since the bias consistently changes the output tokens, the observed sample will be shifted compared to the original sample. Given the equation in (4.7), we can calculate the probability distribution of the output values after the application of the bias as follows (where f_i is the probability density function for category i and $\text{bias}_j(\bullet)$ is the logistic function representing bias j):

$$p(\mathbf{x}|c_i, \text{bias}_j(\bullet)) = f_i(2\mathbf{x} - \text{bias}_j(\mathbf{x})) \quad (4.8)$$

As an illustration, consider Figure 4.5, where the black continuous line shows the original distribution and the black dashed line the transformed one ($b_i = 0.8$; $s_i = 0.5$, which counts as unusually high and is only used for expository reasons here). The simulations in this chapter do not rely on these transformations, but they will be prove useful in the next two chapters.

Let us now turn to the notion of bias proportion. This concept will be discussed in substantial detail in Chapter 6, so I only provide a brief outline here. Most of the phonetic biases discussed in this thesis can be viewed as contextual effects: they only apply to tokens of a given category in the appropriate environment. Thus, for instance, the phonetic bias responsible for [u]-fronting only affects vowels in coronal contexts (at least according to Harrington et al. 2011). Some authors (including Harrington et al. 2008, 2011) suggest that

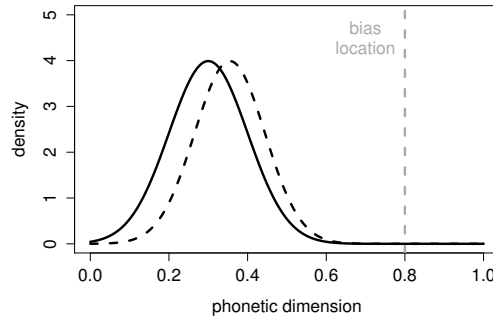


Figure 4.5: *The distribution representing a given category (black solid line) and the expected distribution of surface tokens (black dashed line) after the application of a phonetic bias (grey dashed line). Parameter settings: $b_i = 0.8$; $s_i = 0.5$.*

such contextual effects can cause category-wide shifts, and relate this to the high lexical frequency of the biasing environment in categories that undergo the change. For instance, coronal consonants are particularly frequent among the contexts in which [u] can occur in English, which might explain why it has undergone fronting in so many different dialects (Harrington et al. 2008). Other languages might show different distributions, and in some of them [u] will likely occur less frequently in coronal contexts. This type of lexical frequency is what I will refer to as bias proportion in this thesis. While Chapter 6 looks at the influence of bias proportion in detail, the simulations in this chapter and the following one take an admittedly crude approach to this variable. I will simply assume that the influence of a phonetic bias is negligible in categories with a low bias proportion, and model this by exempting them from the application of the bias. On the other hand, categories with a high bias proportion will not be exempted, and phonetic biases will affect them in the way described above.

Although this thesis focuses on weak biases like coarticulation, it is necessary to include a certain class of strong biases as well: the boundaries of phonetic space (cf. Section 3.4). It would be unreasonable to assume that phonetic dimensions have no upper and lower bounds: the articulatory and perceptual systems are subject to physical limitations which determine the range of possible speech sounds. For example, there is an upper limit on tongue height in vocalic sounds determined by the position of the palate, and a lower limit determined by a number of factors including the maximal degree of opening of the jaw. These

limits are present in the simulations as well. Specifically, production targets that are outside the range of possible values for a given phonetic dimension are simply discarded. For simplicity's sake, the limits of the phonetic dimensions will be $[0, 1]$ in all the simulations in this thesis.

Let us summarise the implementation of phonetic biases presented in this section.

(4.9) PHONETIC BIASES:

Phonetic biases are implemented as point-like attractors that displace target productions towards themselves. This is modelled by using a logistic function (equation (4.7)).

(4.10) BIAS PROPORTION:

The relative frequency of tokens belonging to a given category in contexts embodying a specific bias. In Chapters 4 and 5, bias proportion is implemented by exempting categories with a low bias proportion from the effects of the bias. Chapter 6 presents a more nuanced implementation.

(4.11) THE BOUNDARIES OF PHONETIC SPACE:

Phonetic dimensions are bounded. Productions that fall outside the boundaries of any dimension are simply discarded.

4.1.3 Category update and ambiguity

This section discusses the last two steps in the diagram in Figure 4.1: the ambiguity filter and feedback (i.e. category update). In the discussion below, I reverse the order of these two steps for expository reasons. Since both of these notions have been argued for at length in the previous chapter, the discussion below focuses on technical aspects of their implementation.

The mechanism for category update is slightly different for prototype and exemplar models, owing to the different probability distributions used by these models. Let us first look at prototype models. As it has already been noted in Section 3.3, category update consists in changing the parameters that define the category representations. The implementation of prototype models in this thesis is based on normal distributions, which means that the parameters that

change during update are the mean and the standard deviation. Figure 3.7 in the previous chapter illustrates how these modifications can be visualised. Crucially, the mean will shift towards the perceived stimulus, while the standard deviation will either decrease or increase depending on how far the stimulus is from the mean. The formulae below (derived from maximum likelihood estimators for the parameters of normal distributions) capture both of these dynamics. I present the univariate case first (n refers to the category before update and $n + 1$ after):

$$\mu_{n+1} = \frac{c \mu_n + x}{c + 1} \quad (4.12)$$

$$\sigma_{n+1}^2 = \frac{c [(\mu_{n+1} - \mu_n)^2 + \sigma_n^2] + (x - \mu_n)^2}{c + 1} \quad (4.13)$$

Again, the multivariate case is somewhat more complicated:

$$\boldsymbol{\mu}_{n+1} = \frac{c \boldsymbol{\mu}_n + \boldsymbol{x}}{c + 1} \quad (4.14)$$

$$\boldsymbol{\Sigma}_{n+1}^2 = \frac{c [(\boldsymbol{\mu}_{n+1} - \boldsymbol{\mu}_n)(\boldsymbol{\mu}_{n+1} - \boldsymbol{\mu}_n)^T + \boldsymbol{\Sigma}_n^2] + (\boldsymbol{x} - \boldsymbol{\mu}_n)(\boldsymbol{x} - \boldsymbol{\mu}_n)^T}{c + 1} \quad (4.15)$$

Most of the terms in these equations have already been introduced in Section 4.1.1. There are two main differences. First, in this case the terms x and \boldsymbol{x} stand for the incoming stimuli represented as real numbers and vectors of real numbers, respectively. More importantly, a new constant is introduced: c . This will be referred to as the constant of update, and it is inversely proportional to the amount of influence that a single stimulus can have on the parameters of the category representation. If c is low, the category representation becomes very sensitive to incoming stimuli, and undergoes quick changes if these stimuli deviate from the speaker's own representations. If c is high, the category representation becomes more resistant to incoming stimuli, which makes potential changes slower.

The update mechanism for exemplar models is somewhat simpler, as it only requires the addition of a new exemplar to the distribution. Since exemplar models are non-parametric, the equation below refers directly to a change in

the probability distribution representing category i (instead of changes to a limited number of parameters, as in the previous case):

$$p(x|c_i, n+1) = \frac{c p(x|c_i, n) + K_h(x - x_i)}{c + 1} \quad (4.16)$$

The parameter c plays the same role as in prototype models: it determines the relative importance of new stimuli (represented by x_i).

Importantly, the distribution resulting from several iterations of the production-perception feedback loop using equation (4.16) can still be described by using the weighted kernel density estimate in equation (4.3). In fact, the weights will be predictable if we know (i) when a given token was produced and (ii) how many tokens have been produced on the whole. Older tokens will have a lower weight associated with them and newer tokens a higher weight. This is analogous to the notion of memory activation in Pierrehumbert (2001), which starts at a given value for each exemplar when it is added to the category representation, and steadily declines over time according to an exponential decay function.

The implementation of the filtering listener hypothesis used in this thesis is taken from Wedel (2006). The specific filtering mechanism in the models is called ‘variant pruning’: a certain percentage of misperceived tokens is excluded from the production-perception loop by not being fed back into the original category representations. In order to decide whether a given production is misperceived or not, I will use a simple stochastic mechanism based on categorisation probabilities. Recall from Section 3.1 that categorisation probabilities can be calculated using Bayes’ theorem. Since this thesis does not investigate the role of the absolute frequencies of different categories (e.g. observations like ‘[i:] is a more frequent sound category than [ɔɪ] in English’), we can use equation (3.2), which assumes that the prior probabilities of the categories are equal. This equation is repeated below for convenience:

$$p(c_i|x) = \frac{p(x|c_i)}{\sum_{j=1}^n p(x|c_j)} \quad (4.17)$$

The probability of misperception can be calculated simply by subtracting the categorisation probability for the intended category from the sum of all probabilities (i.e. 1):

$$p(\neg c_i|x) = 1 - \frac{p(x|c_i)}{\sum_{j=1}^n p(x|c_j)}, \quad (4.18)$$

where $p(\neg c_i|x)$ simply refers to the probability that a target production will *not* be categorised as c_i .

To give an example, let us assume that a language has a three-member [i]–[u]–[a] vowel inventory. If a production of [i] (x) is slightly closer to [u] than most typical productions, the categorisation probabilities could be as follows: $p(c_{[i]}|x) = 0.7$, $p(c_{[u]}|x) = 0.25$ and $p(c_{[a]}|x) = 0.05$. Since misperception in this situation means that the token is not perceived as [i], the probability of misperception is $p(\neg c_{[i]}|x) = 1 - 0.7 = 0.3$. The simulations in this thesis rely on the probabilistic response rule described in (3.4) in Section 3.1: categorisation probabilities are used in a stochastic way, yielding misperception 30% of the time when a token like x is produced.

It might be somewhat unrealistic to assume that misperceived tokens of a given category never contribute to category representations. Indeed, it is quite likely that various non-phonetic cues (e.g. pragmatic, semantic, syntactic, etc.) can help to identify the intended lexical item even if one or more of the sound categories it contains are misperceived. In such situations, the identity of the misperceived category might also be restored, and category update will proceed as usual. Therefore, following Wedel (2006) a further parameter is added to the model, which determines the extent to which lexical feedback can correct erroneous category decisions: r , or the rate of ‘hopeless’ misperception (for the sake of brevity, I will refer to r as misperception rate). For instance, if r is set to 0.5, a token with a misperception probability of 0.3 will be excluded from category update with a probability of $0.3 \times 0.5 = 0.15$. This parameter is set relatively low in most of the simulations in order to get a conservative estimate of the effects of misperception.⁷

7. Another possible scenario involves the misperceived token being fed back into the wrong category representation, a process termed variant trading in Blevins & Wedel (2009). As I have already noted in Section 3.5, this mechanism will not be investigated here. It should be noted, however, that even if it was included in the simulations, it would only enhance the contrast-preserving effect of misperception, and not hinder it (cf. Blevins & Wedel 2009).

Here is a brief summary of the points discussed above:

(4.19) **CATEGORY UPDATE:**

In prototype models, category update corresponds to a small change in the category mean and standard deviation (equations (4.12), (4.13), (4.14), (4.15)). In exemplar models, category update consists in adding a new kernel to the probability distribution representing the category (equation (4.16)).

(4.20) **MISPERCEPTION:**

Category update is inhibited when an ambiguous token is misperceived and the category decision cannot be salvaged through lexical feedback. This occurs with a probability $r \times p(\neg c_i | x)$ (equation (4.18)).

4.1.4 *Summary of simulation architecture*

In the preceding sections, the following general simulation architecture was proposed. Production proceeds by sampling exemplar or prototype-based probability distributions (Section 4.1.1). The resulting production targets are displaced under the influence of phonetic biases modelled as point-like attractors (Section 4.1.2). The outcomes are then fed back into the original category representations, unless they are misperceived, in which case there is a small chance that they will be discarded (Section 4.1.3). These processes are iterated over and over again in the same order to simulate the production-perception loop described in Pierrehumbert (2001). We are now in a position to look at some concrete simulation results. The following sections will investigate the evolution of categories in models relying on the architecture described here.

4.2 **COMPARING PROTOTYPE AND EXEMPLAR MODELS**

The motivations for a systematic comparison between exemplar and prototype models have partly been explained in the introduction to this chapter. The two arguments given there can be summarised as follows. First, if prototype and exemplar models can be shown to behave identically, we can conclude that the results in this thesis follow directly from the idea of probabilistic

representations, and not the particular way they are implemented. Second, this comparison can help to bring the main arguments of usage-based modelling closer to theorists who might be skeptical about the validity of exemplar-based representations. There is also a third and more practical reason for attempting to show that exemplar and prototype models are equivalent on an abstract level. Exemplar models place a huge burden on computational systems by requiring all exemplars to be used during production and perception. While this might be less of an issue for a system like the brain, which relies on parallel processing, it imposes serious restrictions on computer implementations of exemplar models. Prototype models can be implemented much more efficiently, and are therefore preferable in situations where their parametric nature does not affect the outcome of the simulations.

The comparison between the models will be conducted by simulating two ‘benchmark’ phenomena: (i) large shifts in the position of a category under the influence of a weak bias (Section 4.2.1) and (ii) ambiguity-driven dispersion (Section 4.2.2). I will show that exemplar and prototype models produce nearly identical results with respect to these two phenomena. Note that both of these effects have been discussed in the literature: the phenomenon in (i) has been described by Pierrehumbert (2001), and (ii) by Wedel (2006) and Blevins & Wedel (2009). Since the main argument of this thesis builds on some of the results presented in these papers, the replication of these simulations is particularly useful in the present context.

4.2.1 Convergence towards the bias

In this section, I look at the effects of a single bias on a single category in a one-dimensional phonetic space. To give a concrete example, this situation is not unlike having a single laryngeal category with lead VOT (i.e. [+voice]), which is affected by a weak pressure towards a more neutral production with short-lag VOT (i.e. [−voice]). The simulations implement an abstract version of the production-perception feedback loop, in which there is only a single ‘soliloquising’ agent. This means that all productions are fed back into the same category representation from which they originate. While this is certainly not a realistic assumption, the next section will show that multi-agent simulations

produce exactly the same results. It should also be noted that no misperception can take place, given that there is only a single category.

In what follows, I provide a brief overview of the parameter settings used in the simulations. Each simulation starts with a single category initialised with a mean of $\mu = 0.3$ and a standard deviation of $\sigma = 0.1$.⁸ All productions of this category are affected by a bias located at $b = 0.7$ with strength $s \in \{0.05, 0.1, 0.15, 0.2\}$ (multiple values are given since this parameter is varied across the simulations). The constant of update is set to $c = 2000$, which means that individual stimuli have a very small effect on the overall category representation. The parameter that determines the amount of random noise in production is set to $h = 0.013$. Each simulation consists of 500,000 iterations of the production-perception loop.

100 simulations were run at each of the four different values of s for each model (i.e. exemplar vs. prototype), yielding 800 simulations on the whole. Such a large number of runs was necessary in order to ensure that the results reflect the general behaviour of the models rather than the potentially misleading outcomes of individual simulations. During each simulation, the mean and the standard deviation of the category was recorded after every 2000 iterations. Most of the results presented below are based on the trajectories for the mean obtained by looking at changes in the recorded values as a function of time.

To give the reader a better idea of what the simulations look like, I will briefly describe two example runs: one based on the exemplar model and the other one on the prototype model ($s = 0.1$ in both cases). Figure 4.6 shows the two category representations at a number of different points in their evolution. Several important observations can be made on the basis of this diagram. First of all, it is clear that the categories are converging towards the bias attractor in both simulations, and essentially reach it after about 100,000 iterations. Second, the standard deviations seem to be decreasing

8. Both exemplar and prototype-based simulations are initialised with a normal distribution. This is done to ensure comparability across the simulations. Although the assumption of an initial normal distribution for the exemplar-based simulation might be seen as problematic (since such a distribution is unlikely to emerge naturally in exemplar models), the effect of the original distribution becomes negligible after a few hundred iterations.

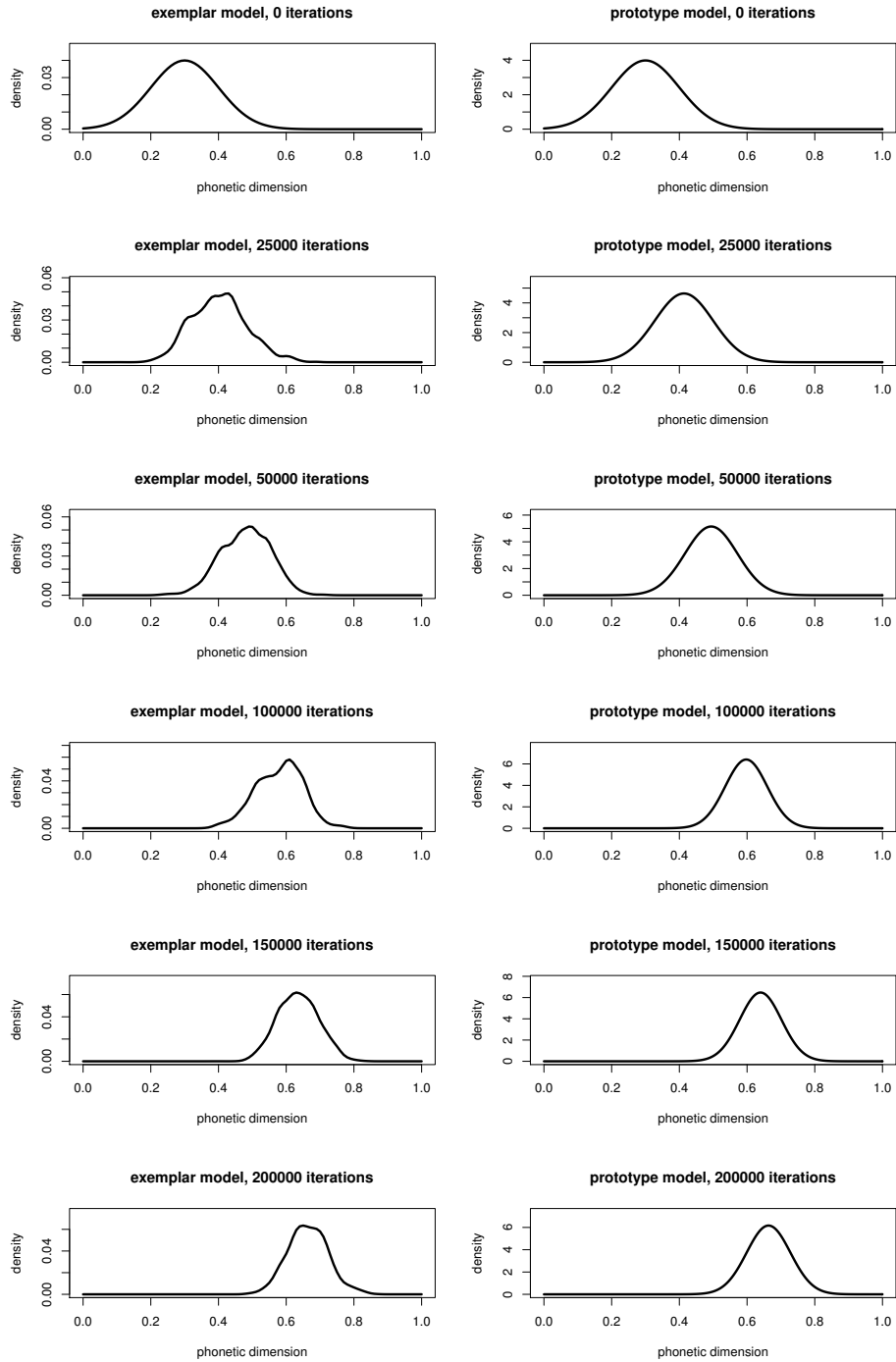


Figure 4.6: Snapshots of the evolution of a category distribution under the influence of a bias in an exemplar model (left) and a prototype model (right).

in both cases.⁹ Third, the category distributions do not change much after a certain point: while there are very clear changes in the distributions between 0–100,000 iterations, they seem to remain more or less the same after about 100,000 iterations. Thus, the bias appears to define a stable state for the category, and the category can be seen as evolving towards this state. The idea of stable states will be described in much more detail in the next chapter.

Although all of these observations are important, the crucial thing to note in Figure 4.6 is the close similarity between the exemplar and the prototype models. The categories seem to be evolving in nearly the same way, with the means moving towards the bias attractor at the same pace, and the standard deviations shrinking gradually. Of course, the exemplar model produces a somewhat irregular distribution, whereas the prototype model stays normally distributed throughout the whole simulation, but this does not seem to affect the general dynamics of the models.

This demonstration suggests a clear parallel between exemplar and prototype models, but it cannot be regarded as a systematic comparison, given that only two examples are examined. Figure 4.7 provides a more comprehensive view of the simulation results by plotting changes in the category means for four exemplar-based and four prototype-based simulations, each of them with different bias strengths. Unsurprisingly, stronger biases lead to faster convergence both in exemplar and prototype models. Once again, the two types of models perform very similarly. However, this claim is still based on isolated examples, and is therefore not sufficiently general.

A more exhaustive comparison is presented in Figure 4.8, which is essentially an extended version of the previous figure. Figure 4.8 illustrates the general behaviour of each set of 100 simulations by plotting 90 per cent confidence intervals for the means at every point in time. In other words, the bands

9. This goes against accounts of sound change which claim that category variance increases during ‘new and vigorous’ changes (see e.g. Labov 1994: p. 457). While this might seem problematic, the next chapter will argue explicitly against the type of overly simplistic simulations presented here, which means that there is no reason to expect realistic predictions from our models at this point. It should also be noted that the lack of increase in variance is likely a result of the absence of independent representations for different phonetic environments. Chapter 6 presents simulations where different phonetic environments are represented through separate sub-distributions. These simulations show the expected increase in the variance of a category under the influence of phonetic biases.

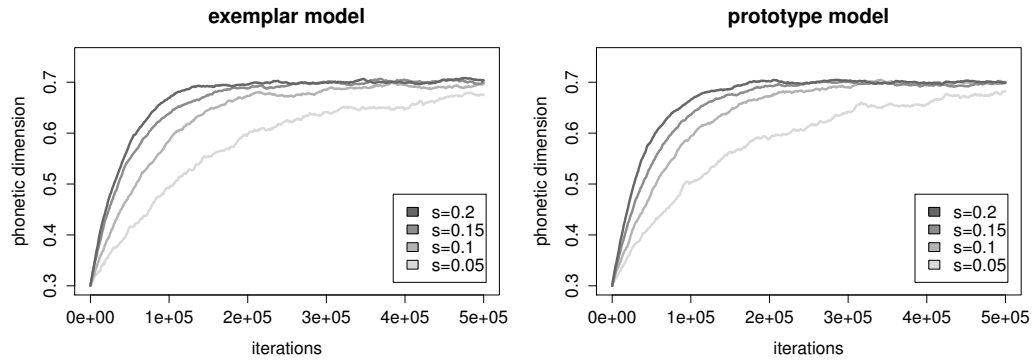


Figure 4.7: The means of the category representations in different models plotted against the number of iterations. Darker lines indicate higher values of s (i.e. bias strength).

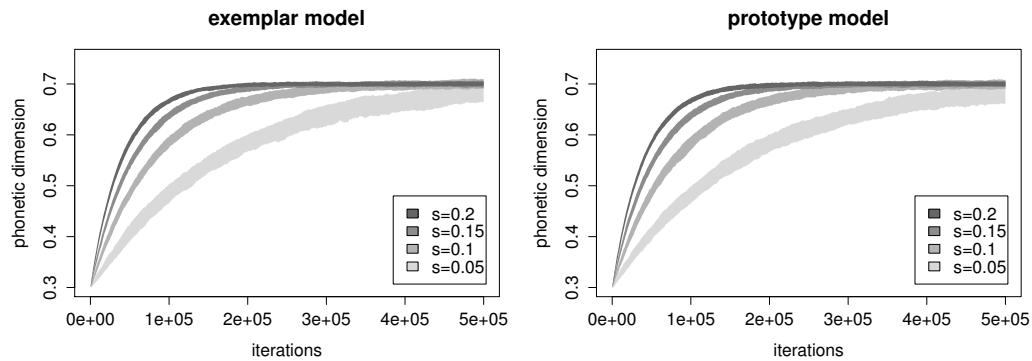


Figure 4.8: 90% confidence intervals for the mean values in simulations with different values of s plotted against the number of iterations. Darker lines indicate higher values of s .

shown in the figure represent 90 per cent of all the simulations with a given parameter setting, with the outlines corresponding to extreme values. The similarity between exemplar and prototype models still holds: in fact, it is hard to discern any difference between the two sets of simulations. Since this comparison is based on a large set of simulations rather than just individual examples, it is a strong argument in favour of the claim that the dynamics of exemplar and prototype models are essentially the same – at least with respect to the influence of phonetic biases.

While the generality of the last comparison is clear, it is still based on eyeballing the bands delineated by the confidence intervals, which cannot be

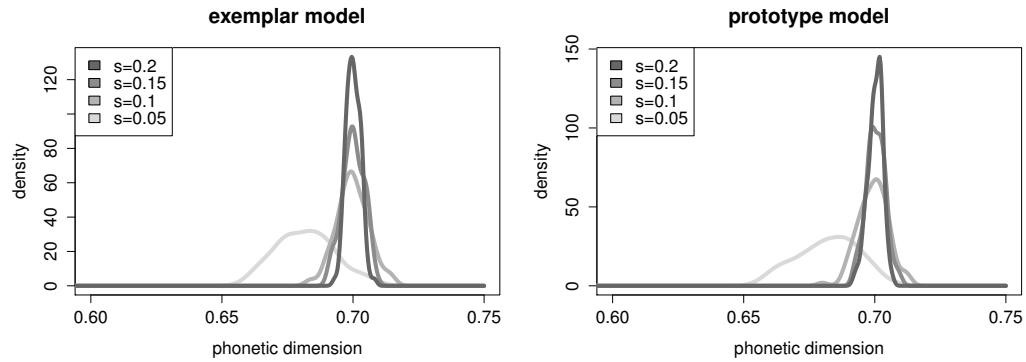


Figure 4.9: The distributions of the category means after 500,000 iterations at different values of s . Darker lines indicate higher values of s .

regarded as a statistically reliable test. In order to be able to use standard statistical tests on the data, I will focus on the simulated category representations after 500,000 iterations. Specifically, I will compare the distributions of category means across exemplar and prototype models at each value of bias strength. The distributions are illustrated in Figure 4.9 (note that these are not the category representations themselves, but summary representations of all the category means in each set). As in the previous graphs, the similarity between the performance of the exemplar model and that of the prototype model is striking. To confirm the identity of the distributions across the two models, simple unpaired t -tests are performed. Since four different comparisons are made (one for each value of s), it is necessary to control for multiple comparisons. For simplicity's sake, I use Bonferroni's correction. As expected, none of the comparisons turned out significant at $\alpha = 0.05$ (note that this holds even without Bonferroni's correction). This constitutes statistical evidence for the claim that exemplar and prototype models perform identically under the influence of phonetic biases.

Bias strength has an important effect on the distribution of possible mean values. In simulations with higher bias strength, the range of possible outcomes after 500,000 iterations becomes strongly limited, while a lot more variation is seen at lower bias strength values. This is easy to see in Figure 4.9, where the distribution of category means is much more disperse in simulations with low bias strength. The implication of this finding is that bias strength affects the extent to which the simulated change is deterministic: although categories will always tend towards the bias attractor regardless of the strength of the

bias, they can occupy a much wider range of positions in phonetic space when the bias is weak.

To sum up, exemplar and prototype models exhibit near-identical dynamics with respect to the evolution of sound categories affected by phonetic biases. The two most important patterns observed in the simulations were as follows: (i) gradual convergence towards the bias attractor and (ii) the existence of stable states. Both models show these patterns and they realise them in strikingly similar ways. In some sense, this result is not surprising: the simulation architectures for the two models were purposely designed in a way that ensures maximal comparability. However, the models were not artificially modified to produce identical results. They are both valid and straightforward implementations of existing models of production and perception, with clear precedents for each of them (e.g. Pierrehumbert 2001 for the exemplar model and Feldman et al. 2009 for the prototype model). Having seen the behaviour of exemplar and prototype models under the influence of a phonetic bias, we can now turn to the phenomenon of contrast maintenance driven by ambiguity.

4.2.2 *Ambiguity-driven dispersion*

The simulations in this section demonstrate the phenomenon of ambiguity-driven dispersion in exemplar and prototype models. The simulated systems consist of two sound categories, which are initialised relatively close to each other, so that the effects of dispersion can be seen clearly. By way of illustration, consider a language with two overlapping laryngeal categories, one of them with a VOT of 10 ms and the other one 20 ms (both of these values can be considered short-lag VOT). Since this contrast is likely to lead to frequent misperceptions, it is predicted to undergo enhancement in the present account. As in the previous section, the focus of this investigation is on the evolution of the categories over time. Note that no biases were included in the simulations.

The drive behind the dispersion that takes place in the simulated sound systems is the frequent failure to correctly perceive ambiguous tokens. However, these simulations are rather unrealistic in the way they implement ambiguity-based misperception, given that both the speaker and the listener are the same. It would be very surprising to see someone misperceive one of their

own utterances, and yet this is exactly what happens below. Although this simulation architecture is not without precedent (see e.g. Wedel 2006), this unnaturalness has to be addressed in some way. First, I would point out that the object of enquiry in these simulations is not a single speaker or listener, but the production-perception feedback loop itself. Consequently, the fact that these misperceptions would seem counterintuitive in the context of real speech interactions does not necessarily imply that they should not be allowed in the more abstract simulations presented here. This, of course, leads to the question of whether this abstraction is sufficiently motivated. From a pragmatic point of view, it certainly is: simulations with only a single agent are computationally much more tractable than simulations with multiple agents. But this in itself does not justify using the former type of simulation in the place of the latter. The next section provides systematic evidence showing that the dynamics of abstract single agent models are the same as those of multi-agent models, which suggests that the production-perception feedback loop is a plausible abstraction.

The parameters of the simulations in this section are as follows. The two categories are initialised at $\mu_1 = 0.45$ and $\mu_2 = 0.55$, both of them with $\sigma_1 = \sigma_2 = 0.1$, which yields a substantial amount of overlap between them. The misperception rate is varied in three steps between 0.05 and 0.2: $r \in \{0.05, 0.1, 0.2\}$. The rest of the parameters are the same as in the simulations in the previous section, the only difference being that there are no phonetic biases in this case. Thus, $c = 2000$ and $h = 0.013$. A further small difference is that these simulations are only run for 250,000 iterations, as this is sufficient to illustrate all relevant aspects of their behaviour. Note, however, that each of these iterations consists of the production and the perception of *both* categories, which means that the total number of production-perception events is actually $2 \times 250,000$. Similarly to the previous section, 100 simulations were run at each value of r for each model, yielding 600 simulations overall. The mean and the standard deviation were recorded after every 1000 iterations.

A summary of two representative simulations is presented in the two series of graphs in Figure 4.10. The developments illustrated in this figure are in many ways similar to those described in the previous section. Both simulations follow the same general dynamic, which in this case is a tendency towards a higher degree of separation in phonetic space. However, in this case the match

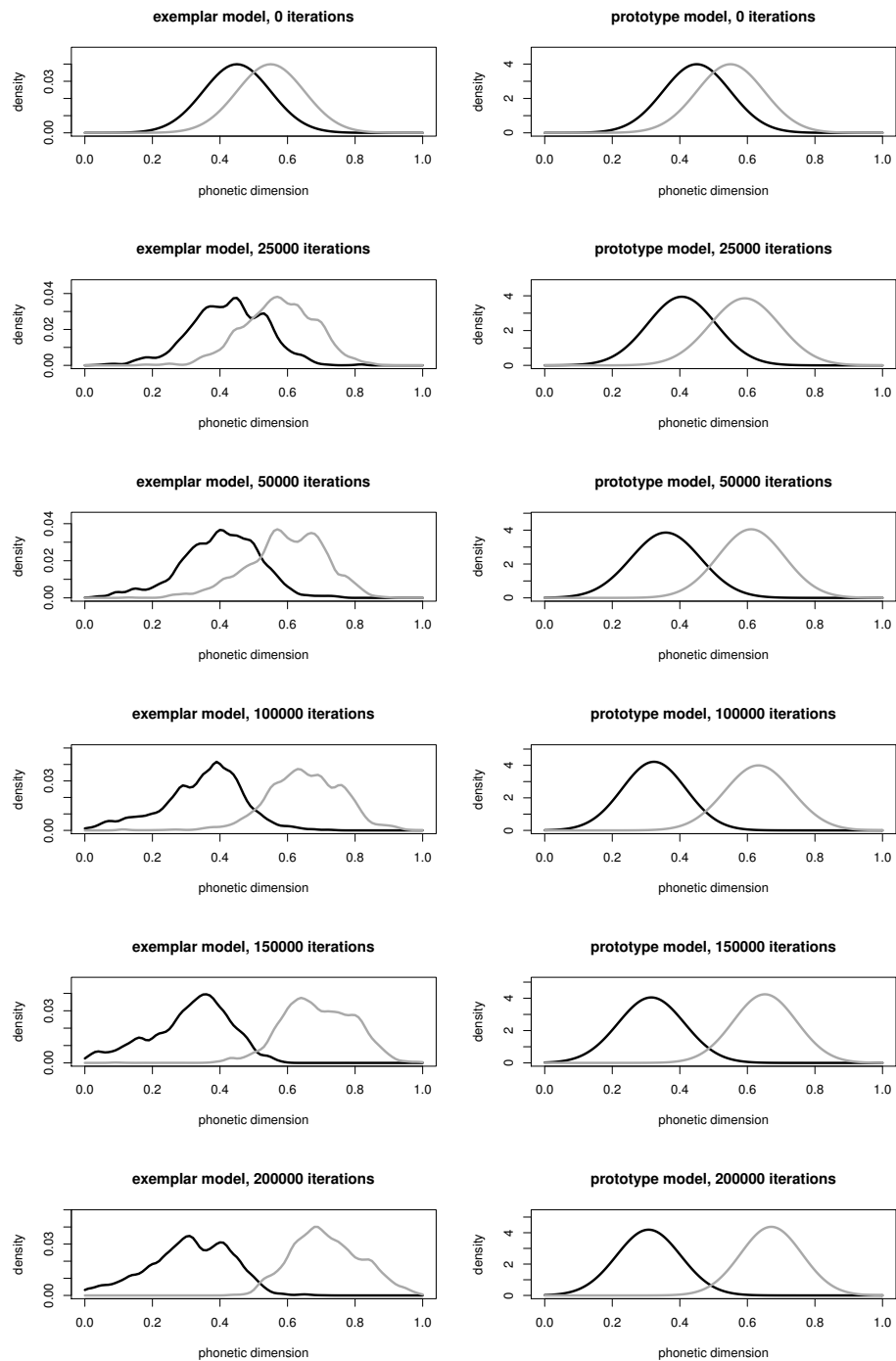


Figure 4.10: Snapshots of the dispersion of two category distributions in the exemplar model (left) and the prototype model (right).

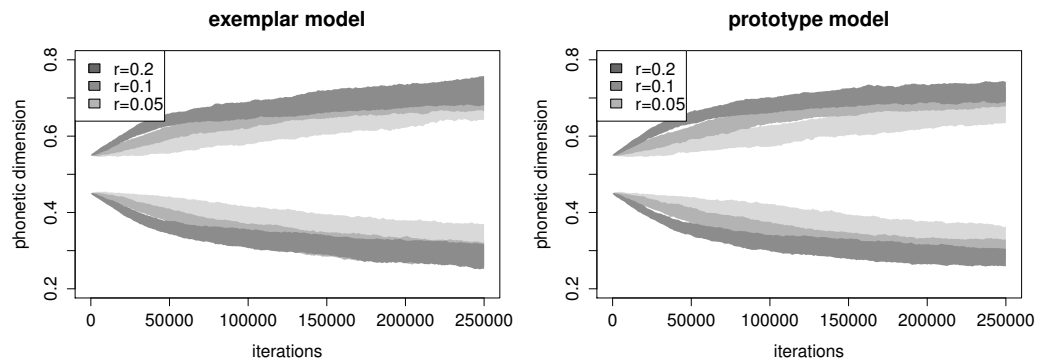


Figure 4.11: 90% confidence intervals for the mean values in simulations with different values of r plotted against the number of iterations. Darker lines indicate higher values of r .

between the realisation of this tendency in the two models is not as close as in the previous section. For instance, the exemplar model seems to tolerate more overlap between the categories in the first 50,000 iterations than the prototype model. Moreover, the shapes of the final distributions are somewhat different in a way that goes beyond the small discrepancies observed in the previous section (where the output of the exemplar model was essentially a noisy normal distribution). In the simulations in Figure 4.10, the exemplar-based distributions develop long tails on the sides opposite the other category, and their modes seem to be closer together than in the prototype model. This being said, the evolution of the two categories still appears sufficiently similar to tentatively conclude that their overall behaviour is the same.

Figure 4.11 plots the general pattern shown by the larger sets of simulations through 90 per cent confidence intervals calculated for the means. The results for the two models appear nearly the same, although this is difficult to establish with a sufficient degree of certainty given the large amount of overlap among the bands representing different values of r . A clearer demonstration is provided in Figure 4.12, which shows the distributions of category means in each of the two models at different values of r . The two peaks in these diagrams correspond to the two different categories. While the similarities are obvious, there are clear differences as well. This is confirmed by unpaired t -tests comparing the means of corresponding categories across the two models at each value of r : significant differences are found between the distributions when $r = 0.05$

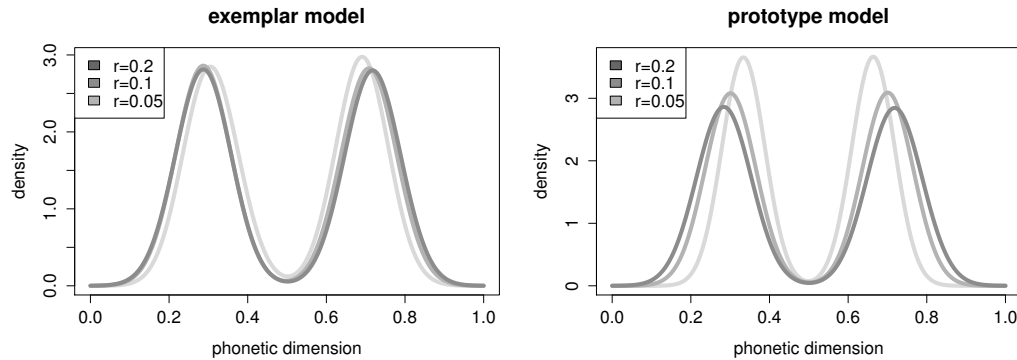


Figure 4.12: *The distributions of the category means after 250,000 iterations at different values of r . Darker lines indicate higher values of r .*

and when $r = 0.1$ (but the distributions seem not to differ significantly at $r = 0.2$; $\alpha = 0.05$ in all the tests). It is not clear what the implications of these differences are. One possible interpretation might be that the categories are all evolving towards the same stable states, but the speed of this evolution is slightly different in the two models. If we assume that the categories have not yet reached their final state, we might observe small differences across the two models as a result of the different speeds at which they evolve. In any case, these differences only concern minor details of the emergence of disperse sound systems, and do not constitute a strong argument against the claim that the general behaviour of the models is the same.

It might also be instructive to look at changes in the variance of the categories. Figure 4.13 shows these changes by plotting confidence intervals against time at different values of r . The difference between exemplar and prototype models is much more obvious here: the prototype model appears to tolerate a smaller range of possible standard deviations, and these are considerably lower than those for the exemplar model. This is confirmed by separate unpaired two-tailed t -tests comparing corresponding distributions from the two models at each value of r , which are all significant at a level of $p < 0.05$ even after Bonferroni's correction. These differences should not come as a surprise. While prototype models can only increase the separation between the two categories by moving the means apart or decreasing the standard deviation, exemplar models have a wider range of strategies at their disposal. One such strategy is to skew the category distributions in a way that fewer tokens fall in the area between the two categories, as in Figure 4.10.

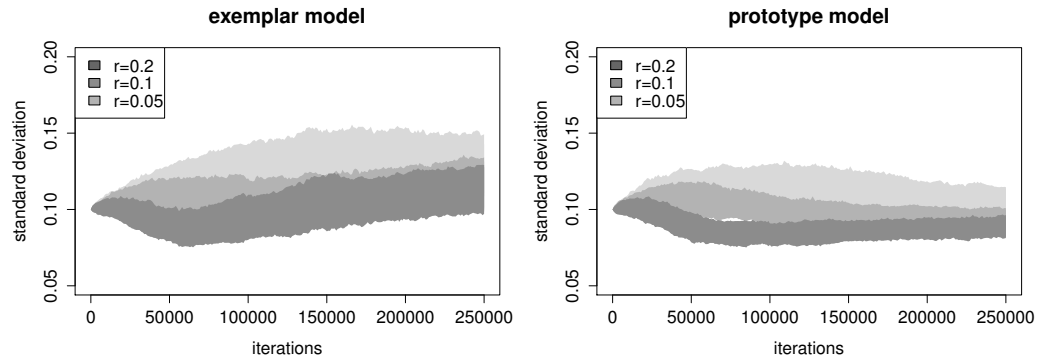


Figure 4.13: 90% confidence intervals for the standard deviations in simulations with different values of r plotted against the number of iterations. Darker lines indicate higher values of r .

Another such strategy is illustrated in Figure 4.14, which shows two exemplar-based category distributions in a simulation where the categories are initialised in exactly the same position ($\mu_1 = \mu_2 = 0.5$, $\sigma_1 = \sigma_2 = 0.1$; $r = 0.2$). In this case, one of the categories (grey line) falls in the middle of a discontinuous category representation (black line). These representations guarantee unambiguous productions, but they look extremely unnatural. Indeed, as far as I am aware no such categories have ever been reported in natural languages. Note, however, that this type of situation only emerges if the amount of initial overlap between the two categories is extremely high (i.e. if the categories have undergone a near-merger). Moreover, they represent transitional states in the simulations, in the sense that they always disappear after a small number of iterations, giving place to more stable distributions like the ones in Figure 4.10. Thus, while such situations are not ruled out by exemplar models, the probability that they would be observed in a natural situation is diminishingly small (due to their transient nature and the infrequency of near-mergers).

In summary, the general dynamics of dispersion in exemplar and prototype models appear very similar, but the parallels between the models are not as close as they were in the case of phonetic biases. While both types of systems are evolving towards well-separated category representations, the realisation of this tendency is slightly different across the models. In particular, the flexibility of exemplar-based representations puts less pressure on the standard deviations and the means of the categories by allowing the emergence of skewed and irregular distributions. Although it is important to note the existence of these

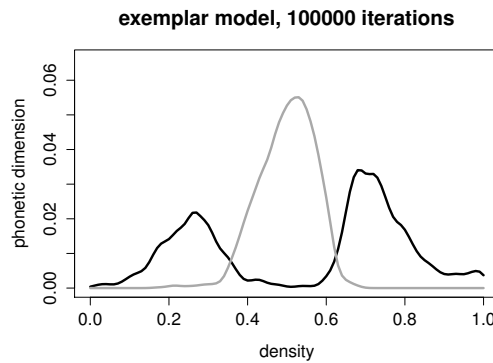


Figure 4.14: *Two exemplar-based category representations illustrating an unusual contrast maintenance strategy that can emerge in exemplar models (note that these are individual category representations and not summary representations of multiple means).*

differences, they do not affect the general argument of this section, namely that the overall behaviour of exemplar and prototype models is similar enough to use them interchangeably in simulations.

4.2.3 Summary

The findings reported in Sections 4.2.1 and 4.2.2 have two important consequences, one of them specific to this thesis and the other one more general. Let us start with the more specific one. Exemplar and prototype models have been found to produce exactly the same behaviour under the influence of phonetic biases (Section 4.2.1), and they exhibited only minor differences in the way they implemented ambiguity-driven dispersion. The fact that the two models lead to the emergence of the same phenomena is a strong indication that these dynamics follow directly from the theoretical assumptions described in the previous chapter, and not the way they are implemented. Thus, we have successfully linked the behaviour of the different models to deeper principles underlying both of them, showing that the choice between exemplar and prototype-based representations is a modelling question, not a theoretical one. This has an important practical implication for this thesis. Since any general results produced by prototype models are likely to be replicated by exemplar models as well, there

is no reason to test both types of representations in each simulation. Therefore, only prototype-based simulations will be used in the rest of this thesis.

The more general consequence of these findings is that arguments directed against exemplar-based representations do not necessarily affect the main results of simulations like the ones presented in Pierrehumbert (2001), Wedel (2006) and Blevins & Wedel (2009). In fact, rather conservative models of phonetic realisation based on learnt phonetic targets with some amount of variation (e.g. Keating 1990b) will likely produce the same results, as long as the well-supported assumptions presented in the previous chapter are incorporated into them. This means that the phenomena described in this section are quite possibly relevant to a much larger range of theoretical approaches than traditionally assumed. They should be given serious consideration even by theorists committed to a more abstract approach to sound change and phonology, especially when the objects of enquiry are phonologisation or dispersion (see e.g. Bermúdez-Otero 2007 for phonologisation and Flemming 2004 for dispersion).

4.3 SINGLE AND MULTI-AGENT SIMULATIONS

The reasons for conducting a systematic comparison of single and multi-agent simulations have been detailed in the previous sections, so only a brief overview is given here. While simulating the production-perception feedback loop without real interactions among multiple agents is certainly a useful abstraction, it needs to be shown that this simplified setup does not crucially alter the behaviour of the model. Therefore, we need to see whether the predictions of single-agent models are also borne out in simulations with multiple agents. In the demonstration that follows, the same two benchmark phenomena will be used as in the previous section: the convergence towards phonetic bias attractors and ambiguity-driven dispersion.

Before discussing the simulation results, it will be useful to overview the general structure of multi-agent simulations. In order to model interactions between agents, each iteration starts by randomly selecting two different agents from the population. One of these agents acts as the speaker and the other one as the listener. The modelling of production is the same as in the previous

simulations: the speaker samples their category representations. These category representations are subsequently displaced by phonetic biases (if there are any). At this point, the focus moves to the listener. Whether the token is misperceived is decided on the basis of the listener's own category representations. If it is correctly identified, the listener's category representations are updated accordingly, but update might not take place if the identification fails (in the same way as in the single-agent model).

Note that these simulations do not incorporate any social factors: the agents produce and perceive tokens in exactly the same way regardless of their conversational partners. This is, of course, a simplification. Real speech interactions depend to a large extent on the identities of the conversational partners (see e.g. Pardo 2006, Babel 2009). This is not a problem: the simulations below attempt to establish that single and multi-agent implementations do not differ simply as a function of the number of agents. Social factors could be added to the multi-agent simulations, and this would undoubtedly change the overall behaviour of the model, but this is orthogonal to the question investigated in this section.

The fact that during one iteration only a single agent's representations are updated means that these systems will evolve slower than single-agent simulations (at least in terms of computer time). Intuitively, if there are n agents, the simulations are likely to change at $1/n$ times the speed of single-agent simulations. This is because it takes at least n iterations for any small change to apply to all the speakers in the population. As we will see below, this intuition is correct. However, this should not be interpreted as a prediction about the speed of change within real communities as a function of population size for at least two reasons. First, real speech interactions can involve more than two speakers, while the interactions modelled here always have two participants. Second, the speakers in these simulations always wait for the previous speech interaction to end before their own interaction starts, whereas in real speech communities conversations can take place simultaneously. As a result, changes will propagate much faster in real communities than in the simulations here. This means that the speed of the changes observed in these simulations does not necessarily make any predictions about the speed of similar changes in real communities.

All the simulations in this section are based on a population with six agents. Importantly, all the agents' production and perception are controlled by the

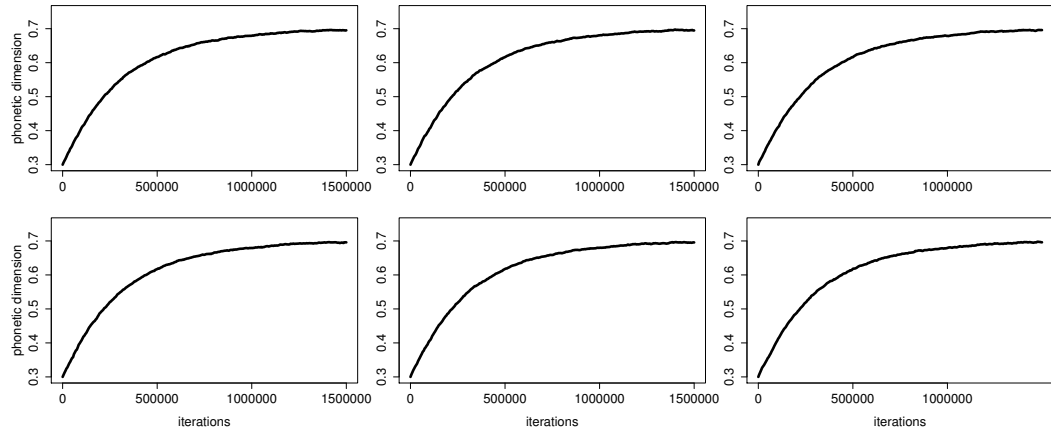


Figure 4.15: *Changes in the category means of six agents plotted against the number of iterations in a simulation with a phonetic bias.*

same set of parameters, which are mostly identical to those in Sections 4.2.1 and 4.2.2. Thus, they can only differ in their category representations (which are prototype-based). Two series of simulations are run: one investigating the effect of a single phonetic bias, and the other one looking at ambiguity-driven dispersion. The initial category representations are identical for all the speakers, with $\mu = 0.3$ in the first set of simulations and $\mu_1 = 0.45; \mu_2 = 0.55$ in the second set ($\sigma = 0.1$ for all simulations and categories). No parameters are varied, which means that only one value of bias strength ($s = 0.15$) and one value of misperception rate ($r = 0.1$) is investigated. The bias-based simulations are run for 1,500,000 iterations, while the dispersion-based simulations for 750,000 iterations. The remaining parameters need not be stated here, as they are exactly the same as in the previous sections. Each of the two simulations was repeated 50 times, yielding 100 simulations on the whole.

Figures 4.15 and 4.16 show changes in the category representations of each of the six agents in two example runs illustrating the two different types of simulations. The results look very similar to those from the simulations in Sections 4.2.1 and 4.2.2: the categories converge towards the bias in the first case, and they exhibit dispersion in the second case. Even more importantly, all the agents' representations develop in the same way, with some minute individual differences (these are more visible in Figure 4.16). Thus, it appears that multi-agent simulations produce the same dynamics as their single-agent counterparts. However, this statement needs to be investigated more systematically.

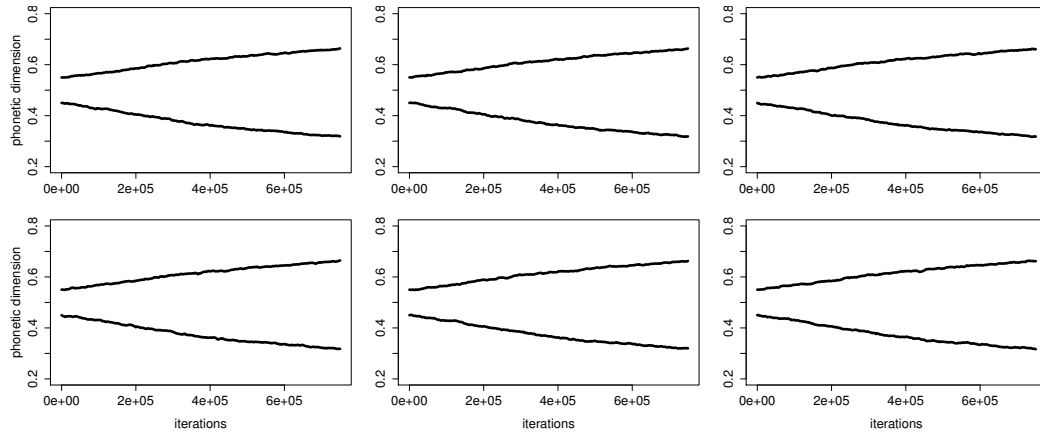


Figure 4.16: Changes in the category means of six agents plotted against the number of iterations in a simulation with ambiguity-driven dispersion.

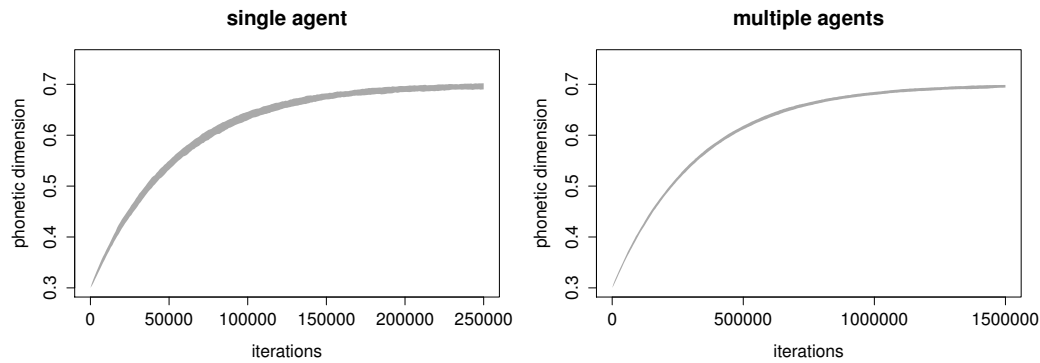


Figure 4.17: 90% confidence intervals for the mean values in single-agent and multi-agent simulations investigating the influence of a single phonetic bias ($s = 0.15$).

The pairs of diagrams in Figures 4.17 and 4.18 provide a more comprehensive comparison by plotting 90% confidence intervals for single agent and multi-agent simulations. Figure 4.17 represents bias-based simulations and Figure 4.18 dispersion-based simulations. The panels on the left are taken from the prototype-based simulations in Sections 4.2.1 and 4.2.2.¹⁰ In order to make the single and the multi-agent simulations comparable, the number of iterations shown on the left is precisely one sixth of the number of iterations on the

10. Specifically, they illustrate single bands with specific parameter settings taken from Figures 4.8 and 4.11. The parameter settings correspond to those used in the simulations in the present section: $s = 0.15$ for the bias-based simulations and $r = 0.1$ for the dispersion-based simulations.

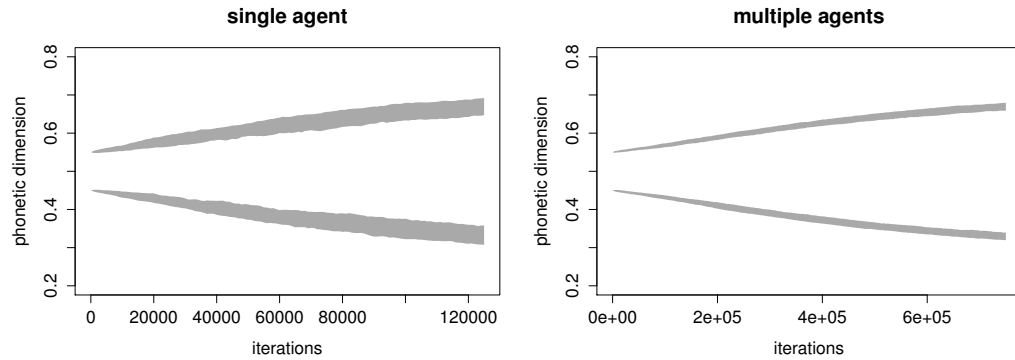


Figure 4.18: 90% confidence intervals for the mean values in single-agent and multi-agent simulations investigating ambiguity-driven dispersion ($r = 0.1$).

right ($250,000 = 1,500,000/6$ in Figure 4.17 and $125,000 = 750,000/6$ in Figure 4.18). This adjustment was necessary since – as it has been noted above – multi-agent simulations evolve n times slower than single-agent ones, where n is the number of agents. Turning to the comparisons now, the curves on the left and those on the right look remarkably similar. The only difference is in the width of the bands: single-agent simulations seem to produce more variable outcomes than multi-agent ones. The most likely explanation for this difference is that multi-agent simulations are more resistant to random fluctuations due to the fact that every small change has to diffuse through the whole population. Conversely, consistent factors like phonetic biases and misperception do not need to diffuse, since they have a chance to apply during each speech interaction regardless of the identity of the agents.

As in the previous sections, it will be useful to provide a statistical comparison based on the category means at a given point in time. Figure 4.19 illustrates the distribution of category means after 250,000 (left) and 125,000 (right) iterations in the single-agent models and after 1,500,000 (left) and 750,000 (right) iterations in the multi-agent models (once again, relying on the observation that the multi-agent model evolves six times slower than the single-agent one). The similarities between the distributions for single and multi-agent simulations are confirmed by Welch's t -test for samples with unequal variances: there are no significant differences between the means that emerge from single *versus* multi-agent simulations (i.e. $p > 0.05$). There is only one exception: the category on the left in the dispersion-based simulations shows a slight difference across

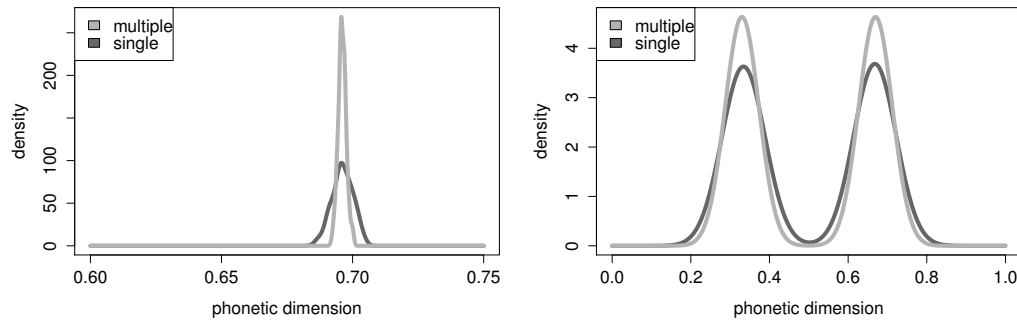


Figure 4.19: A snapshot of the distributions of the category means at a specific point in time in single (dark grey) and multi-agent simulations (light grey) looking at the influence of biases (left) and dispersion (right).

the two types of simulations, which comes out as statistically significant with $p = 0.025$. This is likely a false positive that results from random fluctuations in the simulations, given that there is no principled explanation for why only one of the categories should be different. However, this intuition could only be confirmed by re-running the simulations. I believe this is unnecessary in the present case, given the otherwise clear parallels between the simulations and the extremely small size of the difference.

In conclusion, multi-agent models perform nearly identically to single-agent models both in bias and dispersion-based simulations. The only consistent difference lies in the amount of randomness shown by the simulations: multi-agent simulations seem to produce more consistent results than single-agent ones. In other words, larger populations exhibit more deterministic behaviour in implementing simulated changes. As for our present goal, the simulation results presented above provide ample evidence for the claim that the production-perception feedback loop is a valid abstraction. Since there are no major differences in the dynamics of models with different numbers of agents, the rest of this thesis will continue to rely on single-agent simulations.

4.4 CONCLUSIONS

This chapter has given a detailed outline of the model that forms the basis of the arguments in the next two chapters. Section 4.1 showed how each of the

theoretical assumptions in Chapter 3 find a direct expression in the simulations in this thesis and presented all the relevant mathematical formalisms used in implementing these assumptions. Section 4.2 demonstrated that the choice between exemplar and prototype-based representations does not crucially alter the outcome of the simulations, and can therefore be considered a modelling assumption in the context of this thesis. This is an important result in its own right, since it suggests that the results of exemplar-based simulations generalise to other, more traditional frameworks as well. Finally, Section 4.3 compared single and multi-agent simulations in an attempt to show that simulating an abstract production-perception feedback loop instead of a more realistic community of agents is a valid theoretical simplification. While this comparison found that multi-agent simulations do behave differently in the sense that their evolution tends to be more deterministic, the overall dynamics of the simulations were the same as those of single-agent simulations. This also confirms the validity of previous work that has relied on the abstract version of the production-perception feedback loop.

While the main goal of this chapter was to explain how the theoretical assumptions in the previous chapter can be implemented, the simulations in Sections 4.2 and 4.3 have a further important role in the context of this thesis. In Chapter 2, I presented an argument to the effect that existing models based on phonetic biases seem unable to answer the question of how sound systems can remain stable. One of the pivotal points of this argument is the observation that simple bias-based approaches cannot resist the force of phonetic biases – they predict that a sound change will take place wherever there is an appropriate phonetic bias. The simulations in this chapter provide a strikingly clear confirmation of this point: even though all of the simulations are stochastic in the sense that production targets are chosen at random, the category means always seem to settle around the bias attractor. The same is true for dispersion: when two categories get close to each other, misperception will always ensure that they start drifting apart. As it has been noted in Chapter 2, this prediction is clearly wrong: only a certain subset of phonetic biases seem to serve as the basis of larger changes in any given language.

This issue is addressed in the next chapter, which argues that the problems associated with bias-based models stem not from their theoretical assumptions,

but from the way these assumptions tend to be used in analyses of concrete phenomena. Specifically, the inevitability of sound change in such models is a consequence of their narrow focus on categories in a vacuum. That is, such accounts tend to look at interactions between a single category and a single bias, without giving any consideration to other categories or biases. The next chapter will show that these issues disappear when the object of enquiry is not a change in a single sound category, but a change in a sound system.

A SYSTEMIC VIEW OF SOUND CHANGE

5

The purpose of the previous chapters was to set the scene for the arguments in this chapter. Let us briefly review the main points of the discussion so far. Chapter 2 provided an overview of a number of different approaches to sound change and situated the central research question of this thesis in the context of the debate about the actuation problem. The main conclusion of Chapter 2 can be summarised as follows. While there is a clear parallel between phonetic biases and observed sound changes, bias-based approaches seem unable to account for the stability of sound systems and therefore cannot give a satisfactory answer to the actuation problem. The question, then, is whether it is possible to formulate a theory of sound change that retains the ability to capture the parallels between phonetics and sound change, but does not run into the same problems with the regard to the actuation of sound change. Although the chapters so far have not made an attempt to tackle this issue, the main directions along which a potential resolution can be formulated have already been sketched out. I have suggested that the bias-based approach does not have to be abandoned altogether in search of a more plausible theory; it needs only to be refocused. Specifically, instead of looking at sound categories in a vacuum, we should investigate the influence of phonetic biases on sound systems.

At this point, the general argument of the thesis had to be put aside temporarily, since the system-based investigation proposed in Chapter 2 raised important methodological issues that needed to be addressed. The most worrying of these is the difficulty of establishing the exact predictions of the bias-based approach with respect to sound systems. When sound categories are investigated in a vacuum, these predictions can be mapped out through simple thought experiments with a relatively high level of confidence. However, when more complex systems

are examined, thought experiments become insufficient. Chapter 2 suggested that this problem can be resolved by using computer simulations. Crucially, the computer simulations have to be based on a clear and explicit statement of the theoretical premises of the bias-based approach. This was provided in Chapter 3. Moreover, it has to be shown that the behaviour of the simulations follows directly from the theoretical assumptions, and not additional details of implementation. This task was accomplished in the previous chapter (Chapter 4). With these methodological issues out of the way, we are now ready to return to the main argument.

This chapter presents a systematic investigation of the predictions of bias-based approaches with respect to complex sound systems. Since this requires a more expressive theoretical vocabulary, Section 5.1 will discuss how the notion of ‘adaptive landscape’ from evolutionary biology can help us appreciate the complexity of system-wide changes. Adaptive landscapes also put the fundamental theoretical questions in this thesis in a new light, which will help us see the logic behind the simulations in the following sections more clearly. The simulation results are presented in Section 5.2, where the main focus will be on how complex sound systems evolve towards stable states in an adaptive landscape. These simulation results form the basis of the discussion in the rest of the chapter. Section 5.3 interprets the results against the background of the seemingly problematic determinism of sound change in bias-based accounts. It will be shown that this problem turns out to be illusory if we shift our attention from individual categories to sound systems. Section 5.4 then develops this argument further, suggesting that the solution to the other half of the actuation problem (the underapplication problem) also follows from the idea of adaptive landscapes: if the landscape changes due to external reasons, the sound system might react with changes too. This is followed by a short discussion of how this relates to social factors in sound change, and a tentative proposal that the system-based view predicts the existence of a distinction between changes from below and above (Section 5.5).

5.1 ADAPTIVE LANDSCAPES IN SOUND CHANGE

The notion of adaptive landscape is a powerful visual metaphor used in evolutionary biology (and, more generally, in any area where the focus is on dynamic

systems) that makes it relatively easy to reason about the forces that shape evolution. It was originally proposed by Wright (1932), and has continued to shape the way we think about evolution to the present date. To understand adaptive landscapes, we first have to introduce the notion of fitness. Fitness in biology corresponds to the relative success with which a given characteristic (a genotype or a phenotype) is able to reproduce itself. A characteristic with low fitness will quickly die out in a biological population, while a characteristic with high fitness will likely survive for many generations (this, of course, is a tautology, given that fitness is defined as a function of reproductive success). Adaptive landscapes relate fitness to the parameters we can use to describe a specific characteristic (e.g. size, shape, colour, etc.). The idea is that the parameters define a map of possible characteristics, and fitness provides a topography for this map.

At this point, it will be useful to look at a simple example: aerial locomotion, or flight in the animal kingdom (this example is based partly on Dawkins 1982: p. 45). It is intuitively clear that the fitness of flight as a character should be quite high under the right circumstances. However, there are numerous ways to fly, and not all of them are equally efficient. Dawkins provides two examples: the feathered wings of birds and the skin flap wings of bats. He suggests that feathers are a high-fitness solution to the problem of flight, while skin flaps might have slightly lower fitness (probably because skin flaps are not particularly well-suited to soaring and gliding, although this is not stated explicitly). The map over which fitness is defined in this case is wing type, which is likely a complex multi-dimensional characteristic. There are numerous wing types that are not suitable for flight at all – for instance, stubby fur-covered wings are not particularly useful when it comes to flying. These have low fitness in this example, and they constitute the low-lying plains and valleys of the adaptive landscape. High-fitness solutions like feathered wings and skin flaps are the hills and mountains. Since skin flaps are less efficient than feathered wings, they constitute slightly lower peaks. Such lower peaks are called local optima, while the highest peak is referred to as the global optimum. Importantly, a given population might get stuck at a local optimum during its evolution, since the evolutionary steps it would need to take towards the global optimum would lead through valleys of low fitness. This is likely the reason why we often find species relying on non-optimal solutions to a given problem.

Adaptive landscapes have been used before in the description of language change. For instance, Lass (1997) relies heavily on this notion (in his terminology, ‘epigenetic landscapes’) in describing various types of development in natural languages. He suggests that language can be viewed as a dynamical system which moves in a multidimensional phase-space. This phase space can be characterised by using the terminology of adaptive landscapes. It contains attractors (peaks), narrow paths linking high-lying areas (ridges) and various cyclical attractors consisting of multiple states that tend to be recycled over and over again in the history of a language.¹ One of Lass’s (1997) examples for this approach is the process of grammaticalisation, whereby free-standing content words become grammatical morphemes. For instance, case markers often develop ‘along the pathway Noun > Postposition > Clitic > Case marker’ (Lass 1997: p. 295). Thus, the Hungarian relative case marker *-ból/-ből* ‘from within’ comes from an earlier postposition *belől*, which in turn is a suffixed form of the noun *bél* ‘inside’ (Lass 1997: p. 296). Importantly, the end point of this development is an attractor, or a peak in the adaptive landscape. The resulting state is stable, since grammaticalised elements hardly ever start walking backwards along the path described above to become nouns again.

This thesis uses the term adaptive landscape in a similar way. For example, peaks in the landscape correspond to stable states just as in Lass’ account. However, the use of simulations allows us substantiate this notion in a more rigorous way. Consider Figure 5.1, which shows a single distribution from Figure 4.9 in the previous chapter. The grey distribution represents the category means after 500,000 iterations in 100 different runs of the same bias-based simulation (with $b = 0.7$ and $s = 0.1$). The distribution is centred around 0.7, which is the location of the bias attractor. As we move away from this point in either direction, the distribution falls off, meaning that the category mean typically ends up in the vicinity of the bias attractor by the end of the simulation. In the previous chapter, I also demonstrated that this is a stable state: once a category mean gets close to 0.7, it will not stray very far from it. This behaviour results from the fact that category representations with a mean of 0.7 remain essentially

1. Note that Lass’s (1997) original terminology is the inverse of that used in this thesis. Thus, the peaks and ridges that I refer to are in fact basins and valleys in his account. Since this is by no means an essential point, I have changed his terminology to make it more compatible with the present account.

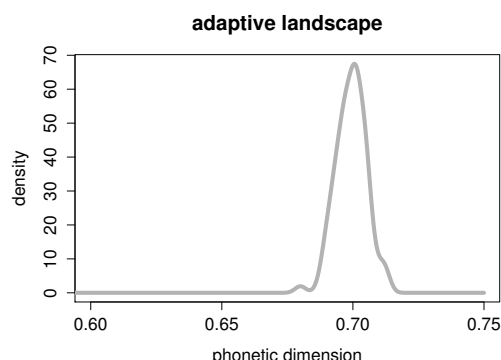


Figure 5.1: *The distribution of category means in simulations with a single category and a single bias that have reached a stable state. This distribution can be interpreted as an adaptive landscape, since the most frequent values are also the ones that are the most likely to be preserved faithfully in later iterations.*

unaffected by the bias, while those with a different mean will necessarily move towards it (this follows from the definition of biases as a logistic function with a value of 0 when $x = b$). To put it slightly differently, categories with a mean of 0.7 are reproduced faithfully in later iterations, while categories with different means are reproduced less faithfully. If we look at a large number of simulations – as in Figure 5.1 – the most frequent outcomes will necessarily be those that are reproduced the most faithfully. In a sense, then, the diagram in Figure 5.1 shows the adaptive landscape for this particular simulation, where the category means evolve towards the peak over many iterations. Importantly, the adaptive landscape in this simulation is defined by two factors: the location of the bias attractor and the strength of the bias (the relevance of the latter becomes clearer when looking at simulations with different bias strengths; cf. Figure 4.9).

A similar argument could be made for the dispersion-based simulations in Section 4.2.2 as well: the category means (and standard deviations) are evolving towards particular stable values that correspond to a peak in an adaptive landscape defined by the parameters of the simulation (in this case, r , or misperception rate). Note that I refer to a single peak in the adaptive landscape, even though Figure 4.12 shows two peaks corresponding to the two category means. This is because the notion of adaptive landscape is used in relation to sound systems and not individual categories. Thus, the adaptive landscape for category means in dispersion-based simulations with two categories is drawn

over a two-dimensional map, where each dimension corresponds to one of the category means. Indeed, in the case of dispersion it would be impossible to talk about the stability of one of the categories without referring to the other one, given that this phenomenon emerges from the interaction between categories.

The two examples above provide a clear illustration of the notion of adaptive landscape as it applies to sound systems, but it will be useful to give a more explicit definition as well. In this thesis, adaptive landscapes are defined over sound systems or parts of sound systems. These sound systems will mostly be represented in terms of category means (as in the examples above).² The topology of the landscape is determined by the extent to which different systems satisfy the various pressures acting on them (such as the effects of phonetic biases and dispersion). In a sense, these pressures define a measure of optimality over sound systems, and it is this optimality that is represented by the hills and valleys of the landscape. The global optimum is represented by a configuration where all these pressures are satisfied to the extent possible in a given type of simulation (e.g. the category affected by the bias reaches the bias attractor, or there is so little overlap among the categories that misperception hardly ever takes place). There usually exist numerous local optima as well, where only a subset of the pressures are satisfied, but the journey towards a more optimal system would lead through a valley of undesirable configurations. As it will be shown, both global and local optima are characterised by stability, which follows from the fact that the highest peaks in the landscape are isolated from each other. Moreover, most of the optimal states can be seen as a balance among the pressures affecting the system: the system stays stable because the different forces acting on it cancel each other out. These points will be illustrated in detail in the next section.

2. This is a somewhat arbitrary choice, given that there are many alternative measures that could also be used to characterise sound systems. For instance, instead of the means, we could look at the standard deviations of the categories, or some more abstract measure of the overall dispersion of the system (e.g. Liljencrants & Lindblom's 1972 measure of total energy). However, this arbitrariness is not a real problem in the present context. Adaptive landscapes are merely a means of making the discussion of complex sound systems more lucid, and not a crucial theoretical assumption in this thesis. The reason for choosing category means as the characters over which adaptive landscapes are defined is therefore mostly didactic: they illustrate changes in sound systems in an abstract but still sufficiently detailed way.

Note that this approach to adaptive landscapes is not identical to that usually taken in evolutionary biology. As it has been explained above, the biological definition of adaptive landscape is built on the notion of fitness, which can be formalised in a mathematically rigorous way. In contrast, I rely on a much more intuitive measure of optimality in this thesis. This difference stems from the fact that the notion of fitness is not directly applicable to sound systems – or at least not when the object of enquiry is the abstract production-perception feedback loop with a single sound system and not a collection of sound systems in a population of agents. The fact that the present use of the term differs from its common usage has no impact on the main argument of this thesis. The notion of adaptive landscape is used in a metaphorical way to make the argumentation easier to follow, but the simulation results do not crucially rely on this metaphor. A small amount of inconsistency in the descriptive language used to make the results more transparent will not deduct from the force of the argument.

We are now in a position to restate the criticism of bias-based approaches in terms of adaptive landscapes. The main problem with the types of simulation illustrated in the previous chapter is that the adaptive landscapes defined by their parameters are too simple. In each simulation, there is only a single global optimum corresponding to the complete satisfaction of a single pressure. The systems inevitably evolve towards these stable states, given that there is no alternative. As I have suggested above, this is a consequence of the fact that these simulations look at categories in a vacuum. The question, then, is whether this behaviour changes in more complex systems.

The next section is an enquiry into the effects of complexity on adaptive landscapes. As in the previous chapter, the main tools for this investigation are computer simulations, but this time the simulated systems contain a higher number of categories ranging between three and seven. As we will see, this creates a much more interesting situation with multiple peaks and valleys in the adaptive landscape. The simulations will look at two central questions related to these complex adaptive landscapes: (i) whether the systems can really be seen as evolving towards stable states and (ii) what the main characteristics of these stable states are.

5.2 SIMULATING COMPLEX SOUND SYSTEMS

Before giving an outline of the simulation architecture, let me briefly explain how the simulations relate to the main argument of this thesis. A more detailed explanation has already been given in Chapter 3, so only the main points are repeated here. In a description of the role of simulations in investigating self-organising systems, Wedel (2011) argues that simulations are typically used ‘either as an existence proof that a given structure can arise through interactions between some defined set of system properties, and/or as a supporting illustration for verbal or analytic arguments’ (Wedel 2011: p. 135). The simulations in this thesis have an element of both in them, but they are also somewhat more ambitious in their scope: they attempt to show that some of the behaviours in the simulated systems are not simply possible, but follow *necessarily* from the theoretical assumptions behind the models. Of course, this is somewhat harder to achieve than simply showing that a given behaviour can arise under the right conditions. For this reason, I have gone to great lengths to make the theory behind the simulations as clear as possible and to link the implementation and the results of the simulations directly to this theory. Moreover, instead of looking at the results of one or two runs, as is typical in simulation-based research, a large number of simulations are run with varying parameter settings. By exploring the parameter space of the model, we can get an idea of how robust the general dynamics of the system are, and whether they persist despite variations in the details of implementation. I do not wish to claim that the dynamics discussed in this chapter arise under every possible combination of parameter settings – in fact, I am sure that with enough ingenuity and patience, it is possible to tweak the model into producing counterintuitive outcomes. However, by looking at a range of different parameter settings and finding the same behaviour, I hope to convince the reader that – at least under conservative assumptions about the model parameters – this investigation tells us something important about the systems themselves and not just about the simulations.

I first introduce the details of the simulations through an example in Section 5.2.1. This example is also used to demonstrate a serious shortcoming of the modelling architecture employed so far in this thesis: the large-scale investiga-

tion of complex systems proposed above is not feasible from a computational perspective unless the simulations are simplified in some way. Section 5.2.2 presents an alternative simulation technique that is indistinguishable from the model used so far in terms of its dynamics, but is much less time-consuming. Finally, Section 5.2.3 discusses the results of a large set of simulations, and summarises their general behaviour with the help of the notion of adaptive landscape described above.

5.2.1 *An example simulation*

In order to model complex sound systems, several modifications are made to the simple setup used in the simulations of the previous chapter. First of all, all the sound systems investigated in this chapter are located in a multidimensional space with two axes. These two axes could correspond to a variety of different phonetic features, but perhaps the most straightforward analogue is the F1-F2 vowel space. This parallelism should not be interpreted too literally, though, since these abstract systems are not modelled after any particular subsystem in natural languages. The phonetic space contains a number of phonetic categories ranging between three and seven; in the example simulation presented below, this number is five. To make the discussion easier to follow, there is only a single phonetic bias, which affects a single category. As it has been noted in Section 4.1.2 of the previous chapter, this setup corresponds to a situation where a sound category exhibits a particularly high bias proportion with respect to a given bias. This is like the case of [u]-fronting in English, where an unusually high proportion of the tokens of [u] occur in contexts that favour fronting (Harrington 2007). The model implements misperception-based asymmetries in category update in the same way as the simulations in Section 4.2.2.

The list on the next page provides an overview of the initial setup and the general parameter settings of the example simulation. Note that the simulation uses prototype-based category representations.

(5.1) NO. OF CATEGORIES: 5

CATEGORY MEANS: $\mu = (\mu_1, \mu_2)$ for each category, where μ_1 and μ_2 are random numbers between $[0.1, 0.9]$

COVARIANCE MATRICES: $\Sigma = \begin{bmatrix} 0.01 & 0 \\ 0 & 0.01 \end{bmatrix}$ for all the categories

VARIANCE INCREASE: $h^2 = (0.013, 0.013)$

BIAS: $b = (0.85, 0.85)$, $s = 0.005$

CONSTANT OF UPDATE: $c = 2000$

MISPERCEPTION RATE: $r = 0.1$

Let me briefly go through these settings. As it has been noted above, there are five sound categories. Each of the categories is initiated with randomly chosen means along both phonetic dimensions, but the standard deviations along the axes and the covariance are fixed (resulting in globular category distributions at the beginning of the simulation). A small amount of production noise is added to the production targets, represented by two-dimensional Gaussians whose variance is given by h^2 . There is a single bias located in the upper right-hand corner of phonetic space (provided that it is plotted as a traditional Cartesian coordinate plane). The constant of update is set to $c = 2000$ as in all previous simulations, and the misperception rate to $r = 0.1$.

Figure 5.2 shows the evolution of the example system over 1 million iterations. The ellipses represent 95 per cent confidence regions for each of the categories. The cross shows the location of the bias, while the shading marks the category affected by the bias (i.e. the category with an unusually high bias proportion). The following general trends can be observed in the evolution of the system. First of all, the amount of overlap among the categories decreases steadily over the course of the simulation. This is not surprising in light of the simulation results demonstrating ambiguity-driven dispersion in the previous chapter, and can be considered yet another replication of the effect discussed in Wedel (2006) and Blevins & Wedel (2009). Second, the categories seem to use up most of the available phonetic space, but they do not grow beyond a certain limit. This follows from the interaction of the variance-inflating effect

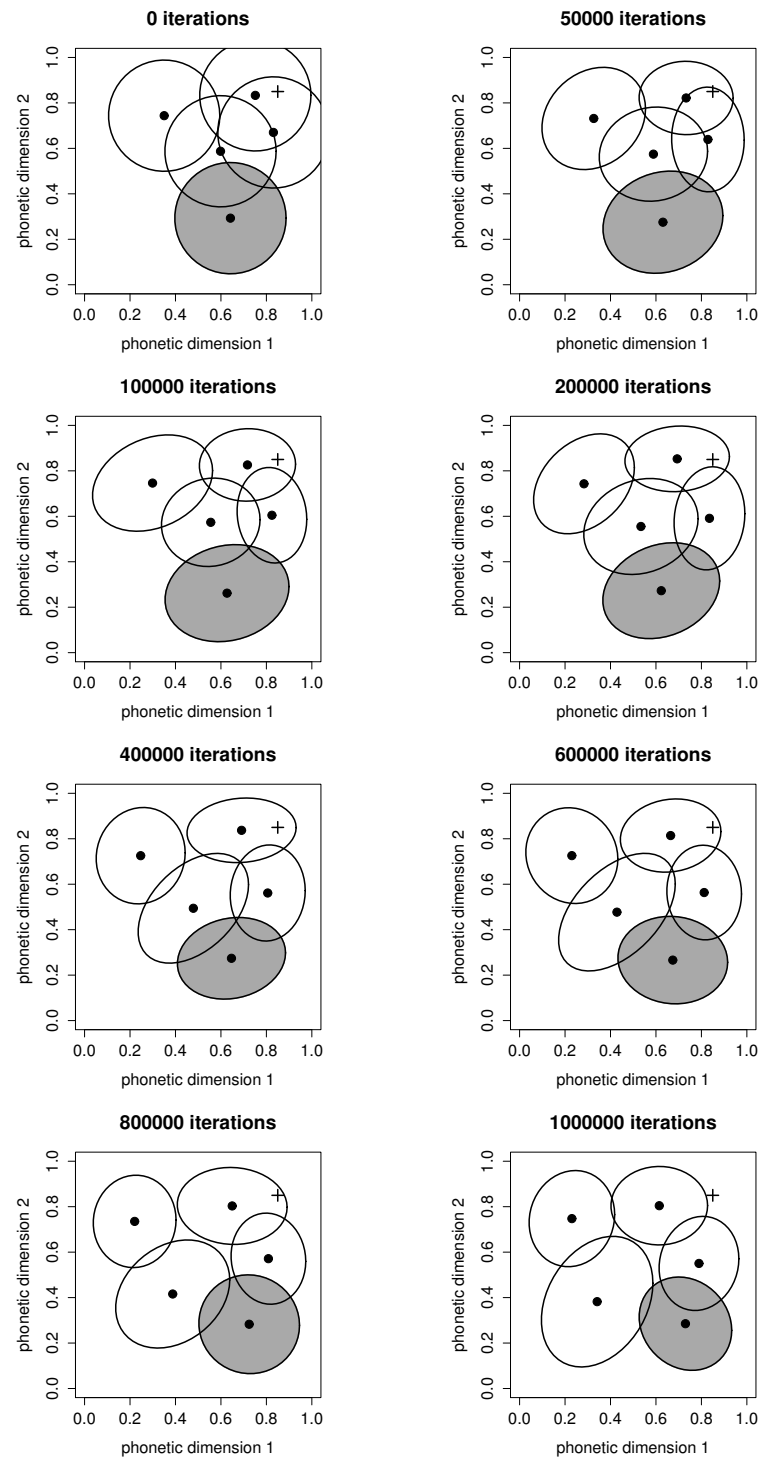


Figure 5.2: The evolution of a complex sound system consisting of five sound categories with a single bias. The ellipses illustrate the category distributions and the cross the location of the bias. The shaded ellipse represents the category affected by the bias.

of production noise and the boundedness of phonetic space. Third, the speed of the changes in the simulation seems to decrease gradually. To get a sense of this deceleration, compare the changes between 0–200,000 iterations with the changes between 800,000–1,000,000 iterations. It is clear that the initial changes are more dramatic than the ones that take place in later stages of the simulation. Fourth, and perhaps most important, the biased category never manages to get close to the bias attractor: it approaches it along the *x* axis, but it stays relatively far from it along the *y* axis. This observation constitutes the first confirmation of the idea that complex systems behave differently from simple systems with respect to consistent phonetic biases.

Intuitively, it seems clear that the sound system is navigating a complex adaptive landscape during its evolution. This adaptive landscape is determined by a variety of different factors, including the tendency towards dispersion introduced by misperception, the location and strength of the phonetic bias, the bias proportions of the different categories (even if bias proportion is represented in a rather crude way), the limits of phonetic space and the variance-inflating effect of production noise. The initial random configuration is, of course, highly unstable: there is a large amount of overlap among the categories, some parts of the category distributions are ‘cut off’ by the boundaries of phonetic space, and the biased distribution is not particularly close to the bias attractor. As the simulation proceeds, the system begins to climb towards higher-altitude areas of the adaptive landscape, which also brings about an increase in its stability. This is why the rate of change seems to be falling gradually.

The range of potential manoeuvres through which the system can increase its optimality appears highly constrained at every point during the simulation. Only small changes can take place between two iterations, and the set of movements available to a given category is limited by the positions of the categories around it and the boundaries of phonetic space. As a result, the system seems to follow a narrow ridge of local optima in the adaptive landscape, and cannot simply jump to the global optimum. Presumably, the most optimal configuration would be one where the biased category is located in the upper right corner of the phonetic space near the bias attractor. However, it appears that as the categories disperse and fill up the available space, the system gets stuck in a local optimum from which this global optimum is simply not accessible.

The description above suggests that the complex structure of the adaptive landscape can indeed create a situation where a phonetic bias does not lead to sound change. It would be useful to see if this observation can also be made about the limiting behaviour of the system – that is, whether the system is permanently stuck in a local optimum, or if it can eventually reach the global optimum. In order to get a sense of how the system behaves over a longer stretch of time, the simulation was run for another 3,000,000 iterations. Figure 5.3 shows eight snapshots of the long-term evolution of the system. Two things should be noted about this diagram. First, the system evidently continues to change after the initial 1,000,000 iterations. Second, there are no significant changes to the category distributions after about 3,000,000 iterations (note that Figure 5.3 illustrates changes after 3,000,000 iterations in slightly more detail than changes before 3,000,000 iterations). While the system continues to undergo small random shifts, these shifts are not consistent inasmuch as they do not result in larger changes over time. This contrasts with the initial behaviour of the system, where the changes are clearly goal-oriented (even if this goal-orientation is not ‘built into’ the simulations in the form of an explicit optimisation algorithm). Returning to the language of adaptive landscapes, the system seems to reach a local optimum which stands as an isolated peak in the landscape. Any movement away from this state would result in a significantly less optimal system, and therefore it is blocked. Again, this blocking follows from the implicit dynamics of the system, and is not hardwired into the model. In sum, the complex nature of the adaptive landscape seems to preclude certain phonetically-driven changes. This is in line with what I suggested earlier in this chapter, namely that the dynamics of complex sound systems are different from those of simplified sound systems with a single category. However, this conclusion needs to be investigated more systematically.

Unfortunately, this type of simulation raises a serious practical problem. In order to be able to make an educated guess about the limiting behaviour of the system, I had to run the simulation for 4,000,000 rounds. This took approximately three and a half hours on my own computer with a 2 GHz processor and 2 GB RAM.³ Although this, in itself, cannot be regarded as a

3. I have no doubt that many future readers of this thesis will smile and perhaps shake their head in disbelief. While this computational problem might not present a serious issue for tomorrow’s computers, this does not diminish the need for a more efficient implementation.

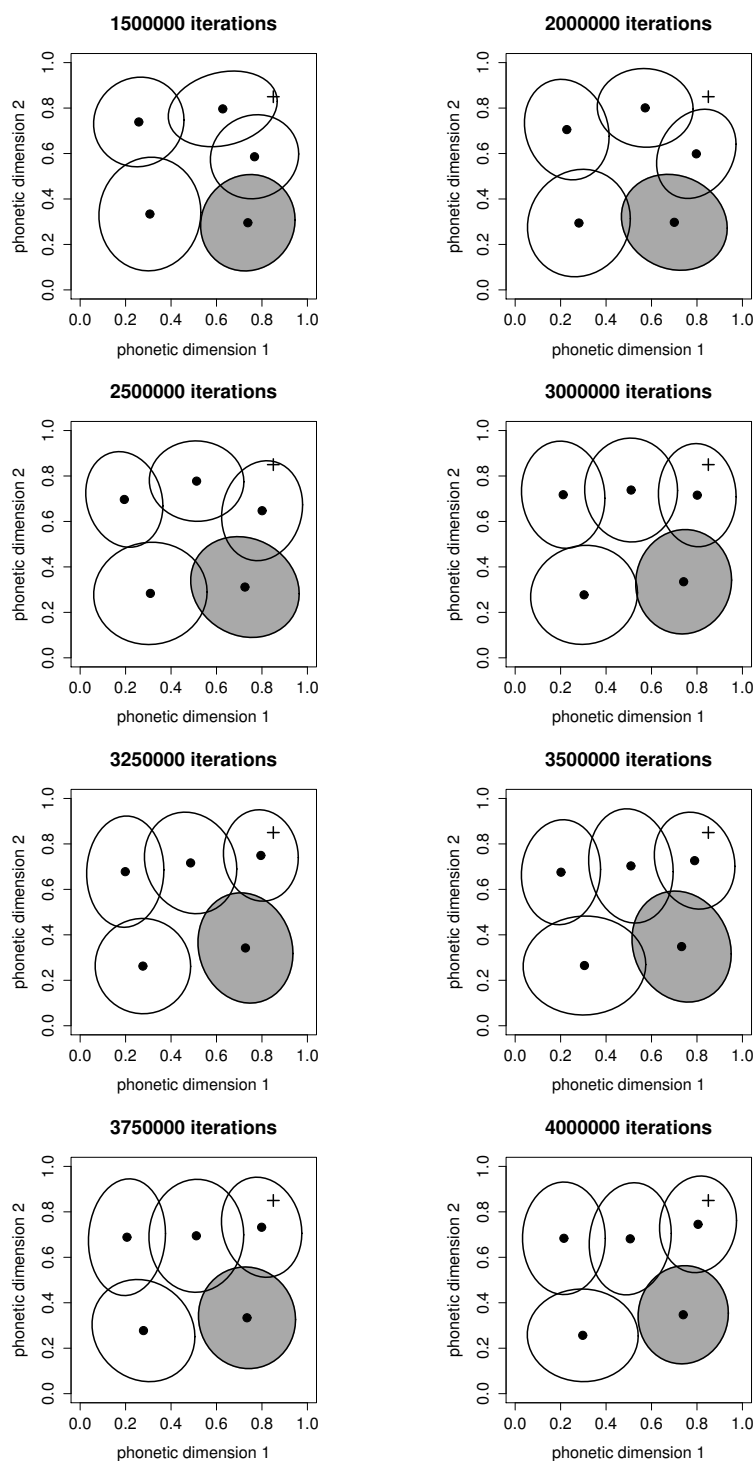


Figure 5.3: *The evolution of the complex sound system in Figure 5.2 over a further 3,000,000 simulations.*

particularly time-consuming simulation, it is clearly not well-suited to a larger investigation with many runs. Section 5.2.3 presents results from 6000 similar simulations, which would take nearly two years to run on the same computer using the current model. While this issue could be avoided by using an expensive computer cluster, the next section will show that there is no need to use up such valuable computing resources, given that a much more efficient implementation is also available.

5.2.2 *An alternative simulation framework*

The main insight behind the alternative framework presented here comes from one of the observations in the previous section: the different pressures affecting the sound system create ridges in the adaptive landscape that allow the system to move towards areas of higher altitude. Although the movement of the systems has an element of randomness, the fact that they seem to become stuck at a certain point in their evolution suggests that consistent changes only occur along these ridges (otherwise the system should be able to find a way to the global optimum even after it has reached a local optimum). If this is the case, the changes in the system are essentially deterministic, in that they follow paths that depend entirely on the parameters of the simulation. The method described below makes it possible to explore these paths directly by eliminating the randomness of the system.

Since one of the fundamental properties of the model used in the simulations so far is its stochastic nature, it might be difficult to see how it could be described in deterministic terms. The key step in bridging this conceptual gap involves increased reliance on probability distributions. Specifically, I will introduce a distinction between the underlying distribution representing a category and the surface distribution of tokens that form the basis of category update (the latter will be referred to as the ‘observed distribution’). Importantly, the model outlined in the previous chapter allows these two distributions to be different. These differences come from four main sources: production noise, the consistent application of phonetic biases, misperception-based asymmetries in category update and constraints imposed by the boundaries of phonetic space. Each of these mechanisms can introduce discrepancies between the underlying

distribution and the observed distribution, and these discrepancies will be consistent given any specific configuration of sound categories. In fact, the relationship between underlying and observed distributions is so transparent that the observed distribution can readily be calculated from the underlying one using a number of simple mathematical transformations. The details of these transformations are explained in Appendix B.

The ability to calculate the differences between the underlying and the observed distributions is extremely useful in the present context. This is because these differences are the sources of all consistent changes in the system. If the speaker's category representations are not consistent with the input they receive (which is determined by the observed distribution), they will necessarily undergo changes as a result of category update. In the present model, the input to the speakers is determined by the observed distribution, which means that knowledge of the latter will allow us to make predictions about the direction of change.

We now have all the components necessary to construct a deterministic simulation architecture where the sound system travels the same path as in the stochastic simulations used so far. In order to model the evolutionary dynamics of the system, we need to use the transformation of the underlying distribution into the observed distribution as a proxy for the production-perception feedback loop. This can be done by iterating the transformation in a way that the output of one iteration (i.e. the observed distribution) serves as the input of the next one (i.e. the underlying distribution).⁴ To give a concrete example, I repeat Figure 4.5 from Section 4.1.2 as Figure 5.4 here. This diagram shows the observed distribution (black dashed line) corresponding to a category representation (black solid line) after the application of a phonetic bias (note that none of the other transformations are included). In the simulation framework proposed here, the distribution marked with a dashed line would become the underlying distribution in the next iteration, and would produce a new observed distribution that is even closer to the bias attractor. Iterating this procedure many times would result in convergence to the bias (similarly to the simulations in Section 4.2.1).

4. Note that since the model used here assumes normally distributed category representations, the observed distribution has to be recoded as a Gaussian function once all the transformations have been performed. This is simply a matter of calculating the first moment (i.e. the mean) and the second central moment (i.e. the variance) of the observed distribution.

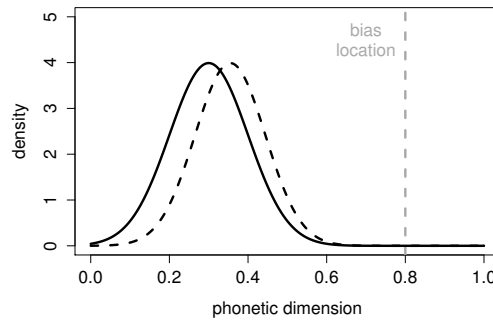


Figure 5.4: *The distribution representing a given category (black solid line) and the observed distribution (black dashed line) after the application of a phonetic bias (grey dashed line).*

Of course, it is not sufficient to simply suggest that this framework produces the same results as the original model: a demonstration is in order. Therefore, I have rerun the example simulation from the previous section using the new model. All the parameter settings were exactly the same as previously (including the initial positions of the categories).⁵ The results are shown in Figure 5.5, where the left-hand side panels illustrate the original simulation and the right-hand side panels the new simulation. The similarity between the two simulations is obvious. While there are some small differences in the locations of the categories, the overall evolution of the two systems is identical. The developments in the two systems also seem to be linearly correlated in terms of computational time: the new model seems to change roughly 2000 times as fast as the original one (at least in terms of the number of iterations). It should be noted that this is not an isolated result: all the other parallel simulations I have run have shown exactly the same tendencies (although these will not be presented here for reasons of brevity).

These results confirm the hypothesis proposed above, according to which all consistent changes in the system follow narrow paths that are determined by the pressures acting on the system. While this is exactly what we expected to see, I believe it is still a somewhat surprising finding. It is in many ways counterintuitive to see that the overall behaviour of a fundamentally stochastic

5. Except for the constant of update, which does not apply to the framework introduced in this section. The reason for this is that category update in this case simply consists in replacing the underlying distribution with the observed one.

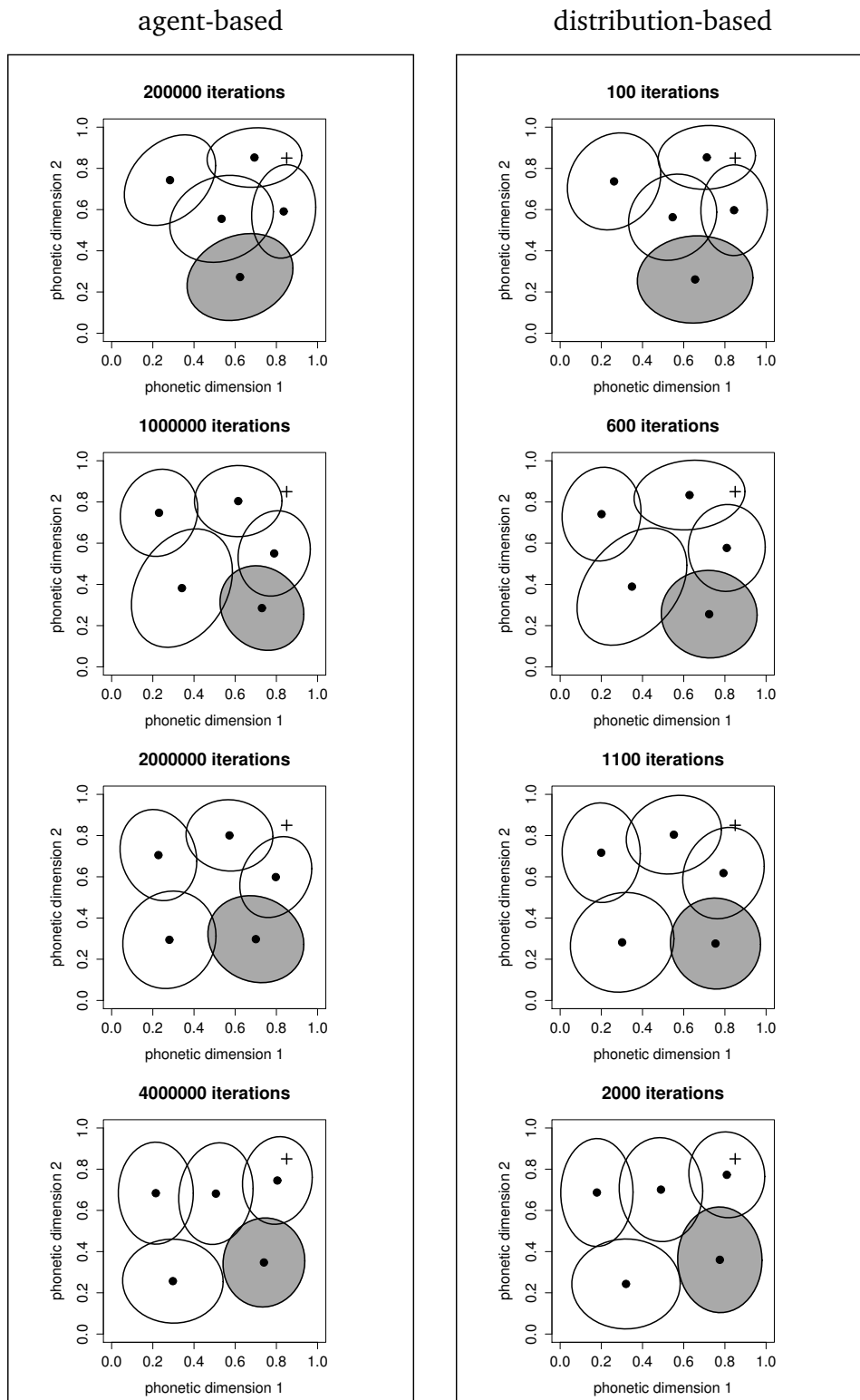


Figure 5.5: A comparison of the original agent-based model described in Section 5.2.1 and the observed distribution-based model introduced in this section.

system is in fact nearly deterministic. However, one should not forget that these dynamics are very typical of all types of evolution. Indeed, this is exactly the reason why we often find the emergence of the same biological trait as a solution to a given problem among unrelated species – a phenomenon termed ‘convergent evolution’. A particularly popular example of convergent evolution is the striking similarity between the structure of the eye in vertebrates and cephalopods (Blevins 2004: p. 48). These two groups of animals have anatomically and functionally similar eyes (so-called ‘camera eyes’), despite the fact that they do not share an ancestor with this anatomical feature or even its precursor. Thus, the same characteristic emerged independently in these two groups. Both in sound change and in biological evolution, the reason for the apparent determinism of such changes lies in the ruggedness of the evolutionary landscape. The systems walk narrow paths towards peaks in the landscape, and cannot significantly deviate from these paths.

Before applying the new simulation technique in a more detailed investigation of the effects of complexity on the adaptive landscape, let us briefly clarify how it relates to the stochastic simulations used so far. Although the two models evidently produce the same results, they should not be equated with each other. The stochastic simulations model the behaviour of simulated agents in a direct way, while the deterministic simulation maps the ridges of the adaptive landscape following a more abstract procedure. If the influence of random variation was increased significantly in the stochastic model (by lowering the value of c , the constant of update), it is likely that the two types of simulation would sometimes diverge. Nevertheless, the most likely outcomes in the stochastic simulations would still evolve along the paths predicted by the deterministic simulations. Since the next section focuses on the structure of the adaptive landscape, and not on potential fluctuations due to random noise, we are justified in using the simulation architecture proposed here. However, it should be borne in mind that this substitution might not be equally suitable to all situations.

5.2.3 *Exploring the adaptive landscape*

This section presents the results of a large computational investigation consisting of 6000 simulations run with a number of different parameter settings. The

aim of this investigation is to see how various internal and external factors affect the evolutionary dynamics of sound systems. The most important factor investigated below is complexity, which is operationalised as the number of categories in the system. Besides complexity, two further factors are examined: bias strength and misperception rate. The structure of the investigation is as follows. After describing the parameter settings and the different experimental conditions used in the simulations, I present a quick tour of the most frequent outcomes, just to give the reader a sense of the general dynamics of the model. This is followed by a detailed look at how the different factors interact with each other in shaping the adaptive landscape, based on large sets of simulations. The remaining sections of this chapter will be devoted to discussing the implications of these findings.

The simulations presented below all rely on the procedure outlined in the previous section. Most of the parameter settings are exactly the same as in the example run in the previous section. The main difference is that (i) certain parameters are varied across the simulations and (ii) each combination of parameter settings is tested in 500 different runs of the simulation with different initial category locations. This random perturbation of initial category locations across otherwise identical simulations is necessary to get a sense of the different pathways of evolution defined by the adaptive landscape. If there were no such perturbations, all simulations would produce exactly the same results, given the deterministic nature of the model.

The following parameter settings are investigated: $s \in \{0.001, 0.003, 0.005\}$ (weak vs. medium-strength vs. strong bias conditions); number of categories $n \in \{3, 5, 7\}$ (small vs. medium-sized vs. large inventory conditions); $r \in \{0.1, 0.5\}$ (low vs. high misperception rate conditions). The conditions representing bias strength and the size of the inventory are crossed with each other, giving a 3×3 matrix. Since the effects of misperception turned out to be relatively weak in preliminary simulations (cf. the results below), I have not fully crossed this factor with all other conditions. The 3×3 matrix of original conditions is only tested at a single misperception rate ($r = 0.1$). This is complemented by a small additional set of simulations with a misperception rate of $r = 0.5$, where the inventory size is kept constant ($n = 5$) and only bias strength is varied. Thus, the overall figure of 6000 simulations can be calculated as

follows: $3 \times 3 \times 500 = 4500$ runs with a low misperception rate and a further $3 \times 500 = 1500$ runs with a high misperception rate. The boundaries of phonetic space, the location of the bias and the amount of production noise are all exactly the same as in the previous section. Each simulation is run for 5000 iterations, and only the final state is recorded.

In Section 5.2.1, I suggested that the simulated systems evolve towards equilibria determined by the locations of local optima in the adaptive landscape. Although I have shown that there are no qualitative changes in the system after a certain number of iterations, I have not yet provided a characterisation of these stable states. Figure 5.6 illustrates a few typical configurations that emerge in the simulations after 5000 iterations (these are all taken from simulations with a low misperception rate $r = 0.1$ and a medium-strength bias $s = 0.003$). The first thing to note is that the systems are all highly symmetrical and organised: the initial random configurations give place to regular geometrical formations as the sound systems settle into equilibria. This is likely due to the fact that such symmetrical and systematic arrangements make the best use of phonetic space both in terms of avoiding category overlap and allowing the categories to stretch as far as they can. Indeed, very similar orderly configurations are observed in the contrast-driven simulations of Liljencrants & Lindblom (1972) and de Boer (2001), which supports the idea that contrast is the main factor behind these particular patterns of self-organisation. In contrast to the works cited above, this thesis does not attempt to draw parallels between the observed outcomes and frequently occurring vowel systems. This is partly because no effort was taken to model the phonetic space in a way that reflects the structure of the vowel space in natural languages (as opposed to the works cited above, which are based on careful implementations of the vowel space). Even more importantly, the simulations in this section make an abstract point about the evolution of sound systems as a function of complexity, phonetic biases and other factors. This point could be translated into more concrete predictions about vowel systems or other subsystems within natural languages, but such a translation is not essential to the argument itself.

Let us briefly describe the details of the overall configurations in Figure 5.6 (ignoring the position of the biased category for the moment). Perhaps the easiest way to do this is to imagine the vowel space as a jewellery box

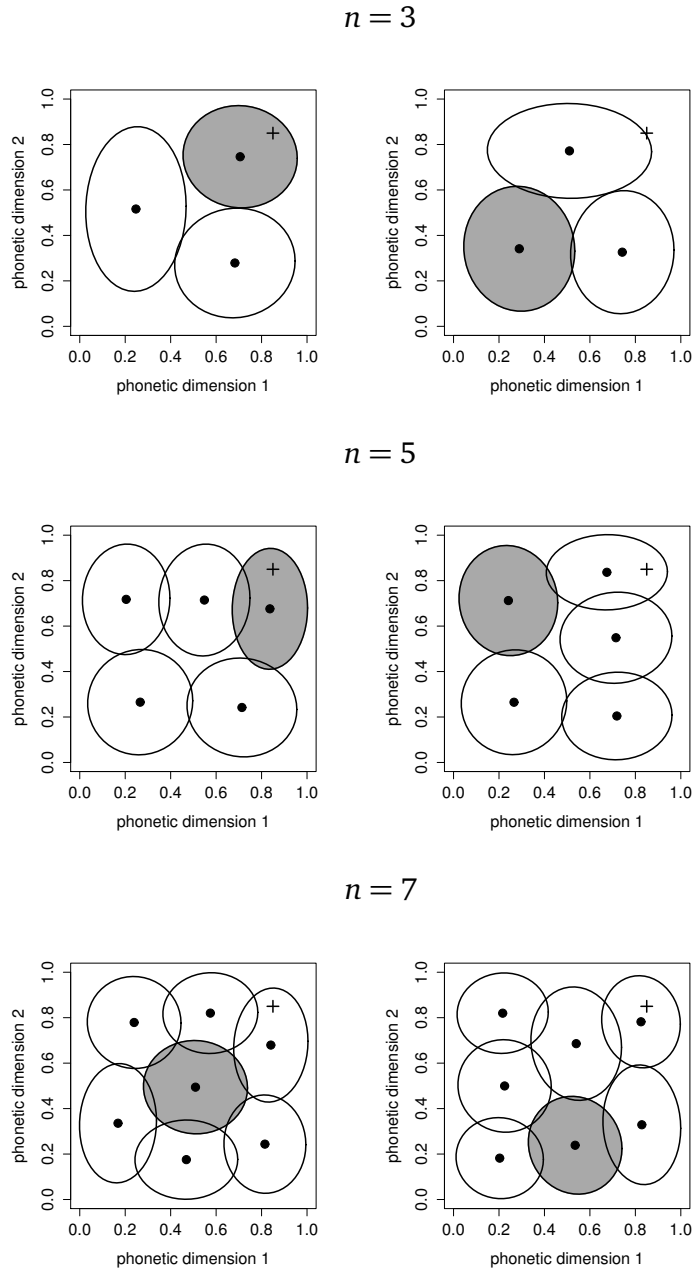


Figure 5.6: Typical configurations of categories in stable sound systems after 5000 iterations. The panels at the top illustrate systems with three categories, the panels in the middle systems with five categories and the panels at the bottom systems with seven categories. The cross indicates the location of the phonetic bias and the shading the biased category.

with a few separate slots in it, and a single category in each slot. In the case of the three-category system on the left, this jewellery box would contain a bigger vertically oriented slot stretching from top to bottom on the left-hand side, and two smaller equal sized slots on the right-hand side. We could fit the same box around the second three-category system as well if we rotated it clockwise by 90 degrees. The situation is very similar in the case of five-category systems: the same jewellery box with a row of three slots and a second row of two slots could accommodate both systems. The seven-category case is a little more complicated: here, a box with three columns of 2–3–2 slots would fit around the first system, but a 3–2–2 box would be needed for the second one. Although this is a weaker case of isomorphism than in the three and five-category cases, the two systems are still remarkably similar. Note that these patterns are extremely robust: in the case of three and five-category simulations, over 90% of the systems ‘fit in the boxes’ described above; this figure is somewhat lower for seven-category systems, at around 50–60%.⁶ The high proportion of systems that can be described in terms of such a small set of configurations confirms the idea that the systems converge towards optima: if these states were not optimal and stable, it would be very surprising to find so many systems exemplifying them.

Having seen some examples of the general trends in the evolution of these systems, we are now in a position to undertake a more systematic investigation of the stable states in the simulations. I begin with a discussion of the influence of the number of categories on the adaptive landscape. Then, I show how this factor interacts with bias strength. Finally, I add in the last factor, namely, misperception rate, and take a quick look at how it changes the overall picture.

Since the focus of this thesis is on the role of phonetic biases in sound change, only one particular aspect of the adaptive landscape will be explored: the extent to which the biased category approaches the bias attractor. Although it would be interesting to look at other factors as well – such as the amount of ambiguity in the systems, or the variance of the individual categories – this simplified measure will be sufficient for our purposes. The easiest way to get a

6. The lower percentage of seven-category systems that conform to the patterns illustrated in Figure 5.6 is a result of the fact that there is a third arrangement as well, which takes care of most of the remaining cases.

sense of the degree to which a given category converges to the bias is to plot the distribution of all the biased category means at the end of the simulation as we did in Sections 4.2.1 and 4.3 in the previous chapter (see Figures 4.9, 4.19 and 5.1). This is slightly more complicated in the present case, given that the phonetic space is two-dimensional. There are a number of different ways to visualise statistical distributions over a two-dimensional space. To ensure that the illustrations are clear, I will use three of these in combination: three-dimensional density plots, heat maps and contour lines. It should be emphasised that each of the graphs below illustrates the distribution of a single category (out of three, five or seven categories) in a set of 500 simulations (with the same parameters settings, but different initial category representations).

Figure 5.7 shows the distribution of the means of the biased category in three sets of simulations with varying numbers of categories (the bias is medium-strength in each case, and the misperception rate is low). Let us first interpret the two panels at the top. Focusing on the three-dimensional graph on the left-hand side, there is a very clear peak around the top right corner, and a small ring of low foothills surrounding it. This means that the biased category (i.e. the category with a particularly high bias proportion that is shaded in the diagrams illustrating the sound systems) nearly always occurs close to the bias attractor at (0.85, 0.85), but in a few exceptional cases it might be located at a different point around the perimeter of the phonetic space. The panel on the right provides exactly the same information but in a slightly different form. The most straightforward way to interpret this diagram is as a topographic map of the distribution: darker colours represent peaks, and the contour lines surround areas that are of higher altitude than the numbers written on them (the numbers themselves are of no importance in the present case).

The five and the seven-category cases look rather different from the three-category case. It is immediately obvious that the distributions shown in the second and the third rows are much more varied than the distribution representing the three-category simulations. In the five-category case, the biased category still appears to end up nearer the bias attractor than it would if its behaviour was guided entirely by chance. However, we see a distinct range of sub-optimal configurations that are not nearly so well-represented in the three-category simulation. There are two peaks in the immediate vicinity of the

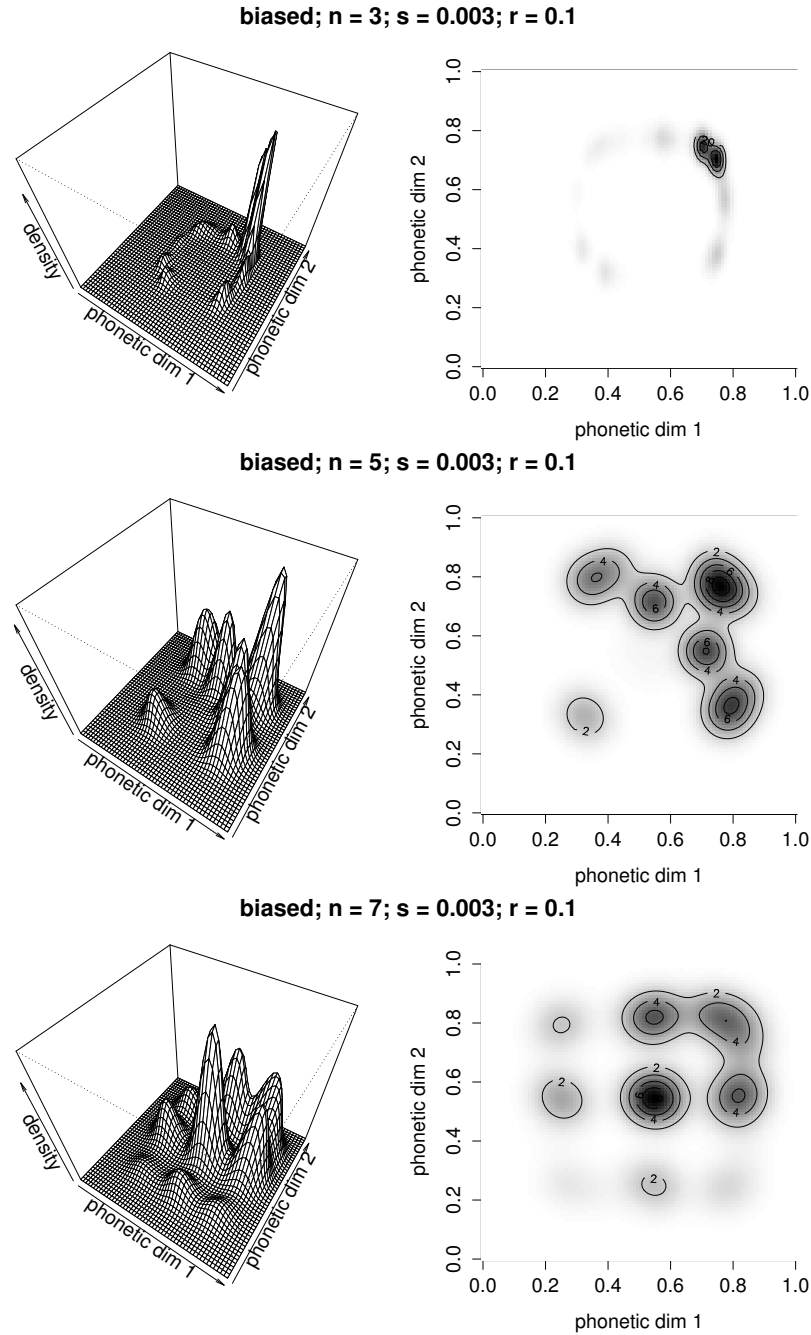


Figure 5.7: The distributions of biased category means in three (top), five (middle) and seven-category (bottom) simulations. The panels on the left visualise the distribution using a three-dimensional density estimate. The panels on the right use a heatmap and contour plot superimposed on it to plot the same information. The location of the bias attractor is (0.85, 0.85).

bias attractor, another two slightly lower peaks a little further from it near the neighbouring corners of phonetic space, and a low hill in the opposite corner. The seven-category situation is even more complicated, with an overall nine peaks. Surprisingly, the highest peak in this case is not the one nearest the bias attractor, but the one in the centre. Nevertheless, the distribution is clearly heavier around the top right corner, which indicates that the bias still has a strong influence on the evolution of the system.

What does this tell us about the influence of complexity on the adaptive landscape? In a somewhat loose sense, the diagrams in Figure 5.7 can be taken as representative of the adaptive landscape itself. The peaks show local and global optima, while the valleys in between mark areas corresponding to non-optimal systems.⁷ As the number of categories increases, the isolated peak in the three-category case turns into a range of peaks with complex features and numerous local optima. In fact, the global optimum in the seven-category case is not even where the bias would predict it. This is exactly what we suggested earlier: the influence of phonetic biases on the landscape is strong when there is only a single category, but as soon as further categories are added in, it becomes significantly weaker. Since the landscape is now shaped by multiple different factors, phonetic biases can no longer have their way all the time. Thus, the implication is clear: complexity brings out the effects of factors that would remain hidden in a single-category scenario, and diversifies the adaptive landscape. To put it more succinctly, complexity can counteract the influence of phonetic biases.

If complexity and biases are indeed antagonistic forces, it will be interesting to see how they interact with each other. Figure 5.8 illustrates the crossed effects of bias strength and inventory size through a set of nine heat maps. The influence of complexity does not appear to change significantly: comparing the rows shows the same diversification from top to bottom that was observed in Figure 5.7. This appears to occur independently of bias strength, although the differences are clearer at lower values of s . Moving on to the effects of bias strength, the results are by no means surprising. At lower values of bias strength, the effects

7. Of course, these diagrams do not show the entire landscape: they focus on the mean of the biased category, and do not look at the standard deviation or any of the other categories. The 'real' adaptive landscape is of much higher dimensionality and therefore more difficult to handle both mathematically and visually.

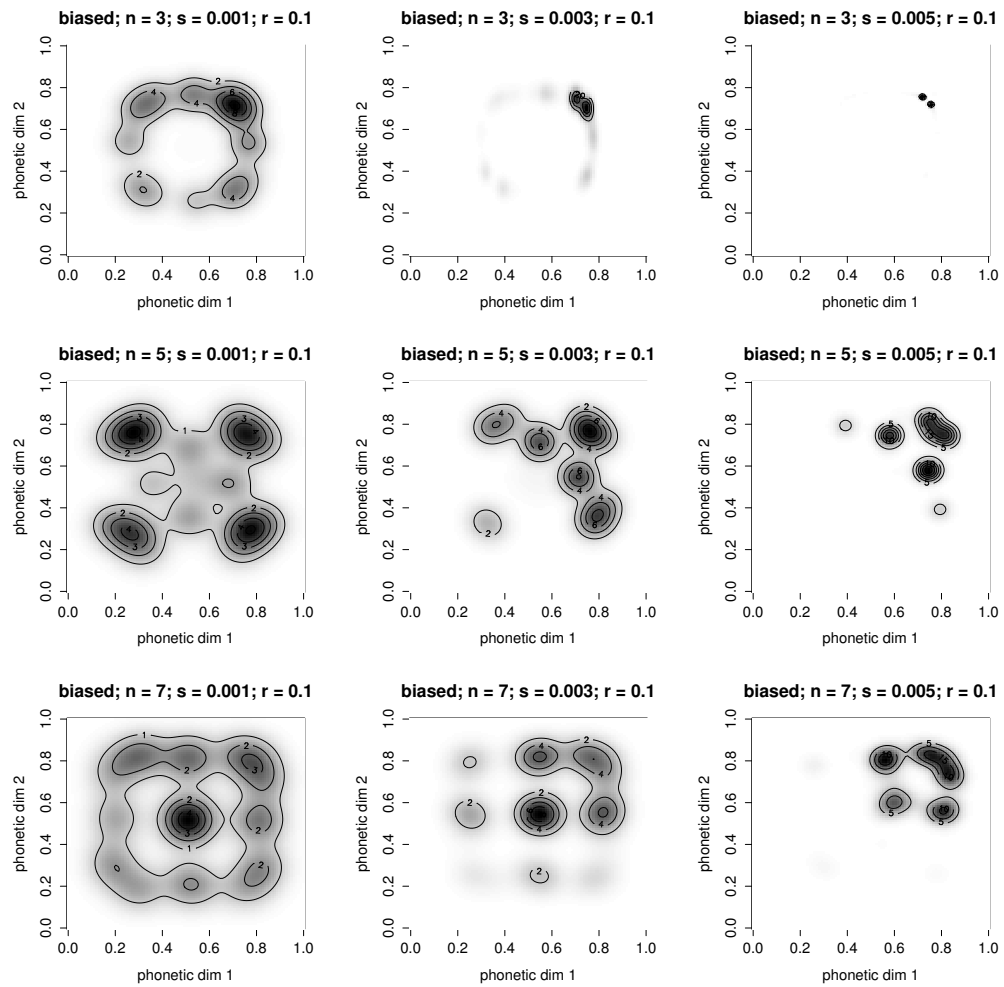


Figure 5.8: The crossed effects of bias strength (varied between low and high horizontally) and inventory size (varied between 3 and 7 vertically). Each diagram shows a heat map illustrating the distribution of biased category means in 500 simulations.

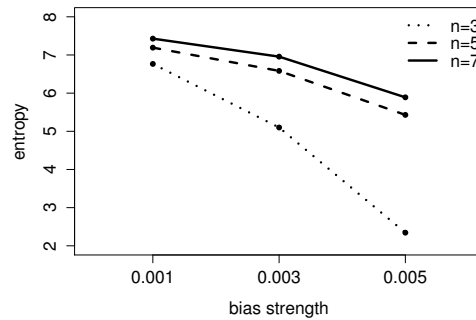


Figure 5.9: *The entropies of the adaptive landscapes of three-category (dotted line), five-category (dashed line) and seven-category (solid line) systems as a function of bias strength.*

of the bias become nearly invisible: only the three-category simulations show a clear pattern, while the five and the seven-category simulations behave almost entirely randomly with respect to the bias. Conversely, when the bias is relatively strong, peaks that are further away from the bias are suppressed. Thus, while the category means in the low bias strength condition fill all the available slots defined by the tendency towards dispersion (cf. the discussion of symmetric and orderly configurations above), the range of possibilities becomes severely restricted in the high bias-strength condition. Note that the peaks themselves remain more or less the same – it is only their relative heights that change.

While the visual illustration of the combined influence of complexity and bias strength in Figure 5.8 is intuitively convincing, there is also a more systematic way to compare these distributions. The key to this comparison lies in the notion of ‘information entropy’ (Shannon 1948). In a nutshell, entropy is a measure of how unpredictable a given random variable is. Thus, if a random variable has a distribution dominated by a single crisp peak (e.g. as in the top right panel of Figure 5.8), the entropy will be low, since the outcomes are highly predictable. If, however, the random variable has a distribution with many peaks, its entropy will be high, since the outcomes are not very predictable. I will not go into the technical details of calculating entropy – it will suffice to say that entropy can easily be obtained for discretised versions of the distributions shown in Figure 5.8 (see e.g. MacKay 2003 for more details). Figure 5.9 plots the entropies of the nine distributions in Figure 5.8. These values support the informal analysis given

above. Entropy is positively correlated with the complexity of the system, which means that more complex systems are less predictable. Conversely, entropy is negatively correlated with bias strength: the stronger a bias is, the more predictable the outcome of the simulation.

Before concluding this section, let us briefly discuss the influence of misperception rate on the simulations. Figure 5.10 illustrates the combined effects of bias strength and misperception rate (recall that these simulations are all based on systems with five categories). Interestingly, the differences across the low and the high misperception rate conditions are relatively minor. The overall shape of the adaptive landscape is more or less the same in the two sets of simulations. The only robust difference seems to lie in the predictability of the results: simulations with higher misperception rates produce more unpredictable outcomes. This is confirmed by a comparison of the entropies in the two different conditions, as shown in Figure 5.11: systems with a lower misperception rate appear to have lower entropy with respect to the distribution of the biased category means. This result might seem a little counterintuitive, as we would expect the behaviour of the system to become more constrained when the pressure towards contrast maintenance is strengthened. However, it should be borne in mind that the diagrams only show one aspect of the adaptive landscape, namely the convergence of the biased category towards the bias attractor. It is quite possible that the adaptive landscape as a whole becomes more rugged (reflecting a higher degree of determinism) when misperception rate is increased, even as those aspects of the landscape that reflect the influence of the bias show the opposite trend. In fact, this observation is highly compatible with the overall picture so far. Ambiguity-driven dispersion is one of the main pressures responsible for the diminished influence of phonetic biases in simulations of larger sound systems, which means that strengthening it should make the influence of biases even weaker. The decreased predictability of the location of the biased category is a direct consequence of this observation.

This concludes our investigation of the role of complexity in shaping the adaptive landscape. Below is a brief summary of the most significant trends we have isolated.

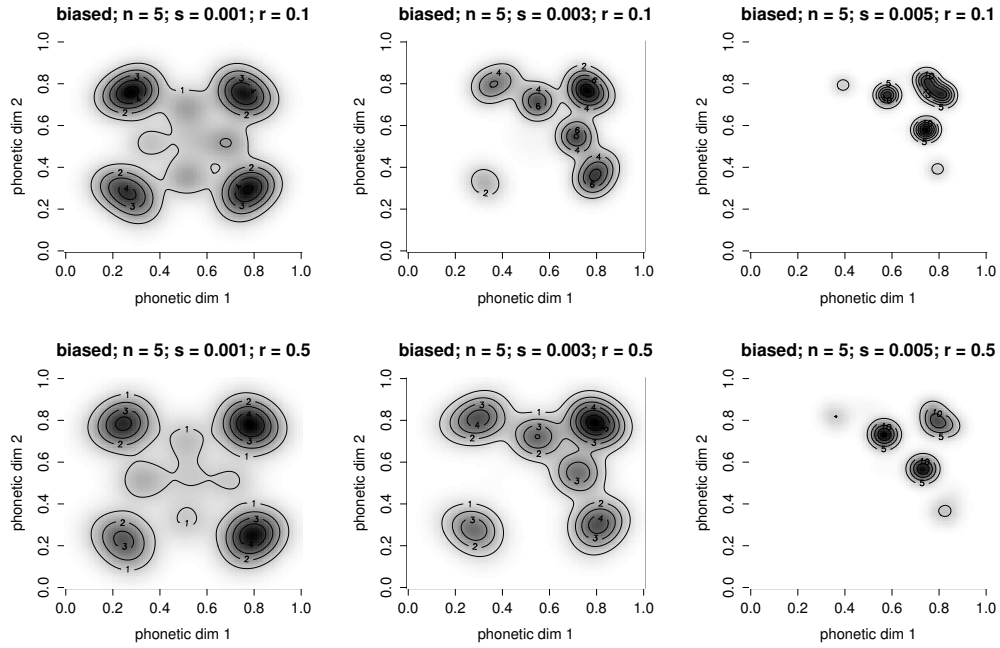


Figure 5.10: The crossed effects of bias strength (varied between low and high horizontally) and misperception rate (varied between 0.1 and 0.5 vertically).

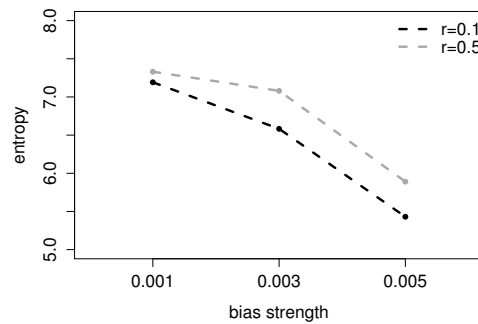


Figure 5.11: The entropies of the adaptive landscapes of five-category systems with a low misperception rate (black line) and a high misperception rate (grey line) as a function of bias strength.

(5.2) SOUNDS SYSTEMS AND EQUILIBRIA:

The simulated systems settle into equilibria at a wide range of parameter settings. The exact nature of these equilibria is determined by the adaptive landscape.

(5.3) COMPLEXITY AND THE ADAPTIVE LANDSCAPE:

The complexity of a sound system – as measured by inventory size – directly influences the adaptive landscape. More complex systems are less predictable with respect to the effects of phonetic biases.

(5.4) BIAS STRENGTH AND THE ADAPTIVE LANDSCAPE:

Bias strength acts against complexity: the stronger a phonetic bias, the more resistant it will be to the entropy-increasing effects of complexity.

(5.5) MISPERCEPTION RATE AND THE ADAPTIVE LANDSCAPE:

Misperception rate and bias strength are antagonistic forces: the influence of phonetic biases becomes less visible as misperception rate is increased.

In the rest of this chapter, I use these results to re-evaluate the issues raised at the beginning of this thesis, and propose a partial solution to the actuation problem.

5.3 WHY SOUNDS DON'T CHANGE

After a lengthy and rather technical argument based on simulation results, it is time to revisit the main question of this thesis. This question is summed up rather succinctly in the title of this section. In the interest of accuracy, perhaps a further clause should be added to this question: 'why sounds don't change when phonetic pressures predict that they should'. The answer to this question has two main components. First, the previous section has shown that when sound categories are investigated in the context of a sound system, the structure of the adaptive landscape becomes quite elaborate. This is important: Section 5.1 argued that the main reason why bias-based models fail to account for cases where sound change does not take place is that their approach to the adaptive landscape is overly simplistic. If an investigation focuses on a single

bias and a single category, it might indeed come to the problematic conclusion that the category will necessarily evolve towards the bias attractor, given that this is the only stable state that exists in the adaptive landscape. When entire sound systems are considered, this problem disappears, since not all of their stable states necessarily correspond to phonetic bias attractors. Note that this argument relies strongly on the assumption of stable states, which comprises the second component of the answer. The complexity of the landscape in itself does not necessarily imply that the system will never find a way to an optimum where a given phonetic bias is satisfied. However, what we observed in the simulations in the previous section was that the local optima stand as isolated peaks in the adaptive landscape, meaning that once a sound system has reached one of them, it will not move any further. Combining these two observations, we can give a clear answer to the original question: a given sound change may fail to take place because sound systems evolve in complex adaptive landscapes and often settle into equilibria that are not optimal in terms of phonetic biases.

This is not to say that phonetic biases have no effect on sound systems at all: bias attractors can skew the adaptive landscape towards themselves, making certain stable states more likely than others. The simulated systems in the previous section provide a clear demonstration of this claim. Although the phonetic bias was not always satisfied, the overall distribution of the outcomes was such that the biased category often ended up close to the attractor. Therefore, this account successfully captures the parallels between phonetic biases and phonological patterns. Since phonetic biases are universal, they will exert the same influence on the adaptive landscapes of all sound systems. This means that – in a statistical sense – we will likely encounter more languages satisfying a given bias than we should expect to if cross-linguistic distributions arose purely by chance.

It should now be clear that the results discussed above provide a resolution to the paradox that inspired this investigation: phonetic biases find a parallel in phonological patterns not because they cause sound change directly, but because they skew the adaptive landscape that determines stable patterns. Since phonetic biases are not the causal drive behind sound change, there is no reason to expect them to be transformed into phonological patterns in every case. Thus, contrarily to what critics of bias-based approaches have suggested (cf. Section 2.3), such models do not make false predictions with respect to the actuation

problem. This, however, crucially hinges on applying this framework to entire sound systems rather than sound categories in a vacuum.

Although this approach is successful in retaining the merits of bias-based analyses while avoiding problematic predictions with respect to the actuation of sound change, it raises two important questions. The first of these is about what kind of changes push a system towards a stable state, while the second one is about how it can move away from it once it is there. In this section, I will only tackle the first one, since it is relatively easy to answer. The second question is, in fact, a restatement of the actuation problem in terms of stable states. As such, it requires a complex answer, which will be provided in the next section.

What kind of changes lead to the emergence of a stable state? Surprisingly, the answer is that it simply does not matter. A stable state guarantees that a system will stay there, but it does not specify how it should get there. The crucial point is that unstable systems will necessarily be evanescent, and therefore not consistently observed across languages, as opposed to stable states, which will frequently be seen due to their resistance to change. This, in itself, accounts for cross-linguistic frequency distributions, and there is no need to appeal to the specific changes that lead to the distributions.

It will be useful to link this argument to existing accounts of sound change as well. The discussion above has demonstrated the emergence of stable states guided by an adaptive landscape in simulations based on a clear set of plausible assumptions about speech production and perception. Any framework that accepts these assumptions will predict the same overall tendencies. In the very least, this encompasses those mathematically explicit implementations of the nudge model (cf. Section 2.3.2) which served as the basis of the simulation architecture developed in this thesis (Pierrehumbert 2001, 2002, Wedel 2004, 2006). The predictions also carry over to broader usage-based frameworks like Bybee (2001), Phillips (2006) and Silverman (2006), even if these works differ from the present thesis in the specific way they conceive of category representations. More generally, the results have implications for other bias-based frameworks as well, such as the leap model (cf. Section 2.3.1; Ohala 1981, Blevins 2004, 2006). Note that in Section 2.3 I criticised these works for their problematic implications with respect to the actuation riddle. Importantly, the discussion above has shown that these problems can be avoided if we translate

the predictions of these models into the language of adaptive landscapes. In other words, much of what is proposed in these models may well be correct, but when claims are made about ‘the likelihood of a given sound change’, they should be reinterpreted as claims about ‘the likelihood of a given stable state’.

The arguments above have been developed through simulations in a somewhat artificial context. Although special care was taken to ground the simulations in concrete proposals about speech production and perception, it will still be useful to illustrate the predictions of the model through an example. This demonstration will also help to make the link between adaptive landscapes and actual sound systems clearer. Since the phenomenon of [u]-fronting has already been discussed to a certain extent, I will continue to use this example below. For reasons of simplicity, the discussion is restricted to high vowels.

Let us start with a scenario where there are three contrastive high vowel categories, which will be referred to as *i*, *y* and *u*. I am purposely using orthographic representations instead of IPA transcriptions, as the following discussion focuses on the predicted realisations of the vowels, which might vary depending on a number of factors. The main question is, what are the stable states predicted by the model? The answer will naturally depend on certain facts about the language under investigation. This is because the adaptive landscape is shaped not only by universal factors such as bias strength and the boundaries of phonetic space, but also by language-specific factors such as bias proportion and misperception rate. Section 4.1.2 has already explained the role of bias proportion: if a given category such as *u* exhibits a particularly high number of forms exemplifying the bias responsible for fronting (in this case, these are forms where *u* appears in a coronal context; cf. Harrington et al. 2008, 2011), it will be more strongly affected by the bias. The influence of misperception rate might be less obvious, given the somewhat artificial way it has been used in the simulations so far. Misperception rate was defined as a general parameter of the simulations, although it should probably be more appropriately interpreted as a characteristic of pairs of sounds. Specifically, misperception rate is likely to be influenced by the extent to which two categories play a role in distinguishing lexical items from each other. To use more established terminology, misperception rate should arguably reflect the ‘functional load’ of a given opposition (see Wedel 2006 and Blevins & Wedel 2009 for a similar argument): ‘hopeless’

misperceptions are more likely to occur when many lexical distinctions hinge on a given contrast.

Returning to the example of [u]-fronting, there are three possible outcomes in a language with three high vowels. First, if the bias proportion is relatively low (i.e. *u* does not often occur in a coronal context), or the misperception rate of the *u*–*y* pair is particularly high, the stable state is predicted to be one where *u* does not front. This is the state that we observe in languages like German, Hungarian or French (at least in terms of the inventory; the lexical distributions referred to above have not been tested systematically for these languages). The second outcome is as follows: if the bias proportion is relatively high while the misperception rate of the *u*–*y* pair is somewhat lower, the stable state will be one where the *u* is fronted, but distinct from *y*. These realisations could be transcribed as [ɥ] and [y], respectively. This situation is exemplified by Swedish and Norwegian, where there are two high rounded vowels in the front and central area of the vowel space (again, it is not clear whether the lexical distributions mentioned above hold in these languages). Finally, if the bias proportion is high and the misperception rate is unusually low, it is possible that the stable state is one where *u* and *y* have fallen together.⁸ Since such cases only involve two contrastive categories, this stable state is indistinguishable from situations where only two initial categories are assumed.

Having seen the possible stable states for three vowel systems, we can now discuss the two vowel case (i.e. only an *i*–*u* contrast is assumed). Since fronting in itself cannot bring about a merger between *i* and *u* (given that they differ both in backness and rounding), there are only two scenarios in this case. First, if the bias proportion is low or if the misperception rate of the *i*–*u* pair is extremely high, the stable state will be one where these two vowels are kept as far apart as possible. An example for such a language is Spanish, where *u* is realised as a back rounded vowel. Conversely, if the bias proportion is higher and/or the

8. Note that this is a situation that the model used in this thesis cannot adequately represent. While it is possible for two category distributions to occupy the same area in phonetic space, the number of categories is fixed for the entire duration of the simulations, which means that mergers and splits cannot occur. This situation could be addressed by allowing the model to merge category representations that are sufficiently close to each other. However, this thesis does not explore this possibility, given that it would greatly increase the complexity of the simulations, making them much less tractable.

misperception rate is relatively low, the stable state will be one where the *u* is fronted (as in English or in Japanese).

Before moving on to the next section, it will be useful to briefly clarify some aspects of this example. First of all, this is meant as an illustration of how the notion of stable states and adaptive landscapes applies to a concrete example, not a substantive analysis of the phenomenon. I am not attempting to derive the observed language types with respect to [u]-fronting from facts about lexical distributions (although I believe that such a project would be worth pursuing), merely fleshing out certain broad predictions of the model. It is quite likely that the case of [u]-fronting is complicated by other factors, which would also need to be incorporated into a full analysis. Second, I have been careful to consistently use the term ‘stable states’ rather than ‘sound change’ when referring to the predicted configurations of high vowels. In line with the discussion above, the claim is not that fronting will be more likely to occur as a sound change under certain conditions, but that a random language observed at a given time is more likely to be in a stable state with a fronted [u] if the adaptive landscape favours such a configuration.

5.4 WHY SOUNDS CHANGE

The previous section identified an important question related to the approach outlined above: how can sound systems escape the pull of stable states? The finding that bias-based models predict stability rather than change was presented as a crucial step forward in a debate where such models have been criticised for predicting too much change. Although this is a legitimate answer to the question of why sound change does not occur, it leads to a different, but equally worrying issue. If all sound systems evolve towards equilibria, and they stop changing once an equilibrium has been reached, why do we see any changes at all? According to the account presented so far, every sound system should eventually become stuck in a stable state. However, exactly the opposite situation seems to hold: as far as I am aware, no known sound system is completely immune to change. It appears, then, that critics of bias-based models have correctly identified the actuation problem as their weak point, but they have been focusing on the wrong issue. The problem is not that bias-based models predict too much sound change, but that they predict none at all.

This section will demonstrate that this problem is merely illusory. While phonetic biases cannot be responsible for the actuation of changes on their own, the bias-based account does not rule out other factors as the sources of change. In the rest of this section, I discuss one such factor: the dynamic nature of the adaptive landscape. We will see that the problematic conclusion that sound change should never occur is an artifact that results from our strongly simplified approach to the environment in which sound systems exist.

In the discussion so far, the adaptive landscape has been treated as entirely static. The parameters of the simulations were kept constant for their entire duration, and I described the map of stable states defined by these parameters as if it was a timeless entity. However, there is every reason to assume that this approach is wrong: the adaptive landscape must change dynamically, since the factors that determine its shape cannot be conceived of as static. For instance, both misperception rate and bias proportion have been treated as essential components of the model. Since these factors depend on lexical distributions, they are necessarily sensitive to changes in patterns of lexical usage. If, for some reason, the functional load of a given opposition becomes diminished, the misperception rate of the relevant pair of categories will also be decreased. Similarly, if there is a rise in the frequency of words where a given category occurs in a biasing context, the bias proportion of the category will be increased. When the pressures that form the basis of the adaptive landscape undergo changes, it is inevitable that the landscape will be affected as well. This has important implications for sound change: any restructuring of the adaptive landscape can give rise to changes in the sound system if the locations of the stable states are altered, or if a local optimum becomes suboptimal. To put it more simply, changes in the factors that determine the adaptive landscape may lead to sound change.

Perhaps an analogy will make the above argument clearer. Imagine a tennis ball dropped on a slope. It will likely keep bouncing downwards as long as it does not encounter any major obstacle. However, at some point it will necessarily be forced to stop either because it reaches the bottom of the slope or because it hits an obstacle that it cannot pass. Once the ball has stopped, the laws of physics guarantee that it will remain stationary, provided that the slope itself remains the same. But this provision might not necessarily hold: although the landscape in which the ball was moving around might seem static, it too can undergo

changes. For instance, an earthquake or a landslide could completely reshape the slope on which the ball was let loose, which might lead to a situation where the ball starts moving again, until it finds a new stable state. This is precisely what happens to the adaptive landscape in which sound systems are located: the factors that determine it can themselves undergo changes, setting off seismic events that might knock the sound system loose. It is quite likely that such seismic events occur with very high frequency within language (and this is where the analogy breaks down), in effect creating a ‘moving target’ for sound systems (to borrow a term from Christiansen & Chater 2008).

Although the dynamic nature of the adaptive landscape might seem entirely trivial, it has rather surprising implications for the actuation problem. Phonetic biases are traditionally seen as the causal drive behind sound change, while ‘external’ factors such as functional load are viewed as secondary variables that can only hinder or facilitate change (but do not cause it). The view advocated here turns this relationship upside down. Phonetic biases are largely static, so they cannot in themselves bring about the shifts in the adaptive landscape that lead to sound change. However, other factors such as lexical distributions can themselves undergo changes, and reshape the adaptive landscape in a way that knocks the sound system out of a stable state. Therefore – in a somewhat loose sense – external factors might have a larger role in initiating changes than phonetic biases.

This view, however, is still not entirely satisfactory. The claim that the relationship between external factors and phonetic biases is reversed in this approach betrays a dogged insistence on identifying a single source for sound changes. Perhaps the most important contribution of the discussion of adaptive landscapes so far is the discovery that it is possible to talk about multiple factors in sound change without singling out any of them as the primary driving force behind change. Phonetic biases and external factors determine the adaptive landscape together, which in turn defines the set of possible changes; this means that neither of them really causes sound change any more than the other. In fact, the best strategy in this situation might be to avoid referring to the ‘cause of sound change’ altogether.

The real difference between phonetic biases and other factors lies in the extent to which their effects are persistent. In this thesis, phonetic biases are

viewed as universal, which means that they always influence sound systems in the same way. Since the effects of phonetic biases are consistent, there is always a possibility that a sound system will arise that satisfies a given bias. For instance, the functional load of an [i]–[u] contrast in a given language might be high for many centuries, but as soon as it starts falling, there is a good chance that the effects of a bias towards fronting will soon become visible. In this sense, phonetic biases are like patient predators: they lie in wait for as long as they have to, but when an opportunity finally arises, they are prepared to take it. On the other hand, the effects of lexical factors are completely accidental. From the point of view of a sound system, we have no more reason to expect a decrease in the functional load of a contrast than an increase. The contrast between phonetic biases and other factors explains both the cross-linguistic frequency of certain patterns and the large amount of variation observed among the languages of the world. The stable states in the adaptive landscape consistently show the influence of phonetic biases (which leads to parallels between phonetic and phonological patterns), but whether a system ends up in a position where it satisfies a given bias depends to a large extent on a combination of accidental factors (which creates cross-linguistic variation).

In the rest of this section, I provide a more detailed discussion of a small subset of the factors that can bring about shifts in the adaptive landscape. I show how the factors themselves might undergo large-scale shifts and – whenever possible – provide examples from the literature supporting the claim that they have an effect on sound change. The main focus is on misperception rate and bias proportion, but I also briefly mention other factors such as the emergence and the loss of sound categories and individual differences in production and perception.

MISPERCEPTION RATE This is one of the most discussed external factors in sound change, although such discussions typically focus not on misperception rate itself but the related notion of functional load. Chapter 3 has already given a summary of the role of misperception in contrast maintenance, which will not be repeated here. In the paragraphs below, I focus on the questions of how misperception rate can change, and whether there is evidence for a connection between misperception rate and sound change.

As it has been noted above, misperception rate likely depends on the functional load of a given contrast. To put it very simply, if two categories distinguish a high number of minimal pairs or near-minimal pairs, a large amount of overlap between them is likely to cause a lot of misperception (the nature and the exact role of this type of misperception has already been discussed in Section 3.5). Conversely, if the contrast does not play a major role in distinguishing lexical items, misperceptions will be less likely even if the two categories overlap with each other. This means that the question of how misperception rate can change is really about how the number of minimal and near-minimal pairs can increase or decrease.

The most straightforward (and, for our purposes, least significant) source of such fluctuations is simply through random changes in the way individual lexical items are used. This could, of course, include a vast number of completely haphazard shifts (like the decline in the use of once fashionable words like *groovy*, and the rise of new expressions like *LOL*), and is therefore unlikely to have a systematic effect on misperception rate. However, given enough time, it is perfectly possible that such random shifts could eventually boost or suppress the misperception rate of a given pair of sound categories. A more systematic source of changes in misperception rate is lexical borrowing. For example, the categories [f] and [v] were in complementary distribution in Old English, with [v] appearing between voiced sounds and [f] everywhere else. For all intents and purposes, the functional load of this pair of sounds can be regarded as zero. In later periods, a relatively large number of lexical items with initial [v] (and other combinations that previously did not occur in English) was borrowed into English from French and Latin, which resulted in a significant increase in the functional load of the [f]–[v] pair. Another, even more systematic source of fluctuations in misperception rate is morphology. If a particular pair of sounds come to be systematically reused in affixes with different morphological functions, its functional load can undergo a significant increase – and if these suffixes are lost, the functional load of the contrast will fall too.

One of the most important sources of shifts in misperception rate is, in fact, sound change itself. When a given pair of sounds undergo a merger, new minimal and near-minimal pairs may be created, increasing the functional load of contrasts between other categories in the words affected by the merger. To

give an example, pairs of words like *thin* and *fun* have become minimal pairs through the merger of [θ] and [f] in certain varieties of English, which boosts the functional load of the vocalic opposition seen in these words. Conversely, when a sound category undergoes a split, former minimal pairs might become more distinct, decreasing the functional load of contrasts involving other sounds within the affected words. For instance, pairs of words like *tool* and *tomb* are no longer minimal pairs in varieties of English where [u]-fronting is inhibited before [l] but not elsewhere (cf. Labov 1994), and this reduces the functional load of the consonantal contrast exemplified by them.⁹

A particularly illuminating example of such shifts is provided by the phenomenon of secondary splits (Hoenigswald 1960; also referred to as ‘transphonologisation’ in Kirby 2010). In this case, two contextual variants of a sound category created through a split become contrastive after the conditioning environments are merged. For instance, Hyman (1976) and Kirby (2010, in press) describe a two-step development that has taken place in a number of southeast Asian languages like Pekinese. Syllables like [pá] and [bá] with a high tone first split into [pá] and [bǎ] (reflecting the influence of a phonetic bias creating F0 differences between voiced and voiceless pairs). In the second stage, the initial voicing contrast was lost, making the opposition between high and rising tones in pairs like [pá] and [pǎ] contrastive. What is particularly important in this case is the subtle interplay between the misperception rates of the initial consonant pairs and the tonal categories: the tonal split reduced the functional load of the contrast in the initial consonants, which subsequently merged, increasing the functional load of the new tonal contrast. Kirby (2010) and Hyman (1976) both note that this is by no means an exceptional scenario: ‘in many instances, phonologization of one feature is accompanied by dephonologization of another’ (Kirby 2010: p. 15). It is almost as if the merger was a consequence of the reduction of the functional load of the original contrast – which is exactly what the present account would predict (a similar analysis is presented in Kirby 2010). Therefore, secondary splits seem to support the idea that changes in the lexical factors shaping the adaptive landscape can lead to sound change.

9. In this section, the term minimal pair refers to words which differ in a single sound at the level of surface realisations. Thus, *tool* [tul] and *tomb* [tʌm] are not considered minimal pairs even if the contrast between [u] and [ʌ] is predictable. The reason for restricting the definition to surface realisations is that this level is the most relevant to misperception.

While the typical patterns of development seen in secondary splits provide strong intuitive support for the role of misperception rate in sound change, there are two recent studies by Wedel et al. (submitted, in press) that argue the same point in a statistically more rigorous way. Wedel and colleagues have looked at pairs of merged and unmerged vowels in phonetically annotated corpora from eight different languages, and compared these two sets in terms of their functional load (and a number of other factors). Both studies have found a significant relationship between functional load (as measured by the number of minimal pairs between two categories) and the likelihood of mergers: on average, the merged vowel pairs have a lower functional load than the unmerged ones (or at least likely had a lower functional load before the merger took place, even though this is difficult to verify on the basis of present-day corpora). Given the link between misperception rate and functional load, this result can be considered as further evidence for the role of misperception rate in sound change.

BIAS PROPORTION The following discussion of bias proportion is necessarily much shorter than that of misperception rate, partly because the potential sources of shifts in bias proportion are essentially identical to those for misperception rate, and partly because there is much less support for the claim that bias proportion influences sound change. Thus, similarly to the case of misperception rate, changes in bias proportion can result from accidental shifts in patterns of lexical usage, extensive borrowing, morphologisation and demorphologisation and sound changes affecting the environments in which the relevant sounds occur. As for the supporting evidence, the next chapter presents a detailed look at how bias proportion can influence splits relying on data from a cross-linguistic production experiment. Since the results of the experiment are clearly in favour of the claim that bias proportion has visible effects on sound systems, there is no need to present a detailed discussion in this section.

There is, however, a series of studies by Jonathan Harrington and colleagues (see e.g. Harrington 2007, Harrington et al. 2008, 2011) that are worth mentioning here. As it has been noted above, Harrington (2007) suggests that the ubiquity of [u]-fronting among various dialects of English might be explained by the high frequency of coronal sounds among the contexts in which [u] can

appear. Based on frequency counts from the CELEX lexical database, Harrington (2007) suggests that [u] is preceded by coronals over 70% of the time (although it is not clear whether this calculation relies on type or token frequencies). A more general investigation in Harrington et al. (2011) finds a related cross-linguistic tendency (based on data from the UCLA Phonological Segment Inventory Database; Maddieson 1984). Looking at languages that have both [i] and [u] in their vowel inventories, it appears that consonants articulated with the tip of the tongue make up a significantly higher proportion of the consonant inventories of these languages than consonants with other active articulators. While this evidence does not necessarily mean that such languages also show a higher bias proportion with respect to [u]-fronting, it is a promising first step towards such a result. Harrington et al. (2011) speculate that the observed tendency might at least partly explain why [u]-fronting is a frequent occurrence among the languages of the world, while other phenomena like [i]-backing are rarely found. These empirical investigations provide support for the claim that bias proportion influences sound systems, even if this support is somewhat weak due to the nature of the evidence they examine.

THE EMERGENCE AND LOSS OF SOUND CATEGORIES The simulations in Section 5.2.3 demonstrated that the number of categories in a given sound system is one of the main determinants of the adaptive landscape. While the computational model in this thesis cannot adequately represent changes in the number of categories, there is no doubt that such changes do occur in natural languages. The fact that the simulations cannot replicate such effects in their present form does not mean that we cannot infer any relevant predictions from them. It is intuitively clear that the addition or the loss of categories should bring about changes in the adaptive landscape, and that these changes can have visible effects on sound systems.

As an illustrative example, consider a set of developments that took place in a number of innovative Canadian English dialects, collectively referred to as the Canadian Shift. According to Clarke et al. (1995), all short front vowels have undergone lowering in these dialects (i.e. [ɪ] → [ɛ] → [æ] → [a]), with some additional backing in the TRAP lexical set and a certain degree of lowering and centring in the STRUT lexical set. Clarke et al. (1995) argue that this followed the

merger of the lexical sets COT and CAUGHT, which left the low central region of the vowel space (previously occupied by the COT vowel) empty, triggering a drag chain. Therefore, it appears that a large set of changes can be traced to the loss of a single contrast (note that Clarke et al. 1995 also make a connection between this merger and a somewhat similar pattern of shifting in California).

The Canadian Shift illustrates a situation where it is the loss of a contrast that triggers changes – but acquiring a new contrast could have similarly dramatic effects. One way such a new contrast could emerge is through a secondary split. Secondary splits have already been exemplified above in the context of misperception rate. Another example comes from the Middle English lengthening of short vowels in open syllables (as described in Labov 1994: p. 332). Since long [a:] had previously undergone backing, the lengthening of short [a] (and the subsequent loss of its conditioning environment through apocope) created a new sound category. Although it is not clear whether the emergence of this category triggered any changes on its own, the new long vowel in words like *name* and *made* certainly had an important role in determining the adaptive landscape navigated by the Great Vowel Shift in later periods. Another potential source of new categories is borrowing, which, however, will not be described in detail here. It will be sufficient to note that a new contrast created through borrowing could have essentially the same effect as a contrast emerging through secondary splits.

INDIVIDUAL DIFFERENCES A relatively new line of research within historical linguistics focuses on individual differences in production and perception in the hope of finding clues to the actuation problem. The rationale behind this approach is that phonetic patterns that are checked by other pressures in the speech community might be amplified in certain individuals, potentially leading to sound change under the appropriate social conditions. Baker et al.'s (2011) study of *s*-retraction in American English has already been mentioned in Section 2.5. Their main finding is that there is a large amount of variability in the extent to which specific individuals exhibit the phonetic bias responsible for retraction. As it has been noted in Section 2.5, Baker et al. (2011) suggest that patterns like *s*-retraction might lead to large-scale sound changes if speakers with an extreme production pattern happen to have a high amount of social influence. This is

because a listener might (i) misinterpret a purely phonetic pattern as a target for production and (ii) imitate this target in their own production due to the social influence of the speaker. Yu (in press) presents a slightly different account, where individual differences lie in perception: certain listeners are better at ‘undoing’ the effects of phonetic biases through perceptual compensation, while others seem to perform worse at the same task. Interestingly, the ability to compensate for the effects of biases correlates significantly with a set of personality traits (notably the Autism Quotient, or AQ): speakers who compensate less also tend to have the personality profiles of leaders in sound change.

Although the accounts summarised above differ from the present approach in many of their assumptions, the general idea of individual differences in production and perception can be translated into the framework in this thesis as well. For instance, the differences reported by Baker et al. (2011) find a straightforward expression in the notion of bias strength. While the simulations in this and the previous chapter did not explore the possibility that the agents might differ in their parameter settings, we can still speculate about the potential effects of such differences. Having two agents with different properties in terms of production and perception breaks the unity of the adaptive landscape: the pressures that influence the development of the sound system will not be the same for the agents. Unfortunately, the precise influence of this type of heterogeneity cannot be established without rigorous testing through simulations and experimental methods. However, it is likely that the existence of interpersonal variability will in itself keep the sound system in flux, especially if the proportions of speakers with different individual patterns are also changing. Therefore, although we might not be sure how exactly the existence of individual differences affects the location of stable states, the overall effects will be the same as for the other factors discussed above: the stability predicted by the bias-based model can be broken by individuals with different patterns.

There are likely many more factors that could have a similar influence on the adaptive landscape, but this short list will suffice for our present purposes. Before presenting the conclusions of this section, I would like to highlight a theme that came up repeatedly in the discussion above. In analysing the potential influence of different factors, it was often noted that changes can be

triggered not only by shifts at different levels such as the morphology and the lexicon, but even within the sound system itself. Secondary splits and changes that occur in reaction to mergers like the Canadian Shift are particularly clear illustrations of this point. This is perfectly in line both with the simulation results in this chapter and with the verbal arguments about the systemic view of sound change that have been put forward in this section and others. Although it is possible to describe changes affecting specific sound categories as isolated events (e.g. [æ] > [a]), such descriptions will often come short of capturing the rich set of interactions that take place within the sound system. Splits, mergers and shifts will often have far-reaching consequences for the development of a sound system, either by creating non-optimal configurations or by reshaping the adaptive landscape in less predictable ways (e.g. by changing the functional loads of different oppositions). Therefore, a fully explanatory account of sound change will focus not on individual categories, but on the sound system as a whole. If such a view is adopted, changes that set off chain reactions within the sound system will be seen as natural pathways for sound change.

To sum up, the bias-based view of sound change can capture both of the main types of development seen in the evolution sound systems: stasis and change. Stasis is accounted for by the discovery of stable states within the adaptive landscape. Sound change, on the other hand, is predicted to occur when the adaptive landscape itself undergoes restructuring under the influence of external factors. The main significance of these findings is that they provide a satisfactory solution to the actuation problem, by identifying a mechanism of sound change that does not predict too much or too little sound change (cf. (2.1) in Section 2.2). Moreover, the solution proposed here has two further important advantages. First, it can account for the parallels seen between phonetic biases and more robust patterns by suggesting that biases can skew the adaptive landscape, making the emergence of certain stable states statistically more likely. Second, it clarifies the role of external factors in sound change by suggesting that they are not secondary variables that can only inhibit or facilitate already existing tendencies, but fundamental forces that can in themselves lead to change. I have also provided a brief review of some of these factors: misperception rate (or functional load), bias proportion, the emergence and loss of sound categories and individual differences in production and perception.

We have seen how each of these can change the adaptive landscape and thereby contribute to the actuation of sound change. However, one external factor is still missing from this list, despite its widespread recognition as a basic element of sound change: the social aspect of language. The last section of this chapter addresses this omission by relating the arguments presented so far to sociolinguistic accounts of sound change.

5.5 THE SOCIAL ASPECTS OF SOUND CHANGE

There is a rich body of evidence showing that sound changes in progress tend to be conditioned by social variables such as age, gender, social class, and finer properties of social networks (see e.g. Labov 1994, 2001, Milroy & Milroy 1985). For example, Labov (2001: ch. 5) demonstrates that the fronting of the vowel in the MOUTH lexical set in Philadelphia is highly sensitive to social factors. In the speech community under investigation, a range of phonetic values are observed between [æo] and [e:o]. The choice of a given variant is conditioned both by age and social class. There is a strong negative correlation between age and the extent of fronting (i.e. younger speakers produce more fronted variants). A more complex non-linear correlation is observed between social class and fronting. Working class speakers produce generally more fronted variants with a slight increase from lower working class towards upper working class. The extent of fronting falls sharply between working class and middle class, and shows a steady decline as we move from lower middle class to upper class. Similar relationships have been demonstrated for a wide range of different changes from numerous languages.

Although social factors are clearly an integral part of sound change, so far I have made no effort to link the predictions of the approach pursued in this thesis to sociolinguistics. This is not a problem in itself: even if social factors cannot be omitted from a comprehensive account of sound change, this does not mean that all aspects of sound change are equally dependent on them. However, the present account has a number of non-trivial implications for sociolinguistic approaches to sound change that should be discussed in some detail. Specifically, the view presented in this thesis implies that not all changes behave in the same way with respect to social factors. Two different types of change are predicted:

convergent changes, where a whole speech community is moving towards a given pattern and *divergent changes*, where different parts of the same speech community are evolving in different directions. Since the focus of this thesis is not on the social aspects of sound change, the discussion below will necessarily be brief and speculative. Nevertheless, the arguments presented here could serve as important pointers for future research.

As it has been noted in Section 2.5, sociolinguistic approaches propose that the actuation problem can only be solved if we focus on the propagation of sound change (see e.g. Weinreich et al. 1968, Milroy 1992). Accounts based solely on phonetic biases predict that sound change will overapply. However, the process through which a given change spreads through a community provides an extra layer of control that can potentially check the effects of blind phonetic variation. The view of sound change developed in this thesis challenges this assumption, inasmuch as it removes the need to include a sociolinguistic component in the solution to the actuation riddle. When the bias-based approach is applied to sound systems rather than individual categories, both stasis and sound change are predicted to occur. Thus, this account provides a solution to the actuation problem without referring to social dynamics.

The claim here is not that social factors have no role in determining sound change, only that they are not solely responsible for the actuation of changes. The question, then, is how such factors interact with the system-based view presented above. In this section I review two possible scenarios, which only differ in the way the adaptive landscape is correlated with social factors. In both cases I will assume that the social structure of the speech community is complex. Members of the community can vary along different social dimensions and they form networks that determine the extent to which different parts of the community are connected to each other. The difference is that in one case the factors that define the adaptive landscape are identical for all speakers, while in the other case they may differ across sub-groups within the speech community. As we will see, sound change is predicted to proceed along rather different lines in these two hypothetical scenarios.

Let us first look at the case where the adaptive landscape is uniform across the entire speech community. This means that all speakers share the same sound system and the same patterns of lexical usage. Moreover, there are no

particularly striking patterns of individual variation – or if there are, these do not correlate with social structures. Since the adaptive landscape is identical for all speakers, the stable states are also shared. Therefore, if a sound change is underway in the community, its direction will be the same for all the speakers. For this reason, such changes will be termed ‘convergent’. The reason why all speakers evolve along the same lines is that the source of sound change in the model advocated in this thesis is language use. That is, changes emerge from speech interactions among speakers through the application of phonetic biases, misperception and other factors. Since these factors are shared among all speakers, their effects will be largely the same across the speech community. Indeed, this is what we saw in the multi-agent simulations in Section 4.3.

Importantly, the fact that the entire speech community is evolving towards the same stable state does not necessarily mean that it will be homogeneous with respect to the speech patterns of individuals. The simulations in Section 4.3 are somewhat misleading in this respect, since they illustrate changes in a community without any social structure. If a more realistic network structure was imposed on the simulations, it is likely that a certain amount of diversification would emerge within the community. This would not affect the stable states within the adaptive landscape, but the speed at which different subgroups converge towards these states would likely differ. Moreover, another small modification might also enable the model to account for the age-grading often seen in sound changes in progress. Baker (2008) discusses a simulation where the agents can differ in terms of their age, and where older speakers – while capable of changing their speech patterns – are more resistant to innovations. This model produces realistic results with respect to age grading. Thus, the assumption that speech representations become somewhat more rigid as speakers age (which is relatively uncontroversial) can lead to the appearance of age-grading even without any social factors.¹⁰

To sum up, convergent sound change may show sociolinguistic patterning as a function of social network structure and age. Note that these patterns arise automatically from language use, and not from the speakers’ desire to express their social identity through their speech. Moreover, even if there are

10. Note that the claim is not that older speakers do not change their representations, only that such changes are somewhat more pronounced in younger speakers.

differences within the speech community, the direction of the change is uniform: eventually, all the sub-groups within the community will converge to the same stable state.

Although this discussion is purely hypothetical, convergent sound change finds an important parallel in empirical studies: *changes from below* (Labov 1994). Labov provides the following description:

Changes from below are systematic changes that appear first in the vernacular, and represent the operation of internal, linguistic factors. At the outset, and through most of their development, they are completely below the level of social awareness. [...] Changes from below may be introduced by any social class, although no cases have been recorded in which the highest-status social group acts as the innovating group.

(Labov 1994: p. 78)

Perhaps the most important element of this description is the statement that changes from below are driven by internal, linguistic factors. This echoes the observation regarding convergent changes according to which they emerge from language use and are not sociolinguistically motivated (although they may show a certain amount of sociolinguistic patterning). Therefore, it appears that the model presented in this thesis correctly predicts the existence of a particular type of change that is widely observed in natural languages.

Let us now turn to the second scenario, in which different sub-groups within the speech community differ with respect to the shape of the adaptive landscape. It is not difficult to imagine how such a situation may emerge. For instance, a certain group within the community might start using a given set of content words (e.g. slang expressions), or even function words and fillers (e.g. the words *like* and *so* in some varieties of American English), which are not used by other groups. Alternatively, certain groups of speakers may have stronger contacts with other languages and dialects, which could result in group-specific patterns of borrowing. A third source of such differences is individual variation: following Baker et al.'s (2011) account, it is possible that unusual patterns of individual variation are better represented in a given sub-group. Such intergroup

differences may all contribute to the diversification of the adaptive landscape across the speech community. As a result of this diversification, different sub-groups will have different stable states, which may in turn lead to divergent changes within the same community.

This divergence will likely occur along the same lines that are traditionally observed in sociolinguistic studies (e.g. social class, social network structure). Importantly, the differences that emerge as a result of this diversification may become salient markers within the speech community (see e.g. Rácz to appear for an analysis of the conditions under which a variable can be regarded as sociolinguistically salient). Under the right circumstances, such markers may be adopted by other groups as well within the speech community, leading to a *change from above*, to use Labov's (1994) terminology. This will typically arise when one of the groups is seen as more prestigious than the other groups (e.g. when it has higher social class; Labov 1994). Thus, similarly to the case of convergent changes, divergent changes have clear parallels in empirical studies of sound change.

In sum, the model presented in this thesis seems to predict the existence of two different types of change with respect to social conditioning. Convergent changes may be influenced by social factors, but they lead the entire speech community towards the same stable states. Divergent changes, on the other hand, lead to diversification within the speech community, which can serve as the basis of further sociolinguistically motivated shifts. These two types of change find relatively clear parallels in changes from below and changes from above, respectively. Once again, I should emphasise that the discussion above is highly speculative, and should be treated with a certain amount of caution. The arguments presented above should be tested through computer simulations in a rigorous way just like the main argument of this thesis. Moreover, the parallels between the types of change suggested by Labov and the changes predicted by the model advocated here should be made more explicit. Unfortunately, the elaboration of these arguments falls outside the scope of the present thesis. However, even this brief discussion is sufficient to demonstrate that the system-based view of sound change can interface with sociolinguistics in a meaningful way.

5.6 SUMMARY

This chapter presented the main argument of this thesis, which can be summarised as follows. Bias-based models that investigate sound categories in a vacuum make false predictions with respect to sound change: phonetic biases result in deterministic shifts, failing to account for cases where no change occurs. The simulations in this chapter showed that this behaviour is an artifact of the simplified view taken in these approaches, and that a radically different picture emerges when such models take a more realistic approach to sound change. Specifically, shifting the attention from individual categories to sound systems furnishes us with a plausible solution to the actuation problem, without making it necessary to discard the main predictions of the bias-based model. Thus, when the bias-based model is applied to an entire sound system, the pressures that result from the interaction of categories and phonetic biases create a complex adaptive landscape, which determines the evolution of the system. One essential feature of this adaptive landscape is the existence of multiple local and global optima. Once an evolving system finds itself on a peak in the adaptive landscape (corresponding to an optimum), it will stop changing, even if it does not satisfy all phonetic biases. It was also shown that phonetic biases consistently skew the adaptive landscape, thereby making it statistically more likely that the effects of a given bias will be visible in a given language (even if not all languages show them). The emergence of multiple stable states in the system solves the problem of overapplication. Moreover, it was also shown that under the right circumstances the system is predicted to undergo changes. When the adaptive landscape is reshaped through significant changes in the factors defining it, the sound system will likely follow suit. Finally, I argued that the system-based view can make plausible predictions about the sociolinguistic aspects of sound change, although these predictions need to be investigated in more detail.

Since the arguments in this chapter have mostly been framed in abstract theoretical terms, the reader might wonder what their implications are for empirical approaches to sound change. As pointed out earlier, one of the main goals of this thesis is to sharpen the predictions of the bias-based approach. After all, it is difficult to provide an insightful analysis of a phenomenon when we cannot be sure what the theory actually predicts. The simulations and verbal

arguments presented in this chapter bridge the gap between theory and data by demonstrating how different factors can influence the possible outcomes of sound change in a bias-based model. Perhaps the most important implication of this investigation is that it strengthens the case for looking at ‘external’ conditioners in sound change. As Section 5.4 explained, these could include lexical factors, changes in the number of categories or individual differences. Of course, this is not to say that such factors have never been investigated: for instance, Labov (2002) suggests (similarly to what has been proposed in Section 5.4) that mergers can have an effect on chain shifts, and Wedel et al. (submitted, in press) have demonstrated that functional load may have an influence on the likelihood of mergers (these are both empirical investigations; theoretical hypotheses relating to the same issues have existed for a much longer time). However, the present thesis goes even further by claiming that shifts in the factors determining the adaptive landscape can be just as important in predicting changes as phonetic biases themselves. In other words, such factors are not secondary components of explanatory accounts of sound change, but crucial predictors in themselves.

The next chapter presents further empirical support for this view by taking a detailed look at one such factor, namely, bias proportion. I show that bias proportion makes an interesting prediction about allophonic splits. Much in the same way as in the present chapter, this prediction is derived from the main theoretical assumptions introduced in Chapter 2 (and one further assumption described in detail in Section 6.1). The chapter then discusses a cross-linguistic study of the effects of voicing on vowel length, and shows that the prediction about splits is borne out by the data. These results are crucial in that they demonstrate how the abstract arguments developed in this chapter find a direct application in the study of sound change.

LEXICAL FACTORS IN CONTEXTUAL EFFECTS

6

The main goal of this chapter is to substantiate the arguments presented in the rest of this thesis. Chapter 5 demonstrated that the system-based approach offers a plausible solution to the actuation riddle. It was shown that both stasis and sound change can be accounted for if we focus on sound systems rather than isolated categories. The main reason for the success of this approach is that sound systems are affected by a much richer set of pressures than sound categories in a vacuum, and are therefore less vulnerable to the effects of phonetic biases. I suggested that there is a wide variety of factors that could influence the evolution of a sound system, including lexical distributions, an increase or a decrease in the number of categories and individual differences. Although I briefly discussed some studies that have found such factors to have a significant influence on sound systems, I presented no systematic investigation of any one of these factors. That is to say, although the predictions of the system-based approach are clear, none of these predictions have been tested so far.

To show how the approach taken in this thesis can inform empirical investigations, I present an in-depth study of one particular prediction of the model (the prediction itself will be discussed below). This will be done in two steps: I first demonstrate how the prediction derives from the underlying theory and then test its validity through a small cross-linguistic study. The prediction will be shown to be supported by the data. While this finding certainly strengthens the main argument of this thesis, the primary goal of this investigation is not to adduce uncontroversial evidence for the system-based approach. Indeed, such an effort is well outside the scope of this thesis given the wide-range of factors that would need to be investigated. This relatively small study serves only to demonstrate that the predictions of the model can be explored in a systematic

way and that investigations along these lines promise to contribute significantly to our understanding of sound change.

Before discussing the prediction itself, it should be noted that the study described in this chapter will focus on individual categories. This may appear to go against the idea developed in the previous chapter according to which sound change can be discussed more successfully when the focus is on sound systems. As it turns out, this contradiction is only apparent. It is true that individual instances of sound change should not be analysed in a vacuum, as this might result in implausible conclusions with respect to the actuation problem. However, this chapter focuses not on specific cases of sound change, but on a general factor that may contribute to the likelihood of change. Although such factors may sometimes manifest themselves at the level of the sound system, there is no reason to assume that they cannot be specific to individual categories. In fact, the phonetic biases examined in the previous chapter only affected a single category, but this did not defeat the purpose of the system-based approach. Therefore, we are fully justified in focusing on category-specific effects even if such factors will likely be only part of the story when it comes to specific instances of change.

Let us now turn to the prediction that serves as the basis of this chapter. This prediction is about the relationship between lexical distributions and the strength of contextual effects. The lexical factor in the focus of this investigation is bias proportion. In Section 4.1.2, I noted that certain biases may only apply to a subset of the forms exemplifying a given category. Whether a form belongs to this subset or not is determined by the phonetic environments embodied in the form. One example for such a bias is the fronting of high back vowels next to coronal consonants (cf. Harrington et al. 2008). This bias affects a subset of all the forms containing [u], namely those in which it is preceded or followed by [t], [d], [tʃ], [l], [n], [j] or any other coronal consonant; forms in which this condition is not met do not exhibit this type of fronting. I suggested that the proportion of eligible forms within a category partially determines how much influence a given bias can have on it. For instance, a category where only 10 per cent of all forms are eligible is going to be much less affected by the bias than one where 90 per cent of the forms are eligible. It is this quantity that I referred to as bias proportion.

There is another important prediction tied to bias proportion, and this is what the present chapter will focus on. Consider the partitioning of the set of forms containing the vowel [u] induced by the fronting bias. The set divides into two subsets: forms that are affected by the bias and forms that are not. Importantly, the former subset will contain examples of [u] that are articulated further to the front than those in the latter one. To put it more formally, the two sub-distributions within the overall category [u] will have slightly different expected values, with the biased sub-distribution being closer to the bias attractor. The appearance of this gap between the two sub-distributions is a simple consequence of the fact that eligible items are consistently displaced towards the bias attractor, whereas ineligible items are not.

The prediction relates to the size of the gap between the sub-distributions. There are two possible ways to approach this issue depending on how one conceives of the internal structure of category representations. The first approach is exemplified by the simulations in Chapters 4 and 5, where no internal structure is assumed: production and perception are based on a single distribution representing the whole category. To return to the example of [u]-fronting, a model of this type chooses production targets for both the coronal and non-coronal sub-distributions by sampling the distribution of all forms with [u]. The only difference is that the target productions in coronal forms are subsequently displaced by the bias, while no such displacement occurs in non-coronal forms (see Section 4.1 for a detailed description of this model). Therefore, when the two sub-distributions have no representational independence, any difference in their expectation dynamics is due solely to the mechanical application of the bias. To use the terminology introduced in the previous chapter, the difference between the two contexts lies not in the underlying but in the observed distributions. In sum, this model predicts only a small gap between the two sub-distributions, whose size is a simple function of the strength of the bias.

The second approach differs from the first one in that it posits a certain degree of representational independence for the two sub-distributions. In other words, the differences between sub-distributions are already apparent at the level of underlying distributions. Let us take the example of the coronal fronting bias again. In this second type of model, production and perception are both based at least partly on the relevant sub-distributions: the coronal

sub-distribution plays a greater role in determining the target production for a coronal token than the non-coronal sub-distribution and *vice versa* for non-coronal tokens. Moreover, perception is also influenced by sub-distributions, introducing context-specific effects in categorisation (these are discussed at greater length in Section 6.1). Since the sub-distributions are independent to a certain extent, the influence of the fronting bias can accrue over many iterations in the representation of coronal forms, but not in that of non-coronal forms. This leads to an increase in the size of the gap between the two sub-distributions. To put it slightly differently, when the two sub-distributions are represented separately, differences in their expectation dynamics result not only from the mechanical application of the bias but also from the production-perception feedback loop. Therefore, this model predicts a larger gap between the two sub-distributions than the previous one.

There is good evidence from experimental studies that the second approach is more plausible than the first one: it appears that speakers routinely rely on sub-distributions both in their production and perception (these studies are reviewed in the next section). We are now in a position to return to the relationship between bias proportion and contextual effects. When we combine the model outlined in the previous chapter with the idea of independent sub-distributions, the following prediction emerges. The gap between the sub-distributions will be wider when the proportion of biased and non-biased items is balanced within the category, and narrower when either biased or non-biased items are over-represented. Since this predicted correlation is the main topic of this chapter, it is repeated below as Prediction 1:

Prediction 1 *Sub-distributions are further apart in categories with a balanced bias proportion than they are in categories with an unbalanced bias proportion.*

Unfortunately, it is difficult to discuss this prediction in detail until the necessary formalisms have been introduced. However, I will attempt to explain briefly (and in rather impressionistic terms) why such a relationship between lexical factors and the size of the gap within the category should exist. There are two main forces that determine the internal dynamics of a category: the different sets of biases that affect each sub-distribution and the internal cohesion of the category (the source of this cohesion is discussed in Section 6.3). These forces act against each other: the biases pull the sub-distributions apart, while

the internal cohesion of the category draws them closer together. Under the right circumstances, the two forces will balance each other out, leaving the sub-distributions in an arrangement where there is no pressure for them to move either apart or towards each other. The size of the gap between the sub-distributions in such an equilibrium is determined by the relative strengths of the biases and the cohesive forces acting on the category. It is at this point that lexical factors come into play. The degree of cohesion within categories is strongly affected by the bias proportion: when eligible and ineligible items are equally frequent, the sub-distributions have a high degree of independence (and therefore less cohesion), but when there is an imbalance in the frequencies, the less frequent sub-distribution loses its independence (increasing the overall cohesion within the category). As a result, categories with a balanced distribution of eligible and ineligible items show a higher degree of separation between the sub-distributions than do categories with a skewed distribution.

The rest of this chapter looks at Prediction 1 in more detail. As a first step, Section 6.1 provides justification for the assumption of partially independent sub-distributions within categories. This is necessary, since much of the work presented in the rest of the chapter takes this assumption for granted. Section 6.2 then gives a detailed description of vowel length differences before voiced and voiceless obstruents. This voicing effect is used to anchor the discussion of contextual effects in a concrete phenomenon. In Section 6.3, I present a formal model of the separation of sub-distributions within categories, and show how Prediction 1 emerges from this model. Both simulations and mathematical calculations are used to explore the behaviour of sub-distributions. Then, Section 6.4 presents a small cross-linguistic study looking at the voicing effect, which finds support for Prediction 1. Section 6.5 concludes the chapter with a brief discussion of its main points.

6.1 SUB-DISTRIBUTIONS WITHIN CATEGORIES

In the introduction to Chapter 3, I highlighted the importance of carefully stating the theoretical assumptions underlying a given computational model. The necessity for such statements derives directly from the role of computational models: they are created with the aim of validating scientific theories by linking

abstract theoretical concepts to concrete data sets. They can only fulfill this role if the underlying theory is made explicit. It was also noted that the theoretical assumptions of the model have to be supported by evidence from outside the data set to avoid circularity. The present chapter proposes to enrich the set of fundamental assumptions presented in Chapter 3 with a further assumption, namely that sound categories are made up of partially independent sub-distributions. The main goal of this section is to elaborate on this assumption and to present evidence for its validity from the phonetic literature.

When this thesis claims that speakers and listeners rely on sub-distributions, the following are meant. First, speakers store probabilistic representations of certain subsets of tokens from a given category, in much the same way as they form probabilistic representations corresponding to the categories themselves. These representations are learnt and phonetically detailed, just like the overall category representations used in the simulations in the previous chapters. Second, speakers use these sub-distributions in generating production targets. Third, listeners rely on their own stored context-specific sub-distributions in perception.

All three of these claims are investigated in this section. I first attempt to clarify the concept of sub-distributions by relating it to the phonetics/phonology divide that is often assumed in discussions of speech representations. Then I discuss the notions of learning and phonetic detail, arguing that they are both essential in order to understand the behaviour of context-specific sub-distributions. This position receives support from a review of findings from production and perception studies. Finally, I present a brief argument to the effect that sub-distributions within a given category are independent, but only within certain limits.

SUB-DISTRIBUTIONS AND PHONETICS/PHONOLOGY This thesis assumes that information about context-specific sub-distributions is already available at the level of lexical representations. That is, a given category in environment *A* will be represented differently from the same category in environment *B*. The main difference between traditional modular approaches to phonetic realisation (see e.g. Keating 1990a, Bermúdez-Otero 2007) and the present model is that in the latter one context-specific differences are not derived procedurally. As

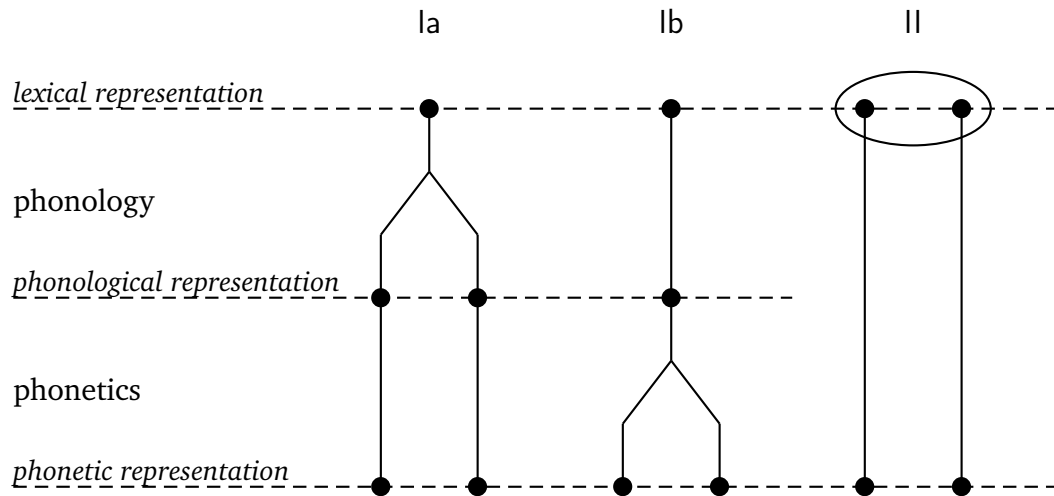


Figure 6.1: Diagrams (Ia) and (Ib): traditional modular approaches to phonetic implementation can represent contextual differences in the realisation of a category in two different ways; (Ia) is a phonological alternation, whereas (Ib) is a phonetic alternation; Diagram (II): the present model represents both phonological and phonetic alternations in the same way.

a result, it makes no sense to distinguish between phonetic and phonological effects in the present model. In what follows, I attempt to make the contrast between these two views clearer, and also to briefly assess their predictions.

Consider the example of the voicing effect. The basic observation is that vowels tend to be longer before voiced obstruents than they are before voiceless ones. This observation holds for a large number of languages from a variety of different language families (see the next section for a more detailed description). Modular approaches to phonetic implementation can represent the voicing effect on two different levels: the level of phonology and that of phonetics. This is illustrated in Figure 6.1 (based partly on Bermúdez-Otero 2007). The diagram in (Ia) shows a phonological alternation, where a single lexical representation is transformed into two different phonological representations before the application of phonetic rules. In the case of the voicing effect, this would mean that phonological rules act on discrete features to create two categorically different representations in voiced and voiceless environments (e.g. [+long] and [−long]). The diagram in (Ib) illustrates a different situation: phonology leaves the original input intact, and it is only at the level of phonetics that contextual differences emerge. Since phonetic rules create gradient output

(cf. Bermúdez-Otero 2007), a phonetic approach allows substantial overlap between vowel length distributions in voiced *versus* voiceless environments.

Some authors (e.g. Hyman 1975) suggest that this two-way distinction is reflected in cross-linguistic data as well: some languages (e.g. English) have phonologised the voicing effect and therefore show a much more pronounced length difference across voiced and voiceless contexts than others (e.g. French), where phonologisation has not taken place. This suggests that a clear binary distinction should be observed between languages where the pattern is phonological and those where it is phonetic. As it turns out, this prediction is not supported by the data. There is gradient cross-linguistic variation in the size of the voicing effect that does not correspond to a phonology/phonetics division in any straightforward way (see Section 6.2). More generally, the review of production patterns presented later in this section shows that such a simple division cannot account for the gradient patterns of cross-linguistic variation characteristic of a wide array of contextual effects. If, however, the phonology/phonetics distinction does not correspond to any systematic differences in the realisation of contextual effects, it becomes unclear why such a distinction should be assumed at all.

These considerations suggest that the simple model illustrated in diagram (II) of Figure 6.1 might be sufficient to account for contextual differences in the realisation of a category (see also Pierrehumbert 2001, 2002 for a similar view). The diagram shows that contextual differences in this type of model emerge simply through linking different sets of forms (the dots) within a category (the ellipsis) to different areas of phonetic space. Since contextual differences do not arise procedurally in this model, there is no branching of representations as in diagrams (Ia) and (Ib). Different groups of forms within a category are already assigned to separate sub-distributions at the level of the lexicon. I do not intend to suggest that every type of sound pattern can be accounted for in this simple framework – and the evidence presented in this section is certainly not sufficient to back such a claim. However, the simple contextual effects investigated in the present chapter do not seem to motivate a multi-levelled model. In the absence of good arguments for such models, the simpler approach advocated in this thesis is preferable.

LEARNING AND PHONETIC DETAIL The foregoing discussion has the following implications with regard to learning and phonetic detail. Sub-distributions have to be learnt in the course of language acquisition, or otherwise there could be no cross-linguistic differences in the size of contextual effects. Moreover, sub-distributions have to be capable of representing phonetically detailed information, as this is necessary to capture gradient differences in the size of contextual effects across languages. In what follows, I give a brief overview of previous approaches to these questions, especially where they differ from the views advocated here.

Learning and phonetic detail tend to be intimately tied together in discussions of contextual effects. Some amount of each clearly has to be allowed in any model of speech: a model without learning would not be able to represent even categorical differences across languages, and a model without phonetic detail would be incapable of accounting for even the smallest amount of gradient variation. However, it has been suggested that there is a crucial disjunction between learning and phonetic detail: learning only occurs at the level of phonology and gradience at the level of phonetics. Models relying on these assumptions were especially popular in the early days of generative linguistics. For instance, Chomsky & Halle (1968) suggest that phonological rules operating on discrete representations account for all language-specific aspects of speech, while phonetics operates in a completely universal and mechanical fashion to translate such representations into continuous signals (see Keating 1984 for a thorough review of this position).¹ This view is not compatible with the position taken in this thesis: I propose that lexical representations are directly mapped to phonetic space, which entails that learning and phonetic detail coexist at the same level.

While the disjunctive approach to learning and phonetic detail still has some proponents (Hale et al. 2007), it has become somewhat of a minority view over the last few decades. There is a long research tradition in phonetics showing that phonetic detail is, in fact, learnt and language-specific (Keating 1984, 1990a,

1. Chomsky & Halle (1968) also propose phonetic detail rules as part of grammar, and these can introduce a certain amount of gradience before the application of universal phonetics. However, they are quite explicit about relegating most contextual effects including coarticulation to the level of universal phonetics (cf. Keating 1984).

Kingston & Diehl 1994), and this view is gaining wide-spread acceptance within the phonological community as well (see e.g. Morén 2003, Bermúdez-Otero 2007, Blaho 2008, Boersma & Hamann 2008). This line of research is perfectly compatible with the present approach in terms of learning and phonetic detail, although it is clearly distinct from it in another area, namely the assumption of multi-levelled representations.

EVIDENCE FOR SUB-DISTRIBUTIONS Let us now turn to the evidence for the claim that learnt and phonetically-detailed sub-distributions are used in speech. I show that (i) contextual effects show cross-linguistic variation and that (ii) this variation is gradient in a way that does not support a simple division between phonetics and phonology. I start by discussing evidence from production studies. This is followed by a brief review of perception studies.

Cross-linguistic differences in production have been investigated extensively, and there is a wealth of evidence suggesting that context-specific effects – such as coarticulation – are learnt and language-specific. Some of this evidence is reviewed in Pierrehumbert (1999), who makes the following conclusion with regard to cross-linguistic variability:

[...] I believe that every thorough study which has looked for a difference between two languages in details of phonetic implementation has found one. These differences concern both detailed outcomes for analogous phonemes in the most analogous available positions, and — to an even greater extent — *principles of allophonic variation in context*.

(Pierrehumbert 1999 p. 114; emphasis mine)

Of particular importance for our present purposes is the statement that allophonic variation also shows differences across languages, which is clearly in line with the view that knowledge of sub-distributions is acquired through learning.

In what follows, I present a few specific examples for such cross-linguistic variation. Studies of vowel nasalisation before nasal consonants have found that the temporal and spatial extent of nasalisation can be vastly different across languages. For instance, Beddor & Krakow (1999) claim that while

nasalisation is present for 80 per cent of the duration of pre-nasal vowels in English,² Thai speakers only exhibit 45 per cent nasalisation in the same context. Montagu (2007) reports even lower numbers for Parisian French, where the temporal extent of nasalisation is between 24–45 per cent, depending on vowel quality. Note that these findings exhibit gradient variation, inasmuch as the average percentages are distinct across the three languages. Such differences cannot simply be explained by assuming a distinction between phonetic and phonological patterns.

Patterns of vowel-to-vowel coarticulation have also been found to vary quite substantially across languages. Beddor et al. (2002) show that there are systematic differences between English and Shona both in the production and the perception of vowel-to-vowel coarticulation. One example is the observation that carryover coarticulatory effects between neighbouring vowels are stronger in English than in Shona. Another relevant investigation of vowel-to-vowel coarticulation is presented in Choi & Keating (1991), who demonstrate that English, Polish, Bulgarian and Russian each show different degrees of coarticulation. A summary of their cross-linguistic findings is presented in Figure 6.2 (taken from Choi & Keating 1991: p. 83). Similarly to the case of vowel nasalisation, the cross-linguistic variation observed here is gradient: the data do not seem to support a binary distinction between phonological *versus* phonetic patterns. Note that some of these differences might be due to the fact that Polish, Bulgarian and Russian all have palatalised consonants, which could block the interactions between neighbouring vowels. However, this still does not explain the differences among these three languages.

Sibilant-vowel interactions also exhibit a considerable amount of variation across languages. Hoole et al. (1993) present the results of an experiment using both articulatory and acoustic measures. They show that German, French and English are all different both in their context-independent realisations of [s] and [ʃ] and the effects that the neighbouring vowels have on these sibilants' articulatory and acoustic properties. Moreover, they find significant differences even between different dialects of the same languages. One finding

2. Based on a study investigating nasals in the speech of speakers from Michigan (Tanowitz & Beddor 1997). Some other studies have found different percentages: for example, Solé (1992) reports 100 per cent nasalisation in American English.

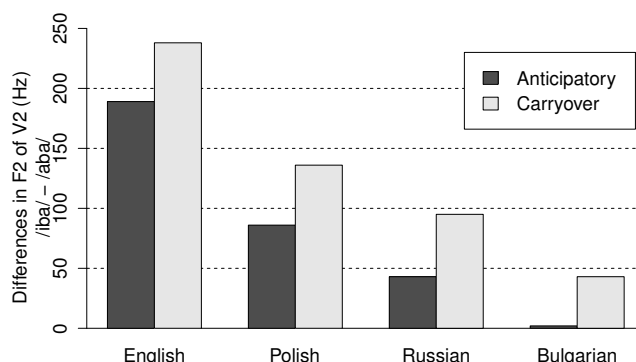


Figure 6.2: *The degrees of anticipatory and carryover vowel-to-vowel coarticulation in English, Polish, Russian and Bulgarian. The measurements are based on differences in F2 between the first vowels in /abi/ vs. /aba/ (anticipatory coarticulation) and the second vowels in /iba/ vs. /aba/ (carryover coarticulation).*

that demonstrates language-specific effects relates to the direction of coarticulation: German and English seem to favour carryover coarticulation, while French favours anticipatory coarticulation. Although Hoole et al. (1993) do not explicitly note this, a glance at their graphs and statistical results also suggests that the amount of anticipatory coarticulation in [s] is much smaller in English than in the other languages.

Let us now turn to the phenomenon of perceptual compensation. Perceptual compensation is a cognitive process, which is part of a more general phenomenon referred to as ‘perceptual constancy’ or ‘subjective constancy’ in cognitive science. Broadly speaking, perceptual compensation is the process whereby humans filter out the influence of the context in which a given object or event occurs, which allows them to see it as unchanged regardless of variation in the external conditions. There are many straightforward examples for this in the field of visual perception (see Palmer 1999 for a thorough review). For instance, it is through perceptual compensation that the reader of this thesis sees the paper that this thesis is printed on as white regardless of whether they are looking at it in a park in broad daylight or in a dimly lit office.³ Visual

3. I urge the reader to test this claim themselves, if broad daylight and a park are available. Unfortunately, readers using an electronic device to display the thesis will have to provide their own piece of white paper.

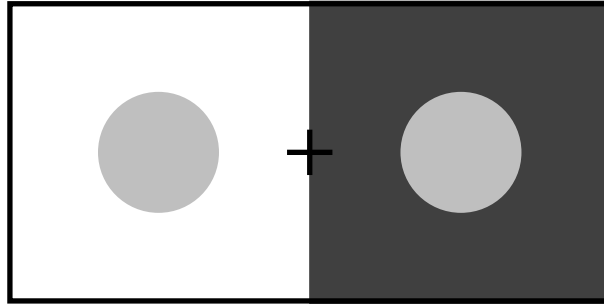


Figure 6.3: *Optical illusion illustrating perceptual compensation in visual perception. The reader should hold the paper relatively close to their eyes and fixate on the cross in the middle. The circle on the right will appear brighter, even though it is exactly the same colour as the one on the left.*

illusions provide another source of illustration for perceptual compensation. Consider the circles in Figure 6.3. They appear to be of different colours, but they are not. The source of the illusion is the context in which they appear. The viewer has implicit knowledge of the fact that everyday objects are usually brighter in a bright environment and darker in a dark environment. Building on this observation, our cognitive system attempts to control for the ambient light levels, and makes the circle on the left appear darker and that on the right lighter. This is the essence of perceptual compensation.

The very same phenomenon also occurs in speech perception. Mann & Repp (1980) show that when speakers of English are played a sound intermediate between [s] and [ʃ], they are more likely to perceive it as [s] when followed by [u], and as [ʃ] when followed by [a]. This situation is completely analogous to the case of the circles in Figure 6.3. The subject has implicit knowledge of the fact that coarticulation makes sibilants sound more [ʃ]-like before an [u] and more [s]-like before an [a] (this can be attributed to the effect of lip-rounding on the spectral centre of gravity). The perceptual system attempts to control for the effects of the phonetic environment, making the stimulus sound more [s]-like before [u] and more [ʃ]-like before [a] (in the same way that it makes the circle appear lighter in a dark environment and darker in a light environment). Crucially, perceptual compensation can only take place if the listener has knowledge of the differences in the acoustic properties of the pre-[a] and pre-[u] sub-distributions. This would not be possible if these differences came by through automatic and universal processes without

any reliance on the listener's previous experience. Therefore, the existence of perceptual compensation is evidence for learnt sub-distributions. Similar effects have been demonstrated for vowel-to-vowel coarticulation (Beddor et al. 2002), the effect of vowel height on fundamental frequency (Hombert 1978) and numerous other context-dependent phenomena (see Sonderegger & Yu 2010 for a brief overview).

Moreover, Sonderegger & Yu (2010) find that there is a close correspondence between speakers' production of different sub-distributions and their behaviour in perceptual compensation: the former can be used to predict the latter with relatively high accuracy. Specifically, phonetic details from the speakers' productions such as the means and dispersions of different sub-distributions are indicative of the amount of perceptual compensation that they display. Thus, the knowledge of contextual effects that forms the basis of perceptual compensation has to include phonetically detailed information as well. Otherwise, the observed parallelism between the amount of perceptual compensation and gradient features of production could not exist.

THE PARTIAL INDEPENDENCE OF SUB-DISTRIBUTIONS At the beginning of this section, the sub-distributions contained within a given category were claimed to be only partially independent. This means that while speakers store information about specific sub-distributions, their behaviour in production and perception can be affected by other sub-distributions as well. For instance, when the speaker produces a vowel in a voiced environment, the sub-distribution corresponding to vowels in a voiceless environment might also play a role in choosing a production target. This assumption is necessary in order to account for the coherence of sound categories: there seems to be a limit to the extent that a single category can be realised differently as a function of the context it appears in. In other words, the sub-distributions corresponding to different environments do not typically drift apart in phonetic space *ad infinitum* (cf. Bermúdez-Otero 2007). The small-scale computational experiment described in Section 3.2 supports this claim. It was shown that an unsupervised clustering algorithm can approximate categories with surprising accuracy. Since the clustering algorithm relies on the assumption that the categories are not

discontinuous, this close approximation is only possible if the categories show a considerable degree of cohesion.

In sum, the results of existing production and perception studies argue strongly for the view that context-specific effects in the realisation of sound categories are learnt and phonetically detailed. Moreover, the data reviewed in this section do not support a categorical distinction between phonological and phonetic patterns. Therefore, this thesis is justified in assuming that contextual effects are encoded at the level of lexical representations in the form of separate sub-distributions. Moreover, I have also presented some arguments to the extent that these sub-distributions are only partially independent. These assumptions are substantiated in the mathematically explicit model presented in Section 6.3.

6.2 THE VOICING EFFECT

In the rest of this chapter, I use the example of vowel length differences across voiced and voiceless contexts to give more substance to the theoretical concepts under discussion. The present section provides some background to this phenomenon and attempts to deal with a number of potential problems so that the central argument of the chapter can be presented without major digressions. I first give a brief description of the phenomenon, clarifying some of its more controversial aspects and reviewing potential explanations for its occurrence. The second half of the section explores the cross-linguistic variability in the size of the voicing effect, and lists some factors that contribute to this variability.

A difference in the average lengths of vowels preceding obstruents that are traditionally spelt or transliterated using the letters ⟨b⟩/⟨d⟩/⟨g⟩ as opposed to ⟨p⟩/⟨t⟩/⟨k⟩⁴ has been observed in a wide variety of languages: Modern Standard Arabic (Hussein 1994), Assamese (Maddieson 1977), Bengali (Maddieson 1977), English (Peterson & Lehiste 1960, Chen 1970, Mack 1982, Laeuffer 1992), French (Mack 1982, Laeuffer 1992), Hindi (Maddieson & Gandour 1976), Lithuanian (Campos-Astorkiza 2007), Spanish (Zimmerman & Sapon 1958),

4. Although this effect extends to fricatives, affricates and stops at different places of articulation as well, I only use ⟨p⟩/⟨t⟩/⟨k⟩ *versus* ⟨b⟩/⟨d⟩/⟨g⟩ for illustrative purposes.

Swedish (Elert 1964, Buder & Stoel-Gammon 2002).⁵ The observed direction of this tendency is always the same: vowels preceding ⟨b⟩/⟨d⟩/⟨g⟩ are longer than those preceding ⟨p⟩/⟨t⟩/⟨k⟩. Note that I have purposely avoided referring to the actual phonetic properties of the conditioning environment. There is good reason to be cautious: the phonetic categories that condition this length difference can be quite different cross-linguistically. Thus, French, Lithuanian, Spanish and Modern Standard Arabic contrast prevoiced and plain voiceless obstruents; English and Swedish contrast optionally prevoiced and voiceless aspirated obstruents; and Hindi, Assamese and Bengali have a four-way laryngeal contrast involving prevoiced and voiceless obstruents that come both in plain and aspirated forms.

ACTIVE/PASSIVE VOICING AND ASPIRATION The latter two language types show that the voicing effect can also appear in languages where the laryngeal contrast is not solely based on prevoicing. This raises the following question: is there a single mechanism underlying the patterns of vowel length variation observed in the languages listed above, or are there several different mechanisms related to different types of laryngeal contrast? I will attempt to answer this question by taking a closer look at languages where voicing and aspiration co-occur. Moreover, I will also briefly discuss the behaviour of vowels preceding passively voiced consonants (e.g. nasals), as this might provide further evidence relating to the question at hand.

As it has been noted above, there are two different types of languages where voicing and aspiration are both used in laryngeal contrasts. The first type comprises languages like English and Swedish, which contrast only two different laryngeal categories, but use different combinations of closure voicing and aspiration to signal this contrast depending on stress, foot structure and position within the word. Both languages show relatively complex patterns with respect to voicing, so I will focus on a smaller subset of forms that are particularly relevant from the perspective of this thesis: monosyllabic items (this is because the stimuli for the experiment described in Section 6.4 are also

5. Maddieson (1977), Hussein (1994) and Campos-Astorkiza (2007) cite many more examples, which are omitted here either because they come from very small studies that are not robust enough in a statistical sense, or because I could not verify them.

all monosyllabic). If the vowel is followed by an obstruent in a monosyllabic form, the obstruent will necessarily be in final position. As it turns out, the role of aspiration is greatly diminished in this position. To see why this is the case, let us compare initial, medial and final obstruents (based on Ladefoged & Johnson 2010, Ringen & Helgason 2004, Helgason & Ringen 2008). Aspiration plays a crucial role in word-initial contrasts in both languages: in this position, English typically contrasts partially prevoiced or plain voiceless consonants with voiceless aspirated ones, and Swedish consistently prevoiced consonants with voiceless aspirated ones. In medial position, the nature of the contrast depends strongly on foot structure. While prevoicing seems to occur more or less invariably, postaspiration is typically only found in cases where the obstruent is in the onset of a stressed syllable. In final position, no postaspiration is seen in either language, while closure voicing appears consistently in Swedish and optionally in English.⁶ Since there is no aspiration in final position, it appears that the source of the vowel length differences in monosyllabic forms is the same in languages with a simple contrast based on prevoicing and languages like English and Swedish. Vowel length in these forms is determined by voicing, not by aspiration. While this analysis is by no means comprehensive, it will be sufficient for the purposes of this thesis, given that the empirical investigation presented in Section 6.4 is also restricted to monosyllabic forms.

The second type includes languages like Hindi, Assamese and Bengali, where both voicing and aspiration can be used to distinguish obstruents from each other, yielding a four-way contrast (e.g. [p]/[p^h]/[p̤]/[b^h]). The presence of voicing and aspiration in these languages is independent of position within the word. Aspirated stops occur in word-final contexts as well, which means that these languages provide us with an opportunity to look at how various combinations of voicing and aspiration affect vowel duration. Maddieson & Gandour (1976) show that both voicing and aspiration increase the length of the preceding vowel in Hindi. Thus, vowels in voiced aspirated contexts are the longest, followed by vowels in voiceless aspirated and voiced unaspirated contexts, with vowels in voiceless unaspirated contexts being the shortest.

6. It should be noted that Swedish exhibits preaspiration in word-medial and word-final position in the ⟨p⟩/⟨t⟩/⟨k⟩ series – however, it is not entirely clear how this phenomenon is related to postaspiration in terms of its articulation and potential phonetic effects on preceding vowels.

Maddieson & Gandour (1976) interpret this observation as follows: the effect of aspiration is independent of the voicing effect, and voicing and aspiration have a cumulative influence on vowel length.

Let us now turn to the issue of passive voicing. If the voicing effect is simply a function of the presence or absence of vocal-fold vibration, vowels preceding sonorants – especially nasals – should also be longer than vowels preceding voiceless obstruents. Indeed, Lisker (1974) claims that nasals and voiced stops behave identically with respect to the voicing effect, although he does not provide any evidence to back this claim. While the effects of passive voicing on vowel length have not been investigated in detail, there are a few studies that report on vowel length in pre-sonorant environments. Zimmerman & Sapon (1958)'s results suggest that vowel length before nasals is intermediate between the durations measured in voiceless and voiced obstruent environments in Spanish. The same paper finds inconsistent effects in English: [m] seems to pattern with [b] as opposed to [p], but alveolar nasals show the same intermediate effect size as the nasals in Spanish. Conversely, Peterson & Lehiste (1960) report that nasals and voiced stops are indistinguishable in terms of their effects on the duration of the preceding vowel. Note that neither Zimmerman & Sapon (1958), nor Peterson & Lehiste (1960) present any statistical analysis of their data. A somewhat earlier paper by House & Fairbanks (1953) presents observations that are also supported by statistics. Their findings suggest that nasals affect vowels in a way that is similar to voiced obstruents, but the amount of incremental duration is somewhat smaller than that for voiced obstruents. In sum, these studies do not provide conclusive evidence for grouping nasals with voiced obstruents.

The following conclusions can be made based on this short review of the factors that could plausibly interfere the voicing effect. It seems that there is, in fact, a single mechanism behind the variation in vowel length in all the languages listed above (at least when the focus is on monosyllabic forms, as in this thesis). In languages where both aspiration and voicing can occur in word-final position, this mechanism is supplemented by another effect related to aspiration. Moreover, it seems useful to restrict the definition of the voicing effect to obstruents, as sonorants behave ambiguously with respect to their influence on the length of the preceding vowel.

THE SOURCE OF THE VOICING EFFECT Having reviewed some of the more problematic aspects of the voicing effect, let us now turn to its phonetic motivation. Two main types of explanation have been proposed for this phenomenon: articulatory and auditory. Both of these locate the source of the voicing effect in low-level phonetics and therefore make it a phonetic bias in the terminology of this thesis (cf. Section 3.4). The next section relies heavily on this assumption, as it describes a simulation based on the idea that the voicing effect has its roots in a phonetic bias.

Since Lisker (1974), Maddieson & Gandour (1976) and Kluender et al. (1988) all reject articulatory accounts after extensive review and criticism, I only consider these approaches briefly here. The present review is based mostly on Kluender et al. (1988). Belasco (1953) proposes that the voicing effect is the result of a tendency to use a constant amount of ‘articulatory force’ within a syllable. In his account, voiceless obstruents require more force than voiced ones, which is compensated for by shortening the preceding vowel before voiceless obstruents. Besides the completely *ad hoc* nature of this explanation, there are many arguments that speak against such an account. For instance, it predicts even greater shortening before aspirated consonants, which arguably require more force than unaspirated ones (Maddieson & Gandour 1976). This prediction is clearly wrong. A different approach is exemplified by Chomsky & Halle (1968). They suggest that ‘[v]owels are lengthened in front of voiced consonants to allow time for laryngeal adjustments needed to maintain glottal vibration during oral constriction or closure’ (Kluender et al. 1988: p. 154). As there is no evidence that such adjustments occur, but there is evidence for adjustments before voiceless stops (Lisker 1974), this approach is also unlikely to be on the right track. The third account suggests that the length difference follows from the greater speed of the closing gesture in voiceless obstruents. This speed difference is argued to result from the fact that more muscular energy is needed to contain the ‘higher intraoral pressure [...] during the production of a voiceless stop or fricative’ (Kluender et al. 1988: p. 155). Unfortunately, this account suffers from the same problem as Belasco’s (1953): aspirated consonants should be preceded by even shorter vowels due to the higher rate of airflow entailed by aspiration (Maddieson & Gandour 1976).

Javkin (1976) and Kluender et al. (1988) attempt to avoid the shortcomings of articulatory accounts by locating the source of the voicing effect at the level of auditory perception. Javkin (1976) suggests that the boundary between a vowel and a voiced obstruent is less clear than that between a vowel and a voiceless obstruent. Consequently, listeners are likely to perceive part of the voiced portion of the consonant as belonging to the vowel. Kluender et al. (1988), on the other hand, argue that the voicing effect serves to enhance another cue to laryngeal contrasts, namely the length difference between voiced *versus* voiceless obstruents. The gist of their proposal is that speakers use the voicing effect to make the short closure interval of voiced obstruents appear even shorter by lengthening the preceding vowel, and *vice versa* for voiceless obstruents. There is some reason to favour Javkin's (1976) account over that of Kluender et al. (1988). Davis & Summers (1989) have shown that closure duration is an unreliable cue to voicing in obstruents following unstressed vowels. Kluender et al. (1988) would predict that in such cases no vowel length differences should be observed either, as there is no cue to enhance. This, however, is wrong: the durations of unstressed vowels are also influenced by the voicing of the following obstruent (Davis & Summers 1989). Therefore, the following tentative conclusion can be made about the source of the voicing effect: the length differences between vowels occurring before voiced *versus* voiceless obstruents are likely to be the result of low-level auditory processes along the lines proposed in Javkin (1976).

VARIATION IN THE SIZE OF THE VOICING EFFECT In the rest of this section, I present an overview of the cross-linguistic and language-internal variation in the size of the voicing effect. As it has been noted above, the voicing effect has been observed in numerous languages. Some authors have used the available data to draw cross-linguistic comparisons with regard to the size of the effect (e.g. Chen 1970, Keating 1984, Hussein 1994). Although the validity of such comparisons has been questioned on the grounds that cross-linguistic investigations typically do not control for all confounding factors (Laeufer 1992, Hussein 1994), there are some findings that seem to emerge consistently from studies of the voicing effect:

1. The size of the voicing effect is greater in English than in most other languages (Chen 1970). This has been demonstrated in a statistically rigorous way for French *versus* English (Mack 1982) and Arabic *versus* English (Hussein 1994).
2. Some languages show very little or no voicing effect. Swedish and Arabic exhibit a relatively small length difference compared to other languages (Elert 1964, Buder & Stoel-Gammon 2002, Hussein 1994), and Polish and Czech show no length difference at all (Keating 1984).

Similarly to the cases cited in the previous section, the observed minimally three-fold distinction in the size of the voicing effect does not support models that attempt to derive such differences from the phonology/phonetics divide. It could be argued that the English pattern is phonologised while the rest of the languages show a phonetic effect. However, it is not clear why the languages with a 'purely phonetic' pattern should show further differences in the size of the effect. This is also the conclusion that Hussein (1994) arrives at after a much more detailed review of cross-linguistic variation in the voicing effect.

Besides these cross-linguistic differences, the size of the effect can also vary within the same language. Two different types of language-internal variation have been reported in connection with the voicing effect: contextual variation and vowel-specific patterns. As for contextual variation, there are numerous segmental and prosodic factors that can interact with the voicing effect in a given language. Only two of these are mentioned here: stress and the phonetic properties of the following obstruent (the interested reader is referred to Laeuffer (1992) and Hussein (1994), who present much more detailed reviews of these factors). Davis & Summers (1989) find that the size of the voicing effect in English is greatly reduced in unstressed syllables as compared to stressed ones. In fact, the voicing effect is weakened to the point of statistical insignificance for some speakers in certain environments. The manner of articulation of the following stop has also been claimed to affect the size of the voicing effect: House & Fairbanks (1953) have found that there is a greater length difference before fricatives than before stops. Since these effects can obscure the influence of voicing on vowel length, experimental investigations have to make sure that they control for them.

The second type of language-internal variation is conditioned by the identity of the target vowel: certain vowels are more strongly affected by the voicing effect than others. For instance, Laeufer (1992) suggests that ‘the effect of obstruent-voicing [is] smaller for inherently shorter vowels than for longer ones’ (Laeufer 1992: p. 412). According to Laeufer, this effect has been found for English and French as well.⁷ Campos-Astorkiza (2007) reports an even more intriguing vowel-specific effect. Lithuanian has a symmetric vowel system with three vowel heights and a distinction between front *versus* back vowels, which yields six different vowel qualities. The high and the low vowels come in contrastive short-long pairs: [ɪ]~[i:], [ʊ]~[u:], [ɛ]~[æ:] and [a]~[ɑ:]. However, the length contrast exists only marginally in the case of mid [ɔ]~[o:] (where the short member of the pair only appears in recent loanwords; cf. Campos-Astorkiza 2007: p. 32), and the front mid vowel [e:] has no short pair at all. Campos-Astorkiza shows that the voicing effect is stronger in mid vowels than it is in high and low vowels, and is particularly robust in the case of [e:]. She suggests a causal link between the structure of the Lithuanian vowel inventory and the vowel-specific differences in the size of the voicing effect: mid vowels allow more contextual variation in duration precisely because they are not involved in the length contrast exhibited by high and low vowels. Although this section does not examine the effects of contrast, vowel-specific patterns of variation are exploited heavily in Section 6.4, where I look at the influence of lexical factors on the size of the voicing effect in different vowels.

Let us briefly go through the main points of this section. It has been shown that both voicing and aspiration can affect vowel length. However, the effects of aspiration are more visible in languages where aspiration and voicing function independently of each other, like Hindi, Assamese and Bengali. The length differences in languages with an English-like system of contrasts seem to be conditioned mostly by voicing, at least in monosyllabic forms. Moreover, there is no strong evidence for grouping sonorants either with voiceless or voiced obstruents, which suggests that the variation in vowel duration is a function of active voicing. A review of potential explanations for the voicing effect suggests

7. Although French has no lexical contrasts based on vowel duration, length differences arise as a function of aperture.

that it stems from low-level properties of auditory perception, and therefore its origins lie in a phonetic bias. There is also a large amount of cross-linguistic variation in the size of the effect, which is evidence for its learnt nature. Finally, the size of the voicing effect is subject to variation even within the same language depending on the context and the target vowel, which means that these factors have to be considered carefully in any investigation of the phenomenon.

6.3 SIMULATING THE BEHAVIOUR OF SUB-DISTRIBUTIONS

We are now in a position to return to the prediction outlined in the introduction to this chapter, which is repeated below for convenience:

Prediction 1 *Sub-distributions are further apart in categories with a balanced bias proportion than they are in categories with an unbalanced bias proportion.*

In order to see how the discussion presented so far in this chapter links to this prediction, let me remind the reader of the general research strategy pursued in this thesis. One of the cornerstones of this strategy is the requirement that predictions should be obtained in a formally rigorous way. Specifically, the researcher has to formalise the assumptions underlying their approach and then derive predictions from the resulting formal system. When the formalisation of the problem is simple and tractable, deductive logic and thought experiments can be sufficient to produce testable predictions. However, when the formal system is complex – as in the present case – more advanced methods are needed. Computer simulations provide one way of linking a theory to its predictions. Mathematical models are another alternative, although such solutions are not always available. The present chapter uses both.

The role of the first half of this chapter was to outline the underlying assumptions in as much detail as possible, and to set the scene for the presentation of the formal system and the simulations. I reviewed a substantial body of evidence suggesting that sub-distributions within categories are used in production and perception, and that they are learnt and phonetically detailed. The present section is the link between the theory and its predictions: it builds a formal model of the behaviour of sub-distributions and uses simulations and mathematical modelling to derive Prediction 1. Although this prediction is applicable to any

situation involving sub-distributions within a category, the model is presented specifically through the example of the voicing effect described in the previous section. This is useful not only because it anchors the argumentation in a familiar phenomenon, but also because it provides directly testable predictions with regard to the voicing effect. Section 6.4 presents a small cross-linguistic study of the voicing effect that attempts to verify these specific predictions.

The structure of this section follows straightforwardly from the logic of the approach described above. Section 6.3.1 explains how the assumption of sub-distributions can be incorporated into the model of speech production and perception developed in Chapter 4. Then, Section 6.3.2 uses a computer simulation to demonstrate how sub-distributions shift apart over time under the influence of different phonetic biases. Finally, Section 6.3.3 presents a simple mathematical model that shows how Prediction 1 emerges from the assumptions underpinning the model.

6.3.1 Modelling sub-distributions

In Chapter 4, I outlined a model of the production and the perception of phonetic categories that relies on probability distributions. The main idea was that production consists in sampling probability distributions, while perception is based on a Bayesian choice rule, which uses the same distributions to derive categorisation probabilities. It was also demonstrated that the dynamics of simulated sound systems are essentially the same regardless of the exact nature of these probability distributions (within certain limits, of course). Since parametric probability distributions are computationally less demanding than non-parametric ones, I continue to use normal distributions in modelling production and perception.

Sub-distributions can be incorporated into production by introducing a small amount of added complexity into category representations and the sampling method. Specifically, I will represent each sub-distribution within a category by a Gaussian mixture component, and the overall category by a mixture distribution (cf. Kirby 2010, *in press*). To make this more concrete, let us compare this type of representation to the one used in Chapters 4 and 5. Consider the distribution of duration values associated with the GOAT vowel in the speech

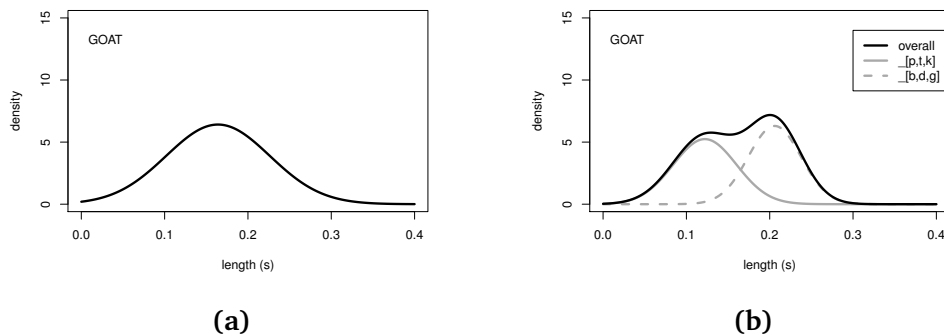


Figure 6.4: (a): a parametric category representation corresponding to the GOAT vowel in SSBE; (b): a mixture distribution corresponding to the GOAT vowel in SSBE, including sub-distributions in voiceless and voiced contexts.

of a speaker of Southern Standard British English (the data are taken from the experiment described in Section 6.4). A production model without sub-distributions would have to use a single normal distribution to represent the whole category, as shown in Figure 6.4a.⁸ The approach based on mixture distributions also relies on normal distributions, but they play a different role: they represent the sub-distributions within the category (Figure 6.4b). Thus, in the present case, the GOAT vowel before voiceless and voiced obstruents is represented by two separate Gaussians (indicated by the solid and the dashed grey lines, respectively). The overall category representation is obtained by summing and renormalising these two Gaussians, which yields the mixture distribution represented by the solid black line in Figure 6.4b.

Similarly to the original model, the selection of a production target is based on sampling the overall distribution corresponding to a given category. This seems to lead to a problematic result: if tokens of the GOAT vowel come from the same mixture distribution regardless of whether they stand before a voiced or a voiceless obstruent, there is no way of accounting for learnt differences in the production patterns associated with different sub-distributions. For this reason, an additional detail has to be introduced. The distribution that serves as the basis of production is not just a simple sum of the mixture components,

8. This is not a full model of production, as it only handles a single phonetic dimension: that of vowel length. However, it is sufficient for our present purposes, given our focus on durational differences among vowels in different phonetic contexts.

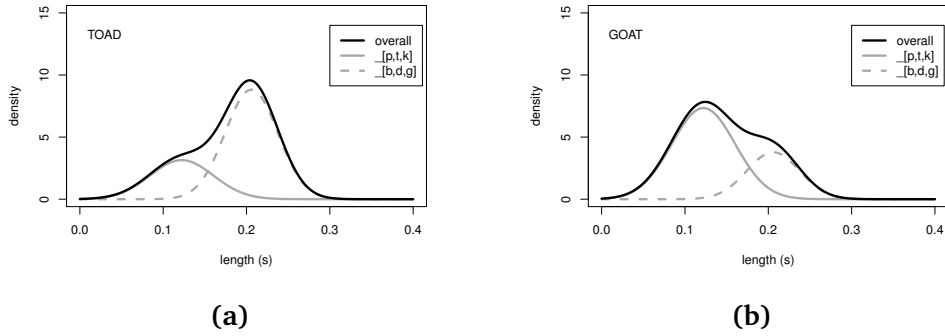


Figure 6.5: (a): a weighted mixture distribution where the Gaussian corresponding to the GOAT vowel in voiced contexts counts more heavily than the Gaussian for the vowel in voiceless contexts (e.g. the word *toad*); (b): a weighted mixture distribution for the same vowel where the voiceless distribution counts more heavily (e.g. *goat*).

but a weighted sum, where the sub-distribution corresponding to the phonetic context of the target vowel (the target sub-distribution) counts more heavily (this solution is inspired by Pierrehumbert 2002). For instance, the voiceless distribution has a greater role in determining the production target for a word like *goat*, and the voiced distribution has a greater role in determining the target for a word like *toad*. This is illustrated in Figure 6.5.

We can now discuss the formula for the sampling distribution that serves as the basis of production in this revised model. The general formula below is valid for any case where category C occurs in environment a :

$$p(x | C, a) = \frac{i p(C, a) \mathcal{N}(x | \mu_a, \sigma_a^2) + \sum_{b \in B} p(C, b) \mathcal{N}(x | \mu_b, \sigma_b^2)}{i p(C, a) + \sum_{b \in B} p(C, b)} \quad (6.1)$$

This is a weighted mixture of $|B| + 1$ Gaussian components (given in the form $\mathcal{N}(x | \mu, \sigma^2)$, where μ is the mean, and σ the standard deviation), where B is the set of all sub-distributions *excluding* the target sub-distribution. The weights are as follows: $i p(C, a)$ for the target sub-distribution, and $p(C, b)$ for any other sub-distribution that is not in the target environment. $p(C, a)$ and $p(C, b)$ are the prior probabilities of the sub-distributions, which correspond

to their relative frequencies within the category. Parameter i represents the extra weight that the target sub-distribution has in determining the production target. To give an example, if $i = 20$, the target sub-distribution will be 20 times more important in calculating the production target than any other sub-distribution. This parameter has a strong effect on the amount of independence that each sub-distribution can have. When $i = 1$, the target sub-distribution plays no special role in predicting the production target, and therefore the model behaves identically to the one presented in Chapter 4. However, as $i \rightarrow \infty$, the target sub-distribution comes to dominate the mixture distribution, and the sub-distributions become fully independent (i.e. they do not interact in predicting production targets). Intermediate values of i result in cases where sub-distributions have a limited amount of independence: productions are more strongly affected by the target sub-distribution, but the other sub-distributions also have a chance to contribute (cf. Figure 6.5).

To make this somewhat more concrete, here are the formulae for calculating production targets (in this case, duration values) for a vowel category (C) in voiced (vd) and in voiceless contexts (vl). Note that the prior probabilities have been replaced by w and $w - 1$, which indicate the proportions of tokens in contexts vd and vl , respectively (i.e. w corresponds to the bias proportion).

$$p(x | C, vd) = \frac{i w \mathcal{N}(x | \mu_{vd}, \sigma_{vd}^2) + (1 - w) \mathcal{N}(x | \mu_{vl}, \sigma_{vl}^2)}{i w + (1 - w)} \quad (6.2)$$

$$p(x | C, vl) = \frac{w \mathcal{N}(x | \mu_{vd}, \sigma_{vd}^2) + i(1 - w) \mathcal{N}(x | \mu_{vl}, \sigma_{vl}^2)}{w + i(1 - w)} \quad (6.3)$$

Equation (6.2) corresponds roughly to Figure 6.5a and equation (6.3) to Figure 6.5b.

Once a target production has been chosen, the relevant phonetic biases apply to it. The implementation of biases in the following simulations is very similar to that in Chapters 4 and 3: a phonetic bias is a displacement of the target production towards a given point in phonetic space. However, this chapter introduces a simplification into the formal definition of biases. Previously, I used logistic curves in their implementation (cf. 4.1.2), while in this model, a bias is simply a linear function. The purpose of this simplification is made clearer in Section 6.3.3, where the new definition is used as part of a mathematical

model to derive Prediction 1. The following formula shows how a given target production x is displaced under the influence of a bias at location b_1 with strength s_1 :

$$\text{bias}(x, b_1, s_1) = x + s_1 (b_1 - x) \quad (6.4)$$

By way of example, consider the case of a vowel produced in a voiced context. Let us assume that the target duration is $x = 0.15$ s, and that a lengthening bias applies because of the following voiced obstruent, with $b = 0.4$ s and $s = 0.1$. The bias moves the target production closer to b by exactly a tenth of the distance between the two of them, which means that the stimulus perceived by the listener is at $x = 0.175$ s.

The modelling of perception is less relevant in the present case, as the simulations in the following sections look at isolated vowel categories. Therefore, I only describe the mechanism for category update, but not the mechanism for making categorisation decisions (cf. Section 4.1.3). The interested reader is referred to Sonderegger & Yu (2010), who provide a detailed discussion of how sub-distributions can be built into models of categorisation.

The main idea for category update remains essentially the same as in Chapters 4 and 3: the perceived token is incorporated into the representation of the appropriate category. Since the model under discussion is parametric, this consists in (i) shifting the mean towards the new stimulus, and (ii) increasing or decreasing the variance, depending on how far the stimulus is from the mean. The crucial innovation in the sub-distribution based model is that the changes only affect the relevant mixture component, and not the others. For example, if the stimulus is a vowel in a voiced environment, only the voiced sub-distribution is updated, but not the voiceless one. The formulae for updating the mean and the variance of a mixture component can be written as follows (n stands for the old distribution and $n + 1$ for the updated distribution):

$$\mu_{n+1} = \frac{c \mu_n + x}{c + 1} \quad (6.5)$$

$$\sigma_{n+1}^2 = \frac{c [(\mu_{n+1} - \mu_n)^2 + \sigma_n^2] + (x - \mu_n)^2}{c + 1} \quad (6.6)$$

In the above equations, x represents the new stimulus, and c is the constant of update that determines the extent to which x can change the representation of the component. A low value of c yields a system that changes quickly under the influence of new stimuli, whereas a high value of c results in a system that is more resistant to changes. Note that these formulae are exactly the same as those provided in Section 4.1.3.

6.3.2 The separation of sub-distributions

Having seen how sub-distributions can be incorporated into our model of production and perception, we can now attempt to simulate changes in sub-distributions over time. The main goal of this section is to see how patterns specific to given phonetic contexts can emerge under the influence of weak phonetic biases. Specifically, I investigate the emergence of the voicing effect in vowel categories. This will help us get a better sense of how Prediction 1 arises from the present model.

The simulations in this section are closely related to those presented in Chapter 4: they mimic the diachronic evolution of sound categories through an abstract production-perception feedback loop (cf. Pierrehumbert 2001, Wedel 2007). Each simulation contains only a single agent, who doubles as both speaker and listener. This agent generates productions based on their category representations. Admittedly, this situation is unrealistic in the sense that real speech interactions typically involve more than one speaker. However, Section 4.3 demonstrated that single and multi-agent models do not differ significantly in terms of their general dynamics and expected outcomes. Since single-agent simulations are computationally more tractable, they are preferable to more complex implementations in cases like the present one. Indeed, the tractability of the abstract production-perception feedback loop will prove crucial for the mathematical calculations presented later in this chapter.

A further parallel with the simulations in previous chapters lies in the idea that patterns like the voicing effect result from the repeated interaction of phonetic biases and the production-perception feedback loop. The phonetic bias responsible for the voicing effect is assumed to be weak: it cannot cause and immediate shift in sound categories. However, when a category or a sub-

distribution is repeatedly exposed to a phonetic bias, more robust changes will occur (cf. the discussion of the nudge model in Section 2.3.2). This means that context-specific patterns like the voicing effect can arise over time due to the fact that different sub-distributions are exposed to different sets of biases.

Let us now look at the structure of the simulations in more detail. As it has been noted above, the simulations contain only a single agent. This agent draws productions from a single one-dimensional category representation with two sub-distributions that are affected by slightly different sets of biases. This setup is analogous to the production of a vowel occurring before voiced and voiceless obstruents. To make this parallel explicit, the following discussion refers to the overall category as the vowel, the sub-distributions as voiced and voiceless, and the predicted values as the duration of the vowel.⁹ As in all previous simulations, the produced tokens are fed back into the agent's category representations (note that no misperception will occur, since there is only a single category).

Production targets are predicted using equations (6.2) and (6.3) for the voiced and voiceless sub-distributions, respectively. The simulated vowel occurs equally frequently in voiced and voiceless environments. Therefore, parameter w – the proportion of tokens followed by voiced obstruents (i.e. the bias proportion) – is set to 0.5. Parameter i is arbitrarily given as 15, which means that the sub-distributions have a limited amount of independence. Both of these parameters are examined in more detail in the next section. The initial values for the means (μ) of the sub-distributions are varied across the simulations, but their initial standard deviations (σ) are always the same: 0.05 s.

There are two phonetic biases in these simulations: a lengthening bias that affects vowels before voiced obstruents, and a shortening bias that affects all vowels. The lengthening bias represents the phonetic origins of the voicing effect. It is inspired by Javkin's (1976) account of the phenomenon, who suggests that the voicing effect arises from a low-level perceptual illusion, whereby listeners interpret the glottal pulsing during the closure phase of voiced obstruents as part of the preceding vowel, making it appear longer than it actually is. The general shortening bias is meant as a reflection of the principle of minimal effort, which

9. Note that there are no sub-distributions corresponding to following sonorants. Sonorants have been excluded as their exact role in the voicing effect is little understood, and they do not clearly pattern with either voiced or voiceless obstruents (cf. Section 6.2).

ensures that vowels do not become arbitrarily long. The tendency to shorten is modelled as a bias attractor at $b_1 = 0.1$ s, and the voicing-related lengthening bias as one at $b_2 = 0.5$ s. Both biases have the same strength: $s_1 = s_2 = 0.05$.

Category update proceeds along the lines specified in equations (6.5) and (6.6). The constant of update is set to $c = 1000$. Unless otherwise specified, the simulations consist of 50,000 iterations of the production-perception loop, which corresponds to the production of 25,000 vowels in a voiced context and 25,000 vowels in a voiceless context.

The first simulation illustrates the separation of the voiced and voiceless sub-distributions over time. Both sub-distributions are initialised at $\mu_{vd} = \mu_{vl} = 0.1$. Figure 6.6 contains visualisations of their evolution. The top six panels are snapshots of the simulation at different points in time, while the bottom panel illustrates the entire simulation by plotting the 5% and 95% quantiles of the two sub-distributions against the number of iterations. There are two important observations that can be made on the basis of these graphs. First, it is clear that the two sub-distributions become separated as the simulation progresses. In fact, the outcome of this simulation is qualitatively quite similar to the distributions of voiced and voiceless tokens shown in Figure 6.4b (although there are clear differences as well). Second, the sub-distributions do not seem to change their positions significantly after about 25,000 iterations. Of course, this is not to say that they remain completely stationary, but the observed changes are small and inconsistent, especially when compared to the shifts seen during the first 25,000 iterations. It seems that the sub-distributions move towards stable states. This is, of course, not surprising, given that similar equilibria emerged in all the simulations discussed in the previous chapters.

To test whether it is reasonable to assume stable states, two further simulations were run. The first of these is simply an extension of the previous simulation, with 500,000 iterations instead of 50,000. This simulation is illustrated in Figure 6.7, which shows quite clearly that the sub-distributions do not change significantly after the initial phase of the simulation. The second simulation is intended as a demonstration of the fact that the initial location of the sub-distributions is not relevant in determining the target locations that they move towards. For this simulation, I changed the initial μ values to $\mu_{vd} = \mu_{vl} = 0.3$ s. Figure 6.8 shows that the stable states remain the same: even

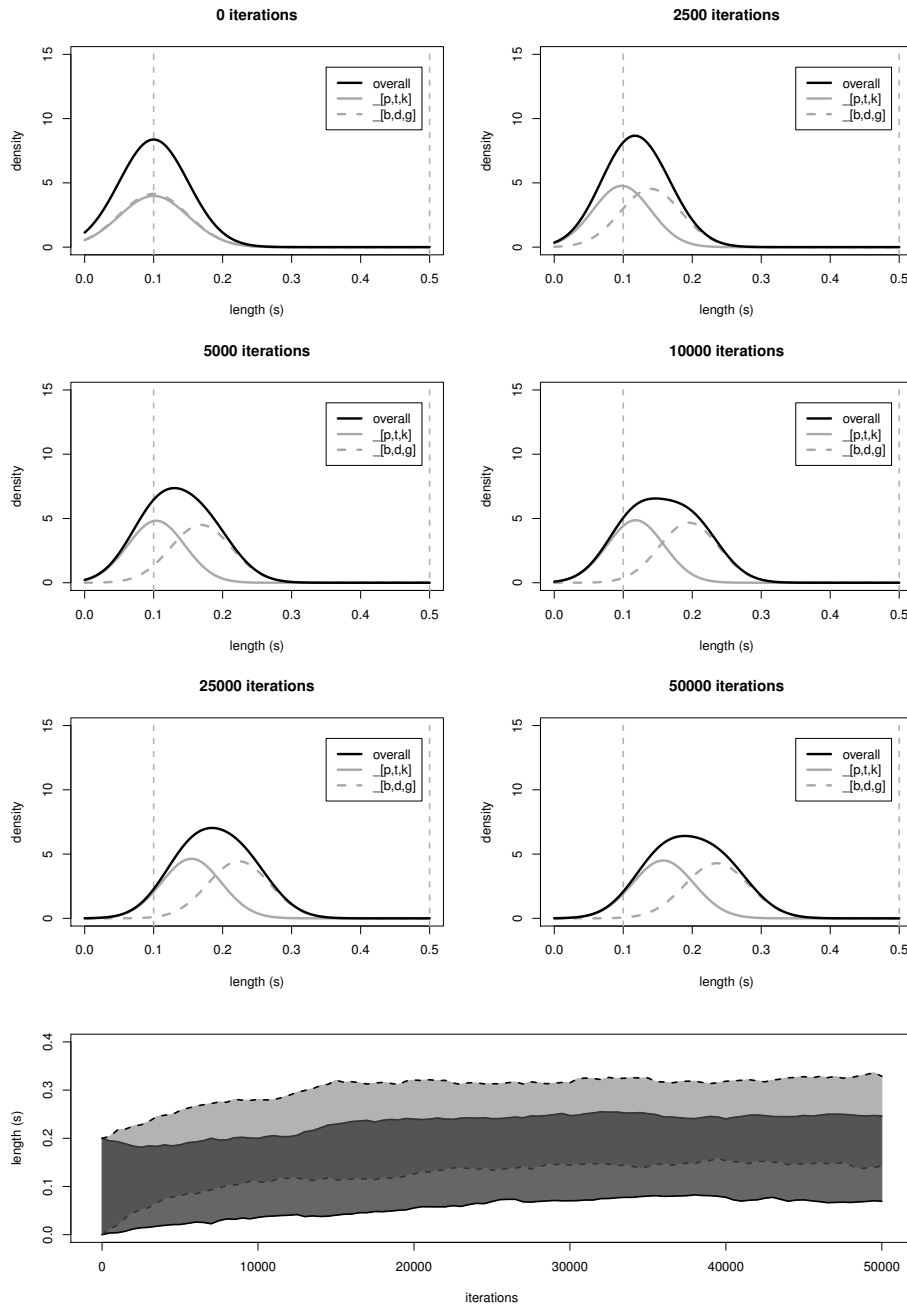


Figure 6.6: The separation of the voiced and the voiceless sub-distributions. The first six panels show the sub-distributions at different points in time. The dashed vertical lines mark the locations of the biases. The bottom panel uses 5% and 95% quantiles to illustrate changes over the course of the simulation (light grey: voiced; medium grey: voiceless; dark grey: overlap).

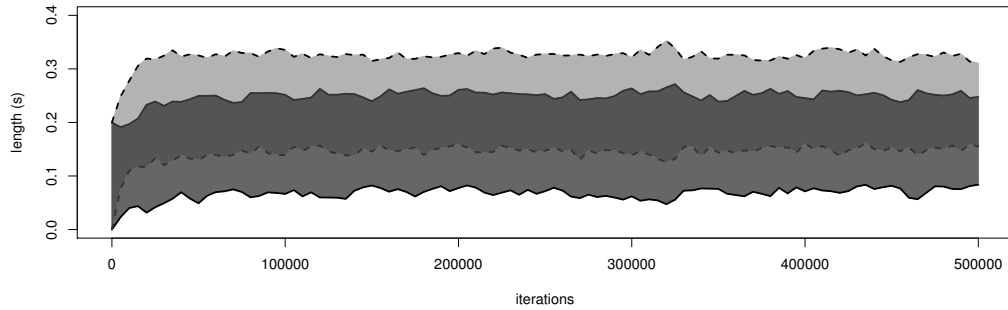


Figure 6.7: *The evolution of the voiced and voiceless sub-distributions over 500,000 iterations.*

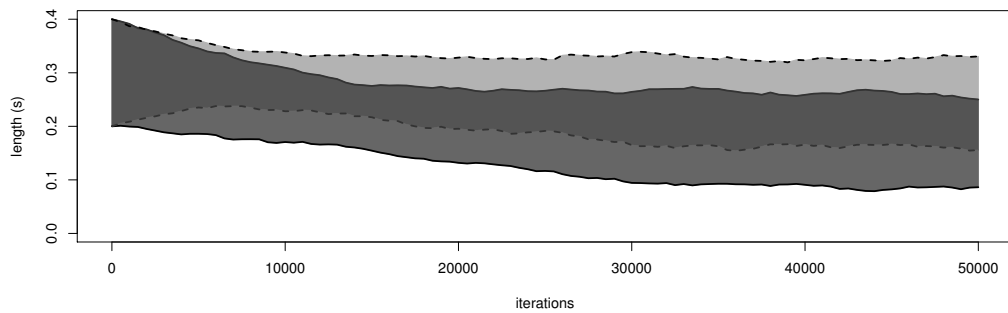


Figure 6.8: *The evolution of the voiced and voiceless sub-distributions under altered initial conditions.*

though the sub-distributions start from different locations, they end up in the same regions as in Figures 6.6 and 6.7.

The similarities among the stable states in different simulations are clearly visible in the figures that I have provided. Nevertheless, it will be useful to compare the simulations in a statistically more rigorous way as well. In order to do this, I re-ran the second simulation ($\mu_{vd} = \mu_{vl} = 0.1$; this will be referred to as the low-initial-value condition) and the third one ($\mu_{vd} = \mu_{vl} = 0.3$; the high-initial-value condition) 100 times each, and recorded the means of the voiced and the voiceless sub-distributions (μ_{vd} and μ_{vl}) after 100,000 iterations for each simulation. The resulting 400 measurements (2 sub-distributions \times 2 conditions \times 100 repetitions) were compared separately for the voiced and the voiceless sub-distributions across the two conditions. The box plots in Figure

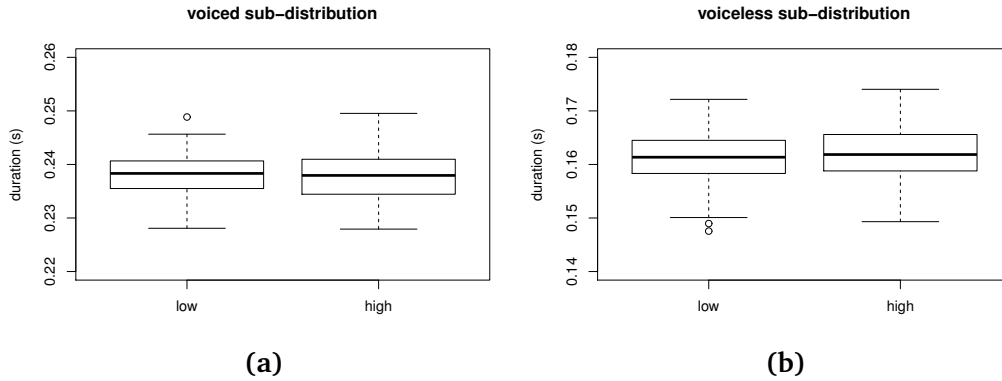


Figure 6.9: (a): a comparison of μ_{vd} values after 100,000 iterations in the low-initial-value and the high-initial-value conditions; (b): the same comparison for μ_{vl} values. Note that vertical view range is not the same in the two box plots.

6.9 serve to illustrate these comparisons. The results are clear: the means of the sub-distributions take on the same values in the high-initial-value and the low-initial-value conditions. This means that the stable states are not affected by the initial locations of the categories. These observations are confirmed by two-tailed t -tests, which found no significant differences between the conditions (voiced: $t = 0.2493$, $p = 0.8034$; voiceless: $t = -0.9675$, $p = 0.3345$).

The previous simulations have shown that (i) sub-distributions drift apart when they are affected by different sets of biases and (ii) regardless of their initial locations, the sub-distributions end up in the same stable states. The first result needs little commentary: this is exactly what we expected to find based on the partial independence of sub-distributions and the fact that only voiced tokens are subject to the lengthening bias. The second result can be explained relatively straightforwardly using the terminology of the previous chapter. Section 5.2.3 showed that complexity has an important influence on the structure of the adaptive landscape that determines the evolution of the sound system. Since the simulations in this section are all based on a single category, the adaptive landscape will be simple, with only a single stable state. This stable state is determined by two main factors: phonetic biases and the internal cohesion of the category (which follows from assumption of mixture-based category representations). Note that these pressures act against each other: phonetic biases push the sub-distributions apart, while the internal cohesion of the category pulls them closer together. The next section will show that there

are a number of further parameters that play an important role in predicting the stable states.

Although this section does not directly address the question of how Prediction 1 emerges from the model, the findings discussed above suggest a way forward. We can use the same general research strategy as we did in the previous chapters: instead of looking at entire simulations, we can try to calculate the stable states for the voiced and voiceless sub-distributions, and see how they are affected by bias proportion. Our expectation is that the gap between the target locations will be large when voiced and voiceless tokens occur equally frequently, and small when one group is significantly more frequent than the other. If this expectation is confirmed, we can conclude that Prediction 1 is a direct consequence of our theoretical assumptions.

6.3.3 *Predicting the effect of bias proportion*

In this section, I investigate the influence of bias proportion on the stable states towards which sub-distributions converge over time. Specifically, I look at the influence of parameter w (introduced in equations (6.2) and (6.3) in Section 6.3.1) on the outcome of the model. There are two possible ways of investigating how changing the value of a parameter affects the behaviour of the model. The first approach was exemplified in the previous section, where a large number of simulations were run to see whether the initial location of the sub-distributions affects their limiting behaviour. Unfortunately, this method can be extremely time-consuming, due to the number of simulations that have to be run, and it can become unfeasible if more than one parameter is investigated. There is, however, a second option. In some cases, mathematical models can be used to predict the outcomes of the simulations without implementing them computationally. In the present case, this would consist in deriving the stable states for the sub-distributions from the formulae in Section 6.3.1. As it turns out, such a prediction is possible. Therefore, this section takes the second, mathematical approach to the question outlined above.

The main idea behind this approach is easy to summarise and requires no mathematical formalism. As it has been noted in the previous section, the voiced and the voiceless sub-distributions are affected by different sets of

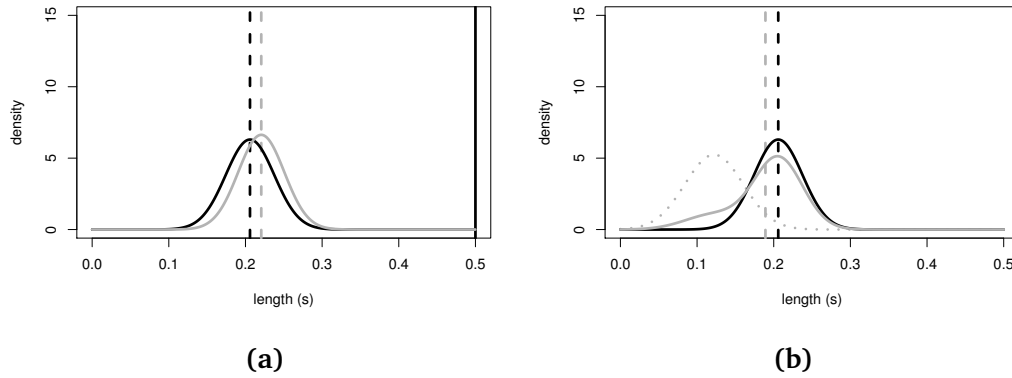


Figure 6.10: (a): the voiced sub-distribution (black) and the observed distribution (grey) after the application of the bias at 0.5 s; (b): the voiced sub-distribution (black) and the corresponding sampling distribution (grey solid), which is a mixture distribution consisting of both the voiced and the voiceless sub-distributions (dotted grey). Note that the expected values (the dashed vertical lines) move in both cases.

biases, which are modelled as small displacements towards given points in phonetic space. Following the method described in Section 5.2.2 and Appendix B, these displacements can also be viewed as transformations of the relevant sub-distributions. The observed distribution of phonetic outcomes differs from the underlying distribution as a result of phonetic biases. This is illustrated in Figure 6.10a, where the black curve shows the original representation of the voiced sub-distribution, and the grey curve the transformed distribution. The horizontal line at 0.5 s represents the lengthening bias. Note that this transformation can also be characterised in terms of the expected values of the distributions (indicated by the dashed lines): the expected value of the voiced sub-distribution shifts towards 0.5 s under the influence of the bias. The cohesive force that keeps the sub-distributions together can also be described in a similar way. Since production is based on a mixture that contains both the target sub-distribution and other sub-distributions within the category, the expected value of the observed distribution is different from the target sub-distribution. For instance, the production of a voiced token involves not only the voiced sub-distribution but also the voiceless one. The expected value of the sampling distribution is intermediate between the expected values of the two

sub-distributions. This is shown in Figure 6.10a, where the black curve shows the voiced sub-distribution, the dotted grey curve the voiceless sub-distribution, and the solid grey curve their mixture. Once again, the vertical lines indicate the expected values of the distributions.

Importantly, the biases and the cohesive forces induce shifts in opposite directions. It is easy to imagine a situation in which the sub-distributions are arranged in such a way that the shift in expected values due to the influence of other sub-distributions is subsequently cancelled by the shift caused by the bias. In such a situation, the expected value of the observed distribution is exactly the same as that of the underlying distribution, which means that the sub-distribution remains stationary. This is, in fact, what we observed in the previous section, where the category settled into an equilibrium after a certain number of iterations. Since both types of shifts can be predicted mathematically, we can use equations to find this specific arrangement of the voiceless and voiced sub-distributions.

To give the reader the option of skipping the technical details of this approach, I have relegated all mathematical formalisms to Appendix C. The formulae for predicting the target locations for the voiced and the voiceless sub-distributions are presented in (C.10) and (C.11). The parameters in these formulae have all been described in Section 6.3.1. Note that target location and stable state are both operationalised as the expected mean of a given sub-distribution – this is the quantity that we are trying to predict (similarly to Chapters 4 and 5). In the following discussion, all results are derived using these formulae and the formula for calculating the size of the gap between the two sub-distributions given in (C.12).

We can now address the question of how bias proportion affects the stable states of the sub-distributions. The most straightforward way of doing this is to explore changes in the value of the parameter corresponding to bias proportion (i.e. w , or the proportion of vowel realisations in voiced environments), while keeping all other parameters fixed. The values of the fixed parameters come from the simulations in the previous section; thus, $i = 15$, $b_1 = 0.1$, $s_1 = 0.05$, $b_2 = 0.5$, $s_2 = 0.05$. The value of w is varied continuously between 0 and 1. Figure 6.11a plots the stable states predicted by (C.10) (voiced sub-distribution; light grey line) and (C.11) (voiceless sub-distribution; dark grey

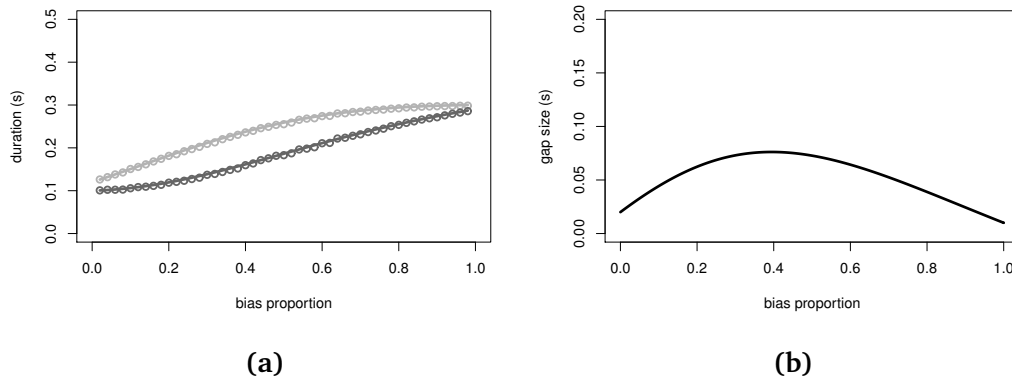


Figure 6.11: (a): mathematically predicted target locations for the voiced (light grey line) and the voiceless sub-distributions (dark grey line), and predictions from simulations (circles), plotted against bias proportion; (b): the size of the predicted gap between the voiced and the voiceless sub-distributions as a function of bias proportion.

line) against bias proportion. To ensure that the predictions are valid, I have also included estimations of the target locations based on simulations (indicated by light and dark grey circles).¹⁰ The simulated values are remarkably close to those predicted by the mathematical model, which suggests that (C.10) and (C.11) can be used reliably in determining the target locations for the sub-distributions. There are several things to note about this graph. First of all, the voiced sub-distribution always has slightly higher duration values regardless of bias proportion. This is because the biases remain active regardless of the amount of independence that the two sub-distributions have. Small differences in the expected values of the sub-distributions arise even when their lack of representational independence does not allow for such differences to be encoded in the underlying probability distributions. Moreover, both sub-distributions move towards higher duration values as the proportion of voiced items is increased. The reasons for this are as follows: (i) the influence of the lengthening bias (which only affects voiced items) grows as the proportion of voiced items

10. These simulations were identical to the first simulation described in the previous section, with the exception of parameter w , which was varied in small steps between 0 and 1. The target locations were estimated by finding the regions of phonetic space that were most frequently visited by each sub-distribution in the course of a given simulation (i.e. the modes of the stationary distributions of the means).

is increased and (ii) the two sub-distributions are never entirely independent, which means that they move together even if the biases affecting them would dictate otherwise. To put it more simply, the voiced sub-distribution pulls the voiceless one along at higher bias proportion values.

Turning now to the gap between the two sub-distributions, the following observation can be made. The size of the gap varies as a function of bias proportion, being larger towards intermediate values of w and smaller at extreme values. This is demonstrated even more clearly in Figure 6.11b, which plots the distance between the target locations against bias proportion. This finding is in agreement with Prediction 1, and thus concludes a lengthy argument, whose aim was to derive this prediction from the theoretical assumptions presented in Chapter 3 and Section 6.1 of the present chapter. To highlight the importance of this result, I repeat it below in a more explicit form.

(6.7) BIAS PROPORTION AND THE SIZE OF CONTEXTUAL EFFECTS

The size of a contextual effect is larger when the proportion of forms that exhibit the bias responsible for the effect is close to the proportion of forms without the bias. This follows from the following theoretical assumptions:

- a. production and perception are probabilistic and category-based (Sections 3.1, 3.2)
- b. categories are continuously updated (Section 3.3)
- c. production and perception are affected by low-level phonetic biases (Section 3.4)
- d. speakers rely on learnt and phonetically detailed sub-distributions (Section 6.1)

Note that this conclusion has to be qualified slightly in view of the fact that the point of greatest separation between the two sub-distributions is not at $w = 0.5$, but instead at $w = 0.394$.¹¹ The reason for this is that the two sub-distributions are affected by the biases in an asymmetric fashion: the voiceless sub-distribution is affected only by b_1 , while the voiced sub-distribution is affected both by b_1 and b_2 . A full explanation of how this asymmetry leads to

11. Readers with an esoteric turn of mind will be disappointed to hear that this proportion is not identical to the golden ratio, although it is certainly very close to it.

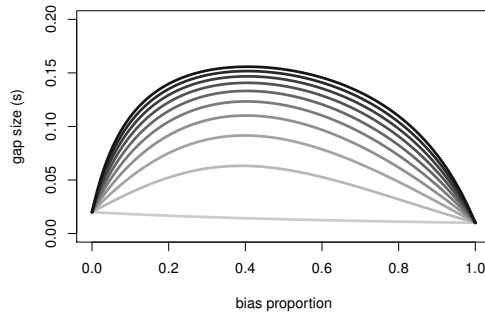


Figure 6.12: *The influence of parameter i on the relationship between bias proportion and the size of contextual effects. Darker lines correspond to higher values of i . The parameter is varied between 1 and 46 in steps of 5.*

the observed results is well beyond the scope of this thesis. Let us only note that the exact shape of this curve can vary as a function of the way the biases are defined, but shows the same overall pattern: an initial increase followed by a decrease in the size of the gap.

The formulae in Appendix C make it easy to investigate the effects of different parameter settings on the sub-distributions. Therefore, let us briefly discuss the role of parameter i in Prediction 1. Figure 6.11b shows how changing the value of i affects the relationship between bias proportion and the size of the gap between the sub-distributions. The darkness of the lines represents the value of i : darker lines correspond to higher values. The value of i is varied between 1 and 46 in steps of 5. Since higher values entail higher degrees of independence, it is not surprising to see that the overall distance between the sub-distributions increases monotonically as a function of i . It should also be noted that increasing the value of i beyond a certain level diminishes the influence of bias proportion on the size of the gap (this is demonstrated by the increasingly wide plateau around intermediate values of w). Again, this is a relatively straightforward result: if the sub-distributions have a high degree of independence to begin with, a balanced bias proportion cannot enhance it much further. In such cases, the effects of bias proportion are only seen at extreme values of w , where it acts against the independence of the sub-distributions.

Although Prediction 1 has implications for any contextual effect that derives from phonetic biases, the present section offers a narrower interpretation as

well. The model presented above has been developed through the example of voiced and voiceless sub-distributions within vowel categories. Therefore, under a very literal interpretation, its predictions can be taken to relate directly to the voicing effect. Prediction 1 becomes a statement about vowel categories with different proportions of voiced *versus* voiceless forms, which can be investigated experimentally by looking at production data from natural languages. The next section takes up this task and presents evidence for the validity of this prediction from a small cross-linguistic production study.

6.4 AN EMPIRICAL STUDY

Based on the simulation results in the previous section, the following prediction can be made about the size of the voicing effect (cf. Figure 6.11):

Prediction 2 *The voicing effect is stronger in vowel categories where the relative frequencies of voiced and voiceless contexts are balanced than in vowel categories where they are unbalanced.*

Note that this prediction does not say anything new: it is simply a more specific instantiation of Prediction 1. The notion of bias proportion is translated into frequencies, and the abstract statements about sub-distributions are related to observable aspects of vowel categories. Since all the individual components of this prediction are easily quantified, it can be tested experimentally. In order to conduct such a test, two types of data have to be collected: the relative frequencies of voiced and voiceless obstruent environments for different vowels and duration measurements for both environments. The first type of data can be obtained relatively easily from phonologically transcribed frequency dictionaries. The second type of data requires a controlled production experiment focusing on the duration of vowel categories in voiced and voiceless environments.

This section presents a small cross-linguistic experiment that includes both frequency counts from corpora and production data. The main goal of this experiment is to test the validity of Prediction 2. The experiment focuses on three languages: English, French and Hungarian. Section 6.4.1 describes the procedure used to obtain the relevant frequency counts for these languages. Then, Section 6.4.2 provides an outline of the production experiment and presents its main results. The implications of the findings will be discussed in

the concluding section of this chapter along with a summary of the chapter's key arguments.

6.4.1 *Estimating bias proportion from corpus data*

In the context of the voicing effect, the bias proportion of a given vowel category refers to the ratio of tokens with a following voiced obstruent to all tokens with a following obstruent. This value can be estimated from corpus data. To give an example, let us look at the token frequencies of forms containing the FACE vowel (realised as [eɪ] in Southern Standard British English) in the English subsection of the CELEX lexical database (Baayen et al. 1993; the frequency counts have been normalised to 1 million words). The overall token frequency of forms in which the FACE vowel occurs before voiced obstruents is 6,284; for voiceless obstruents, the frequency count is 12,211. We can now estimate the bias proportion by solving $6,284 / (6,284 + 12,210)$, which gives 0.34.

This estimation was performed for an overall 23 vowels from the three languages. The choice of vowels is described in more detail in Section 6.4.2 below (the vowels themselves can also be seen in Figure 6.13 below). The frequency measurements were taken from the following sources:

- the CELEX lexical database for English (based on an 18 million word corpus of English)
- the Lexique database for French (New et al. 2001; based on a 31 million word corpus of French)
- a frequency dictionary for Hungarian described in Grimes (2006) (based on a 188 million word corpus of Hungarian)

All three of these databases contain phonological transcriptions, which can be used in a straightforward way to search for vowels in specific environments. To ensure consistency across the frequency measurements and the production experiment, I restricted the frequency counts to items in which the target vowel is followed by an obstruent within the same syllable. This was necessary since all the stimuli in the production experiment are monosyllabic, which means that the target vowels are always followed by tautosyllabic obstruents (see the next Section for more details).

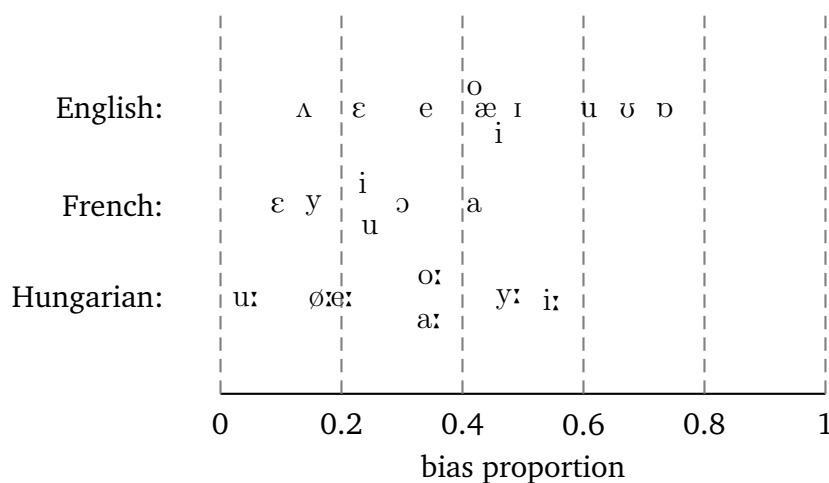


Figure 6.13: The estimated bias proportion values for the vowels from English, French and Hungarian.

Figure 6.13 presents the estimated bias proportions for the vowels grouped by languages. Although the proportions vary quite widely within each language, it is evident that the region representing high values is underrepresented in this sample. This means that any findings about the durations of voiced and voiceless sub-distributions at higher bias proportion values will have to be taken with a pinch of salt, due to the scarcity of relevant data points. The overall proportion of voiced codas is 0.48 in English (an almost balanced proportion), 0.38 in French and 0.33 in Hungarian. Note that these proportions are not particularly meaningful in themselves, as they depend mostly on idiosyncratic facts about lexical distributions. For instance, [ʊ] has a high bias proportion in English. This is likely due to words like *could*, *would* and *should* (with voiced coda consonants), which are very frequent. On the other hand, [Λ] has a low bias proportion. This might be due to the word *but* (with a voiceless coda consonant), which, again, is of very high frequency.

We can now move on to the description of the production experiment. The proportion values estimated in this section will be used to predict the size of the gap between the voiced and voiceless sub-distributions in different vowel categories.

6.4.2 *The production experiment*

Before giving a description of the methods, it is worth looking at the reasons for choosing English, French and Hungarian in a little more detail. The first criterion for selecting the target languages was the availability of phonetically annotated frequency dictionaries, which served as the basis of the estimation of bias proportion values. Clearly, languages without such resources are not suitable for the present experiment. The second criterion was the existence of non-neutralised laryngeal contrasts in coda position. This was motivated by the experiment's focus on vowel length before tautosyllabic obstruents, which, in turn, rests on the observation that length differences are more pronounced in this context than they are before heterosyllabic consonants (Laeufer 1992). The three languages that this experiment focuses on all conform to these criteria. It should be noted that English is a less obvious candidate due to the fact that both voicing and aspiration are used in implementing the contrast between fortis and lenis consonants. However, in Section 6.2 I argued that aspiration plays little role in signalling the laryngeal contrast in coda obstruents. While English does tend to partially devoice final obstruents, prevoicing still seems to be a more reliable cue to the laryngeal contrast than aspiration in this position. Thus, the voicing effect in languages like English is likely to be closely related to the effect seen in French and Hungarian.

PARTICIPANTS The experiment had six participants: two native speakers of Southern British English, two native speakers of French and two native speakers of Hungarian. Special care was taken to recruit participants from the same dialect regions. Thus, both English speakers were from the vicinity of London, both French speakers from the vicinity of Paris and both Hungarian speakers from Budapest. The sample of speakers is relatively balanced in terms of gender: there were two male speakers (one English and one Hungarian) and four female speakers (one English, one Hungarian and two French). All participants had been resident in Edinburgh in the United Kingdom for at least a year at the time of the experiment. There is a possibility that the long-term exposure to English has interfered with the sound patterns of the Hungarian and French speakers. However, it is highly unlikely that this interference would have affected vowel-

specific patterns in the realisation of the voicing effect, given the high degree of dissimilarity among the vowel systems of these languages. It should also be noted that the results of the experiment did not show any obvious signs of convergence (cf. Figure 6.15, which shows the strength of the voicing effect for all the vowel categories from all three languages).

MATERIALS The materials for each language consisted of existing medium-frequency monosyllabic words in which the target vowel was followed by a word-final obstruent. Polysyllabic items were deliberately avoided in order to make sure that stress and the syllabic status of the following consonant (tautosyllabic vs. heterosyllabic) do not interfere with the voicing effect. Although there is no reason to expect that these factors would significantly alter the general patterns reported below, they would likely obscure the results to some extent. The overwhelming majority of the target words was of the shape CVC with a small number of CCVC items. The words were embedded in the following carrier phrases: *I say the word* __ for English, *Je dit le mot* __ for French and *Azt mondom,* __ for Hungarian (the meaning is the same for all three sentences).

As it has been noted in Section 6.4.1 above, there were 23 vowels in the experiment: 10 from English [i, u, ɪ, ʊ, ε, ɒ, e, o, æ, ʌ], 6 from French [i, u, y, ε, ɔ, a] and 7 from Hungarian [i:, u:, e:, a:, o:, y:, ø:]. The choice of vowels in French was relatively straightforward: these are the only six non-nasal stressed vowels that are adequately represented in the environments that this study investigated. The decision procedure was somewhat more complicated for Hungarian. The seven vowels presented above comprise the full set of long vowels in the language. Short vowels were excluded since a previous unpublished study on short [ɒ] and [o] found no consistent length differences. It was expected that long vowels may show more robust contextual effects, since they can lengthen without encroaching on the region of phonetic space occupied by short vowels. Finally, the English set includes both lax [ɪ, ʊ, ε, ɒ, æ, ʌ] and tense [i, u, e, o] vowels. Since length is not the only (and perhaps not even the primary) cue to the tense-lax distinction, it was assumed that the inclusion of both types of vowels would not interfere with the results.

Each vowel category was represented by 20 tokens in the experiment: 10 in voiced contexts and 10 in voiceless contexts. Whenever possible, the list of

		ENGLISH				FRENCH				HUNGARIAN			
		+voi		–voi		+voi		–voi		+voi		–voi	
COR	seed	2	beat	2	vide	2	gîte	2	híd	5	szít	5	
	deed	3	sheet	3	bide	3	site	3					
VEL	league	5	beak	2	figue	3	chic	3	híg	3	sík	3	
			peak	3	ligue	2	pic	2	víg	2	csík	2	

Table 6.1: Example items with an [i] vowel quality from the three languages in the production experiment. COR stands for ‘coronal’ and VEL for ‘velar’. The columns next to the words indicate the number of times the word occurs in the experiment. Note that all the columns add up to ten.

tokens was compiled so that it contained both velar and coronal consonants in postvocalic position. Note that the number of tokens does not equal the number of different words: every word was repeated at least twice in order to obtain a sufficient number of tokens. By way of illustration, Table 6.1 presents the full set of tokens representing the vowels [i], [i] and [i:] from English, French and Hungarian, respectively.¹² In order to make segmentation easier, initial sonorants were avoided. In the few cases where there were no suitable obstruent-initial words, forms beginning with nasals and occasionally laterals were used.

The overall number of target tokens for the different languages is as follows: 200 for English, 120 for French and 140 for Hungarian. The list of items also contained monosyllabic sonorant-final fillers: 100 for English, 60 for French and 70 for Hungarian.

PROCEDURE Participants were seated in a sound-proofed booth with a computer screen and a microphone, and instructed to read out the sentences on the screen at a comfortable pace. Once they indicated that they were ready, they were presented each item in the experiment one by one in random order (the randomisations were different across the participants). Halfway through the experiment, they were asked if they wanted to have a short break. The

12. The IPA transcriptions and the glosses for the French and Hungarian words are as follows: *vide* [vid] ‘empty’, *bide* [bid] ‘belly’, *figue* [fig] ‘fig’, *ligue* [lig] ‘league’, *gîte* [ʒit] ‘shelter’, *site* [sit] ‘site’, *chic* [ʃik] ‘chic’, *pic* [pik] ‘peak’, *híd* [hi:d] ‘bridge’, *híg* [hi:g] ‘thin (e.g. soup)’, *víg* [vi:g] ‘merry’, *szít* [si:t] ‘fan up a fire’, *sík* [ʃi:k] ‘plane’, *csík* [tʃi:k] ‘stripe’.

whole procedure took between 15–25 minutes depending on the language and the subject.

The participants' productions were recorded using a SHURE SM7B cardioid dynamic microphone and saved as 44kHz WAV files.

ANALYSIS In order to take duration measurements, the target vowels had to be extracted from the produced tokens. Segmentation was performed manually in PRAAT (Boersma & Weenink 2009) by inspection of the waveform and the spectrogram. In most cases, both consonants were obstruents, and the beginning and the end of the target vowel could be established with relatively high confidence by looking for (i) the low frequency periodic noise typical of voicing and (ii) clearly visible formant structure. Both of these cues had to be present for identifying a portion of the word as part of the vowel. As a result, the aspiration following the release of voiceless consonants and closure voicing during the following consonant were excluded from the length measurements.¹³ In the case of preceding nasals, the segment boundary could also be seen quite clearly: the appearance of high-amplitude formants after the release of the nasal was abrupt and easy to locate. Laterals presented somewhat more of a challenge, as the major cues for the C–V boundary in this case are gradual changes in formant values and intensity. However, the beginning of the vowel could still be identified with acceptable confidence in most cases (and only a small number of tokens had initial laterals).

Following the manual segmentation of the target tokens, a PRAAT script automatically exported duration values for the vowels and saved them in a database along with further information about each token. This yielded an overall 922 data points (some of the vowel tokens had to be discarded due to production errors or uncertainties in segmentation). The duration values were subsequently standardised for each speaker (i.e. recoded as the number of standard deviations from the sample mean for that speaker) to account for between-speaker differences in tempo. The statistical tests reported in the next section were all performed on these standardised values.

13. In English, a small number of tokens were slightly preaspirated. The aspirated portion was also excluded from the length measurements.

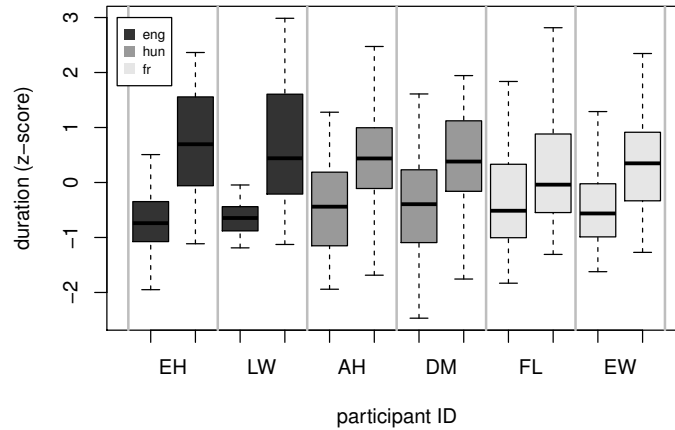


Figure 6.14: *The size of the voicing effect in English, Hungarian and French. The cells separated by vertical lines indicate different speakers. Within each cell, the box on the left corresponds to vowels in a voiceless context, and the box on the right to vowels in a voiced context.*

RESULTS The box plot in Figure 6.14 illustrates the size of the voicing effect in the three different languages. All the subjects show robust differences across the two environments (which are significant by one-tailed t -tests at a level of $p < 0.005$), with the vowels in voiced contexts being longer than the vowels in voiceless contexts. This suggests that the voicing effect is present in all three languages. There are also differences in the robustness of the effect: the size of the gap is much larger in English than it is in Hungarian and French. This is, of course, what one would expect based on the overview in Section 6.2: English is often cited as a language where the voicing effect is particularly robust, and comparisons with languages like French have repeatedly confirmed this observation.

Let us now turn to the interaction between bias proportion and the voicing effect. Figure 6.15 shows the difference in duration between the voiced and the voiceless sub-distributions for each vowel category in the experiment. The direction of the effect is always the same: the vowel is longer in voiced than in voiceless contexts (otherwise the plot would contain negative values as well). The dashed line is a smooth curve based on local polynomial regression (LOESS), and is intended as an indication of the general trend in the data set. A comparison of this curve with those in Figures 6.11b and 6.12 reveals

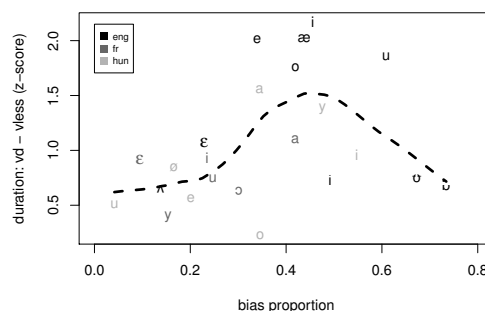


Figure 6.15: The strength of the voicing effect as a function of bias proportion. The x axis corresponds to bias proportion, and the y axis to the durational difference between the voiced and the voiceless sub-distributions. The dashed line is a LOESS-smoothed curve.

that the observed interaction between bias proportion and the voicing effect corresponds closely to the predictions of the model presented in the previous section. The length differences are relatively small when the vowel categories occur predominantly in voiceless contexts, they become greater when the two contexts are balanced in terms of frequency, and begin to fall again as the voiced context comes to dominate the category. Even the location of the turning point ($w = 0.452$) is relatively close to that predicted by the model ($w = 0.394$).

These results strongly suggest that Prediction 2 is valid, but they do not constitute statistical evidence for it. In order to perform a more rigorous test, Prediction 2 has to be translated into a form that allows us to use standard statistical techniques. This is relatively easy: Prediction 2 is a statement about an interaction between the effect of voicing and bias proportion. As such, it can be incorporated into multiple regression analyses as an interaction term. The only challenge in this case is that this interaction is non-linear: between $w = 0$ and $w = 0.45$, an increase in bias proportion enhances the voicing effect, but when $w > 0.45$, this relationship is reversed (as discussed above). One way to account for this reversal in a regression model is to use a technique called breakpoint regression. The details of this method are described in Baayen (2008), so I only give a very broad summary here. The main idea is to fit a different regression line to each half of the data set within a single model. In other words, we need a regression line for vowel productions that have a bias proportion of $w \leq 0.45$ and another regression line for vowel productions with

	ESTIMATE	STD. ERR.	t-VALUE	$p(> t)$
(INTERCEPT)	0.353	0.098	3.586	0.0005*
VOI=yes	1.545	0.090	17.115	< 0.0001*
V.HEIGHT	-0.354	0.032	-10.987	< 0.0001*
PoA=vel	-0.025	0.051	-0.499	0.6182
VOI=no:PROP:w-LOW	-0.066	0.303	-0.219	0.8270
VOI=yes:PROP:w-LOW	3.122	0.304	10.279	< 0.0001*
VOI=no:PROP:w-HIGH	-1.746	0.527	-3.310	0.0010*
VOI=yes:PROP:w-HIGH	-4.829	0.532	-9.081	< 0.0001*

Adjusted R^2 : 0.4193; $p < 0.0001$

Table 6.2: *The results of the multiple regression looking at the effects of bias proportion on the size of the voicing effect.*

$w > 0.45$. This can be achieved by including an interaction with an indicator variable that shows whether $w \leq 0.45$ or $w > 0.45$ (see Baayen 2008: p. 216–217 for more detail). To simplify matters, these regression lines will be marked as *w-LOW* and *w-HIGH*.

The regression model was set up as follows. The dependent variable is DURATION (given in *z*-scores), and the terms in the model are VOI (whether the following consonant is voiced or not), PROP:VOICED:w-LOW/HIGH (the interaction term that represents the influence of bias proportion on the size of the voicing effect), and two variables included for control: V.HEIGHT (a three-valued variable ranging through low–mid–high represented as 1–2–3) and PoA (the place of articulation of the following consonant represented as a nominal variable). The results are presented in Table 6.2. The following observations can be made on the basis of these findings. First, the effect of voicing (VOI) is highly significant. This further strengthens the conclusions made in connection with Figure 6.14. Second, vowel height (V.HEIGHT) has an important role in determining duration values: the lower the vowel, the higher the duration. This result has been reported for a variety of languages in the literature (see e.g. Keating 1984), and is thus not particularly suprising. Third, and most important, the last four rows of the table confirm the suggestions about the validity of Prediction 2: the coefficients that represent the influence of bias proportion on the size of the voicing effect are all significant with the exception of VOI=no:PROP:w-LOW.

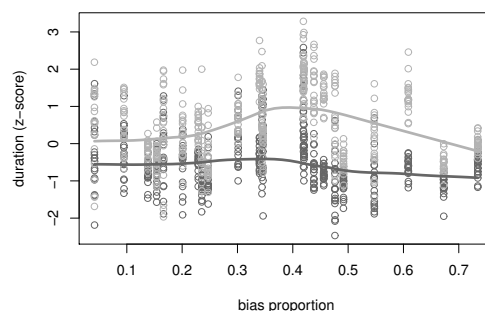


Figure 6.16: The durations of all the tokens in the experiment grouped by voicing. The solid lines are LOESS-smoothed curves for each context (light grey: voiced; dark grey: voiceless).

To make this clearer, Figure 6.16 plots all 922 duration measurements against bias proportion, and shows separate LOESS-smoothed curves for the vowels in voiced (light grey) and voiceless (dark grey) environments. The two lines correspond to $\text{VOI}=\text{yes}:\text{PROP}:w\text{-LOW}/\text{HIGH}$ (light grey) and $\text{VOI}=\text{no}:\text{PROP}:w\text{-LOW}/\text{HIGH}$ (dark grey). Looking at voiceless tokens, Table 6.2 and Figure 6.16 indicate that the duration of vowels before voiceless consonants is not strongly correlated with bias proportion when $w \leq 0.45$, and starts to fall slightly after $w = 0.45$. As for voiced tokens, there is a sharp rise in duration values up to $w = 0.45$, and a similarly sharp fall after $w = 0.45$. The net effect of these tendencies is a wider gap between voiced and voiceless sub-distributions in categories with a balanced bias proportion, and a narrower gap at extreme bias proportion values, in line with Prediction 2.

Figure 6.16 also highlights an area where the results are only in partial agreement with the predictions of the model. Since the curves in the figure illustrate the observed locations of the voiced and voiceless sub-distributions as a function of bias proportion, they can be compared to the predicted curves in Figure 6.11a. The predicted curves can be described as follows.

1. The target durations rise monotonically as a function of bias proportion for both the voiced and the voiceless sub-distributions. This means that vowel categories with a high bias proportion are expected to have higher durations on the whole than categories with a low bias proportion (this is also noted in Section 6.3.3).

2. The curves are not parallel: they start together, become separated around intermediate bias proportion values and then come together again as w approaches 1. This is the behaviour described in Predictions 1 and 2: the size of the gap between the sub-distributions is larger when the bias proportion is balanced.

The empirical curves show the second effect, but not the first one. That is to say, the overall duration of the category does not seem to increase at higher bias proportion values. Note, however, that bias proportion values of over 0.5 are strongly underrepresented in this sample: only 4 vowel categories out of 23 have $w > 0.5$. This means that we cannot make any reliable conclusions about vowel categories with a high bias proportion. Importantly, it is precisely these categories that appear to go against the model predictions outlined above. Thus, we can make the following conclusion about the relationship between the predicted and the observed curves: the model predictions are supported at $w \leq 0.5$, while the results are inconclusive at $w > 0.5$. The fact that the predictions of the model are borne out by lower bias proportion values is an important result in its own right. However, it is clear that this experiment will need to be extended in future research in order to see whether higher bias proportion values are also in agreement with the model.

6.5 CONCLUSIONS

The results presented in the previous section are in agreement with the specific claim about the voicing effect in Prediction 2 and the more general statement in Prediction 1. The strength of the voicing effect was found to vary as a function of how balanced the bias proportion of a given category is. At low bias proportion values, the voiced and the voiceless sub-distributions are relatively close to each other in terms of vowel length. At intermediate bias proportion values the sub-distributions are significantly further apart. Unfortunately, higher bias proportion values are underrepresented in the data set, which means that it is not possible to make any reliable conclusions about their influence on the sub-distributions. While more work is needed to fully confirm the validity of Predictions 1 and 2, the results above are certainly promising.

This brings the main argument of the present chapter to its conclusion. The structure of this argument was as follows. First, I argued that production and perception are influenced by phonetically detailed knowledge about context-specific sub-distributions within categories. This assumption served as the basis of a formal model of the production-perception feedback loop. This model was used to show how the prediction about the influence of lexical factors on the behaviour of sub-distributions (Prediction 1) follows from the theoretical assumptions underpinning the present approach. It also gave rise to a more specific prediction relating to the voicing effect (Prediction 2). The fact that both of these predictions are in agreement with the data demonstrates the viability of this approach and lends support to the underlying theoretical framework.

The argument summarised above has several important implications for the study of sound change. The most specific of these relates to allophonic splits. Although this term was avoided in the above discussion, it is clear that the changes in sub-distributions investigated here are closely related to allophonic splits. In a sense, the predictions discussed in this chapter are about the likelihood of allophonic splits: balanced bias proportion values increase the probability that a given category will undergo a split. While the present chapter did not investigate how this effect is manifested in complex sound systems, it is likely that its overall influence is very similar to that of bias strength. That is, categories with balanced bias proportion values will be more likely to split, but the sound systems containing these categories will often settle into equilibria where other factors make splits impossible.

It should be noted that the present model cannot fully account for splits. First, as in previous simulations, the number of categories is fixed, which means that splits cannot create new contrasts. This is clearly problematic, given that contrasts do sometimes arise as a result of splits. More importantly, the model may not even be able to fully capture the behaviour of allophonic (non-contrastive) splits. The description of the simulation results in Sections 6.3.2 and 6.3.3 clearly indicates that the parameter settings of the model define an upper limit on the amount of separation that is tolerated within a given category. The sub-distributions converge towards this limit over time, but the current model does not predict any further separation once this limit is reached. There

is no indication that allophonic splits are bound in the same way: different allophones of a given sound often go down divergent paths in their evolution and become completely independent over time. However, it was not my goal here to predict every aspect of allophonic splits. The main goal of this chapter was to demonstrate how specific predictions can be derived from the theoretical framework that I argued for in this thesis. The framework could be extended to capture further properties of splits, but this is by no means necessary for our present purposes.

Another implication of the argument in the present chapter relates to the general approach to sound change adopted in this thesis. I have argued throughout the thesis that looking at the evolution of sound systems through a simulated production-perception feedback loop can yield many insights into the nature of sound change beyond those that already exist in the literature. The present chapter is a particularly clear demonstration of this. To my knowledge, the specific relationship between bias proportion and the size of contextual effects assumed here has not been discussed elsewhere in the literature. The fact that the predictions of this model were confirmed by experimental results is a strong argument for its overall validity. I suspect that many more such discoveries can be made by carefully exploring the behaviour of simulated systems and comparing them to patterns observed in natural languages.

Finally, this chapter is a good example of the methodological approach advocated in this thesis. The foundational idea of this approach is that computational simulations can serve as a way of linking abstract theoretical concepts to concrete predictions. All of the studies presented in this thesis rely on this assumption to some extent, but none of them apply it quite as successfully as the present one. The predictions that emerge from the computational model in this chapter can be presented in a very explicit form, and are easy to verify through empirical methods. I believe that this research strategy is a useful model for future investigations of the theoretical predictions of abstract frameworks.

This thesis proposed a solution to the actuation problem in the context of bias-based models. The original issue that motivated this investigation was the inability of bias-based models to account for stability in sound systems, despite their success at capturing cross-linguistic regularities governing sound change. I argued that the problematic predictions of bias-based models derive not from their theoretical assumptions, but from the way they are typically used in studies of sound change. Specifically, the mechanism of sound change in bias-based models tends to overapply because the situations that are usually investigated involve categories in a vacuum. The solution I outlined in this thesis is to shift our attention from single categories to sound systems, where the effects of phonetic biases can be counteracted by systemic pressures.

I suggested that system-based approaches can make more accurate predictions about sound change because of the complexity of the adaptive landscape in which sound systems exist. When the only pressure affecting a sound system (or a single category) is a phonetic bias, it is natural that the outcomes will be simple and deterministic. However, when other pressures such as lexical distributions and functional load are added in, the range of possible outcomes becomes much more diverse, and phonetic biases will not necessarily be satisfied in every case. This was confirmed by the simulations in Chapter 5, which demonstrated that the effects of phonetic biases can be suppressed as the complexity of the system is increased.

I also noted that sound systems move towards peaks in the adaptive landscape, or local optima, where the different pressures balance each other out. As a result, the system-based approach predicts stability. It was shown that this stability can be broken by changes in the pressures that define the adaptive landscape. Thus, an increase or a decrease in functional load or a change in lexical distributions can create a situation where the sound system is knocked

out of an equilibrium and starts evolving towards a new stable state. In essence, the adaptive landscape can create moving targets for the sound system. This ensures that both stability and change are observed.

Although the arguments summarised above describe sound systems in rather abstract terms, the simulations I used to investigate them are rooted in concrete and plausible assumptions about speech production and perception. These assumptions are as follows: (i) speech production and perception are based on probabilistic category representations; (ii) category representations are subject to continuous update throughout the lifetime of an individual; (iii) speech production and perception are affected by low-level universal phonetic biases; and (iv) category update is inhibited in cases where too many ambiguous tokens are produced due to category overlap. All of these assumptions were shown to be supported by independent evidence, which is a strong argument for the general validity of the results derived from them. The simulation architecture used in the thesis is a direct implementation of these principles. To make this point clearer, Chapter 4 demonstrated that certain implementation details of the simulations can be altered without substantially changing their outcomes. Thus, prototype and exemplar models were shown to make essentially identical predictions. Moreover, the abstract model of the production-perception feedback loop used in the thesis did not perform significantly differently from a more realistic model with multiple agents.

Of course, many simplifications had to be made in order to be able to present the main argument of the thesis in a principled and rigorous form. Perhaps the most obvious of these is the decision to keep the number of categories constant for the entire duration of the simulations. The main result of this restriction is that the model cannot provide a fully satisfactory account of mergers and splits (although it can capture some aspects of both; cf. Chapters 5 and 6). It is likely that there are many other situations where this simple architecture cannot be applied without modifications. This, however, is not a major problem. The main strength of this approach lies not in its ability to account for individual changes, but in the rich set of predictions that it makes about cross-linguistic patterns in sound change.

The shift from isolated sound categories to sound systems brings about a change in the way different predictors of sound change are viewed. While

bias-based accounts traditionally focus on phonetic biases, the approach here moves the emphasis to other factors that are often treated as secondary in the literature. The system-based view predicts that lexical distributions, functional load and individual differences in production or perception may be just as important in the actuation of sound change as phonetic pressures. Although some of these factors have been studied as mechanisms that may inhibit or facilitate sound change, the present approach suggests that their role extends beyond that: they may play a fundamental role in initiating changes. This has important implications for the study of sound change. Even if the phonetic roots of sound change may seem easier to study due to their constancy across different languages, non-phonetic factors should be given equal consideration. We will only be able to fully understand sound change if we explore the set of possible predictors and carefully assess their roles.

Chapter 6 of this thesis is a demonstration of how the role of non-phonetic factors can be investigated. This was done by focusing on the influence of bias proportion on the size of allophonic splits within categories. The logical structure of the first half of the chapter was much like that of the rest of the thesis. After an overview of the underlying principles of the model, I presented a simple simulation architecture that was used to explore its predictions. The main prediction that emerged is that allophonic splits will be more robust when the proportion of biased *versus* non-biased environments is balanced for a given category. This prediction was then tested by looking at the magnitude of the effect of voicing on vowel length in categories with different bias proportions. The data that served as the basis of this investigation came from a small production study involving English, French and Hungarian. A statistically significant correlation was found, which means that the predictions of the bias-based approach are borne out by this particular data set. The contributions of this small study are two-fold. First, it finds an interesting effect that has not been discussed elsewhere in the literature. Second, it illustrates how the abstract framework investigated in this thesis can be translated into substantive predictions, and how these predictions can be tested empirically. The success of this preliminary endeavour shows that the approach advocated in the previous chapters can contribute significantly to our understanding of sound change.

APPENDIX: VARIANCE INFLATION



In what follows, I provide mathematical derivations of the claims about the variance-inflating effects of (i) kernel density estimation and (ii) adding Gaussian noise to a parametric distribution.

The discussion of kernel density estimation below is based mostly on Hansen (2009). We start with a sample of observations $X = \{x_1, x_2, \dots, x_n\}$. The kernel density estimate of this sample can be written as follows (using equations (4.3) and (4.4)):

$$\hat{f}(x) = \frac{1}{\sum_{j=1}^n w_j} \sum_{k=1}^n \frac{w_k}{h} \phi\left(\frac{x - x_k}{h}\right) \quad (\text{A.1})$$

We are interested in the variance of this distribution. In order to calculate it, we first have to obtain the first and the second moments of this function. The calculation of the first moment is shown below, using the following change of variables: $u = (x - x_i)/h$.

$$\begin{aligned} \int_{-\infty}^{\infty} x \hat{f}(x) dx &= \int_{-\infty}^{\infty} \frac{1}{\sum_{j=1}^n w_j} \sum_{k=1}^n x \frac{w_k}{h} \phi\left(\frac{x - x_k}{h}\right) dx \\ &= \frac{1}{\sum_{j=1}^n w_j} \sum_{k=1}^n \int_{-\infty}^{\infty} (x_k + uh) w_k \phi(u) du \\ &= \frac{1}{\sum_{j=1}^n w_j} \sum_{k=1}^n w_k \left[\int_{-\infty}^{\infty} x_k \phi(u) du + \int_{-\infty}^{\infty} uh \phi(u) du \right] \end{aligned}$$

$$\begin{aligned}
&= \frac{1}{\sum_{j=1}^n w_j} \sum_{k=1}^n w_k \left[x_k \int_{-\infty}^{\infty} \phi(u) du + h \int_{-\infty}^{\infty} u \phi(u) du \right] \\
&= \frac{1}{\sum_{j=1}^n w_j} \sum_{k=1}^n w_k x_k
\end{aligned} \tag{A.2}$$

Note that $\int_{-\infty}^{\infty} u \phi(u) du = 0$, since the left-hand side of the equation refers to the mean of a standard normal distribution, which, by definition, is 0. If all the weights are the same, the first moment of the distribution is the sample mean; otherwise, (A.2) is equivalent to the mean of a discrete probability distribution, which in this case consists of exemplars with different activation levels (for the notion of activation levels, see Pierrehumbert 2001 and Section 4.1.3 of this thesis). The second moment can then be calculated as follows:

$$\begin{aligned}
\int_{-\infty}^{\infty} x^2 \hat{f}(x) dx &= \int_{-\infty}^{\infty} \frac{1}{\sum_{j=1}^n w_j} \sum_{k=1}^n x^2 \frac{w_k}{h} \phi\left(\frac{x - x_k}{h}\right) dx \\
&= \frac{1}{\sum_{j=1}^n w_j} \sum_{k=1}^n \int_{-\infty}^{\infty} (x_k + uh)^2 w_k \phi(u) du \\
&= \frac{1}{\sum_{j=1}^n w_j} \sum_{k=1}^n w_k \left[x_k^2 \int_{-\infty}^{\infty} \phi(u) du + 2hx_k \int_{-\infty}^{\infty} u \phi(u) du \right. \\
&\quad \left. + h^2 \int_{-\infty}^{\infty} u^2 \phi(u) du \right] \\
&= \frac{1}{\sum_{j=1}^n w_j} \sum_{k=1}^n w_k x_k^2 + h^2
\end{aligned} \tag{A.3}$$

In this case, $\int_{-\infty}^{\infty} u^2 \phi(u) du$ is the variance of a standard normal distribution, which is 1 by definition. We are now in a position to calculate the second central moment (i.e. the variance):

$$\begin{aligned}
&\int_{-\infty}^{\infty} x^2 \hat{f}(x) dx - \left(\int_{-\infty}^{\infty} x \hat{f}(x) dx \right)^2 = \\
&= \frac{1}{\sum_{j=1}^n w_j} \sum_{k=1}^n w_k x_k^2 - \left(\frac{1}{\sum_{j=1}^n w_j} \sum_{k=1}^n w_k x_k \right)^2 + h^2 =
\end{aligned}$$

$$\begin{aligned}
&= \frac{1}{\sum_{j=1}^n w_j} \sum_{k=1}^n w_k x_k^2 - \left(\frac{1}{\sum_{j=1}^n w_j} \sum_{k=1}^n w_k x_k \right)^2 = \\
&= \hat{\sigma}_X^2 + h^2
\end{aligned} \tag{A.4}$$

The term $\hat{\sigma}_X^2$ refers either to the variance of the original sample (if all the weights are equal) or to the variance of the discrete probability distribution representing memory activation values. We can thus conclude that kernel density estimation inflates the variance of the sample by h^2 , which is the variance of the Gaussian kernel used in modelling random noise in production.

The second point that needs to be dealt with in this appendix is the effect of adding random noise to a normal distribution. This is a much more trivial task, given the following identity describing the sums of normally distributed random variables (where $X \sim \mathcal{N}(\mu_X, \sigma_X^2)$ and $Y \sim \mathcal{N}(\mu_Y, \sigma_Y^2)$):

$$X + Y \sim \mathcal{N}(\mu_X + \mu_Y, \sigma_X^2 + \sigma_Y^2) \tag{A.5}$$

That is, the mean and the variance of a random variable that results from adding two independent normally distributed random variables are simply the sums of the means and the variances of the original random variables. Since adding Gaussian noise to a production target coming from another Gaussian distribution (i.e. the category representation) is equivalent to summing two normally distributed random variables, the size of the variance inflation will be equivalent to the variance of the Gaussian noise function, which, again, could be represented by h^2 .

The above derivations should make it clear that there is a clear correspondence between variance inflation in non-parametric density estimates and parametric density estimates with Gaussian noise: in each case, the size of the inflation is the same as the variance of a Gaussian function. Since the simulations in this thesis use Gaussians to represent noise both in exemplar and prototype models, we can conclude that the dynamics of variance inflation are the same in these two models.

APPENDIX: THE OBSERVED DISTRIBUTION

B

This brief section explains how the observed distribution for a given category can be calculated from the underlying distribution if we know the location of all the other categories and the parameters of the model. There are four factors that can affect the observed distribution: variance inflation through production noise, phonetic biases, ambiguity-driven dispersion and the boundaries of phonetic space. Each of these are discussed below. Note that only the univariate case is described, but the method extends straightforwardly to multivariate distributions as well.

The influence of production noise on the observed distribution is easy to model, and has already been discussed in Appendix A. The following equation can be used to calculate the observed distribution based on (i) a normal distribution representing the original category c_i (with parameters μ_i and σ_i^2) and (ii) a Gaussian noise function (with a variance of h^2):

$$p(x|c_i, h^2) \sim \mathcal{N}(\mu_i, \sigma_i^2 + h^2) \quad (\text{B.1})$$

That is, the mean of the observed distribution is the same as that of the original one, while the variance is increased by h^2 .

The shifts caused by phonetic biases have been described in Section 4.1.2. The equation for calculating the observed distribution is repeated below (where f_i is the probability density function for category i and $\text{bias}_j(\bullet)$ is the logistic function representing bias j):

$$p(x|c_i, \text{bias}_j(\bullet)) = f_i(2x - \text{bias}_j(x)) \quad (\text{B.2})$$

When the original category is represented by a normal distribution, this corresponds to a small shift in the mean of the distribution towards the bias attractor, and a small reduction in the variance.

The effects of ambiguity-driven misperception can be predicted by constructing a fitness function from the function that specifies the probability of misperception. As it has been noted in Section 4.1.3, the probability that a token is fed back into the original category representation is higher if it is unambiguous, and lower if it is ambiguous. The function that specifies the ‘survival rate’ of tokens for a given category c_i can be written as follows:

$$f_s(x) = 1 - r p(\neg c_i | x), \quad (\text{B.3})$$

where $f_s(x)$ denotes the survival function, r is the misperception rate and $p(\neg c_i | x)$ is the probability of misperception (cf. Section 4.1.3). The observed distribution is obtained by multiplying the original distribution by the fitness function and re-normalising:

$$p(x | c_i, f_s(\bullet)) = \frac{f_s(x) p(x | c_i)}{\int_{-\infty}^{\infty} f_s(x) p(x | c_i) dx} \quad (\text{B.4})$$

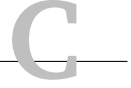
The boundaries of phonetic space are modelled by discarding tokens outside the boundaries, which relies on the same principle as the ambiguity filter. Therefore, equation (B.4) also applies to this case. The fitness function for the boundaries of phonetic space is a simple rectangular window:

$$f_s(x) = \begin{cases} 1 & \text{if } a \leq x \leq b \\ 0 & \text{otherwise,} \end{cases} \quad (\text{B.5})$$

where a is the lower boundary of phonetic space and b the upper boundary.

These functions yield a series of transformations that define the observed distribution, when applied in the following order: (B.1) (variance inflation) \rightarrow (B.2) (phonetic bias) \rightarrow (B.4) using (B.3) (ambiguity filter) \rightarrow (B.4) using (B.5) (boundaries of phonetic space).

APPENDIX: TARGET LOCATIONS



Let us start by deriving the expected values of the sampling distributions for vowels in voiced and voiceless environments (this corresponds to the shift illustrated in Figure 6.10b). The random variables for vowels in voiced/voiceless contexts are denoted X_{vd} and X_{vl} , respectively. Using the formulae in (6.2) and (6.3) and the definition of expected value, the following formulae can be obtained ($E_s[X]$ stands for the expected value of the sampling distribution belonging to a target sub-distribution X):

$$\begin{aligned} E_s[X_{vd}] &= \int_{-\infty}^{\infty} x \frac{i w \mathcal{N}(x|\mu_{vd}, \sigma_{vd}^2) + (1-w) \mathcal{N}(x|\mu_{vl}, \sigma_{vl}^2)}{i w + (1-w)} dx = \\ &= \frac{i w E[X_{vd}] + (1-w) E[X_{vl}]}{i w + (1-w)} \end{aligned} \quad (C.1)$$

$$\begin{aligned} E_s[X_{vl}] &= \int_{-\infty}^{\infty} x \frac{w \mathcal{N}(x|\mu_{vd}, \sigma_{vd}^2) + i(1-w) \mathcal{N}(x|\mu_{vl}, \sigma_{vl}^2)}{w + i(1-w)} dx = \\ &= \frac{w E[X_{vd}] + i(1-w) E[X_{vl}]}{w + i(1-w)} \end{aligned} \quad (C.2)$$

These formulae use the parameters introduced in Section (6.3): i stands for the extra weight that the target sub-distribution has in determining the production target and w is the proportion of tokens in a voiced context within the overall category.

The next step involves deriving the other type of shift in expected values due to phonetic biases (cf. Figure 6.10a). Let us assume a bias b with strength s , and

a hypothetical sampling distribution X with an expected value of $E[X]$. Since the bias is essentially a linear transformation of a random variable, the following can be written (based on equation (6.4); $E_b[X]$ represents the expected value of a sampling distribution X after the application of biases):

$$\begin{aligned}
 E_b[\text{bias}(X, b, s)] &= E[X + s(b - X)] = \\
 &= E[X] + E[s(b - X)] = \\
 &= E[X] + s b - s E[X] = \\
 &= s b + (1 - s) E[X]
 \end{aligned} \tag{C.3}$$

The question we are trying to answer is as follows: what are the values of $E[X_{vd}]$ and $E[X_{vl}]$ at which the expected values of the observed distributions calculated from X_{vd} and X_{vl} (through the application of (C.1)/(C.2) and (C.3)) are also $E[X_{vd}]$ and $E[X_{vl}]$, respectively? In other words, what are the target locations at which the sub-distributions are not predicted to move any further under the influence of the biases and the cohesive forces within the category? To mark the target locations off from simple expected values, they will be represented by $E_t[X_{vd}]$ and $E_t[X_{vl}]$. The equations in (C.4) and (C.5) express the question above in formal terms. Note that the expression related to the voiced sub-distribution contains two biases (b_1 , the shortening bias, and b_2 , the lengthening bias), while the one related to voiceless sub-distributions only a single bias (b_1).

$$E_t[X_{vd}] = s_1 b_1 + s_2 b_2 + (1 - s_1 - s_2) \frac{i w E_t[X_{vd}] + (1 - w) E_t[X_{vl}]}{i w + (1 - w)} \tag{C.4}$$

$$E_t[X_{vl}] = s_1 b_1 + (1 - s_1) \frac{w E_t[X_{vd}] + i(1 - w) E_t[X_{vl}]}{w + i(1 - w)} \tag{C.5}$$

These equations can be rearranged so that $E_t[X_{vd}]$ and $E_t[X_{vl}]$ each appear only on one side.

$$E_t[X_{vd}] = \frac{i w (s_1 b_1 + s_2 b_2) + (1 - w) [s_1 b_1 + s_2 b_2 + (1 - s_1 - s_2) E_t[X_{vl}]]}{i w (s_1 + s_2) + 1 - w} \quad (C.6)$$

$$E_t[X_{vl}] = \frac{i (1 - w) s_1 b_1 + w [s_1 b_1 + (1 - s_1) E_t[X_{vd}]]}{i (1 - w) s_1 + w} \quad (C.7)$$

Crucially, we are seeking a combination of $E_t[X_{vd}]$ and $E_t[X_{vl}]$ values at which both equations hold. Therefore, we need to rewrite each of them in the following way. In equation (C.6), we substitute (C.7) for $E_t[X_{vl}]$, and in equation (C.7), we substitute (C.6) for $E_t[X_{vd}]$. This yields the following equations.

$$E_t[X_{vd}] = \frac{i w (s_1 b_1 + s_2 b_2) + (1 - w)(s_1 b_1 + s_2 b_2)}{i w (s_1 + s_2) + 1 - w} + \frac{(1 - w)(1 - s_1 - s_2) \frac{i (1 - w) s_1 b_1 + w [s_1 b_1 + (1 - s_1) E_t[X_{vd}]]}{i (1 - w) s_1 + w}}{i w (s_1 + s_2) + 1 - w} \quad (C.8)$$

$$E_t[X_{vl}] = \frac{i (1 - w) s_1 b_1 + w s_1 b_1}{i (1 - w) s_1 + w} + \frac{w (1 - s_1) \frac{i w (s_1 b_1 + s_2 b_2) + (1 - w) [s_1 b_1 + s_2 b_2 + (1 - s_1 - s_2) E_t[X_{vl}]]}{i w (s_1 + s_2) + 1 - w}}{i (1 - w) s_1 + w} \quad (C.9)$$

The last step in deriving the formulae presented in Section 6.3.3 is a simple algebraic transformation: the equations are rearranged so that their subjects – $E_t[X_{vd}]$ and $E_t[X_{vl}]$ – only appear on the left hand side. The resulting formulae are shown on the next page in landscape format, along with a derived formula for calculating the gap between the two sub-distributions.

$$E_t[X_{vd}] = \frac{[i(1-w)s_1 + w](i w + 1 - w)(s_1 b_1 + s_2 b_2) + (1-w)(1-s_1-s_2)(i(1-w) + w)s_1 b_1}{[i(1-w)s_1 + w][i w (s_1 + s_2) + 1 - w] - w(1-w)(1-s_1-s_2)(1-s_1)} \quad (C.10)$$

$$E_t[X_{vl}] = \frac{[i w (s_1 + s_2) + 1 - w](i(1-w) + w)s_1 b_1 + w(1-s_1)(i w + 1 - w)(s_1 b_1 + s_2 b_2)}{[i(1-w)s_1 + w][i w (s_1 + s_2) + 1 - w] - w(1-w)(1-s_1-s_2)(1-s_1)} \quad (C.11)$$

$$E_t[X_{vd}] - E_t[X_{vl}] = \frac{[i(1-w) + w](i w + 1 - w)s_1 s_2 (b_2 - b_1)}{[i(1-w)s_1 + w][i w (s_1 + s_2) + 1 - w] - w(1-w)(1-s_1-s_2)(1-s_1)} \quad (C.12)$$

BIBLIOGRAPHY

- Ashby, F. G. & L. A. Alfonso-Reese (1995). Categorization as probability density estimation. *Journal of Mathematical Psychology* 39, pp. 216–233.
- Ashby, F. G. & W. T. Maddox (1993). Relations between prototype, exemplar and decision bound models of categorization. *Journal of Mathematical Psychology* 37, pp. 372–400.
- Baayen, R. H. (2008). *Analyzing linguistic data: A practical introduction to statistics using R*. Cambridge University Press, Cambridge.
- Baayen, R. H., R. Piepenbrock & H. van Rijn (1993). The CELEX Lexical Database. University of Pennsylvania, Philadelphia, PA: Linguistic Data Consortium.
- Babel, M. E. (2009). *Phonetic and social selectivity in speech accommodation*. Ph.D. thesis, University of California, Berkeley.
- Baker, A. (2008). Addressing the actuation problem with quantitative models of sound change. *University of Pennsylvania Working Papers in Linguistics* 14:1, pp. 29–41.
- Baker, A., D. Archangeli & J. Mielke (2011). Variability in American English s-retraction suggests a solution to the actuation problem. *Language Variation and Change* 23, pp. 347–374.
- Bane, M., P. Graff & M. Sonderegger (2010). Longitudinal phonetic variation in a closed system. *Proceedings of the 46th Annual Meeting of the Chicago Linguistics Society*.

- Beddor, P. S. & R. A. Krakow (1999). Perception of coarticulatory nasalization by speakers of English and Thai: Evidence for partial compensation. *The Journal of the Acoustical Society of America* 106:5, pp. 2868–2887.
- Beddor, P. S., J. D. Harnsberger & S. Lindemann (2002). Language-specific patterns of vowel-to-vowel coarticulation: acoustic structures and their perceptual correlates. *Journal of Phonetics* 30:4, pp. 591–627.
- Belasco, S. (1953). The influence of force of articulation of consonants on vowel duration. *The Journal of the Acoustical Society of America* 25:5, pp. 1015–1016.
- Bermúdez-Otero, R. (2007). Diachronic phonology. de Lacy, P. (ed.), *The Cambridge handbook of phonology*, Cambridge University Press, Cambridge, pp. 497–517.
- Bermúdez-Otero, R. (to appear). Amphichronic explanation and the life cycle of phonological processes. *The Oxford handbook of historical phonology*, Oxford University Press, Oxford.
- Blaho, S. (2008). *The syntax of phonology: A radically substance-free approach*. Ph.D. thesis, University of Tromsø.
- Blevins, J. (2004). *Evolutionary Phonology: The Emergence of Sound Patterns*. Cambridge University Press, Cambridge.
- Blevins, J. (2006). A theoretical synopsis of evolutionary phonology. *Theoretical Linguistics* 32:2, pp. 117–166.
- Blevins, J. & A. Garrett (2004). The evolution of metathesis. Hayes, B., R. Kirchner & D. Steriade (eds.), *Phonetically driven phonology*, Cambridge University Press, Cambridge, pp. 117–156.
- Blevins, J. & A. B. Wedel (2009). Inhibited sound change: An evolutionary approach to lexical competition. *Diachronica* 26:2, pp. 143–183.
- de Boer, B. (2001). *The origins of vowel systems*. Oxford University Press, Oxford.
- Boersma, P. & S. Hamann (2008). The evolution of auditory dispersion in bidirectional constraint grammars. *Phonology* 25:2, pp. 217–270.

- Boersma, P. & D. Weenink (2009). Praat: doing phonetics by computer (version 5.0.38) [computer program]. Version 5.1.17; Retrieved on 29/09/2009 from <http://www.praat.org/>.
- Browman, C. P. & L. M. Goldstein (1992). Articulatory phonology: An overview. *Phonetica* 49:3-4, pp. 155–180.
- Buder, E. H. & C. Stoel-Gammon (2002). American and Swedish children's acquisition of vowel duration: Effects of vowel identity and final stop voicing. *The Journal of the Acoustical Society of America* 111:4, pp. 1854–1864.
- Bybee, J. (2007). *Frequency of use and the organization of language*. Oxford University Press, Oxford.
- Bybee, J. L. (2001). *Phonology and language use*. Cambridge University Press, Cambridge.
- Campbell, L. (1998). *Historical linguistics: an introduction*. University of Edinburgh Press, Edinburgh.
- Campos-Astorkiza, J. R. (2007). *Minimal contrast and the phonology-phonetics interaction*. Ph.D. thesis, University of Southern California.
- Chambers, J. K. & P. Trudgill (1980). *Dialectology*. Cambridge University Press, Cambridge.
- Chen, M. (1970). Vowel length variation as a function of the voicing of the consonant environment. *Phonetica* 22:3, pp. 129–159.
- Choi, J. D. & P. Keating (1991). Vowel-to-vowel coarticulation in Slavic languages. *UCLA Working Papers in Phonetics* 78, pp. 78–86.
- Chomsky, N. (1980). *Rules and Representations*. Basil Blackwell, Oxford.
- Chomsky, N. & M. Halle (1968). *The Sound Pattern of English*. Harper & Row, New York, Evanston, London.
- Christiansen, M. & N. Chater (2008). Language as shaped by the brain. *Behavioral and Brain Sciences* 31, pp. 489–558.

- Clarke, S., F. Elms & A. Youssef (1995). The third dialect of English: Some Canadian evidence. *Language Variation and Change* 7:02, pp. 209–228.
- Coleman, J. (2000). Commentary: Where is coarticulation? Broe, M. B. & J. B. Pierrehumbert (eds.), *Papers in Laboratory Phonology V: Acquisition and the Lexicon*, Cambridge University Press, Cambridge, pp. 102–117.
- Croft, W. (2000). *Explaining language change: An evolutionary approach*. Pearson Education, Harlow, UK.
- Davis, S. & W. V. Summers (1989). Vowel length and closure duration in word-medial vc sequences. *The Journal of Phonetics* 17, pp. 339–353.
- Dawkins, R. (1982). *The extended phenotype*. Oxford University Press, Oxford.
- Devroye, L. (1986). *Non-uniform Random Variate Generation*. Springer Verlag, New York.
- Elert, C.-C. (1964). *Phonologic studies of quantity in Swedish*. Skriptor, Uppsala.
- Evans, B. & P. Iverson (2007). Plasticity in vowel perception and production: A study of accent change in young adults. *The Journal of the Acoustical Society of America* 121:6, pp. 3814–3826.
- Feldman, N. H., T. L. Griffiths & J. L. Morgan (2009). The influence of categories on perception: Explaining the perceptual magnet effect as optimal statistical inference. *Psychological Review* 116:4, pp. 752–782.
- Flemming, E. (2001). Scalar and categorical phenomena in a unified model of phonetics and phonology. *Phonology* 18, pp. 7–44.
- Flemming, E. (2004). Contrast and perceptual distinctiveness. Hayes, B., R. Kirchner & D. Steriade (eds.), *Phonetically-based Phonology*, Cambridge University Press, pp. 232–276.
- Forster, K. I. (1994). Computational modeling and elementary process analysis in visual word recognition. *Journal of Experimental Psychology: Human Perception and Performance* 20, pp. 1292–1310.

- Fowler, C. A. (2010). The reality of phonological forms: a reply to Port. *Language Sciences* 32:1, pp. 56–59.
- Frisch, S. A. & R. Wright (2002). The phonetics of phonological speech errors: An acoustic analysis of slips of the tongue. *Journal of Phonetics* 30:2, pp. 139–162.
- Fromkin, V. (1971). The non-anomalous nature of anomalous utterances. *Language* pp. 27–52.
- Garrett, A. & K. Johnson (2013). Phonetic bias in sound change. Yu, A. C. L. (ed.), *Origins of sound change: Approaches to phonologization*, Oxford University Press, Oxford, pp. 51–97.
- Gendrot, C. & M. Adda-Decker (2007). Impact of duration and vowel inventory size on formant values of oral vowels: an automated formant analysis from eight languages. *Proceedings of ICPHS XVI, Saarbrücken, Germany*, pp. 1417–1420.
- Goldinger, S. D. (1996). Words and voices: Episodic traces in spoken word identification and recognition memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 22:5, pp. 1166–1183.
- Goldinger, S. D. (1998). Echoes of echoes? an episodic theory of lexical access. *Psychological Review* 105, pp. 251–279.
- Goldsmith, J. A. (1979). *Autosegmental phonology*. Garland, New York.
- Greenlee, M. & J. J. Ohala (1980). Phonetically motivated parallels between child phonology and historical sound change. *Language Sciences* 2, pp. 283–308.
- Gregory, S. W., S. Webster & G. Huang (1993). Voice pitch and amplitude convergence as a metric of quality in dyadic interviews. *Language and Communication* 13, pp. 195–217.
- Grimes, S. (2006). On the creation of a pronunciation dictionary for Hungarian. Midwest Computational Linguistics Colloquium 2006. Urbana, Illinois.

- Guy, G. R. (2003). Variationist approaches to phonological change. Joseph, B. D. & R. D. Janda (eds.), *The handbook of historical linguistics*, Blackwell, Malden, MA & Oxford, pp. 343–368.
- Hale, M. (2003). Neogrammarian sound change. Joseph, B. D. & R. D. Janda (eds.), *The handbook of historical linguistics*, Blackwell, Malden, MA & Oxford, pp. 343–368.
- Hale, M., M. Kissock & C. Reiss (2007). Microvariation, variation and the features of universal grammar. *Lingua* 117, pp. 645–665.
- Hansen, B. E. (2009). Lecture notes on nonparametrics. Unpublished lecture notes, University of Wisconsin.
- Hansson, G. Ó. (2008). Diachronic explanations of sound patterns. *Language and Linguistics Compass* 2:5, pp. 859–893.
- Harrington, J. (2006). An acoustic analysis of ‘happy-tensing’ in the Queen’s Christmas broadcasts. *Journal of Phonetics* 34:4, pp. 439–457.
- Harrington, J. (2007). Evidence for a relationship between synchronic variability and diachronic change in the Queen’s annual Christmas broadcasts. Cole, J. & J. Hualde (eds.), *Laboratory Phonology* 9, Mouton de Gruyter, pp. 125–143.
- Harrington, J., S. Palethorpe & C. Watson (2000). Monophthongal vowel changes in Received Pronunciation: An acoustic analysis of the Queen’s Christmas broadcasts. *Journal of the International Phonetic Association* 30:1-2, pp. 63–78.
- Harrington, J., F. Kleber & U. Reubold (2008). Compensation for coarticulation, /u/-fronting, and sound change in Standard Southern British: an acoustic and perceptual study. *Journal of the Acoustical Society of America* 123, pp. 2825–2835.
- Harrington, J., P. Hoole, F. Kleber & U. Reubold (2011). The physiological, acoustic, and perceptual basis of high back vowel fronting: Evidence from German tense and lax vowels. *Journal of Phonetics* 39, pp. 121–131.

- Hawkins, S. (2003). Roles and representations of systematic fine phonetic detail in speech understanding. *Journal of Phonetics* 31:3-4, pp. 373–405.
- Hay, J. & A. Sudbury (2005). How rhoticity became /r/-sandhi? *Language* 81, pp. 799–823.
- Helgason, P. & C. Ringen (2008). Voicing and aspiration in Swedish stops. *Journal of Phonetics* 36:4, pp. 607–628.
- Hillenbrand, J., L. A. Getty, M. J. Clark & K. Wheeler (1995). Acoustic characteristics of American English vowels. *Journal of the Acoustical Society of America* 97, pp. 3099–3111.
- Hoenigswald, H. M. (1960). *Language change and linguistic reconstruction*. University of Chicago Press, Chicago.
- Hombert, J.-M. (1978). Consonant types, vowel quality, and tone. Fromkin, V. (ed.), *Tone: A linguistic survey*, New York: Academic Press, pp. 77–111.
- Honeybone, P. (2006). Review of *Optimality Theory and language change* [book by D. Eric Holt (ed.)]. *Journal of Linguistics* 42, pp. 726–731.
- Honeybone, P. & J. Salmons (to appear). Introduction: key questions for historical phonology. Honeybone, P. & J. Salmons (eds.), *The Oxford handbook of historical phonology*, Oxford University Press, Oxford.
- Hoole, P., N. Nguyen-Trong & W. Hardcastle (1993). A comparative investigation of coarticulation in fricatives: Electropalatographic, electromagnetic, and acoustic data. *Language and Speech* 36:2-3, pp. 235–260.
- House, A. S. & G. Fairbanks (1953). The influence of consonant environment upon the secondary acoustical characteristics of vowels. *The Journal of the Acoustical Society of America* 25:1, pp. 105–113.
- Hussein, L. (1994). *Voicing-dependent vowel duration in standard Arabic and its acquisition by adult American students*. Ph.D. thesis, Ohio State University.
- Hyman, L. M. (1975). *Phonology: theory and analysis*. Holt, Rinehart & Winston, New York.

- Hyman, L. M. (1976). Phonologization. Juilland, A. (ed.), *Linguistic studies offered to Joseph Greenberg*, Anma Libri, Saratoga, CA, vol. 2, pp. 407–418.
- Iverson, P. & P. K. Kuhl (1995). Mapping the perceptual magnet effect for speech using signal detection theory and multidimensional scaling. *The Journal of the Acoustical Society of America* 97:1, pp. 553–562.
- Jakobson, R. (1942/1968). *Child language, aphasia, and phonological universals* (A. R. Keiler, Trans.). Mouton, The Hague.
- Jakobson, R., C. G. M. Fant & M. Halle (1952). *Preliminaries to speech analysis: The distinctive features and their correlates*. MIT Press, Cambridge, MA.
- Javkin, H. R. (1976). The perceptual basis of vowel duration differences associated with the voiced/voiceless distinction. *Report of the Phonology Laboratory, UC Berkeley* 1, pp. 78–92.
- Johnson, K. (1997). Speech perception without speaker normalization: An exemplar model. Johnson, K. & J. W. Mullennix (eds.), *Talker variability in speech processing*, Academic Press, San Diego, CA, pp. 145–165.
- Kahn, D. (1976). *Syllable-based generalisations in English phonology*. Ph.D. thesis, Massachusetts Institute of Technology.
- Keating, P. (1984). Universal phonetics and the organization of grammars. *UCLA Working Papers in Phonetics* 59, pp. 35–49.
- Keating, P. (1990a). Phonetic representations in a generative grammar. *Journal of phonetics* 18:3, pp. 321–334.
- Keating, P. (1990b). The window model of coarticulation: articulatory evidence. Kingston, J. & M. Beckman (eds.), *Papers in laboratory phonology I*, Cambridge University Press, Cambridge, pp. 451–470.
- Kenstowicz, M. (1994). *Phonology in Generative Grammar*. Blackwell, Oxford & Cambridge, Mass.
- Kess, J. F. (1992). *Psycholinguistics: psychology, linguistics, and the study of natural language*. John Benjamins, Amsterdam.

- Kingston, J. & R. L. Diehl (1994). Phonetic knowledge. *Language* 70:3, pp. 419–454.
- Kirby, J. (2010). *Cue selection and category restructuring in sound change*. Ph.D. thesis, University of Chicago.
- Kirby, J. (in press). The role of probabilistic enhancement in phonologization. Yu, A. (ed.), *Origin of Sound Patterns: Approaches to Phonologization*, Oxford University Press, Oxford.
- Kirchner, R., R. K. Moore & T.-Y. Chen (2010). Computing phonological generalization over real speech exemplars. *Journal of Phonetics* 38, pp. 540–547.
- Kluender, K. R., R. L. Diehl & B. A. Wright (1988). Vowel-length differences before voiced and voiceless consonants: An auditory explanation. *Journal of Phonetics* 16:2, pp. 153–169.
- Kornai, A. (1996). Analytic models in phonology. Durand, J. & B. Laks (eds.), *Current Trends in Phonology: Models and Methods*, European Studies Research Institute, University of Salford, Salford, UK, pp. 395–418.
- Kuhl, P. K. (1991). Human adults and human infants show a perceptual magnet effect for the prototypes of speech categories, monkeys do not. *Attention, Perception, & Psychophysics* 50:2, pp. 93–107.
- Labov, W. (1994). *Principles of Linguistic Change. Vol. 1: Internal Factors*. Blackwell, Oxford.
- Labov, W. (2001). *Principles of Linguistic Change. Vol. 2: Social Factors*. Blackwell, Oxford.
- Labov, W. (2002). Driving forces in linguistic change. *Proceedings of the 2002 International Conference on Korean Linguistics, Seoul National University*.
- Ladefoged, P. & K. Johnson (2010). *A course in phonetics*. Wadsworth, Boston.
- Laeuffer, C. (1992). Patterns of voicing-conditioned vowel duration in French and English. *Journal of Phonetics* 20:4, pp. 411–40.

- Lass, R. (1997). *Historical linguistics and language change*. Cambridge University Press, Cambridge.
- Levelt, W. J. M., A. Roelofs & A. S. Meyer (1999). A theory of lexical access in speech production. *Behavioral and Brain Sciences* 22, pp. 1–75.
- Liberman, A. M., K. S. Harris, H. S. Hoffmann & B. C. Griffith (1957). The discrimination of speech sounds within and across phoneme boundaries. *Journal of Experimental Psychology* 54:5, pp. 358–368.
- Liljencrants, J. & B. Lindblom (1972). Numerical simulation of vowel quality systems: the role of perceptual contrast. *Language* 48, pp. 839–862.
- Lindblom, B. (1963). Spectrographic study of vowel reduction. *The journal of the Acoustical society of America* 35:11, pp. 1773–1781.
- Lindblom, B. (1990). Explaining phonetic variation: a sketch of the H&H theory. Hardcastle, W. J. & A. Marchal (eds.), *Speech Production and Speech Modelling*, Kluwer, Dordrecht, pp. 403–439.
- Lindblom, B. (2000). Developmental origins of adult phonology: The interplay between phonetic emergents and the evolutionary adaptations of sound patterns. *Phonetica* 57, pp. 297–314.
- Lisker, L. (1974). On “explaining” vowel duration variation. *Glossa* 8, pp. 223–246.
- Lu, R.-P., E. P. Smith & I. J. Good (1989). Multivariate measures of similarity and niche overlap. *Theoretical Population Biology* 35:1, pp. 1–21.
- Luce, P. A. & D. B. Pisoni (1998). Recognizing spoken words: The neighborhood activation model. *Ear and Hearing* 19:1, pp. 1–36.
- Mack, M. (1982). Voicing-dependent vowel duration in English and French: Monolingual and bilingual production. *The Journal of the Acoustical Society of America* 71, p. 173.
- MacKay, D. J. C. (2003). *Information Theory, Inference, and Learning Algorithms*. Cambridge University Press, Cambridge.

- Maddieson, I. (1977). Further studies on vowel length before aspirated consonants. *UCLA Working Papers in Phonetics* 38, pp. 82–90.
- Maddieson, I. (1984). *Patterns of sounds*. Cambridge University Press, Cambridge.
- Maddieson, I. & J. Gandour (1976). Vowel length before aspirated consonants. *UCLA Working Papers in Phonetics* 31, pp. 46–52.
- Maddox, W. T. & F. G. Ashby (1993). Comparing decision bound and exemplar models of categorization. *Attention, Perception, & Psychophysics* 53:1, pp. 49–70.
- Mann, V. A. & B. H. Repp (1980). Influence of vocalic context on perception of the [ʃ]-[s] distinction. *Attention, Perception, & Psychophysics* 28:3, pp. 213–228.
- Martinet, A. (1952). Function, structure and sound change. *Word* 8:1, pp. 1–32.
- Matusita, K. (1966). A distance and related statistics in multivariate analysis. Krishnaiah, P. R. (ed.), *Multivariate Analysis: 1st International Symposium Proceedings*, Academic Press, New York, pp. 187–200.
- Menn, L. & M. M. Vihman (2011). Features in child phonology: inherent, emergent, or artefacts of analysis? Clements, G. N. & R. Ridouane (eds.), *Where do phonological features come from? The nature and sources of phonological primitives*, John Benjamins, Amsterdam, pp. 261–301.
- Milroy, J. (1992). *Language variation and change: on the historical sociolinguistics of English*. Blackwell, Oxford.
- Milroy, J. & L. Milroy (1985). Linguistic change, social network and speaker innovation. *Journal of linguistics* 21:2, pp. 339–384.
- Mohanan, K. P. & T. Mohanan (1984). Lexical phonology of the consonant system in Malayalam. *Linguistic Inquiry* pp. 575–602.
- Montagu, J. (2007). Temporal extent of nasalization relative to the tongue articulation in French nasalized vowels. Proceedings of the 16th International Congress of Phonetic Sciences, Saarbrücken.

- Morén, B. (2003). The parallel structures model of feature geometry. *Working Papers of the Cornell Phonetics Laboratory* 15, pp. 194–270.
- Moreton, E. (2008). Analytic bias and phonological typology. *Phonology* 25, pp. 83–127.
- Mowrey, R. A. & I. R. A. MacKay (1990). Phonological primitives: Electromyographic speech error evidence. *The Journal of the Acoustical Society of America* 88:3, pp. 1299–1312.
- Natale, M. (1975). Convergence of mean vocal intensity in dyadic communication as a function of social desirability. *Journal of Personality and Social Psychology* 32:5, pp. 790–804.
- New, B., C. Pallier, L. Ferrand & R. Matos (2001). Une base de données lexicales du français contemporain sur internet: LEXIQUE. *L'Année Psychologique* 101, pp. 447–462.
- Nielsen, K. Y. (2008). *Word-level and feature-level effects in phonetic imitation*. Ph.D. thesis, University of California, Los Angeles.
- Norris, D. (2005). How do computational models help us develop better theories? *Twenty-first century psycholinguistics: Four cornerstones*, Lawrence Erlbaum Associates, Mahwah, New Jersey, pp. 331–346.
- Nosofsky, R. M. (1986). Attention, similarity and the identification-categorization relationship. *Journal of Experimental Psychology: General* 115, pp. 39–57.
- Nosofsky, R. M. (1988). Similarity, frequency and category representations. *Journal of Experimental Psychology: Learning Memory and Cognition* 14, pp. 54–65.
- Ohala, J. J. (1981). The listener as a source of sound change. Masek, C. S., R. A. Hendrik & M. F. Miller (eds.), *Papers from the Parasession on Language and Behavior*, Chicago Linguistic Society, Chicago, pp. 178–203.
- Ohala, J. J. (1989). Sound change is drawn from a pool of synchronic variation. Breivik, L. E. & E. H. Jahr (eds.), *Language change: Contributions to the study of its causes*, Mouton de Gruyter, Berlin, New York, pp. 173–198.

- Ohala, J. J. (1993). The phonetics of sound change. Jones, C. (ed.), *Historical linguistics: problems and perspectives*, Longman, London, pp. 237–278.
- Padgett, J. (2003). The emergence of contrastive palatalization in Russian. Holt, D. E. (ed.), *Optimality Theory and language change*, Kluwer Academic Publishers, Dordrecht, pp. 307–335.
- Padgett, J. (2011). On Dispersion Theory and contrast. Talk given at the 19th Manchester Phonology Meeting.
- Palmer, S. E. (1999). *Vision science: Photons to phenomenology*. MIT Press, Cambridge, MA.
- Pardo, J. S. (2006). On phonetic convergence during conversational interaction. *The Journal of the Acoustical Society of America* 119, pp. 2382–2393.
- Peterson, G. E. & H. L. Barney (1952). Control methods used in the study of vowels. *Journal of the Acoustical Society of America* 24, pp. 175–184.
- Peterson, G. E. & I. Lehiste (1960). Duration of syllable nuclei in English. *The Journal of the Acoustical Society of America* 32:6, pp. 693–703.
- Phillips, B. S. (2006). *Word Frequency and Lexical Diffusion*. Palgrave Macmillan, Basingstoke, UK & New York, NY.
- Pierrehumbert, J. B. (1999). What people know about sounds of language. *Studies in the Linguistic Sciences* 29:2, pp. 111–120.
- Pierrehumbert, J. B. (2001). Exemplar dynamics: Word frequency, lenition, and contrast. Bybee, J. L. & P. Hopper (eds.), *Frequency effects and the emergence of lexical structure*, John Benjamins, Amsterdam, pp. 137–157.
- Pierrehumbert, J. B. (2002). Word-specific phonetics. Gussenhoven, C. & N. Warner (eds.), *Laboratory phonology, Vol. VII*, Mouton de Gruyter, Berlin.
- Pierrehumbert, J. B. (2003). Probabilistic phonology: discrimination and robustness. Bod, R., J. Hay & S. Jannedy (eds.), *Probabilistic Linguistics*, MIT Press, Cambridge, MA, pp. 177–228.

- Port, R. F. (2010a). The reality of phonological forms: a rejoinder. *Language Sciences* 32:1, pp. 60–62.
- Port, R. F. (2010b). Rich memory and distributed phonology. *Language Sciences* 32:1, pp. 43–55.
- Posner, M. I. & S. W. Keele (1968). On the genesis of abstract ideas. *Journal of Experimental Psychology* 77, pp. 353–363.
- Rácz, P. (to appear). *Salience in Sociolinguistics*. De Gruyter Mouton, Berlin, New York.
- Repp, B. H. (1984). Categorical perception: Issues, methods, findings. Lass, N. J. (ed.), *Speech and language: Advances in basic research and practice*, Academic Press, New York, vol. 10, pp. 243–335.
- Ringen, C. & P. Helgason (2004). Distinctive [voice] does not imply regressive voicing assimilation: evidence from Swedish. *International Journal of English Studies* 2, pp. 53–71.
- Rumelhart, D. E. & J. L. McClelland (1986). *Parallel Distributed Processing*. MIT Press, Cambridge, MA.
- Sancier, M. L. & C. A. Fowler (1997). Gestural drift in a bilingual speaker of Brazilian Portuguese and English. *Journal of Phonetics* 25:4, pp. 421–436.
- Sankoff, G. (2004). Adolescents, young adults and the critical period: Two case studies from “Seven Up”. Fought, C. (ed.), *Sociolinguistic variation: Critical reflections*, Oxford University Press, Oxford, pp. 121–139.
- Selkirk, E. (1980). The role of prosodic categories in English word stress. *Linguistic Inquiry* 11, pp. 563–605.
- Shannon, C. E. (1948). A mathematical theory of communication. *Bell System Technical Journal* 27, pp. 379–423, 623–656.
- Shattuck-Hufnagel, S. (1983). Sublexical units and suprasegmental structure in speech production planning. MacNeilage, P. F. (ed.), *The Production of Speech*, Springer-Verlag, New York, pp. 109–136.

- Silverman, B. W. (1986). *Density estimation for statistics and data analysis*. Chapman and Hall, London.
- Silverman, D. (2006). *A critical introduction to phonology: Of sound, mind, and body*. Continuum, London, New York.
- Silverman, D. (2012). *Neutralization*. Cambridge University Press, Cambridge.
- Solé, M.-J. (1992). Phonetic and phonological processes: the case of nasalization. *Language and Speech* 35:1-2, pp. 29–43.
- Sonderegger, M. & A. Yu (2010). A rational account of perceptual compensation for coarticulation. Ohlsson, S. & R. Catrambone (eds.), *Proceedings of the 32nd Annual Conference of the Cognitive Science Society*, Cognitive Science Society, Austin, TX, pp. 375–380.
- Szeredi, D. (2010). Vowel centralization and vowel harmony in Hungarian. Sóskuthy, M. (ed.), *The Odd Yearbook* 8, Eötvös Loránd University, pp. 111–137.
- Tanowitz, J. & P. S. Beddor (1997). Temporal characteristics of coarticulatory vowel nasalization in English. *The Journal of the Acoustical Society of America* 101, p. 3194.
- Tilsen, S. (in press). Inhibitory mechanisms in speech planning maintain and maximise contrast. Yu, A. (ed.), *Origins of sound change: Approaches to phonologization*, Oxford University Press, Oxford.
- Wedel, A. B. (2004). *Self-organization and categorical behavior in phonology*. Ph.D. thesis, University of California at Santa Cruz.
- Wedel, A. B. (2006). Exemplar models, evolution and language change. *The Linguistic Review* 23, pp. 247–274.
- Wedel, A. B. (2007). Feedback and regularity in the lexicon. *Phonology* 24, pp. 147–185.
- Wedel, A. B. (2011). Self-organization in phonology. van Oostendorp, M., C. J. Ewen, E. Hume & K. Rice (eds.), *The Blackwell Companion to Phonology*, Blackwell, Oxford, vol. 1, pp. 130–147.

- Wedel, A. B., A. Kaplan & S. Jackson (submitted). High functional load inhibits phonological contrast loss: A corpus study. *Cognition*.
- Wedel, A. B., S. Jackson & A. Kaplan (in press). Functional load and the lexicon: Evidence that syntactic category and frequency relationships in minimal lemma pairs predict the loss of phoneme contrasts in language change. *Language and Speech*.
- Weinreich, U., W. Labov & M. R. Herzog (1968). Empirical foundations for a theory of language change. Lehmann, W. P. & Y. Malkiel (eds.), *Directions for historical linguistics*, University of Texas Press, Austin, TX, pp. 95–188.
- Wells, J. (1982). *Accents of English. 3 Volumes*. Cambridge University Press, Cambridge.
- Wright, S. (1932). The roles of mutation, inbreeding, crossbreeding and selection in evolution. *Proceedings of the sixth international congress on genetics*, vol. 1, pp. 356–366.
- Yu, A. (in press). Individual differences in socio-cognitive processing and the actuation of sound change. Yu, A. (ed.), *Origin of Sound Patterns: Approaches to Phonologization*, Oxford University Press, Oxford.
- Yu, A. C. L. (2011). On measuring phonetic precursor robustness: a response to Moreton. *Phonology* 28, pp. 491–518.
- Zimmerman, S. A. & S. M. Sapon (1958). Note on vowel duration seen cross-linguistically. *The Journal of the Acoustical Society of America* 30:2, pp. 152–153.
- Zuidema, W. & B. de Boer (2009). The evolution of combinatorial phonology. *Journal of Phonetics* 37, pp. 125–144.