

# Chapter 1

## Introduction

A considerable number of languages use phonetic voicing, the low frequency periodic energy in the speech signal that is produced by vocal fold vibration, to signal a two way lexical distinction between obstruents. For example, Dutch contrasts the voiceless plosives in [pɔl], *tussock (of grass)*, and [tɔl], *spinning top*, with the voiced initial plosives of [bɔl], *round, spherical*, and [dɔl], *crazy (about), foolish*. The two lexical categories identified by voicing in these languages are often described as phonologically *voiceless* vs. *voiced*, but such labels obscure the fact that voicing virtually always acts as part of a cluster of phonetic features when it is used to cue lexical contrast. For example, *ceteris paribus*, contrastively voiceless (aspirated) obstruents are usually relatively long, preceded by somewhat shortened vowels, and cause a slight increase in the  $F_0$  and  $F_1$  of flanking vowels. This is one of the reasons why I will refer to ‘phonologically voiceless’ obstruents as *fortis, tense*, or [+tense], and to their ‘phonologically voiced’ counterparts as *lenis, lax*, or [-tense].

Not all languages that have a [tense] contrast in this sense use the same voicing categories to cue fortis and lenis obstruents. One type of language contrasts voiceless aspirated fortis plosives (e.g., [p<sup>h</sup>, t<sup>h</sup>, c<sup>h</sup>, k<sup>h</sup>, q<sup>h</sup>]) with passively voiced lenis plosives in word-initial and word-medial contexts. If the latter appear utterance initially or after another obstruent, they are generally realised as voiceless and unaspirated, e.g., [b̥, d̥, j̥, ɡ̥], but after a vowel or sonorant consonant they are commonly more or less voiced. I will refer to this type of language, which is exemplified by (standard varieties of) English and German as *aspirating*. A second type of language contrasts plain voiceless fortis plosives ([p, t, c, k, q]) with lenis plosives that are generally prevoiced across phonetic contexts ([b, d, j, ɡ, ɣ]), and will be referred to as *voicing*. Southern and Western varieties of Dutch as well as French and Hungarian are typical voicing languages.

Crucially, the two types of language are consistent in the mapping of [±tense] into durational distinctions and spectral cues other than voicing. For

example despite their differences in (utterance and post-obstruent) voicing the lenis stops of both voicing and aspirating languages are shorter than the corresponding stops, have longer preceding vowels, and act as  $F_0/F_1$  depressors. This justifies the use of the four (and perhaps more) gross phonetic categories introduced in the previous paragraph to describe tense and lax obstruents, rather than the two or three sometimes suggested by the phonological literature: aspirated fortis ([p<sup>h</sup>]), plain voiceless fortis ([p]), passively voiced lenis ([b] utterance initially or after another obstruent), and actively voiced lenis ([b]).

This dissertation investigates the formal and phonetic properties of fortis and lenis obstruents, with a descriptive focus on the Germanic languages and Hungarian. It argues that these properties are best understood in terms of the nature of human speech production and perception, and is therefore broadly *functionalist* in outlook. The following paragraphs outline how the argument is built up.

## 1.1 Synopsis

Chapter 2 starts with a description of the production of voicing distinctions in obstruents and defines the notions *active* and *passive* devoicing in terms of the aerodynamic constraints that supraglottal articulatory settings impose on the initiation, continuation, and termination of vocal fold vibration. The second part of this chapter reviews the literature on the production and perception of the complex of cues that signals [tense] in stops and fricatives and the role of voicing within this complex.

Chapter 3 shifts the focus from the phonetic expression of [tense] to its neutralisation in the form of dynamic ‘final devoicing’ as well as at the lexical level. It discusses two issues that can be regarded as independent, but both of which are important to models of laryngeal neutralisation. The first of these issues is the nature of [tense] neutralisation itself. A long-standing and popular approach is to treat [tense] neutralisation processes as instances of phonological *fortition* or *lenition*, i.e. asymmetric rules that targets lax obstruents only and convert them into their respective tense counterparts, or vice versa. An alternative view regards the neutralisation of fortis-lenis distinctions as a symmetric phenomenon that derives a phonologically and phonetically distinct third category of [0tense] obstruents. From the available phonetic evidence it appears that [tense] neutralisation may not be a phonetically homogeneous phenomenon. Data from different languages and processes is sometimes consistent with the first view, sometimes with the second, and sometimes seems to support a third approach that essentially treats neutralisation as an (extreme) case of contrast *reduction* which leaves residual cues to lexical fortis-lenis contrast.

Neutralisation of [tense] contrasts is not equally probable across contexts

and types of contrast-bearing sound. The second part of chapter 3 identifies the factors behind neutralisation asymmetries and contrasts formalist approaches to these asymmetries with perceptibility-driven functionalist accounts of the type developed by Steriade (1997). Formalist models tend to concentrate on neutralisation asymmetries induced by neighboring sounds and the position of target obstruents within morphemes or words, and propose that such asymmetries be explained in terms of syllabic and/or higher-order prosodic conditions on phonological rules. Cue-based accounts on the other hand, claim that neutralisation is more likely to occur in contexts where the contrast between tense and lax obstruents is relatively imperceptible, and less likely where it is relatively salient. One of the crucial predictions that distinguishes syllable-driven formalist models from a cue-based approach is that in languages with word-final [tense] neutralisation should also suspend the tense-lax distinction in word internal obstruent + sonorant sequences straddling a syllable boundary. Consequently, the observation that the occurrence of laryngeal neutralisation in obstruent + sonorant sequences is neither constrained by syllabification nor by the presence vs. absence of word-final neutralisation constitutes evidence in favour of the cue-based account.

Furthermore, although the scope of the cue-based model proposed by Steriade (1997) appears to be similar to that of syllable-driven formalist models, I will argue that, at least in principle, it extends naturally to the asymmetry between word-initial and word-final contexts and asymmetries between different types of obstruents. Provided that the hypothesised segmental, positional, and stress-based asymmetries in perceptibility are real, this means that a cue-based model is able to account for a range of neutralisation phenomena in terms of a single mechanism, which would make it far superior to any formalist model available in the literature.

Chapter 4 deals with the various forms of voicing assimilation that can be found in the Germanic group and beyond. Drawing on proposals for the analysis of sandhi phenomena more in general, it establishes criteria for distinguishing phonological from coarticulation-based forms of voicing assimilation and then uses these criteria to classify a number of assimilation processes as they are described in the literature. One of the most important generalisations that plays a role in this exercise is the observation that lenis stops only appear to trigger regressive voicing assimilation (RVA) under word sandhi if they belong to the actively prevoiced type, i.e., lenis stops only trigger RVA in voicing languages. This suggests that RVA at word boundaries is a coarticulatory phenomenon, or at least (diachronically) rooted in coarticulation processes. By contrast, voicing assimilation phenomena in morphological paradigms, such as the well-known past tense paradigms of English and Dutch are not phonetically conditioned in this way, and therefore appear to act as phonological rules.

Chapters 5, 6 and 7 report on three experiments designed to test whether RVA across word boundaries is indeed properly regarded as a coarticulation process. The first two experiments examine the phonetic behaviour of velar stop + alveolar consonant sequences in an aspirating variety of English (chapter 5) and Hungarian, a voicing language (chapter 6). Neither of these two languages neutralises [tense] in word-final context and therefore they represent an ideal testing ground for assimilation models. The results of the first experiment indicate that English has a purely coarticulatory form of regressive voicing assimilation at word boundaries. English passively voiced /d/ does not trigger assimilation in a preceding plosive, in contrast to actively voiced /z/, and to a lesser extent to (actively devoiced) /t/ and /s/. Moreover, assimilation only affects the duration of the voiced interval of the velar plosives, but not the duration of their closed phase or the length of the preceding vowels, which is again in full agreement with the predictions of a phonetic approach to RVA.

The results of the second experiment are more complicated. They indicate that as in English, Hungarian RVA is not a phonologically neutralising process. However, unlike the English data, the Hungarian data shows (near-) neutralisation of vowel length distinctions before some obstruent clusters. Although the observed patterns cannot be seen as assimilatory in a straightforward fashion, they contradict a purely articulation-based account and suggest that Hungarian RVA may be (partially) phonologised.

Chapter 7 discusses the results of the third experiment, which was designed to assess the assimilatory effect of Dutch word-initial /p, t, b, d, m, h, V(owel)/ on a preceding /ps/ cluster. The results of this experiment indicate that, as in English, Dutch RVA affects phonetic voicing but not duration features (or  $F_0$ ) and thus support a phonetic account of RVA in Dutch. Moreover, the data from this experiment calls for a revision of the standard conception of Dutch RVA as a [tense]-asymmetric process triggered by lenis but not fortis obstruents: both the lax plosives /b, d/ and the tense plosives /p, t/ cause statistically significant changes in the duration of the voiced interval of preceding /ks/ and /ps/ clusters vis-à-vis /m/. This finding is consistent with the phonetic underspecification approach to Dutch word-final neutralisation proposed by Ernestus (2000).

Chapters 2 through 7 reject the general thrust of formalist approaches to laryngeal phonology and phonetics in favour of an auditory model of laryngeal neutralisation and an articulatory model of RVA. This argument is mainly founded on the distribution of laryngeal contrast and the phonetic manifestation of regressive voicing assimilation. Some might argue that these are insufficient grounds for an outright rejection of formalist approaches because such approaches still have a role to play, for example in defining the set of laryngeal neutralisation and assimilation rules that the human mind is able to represent. More specifically, they might point out that most of the predictive power of

current generative models resides in the detail of modality-neutral segmental representations, and that these models are therefore capable of narrowing the range of phenomena that have to be explained in auditory, articulatory, or other functional terms.

The role of formalist models as a possible source of metaconstraints on functional explanations is investigated in the sixth and final chapter. Two general designs are discussed here: [tense]-based models along the lines of (Lombardi 1994 et seq.), and the VOT-based models proposed by e.g., Harris (1994) and Iverson & Salmons (1995, 1999). Both models are found seriously wanting, because the predicted connections among laryngeal neutralisation rules, regressive assimilation processes, the behaviour of the ‘Germanic’ past tense paradigm, and other phenomena are not borne out by the data. Moreover, under a strict interpretation of monovalent feature representation, both models undergenerate, in particular with regard to the ‘phonologically active’ nature of plain voiceless fortis obstruents. Neither of these models can therefore be regarded as in any sense complimentary or prerequisite to the functional accounts of RVA and [tense] neutralisation developed in the preceding chapters. The failure of the formalist enterprise is further underlined by the observation that representationally richer frameworks are successful to the extent that they approximate continuously-valued feature systems constrained by grammar-external (functional) principles.

The remainder of this chapter outlines the phonological and phonetic transcription conventions used in this study (section 1.2), and more importantly, the descriptive model underpinning chapters 2-8.

## 1.2 Notes on transcription

Lexical contrasts are transcribed with slanted brackets, e.g., /p, b/, whilst the physical/perceptual manifestations of phonological categories are symbolised using square brackets, e.g., [p<sup>h</sup>, p, b̥, b]. Orthographic forms appear in angular brackets (<, >) and in running text, glosses of non-English words are italicised.

With three exceptions, all impressionistic data from the literature are transcribed as in the sources. The same applies to data from specific regional varieties of Dutch and English. However, phonetic data concerning standard Dutch is represented according to my own pronunciation of the standard language: i.e. with [χ] for /x/ and /ɣ/ in all contexts, with diphthongised long mid vowels [e<sup>h</sup>ɪ, ø<sup>h</sup>ɪ, o<sup>w</sup>ɪ:], [au] for the back diphthong /ɔu/, and [ɾ] and [ɻ] for onset and coda /r/ respectively. I have transcribed the lax front rounded vowel that is often analysed as /œ/ with [ɻ]. In the transcription of Dutch underlying forms the IPA diacritic [̄] for long sounds is used to represent the set of ‘tense’ or phonotactically long vowels, even though the high tense vowels are phonetically short in Dutch stan-

dard (and my own) pronunciation (so /i:, y:, u:/ for phonetic [i, y, u]).<sup>1</sup>

Where regional variation is not an issue I have chosen the (southern) British, non-rhotic variety that forms the basis for the pronunciation dictionary of Wells (2000) to represent English data. Apart from the absence of coda /r/ the most notable feature of this variety is a phonemic distinction between the low vowels [æ], [ɑ:], [ɒ]. Finally, standard German data are transcribed according to the conventions in Drosdowski & Eisenberg (1995).

## 1.3 The descriptive framework

### 1.3.1 Linguistic and extralinguistic speech processing

Few phonologists would disagree with the idea that there are peripheral stages in the production and perception of speech that are independent of any form of linguistic knowledge. Take for example the pulsing of the vocal cords during voicing. It is universally accepted that the individual pulses of the glottis do not result from individual instructions (nerve firings) to the vocal folds. Instead, the musculature of the larynx is more or less static during the production of vocal fold vibration (barring changes in pitch or movements of the larynx as a whole), forcing the glottis to be closed but not too tightly adducted. Glottal pulsing then arises through the aerodynamic-myoelectric effects of pushing air from the lungs through the closed glottis (van den Berg, 1958). Similarly, no one would want to describe mechanical interactions between the movement of the tongue root and tongue tip, or the fact that the physiology of the inner ear warps the incoming acoustic signal in various ways, as linguistic knowledge.

In addition, certain short-term adaptations in articulator movements appear to be beyond what most researchers regard as *linguistic* control. For instance, if the closing gesture of the lower jaw is suddenly interrupted during the production of a bilabial constriction, speakers compensate with increased movement of the upper and lower lips. The lag between the interruption of the lower jaw gesture and the onset of compensatory articulations (often  $\leq 30$  ms) cannot be attributed to any sort of mechanical linkage. Given that reaction times (to linguistic tasks) typically run in the hundreds of milliseconds it is not plausible either that short term adjustments of this kind are orchestrated at any level

<sup>1</sup>Dutch [i, y, u] share the phonotactics of long vowels such as [a:] rather than ‘true’ short vowels such as [ɪ, ʏ]. Thus, they can appear in open monosyllables (e.g., [ku], *cow*) and open final syllables. In these contexts they can only be closed by a single consonant (modulo the same exceptions that apply to the other long vowels) whilst they can only occur in open non-final syllables. Characterising [i, y, u] as simply *long* is not wholly unproblematic however, because standard Dutch does allow phonetically long high vowels in loans such as [anali:ze], *analysis*. This has created near-minimal pairs such as [zun], *kiss* vs. [zum], *zoom*. However, in the absence of an agreed IPA diacritic for ‘tenseness’ I have opted to appropriate the length diacritic to mark the class of phonotactically long vowels in Dutch underlying representations.

of (linguistic) planning. Moreover, similar short-term adaptations have been observed in other, non-linguistic forms of motor behaviour, such as hand and finger movements. Consequently, they are normally treated as reflex-like behaviour triggered by proprioceptive feedback (see [Saltzman & Munhall 1989](#) for an overview and references).

Note that all these ‘physical’ and otherwise extralinguistic aspects of speech processing are roughly what is modelled by the articulatory synthesis models of [Ishizaka & Flanagan \(1972\)](#), [Boersma \(1998\)](#), the *task-dynamic model* implementing the *gestural scores* of ([Browman & Goldstein 1986](#) et seq.), or the cochlear model of [Lyons \(1982\)](#).

There cannot be many phonologists either, who would dispute the claim that the information that is exchanged at the interface between the extralinguistic levels of speech processing and linguistic competence is discretised at anything near the granularity of lexical phonological features (this information corresponds to the input and output parameters respectively of the models mentioned in the previous paragraph). Speakers are able to vary the position of their tongue, the pitch of their voice, their speaking rate, and many other speech features on what for all practical purposes are continuous scales. Similarly, although the mechanics of the inner ear (and pre-cortical processing) introduce various non-linearities in the signal, and although e.g., the frequency resolution of the human auditory signal is far from infinite, this resolution is again greater than that of virtually all phonological feature systems. For instance, [Boersma \(1998\)](#) estimates that (cardinal) [i] and [u] are 12 *Just Notable Differences* (JNDs) apart in auditory ( $F_1$ - $F_2$ ) space, but no known language has 11 intermediate vowels between [i] and [u] along the front-back dimension (i.e. vowels with the same auditory  $F_1$ ).

The scalar nature of the information that is exchanged between linguistic competence and the peripheral physical systems is also evinced by the observation that languages that according to ‘broad’ descriptions share sounds or sound inventories, often display subtle but reliable phonetic differences between seemingly equivalent sounds. It is well-known for example, that Danish /i/ is somewhat higher and fronter, on average, than English /i/ ([Disner, 1983](#)), and [Bradlow \(1995\)](#) finds similar differences between English and Spanish vowels. As [Pierrehumbert et al. \(2000\)](#) point out, there is no reason to assume that this sort of crosslinguistic variation is constrained in terms of points on a discrete scale, and so it must be concluded that linguistic competence includes knowledge that is best represented on continuous scales.

Although the topic of gradient but linguistic processing has come to the fore in recent years, its existence is acknowledged by [Chomsky & Halle \(1968\)](#), who conceive of lexical representation in binary terms, but allow features to acquire scalar values at the final stages of a derivation. Lexical Phonology also allows

features with scalar values, at least at the postlexical level (Kaisse & Shaw, 1985; Mohanan, 1986). Other models seek to model all linguistic processing in terms of discrete representations and therefore try to dispense with the ‘systematic’ or ‘linguistic’ phonetic level as a significant level of representation (Pierrehumbert & Beckman, 1988; Kaye, 1989; Coleman, 1992; Harris & Lindsey, 1995). But to the extent that they are intended as (partial) models of human speech production and perception, such frameworks cannot go without a module that translates between discrete feature structures and the continuously-valued information that is supplied and required by the relevant peripheral physical systems.<sup>2</sup>

From here on, I will refer to the collective aspects of speech processing that are guided by linguistic competence, i.e. both categorical and gradient processes, as the *phonetic grammar*. Similarly, following Kingston & Diehl (1994) I will refer to the part of linguistic competence that guides this collection of processes as *phonetic knowledge*. The next sections are devoted to the assumptions this study makes about the organisation of the phonetic grammar.

### 1.3.2 Phonology and phonetics

A conception of the phonetic grammar that is associated with a lot of (early) work in laboratory phonology holds that categorical and gradient processes operate in two separate modules and on fundamentally different feature structures (Keating, 1990a; Gussenhoven, 1996). According to this view, the *phonology* is the module that deals with lexical representations and categorical rules, whereas the (linguistic) *phonetics* takes care of the subsequent gradient processes. At the interface between the two levels, discrete phonological representations are translated into continuously-valued structures of the sort that are produced/understood by the physical levels.

In this type of framework, the Dutch vowel /i/ is represented by the phonology as [+high, -low, -back, -round], or some equivalent (autosegmental) structure. This discrete structure is translated from an auditory representation with  $F_1$  and  $F_2$  values of, say, 3.5 and 14 Bark (339 and 2357 Hz) for a male (Dutch) speaker by the phonetics-phonology interface (or rather from normalised values that filter out the effect of speaker size). The articulatory component of the phonology-phonetics interface translates the discrete representation of /i/ into the corresponding instructions or *targets* for the physical system. In a language

<sup>2</sup>Proponents of the latter type of model often subscribe to a ‘denotational’ view of the relationship between phonetics and phonology. On this view, phonological structures denote real-world articulatory and acoustic ‘events’. Note that, despite appearances to the contrary, this view does not entail that all linguistic structure is discrete: it is technically possible to take a denotational view of a systematic phonetic representation consisting of scalar features. It is difficult to see however, how a denotational phonology-phonetics interface could be embodied by real human language users, whose knowledge about articulatory and acoustic events is mediated by peripheral processing of acoustic, visual, and proprioceptive feedback (cf. Pierrehumbert et al. 2000).

in which /i/ is subject to a categorical labial harmony process, the phonology first converts its structure into [+high, -low, -back, +round] (equivalent to lexical /y/), and the phonology-phonetics interface translates between this structure and the appropriate perceptual (say, 12.5 Bark, 1884 Hz or again, the appropriate value on a normalised scale) and articulatory scales.

An alternative approach, which is embodied in the framework of *Articulatory Phonology* (e.g., Browman & Goldstein 1986 et seq.; Byrd 1996a) and adopted by Flemming (2001) and Pierrehumbert et al. (2000), is to dispense with the phonology as an independent model and represent both categorical and gradient processes in terms of continuously-valued feature structures. On this view, the lexical representation of Dutch /i/ consists simply of  $F_1$ , (value: 3.5 Bark)  $F_2$  (value: 14 Bark), other relevant spectral and durational parameters, and the corresponding articulatory targets. The labial harmony rule referred to above is assumed to act directly on these parameters, changing the  $F_2$  value of /i/ into 12.5 Bark. Since this 'is' the lexical  $F_2$  value of /y/, the harmony rule acts as a categorical, neutralising process even though it is stated in terms of gradient features. In other words, the single-module approach capitalises on the fact continuously-valued features can encode categorical processes (and thus eliminates the duplication of information at the phonetics-phonology interface: see further below).

Strictly speaking these two conceptions of the phonetic grammar are independent from the choice between a formalist or functionalist view on the origin of phonological and phonetic constraints. However, practically speaking, formalist models that aim to explain the nature of phonological constraints depend on a separation of phonetics and phonology. As discussed in 1.4 below, formalist models usually derive the set of possible phonological rules from an alphabet of representational primitives and a severely restricted set of combinatory principles. However, if phonetic knowledge is encoded in terms of continuous representations the number of possible natural classes is infinite or (stipulating that all categories must be at least 1 JND apart) at least very large. Consequently, the number of possible rules that can be derived according to the formalist logic grows very large as well, and the resulting grammars are almost guaranteed to be massively overgenerating.

Therefore, the non-modular<sup>3</sup> conception of the phonetic grammar more or less implies a (partially) functionalist perspective on the origin of phonological and phonetic constraints. Functionalist models derive the set of possible, or rather *probable*, rules from external, 'ecological', factors, such as need for robustly perceptible cues to phonological distinctions. Consequently, functionalist

---

<sup>3</sup>I call this view of the phonetic grammar *non-modular* because it does not distinguish separate phonological and linguistic phonetic modules. However, strictly speaking it is still modularised, because it consists of articulatory and auditory (as well as other perceptual) components.

models are able to rule out e.g., processes that change the  $F_2$  value of a vowel upwards by the equivalent of 1 Hz: the effects of such a process would simply be imperceptible.

Although I ultimately subscribe to the single-module conception of the phonetic grammar, I will refer to *phonological* (categorical) vs. *phonetic* (linguistic gradient) rules, mainly for expository reasons, and using the diagnostics identified by Myers (2000) to distinguish between the two types of processes. For the same reasons, I will refer to (lexical) *phonological categories* and their *phonetic interpretation*. In transcriptions, the former will be indicated by slanted, and the latter by square, brackets. For instance, the contrastively voiced labial and alveolar stops of Dutch will be referred to as phonologically [-tense] and symbolised as /b, d/ if their lexical status is at issue, but (outside neutralisation contexts) as [b, d] where their phonetic properties are relevant to the discussion.

However, because these labels merely serve descriptive convenience, I will not make any specific assumptions about the ‘nature of phonological representation’. [ $\pm$ tense] is used to represent lexical laryngeal contrast rather than the more familiar [ $\pm$ voice] to keep track of the essential distinction between phonetic voicing and what is often known as ‘phonological’ voicing, but nothing of importance hinges on this. Where it is relevant, the representations used by others will be described in what I hope is sufficient detail. Furthermore, I will use the terms *rule*, *process* or *constraint* in a purely descriptive way, without committing to a procedural (derivational) or declarative interpretation.<sup>4</sup>

### 1.3.3 Phonetic rules and representations

Johnson et al. (1993) describe the production of an utterance as a two-stage process. As they subscribe to a modular theory of the phonetic grammar, the first step maps discretely-valued phonological features into phonetic features with values drawn from continuous articulatory scales. The second step modifies the initial values of these features to derive the variation in the realisation of phonological categories that is observable in speech. Because Johnson et al. describe the second step as a mapping between ‘parametric phonetic’ representations, I will assume that they are identical in nature, i.e. that they consist of the same set of features that range over the same scales of values. Moreover, I will assume that the output of the second step represents the instructions to the physical articulatory system, and thus that Johnson et al.’s ‘parametric-to-parametric’ mapping encompasses the full body of linguistic phonetic (i.e. gradient) rules in the sense defined above.

---

<sup>4</sup>Models that maintain separate phonological and linguistic phonetic components are inherently procedural at the interface, where discrete representations are converted into gradient ones. Single-module phonetic grammars however, can be modelled in terms of declarative constraints (Flemming, 2001).

The central topic of Johnson et al.'s paper is the nature of the initial values of the parametric phonetic representations, in a sense the 'underlying' phonetic values. They argue that these values are chosen to optimise auditory contrast among phonological categories within the available phonetic space. Using terminology associated with similar theory of speech production proposed by Lindblom (1990), they label these optimally spaced points *hyper(articulated) targets*. The second, phonetic rule, stage in the production process either maintains these hypertargets, or modifies them in a way that generally speaking results in a diminished amount of contrast between phonetic categories in the resulting utterance. Following Lindblom's theory, I will refer to the latter phenomenon as *hypoarticulation* but it is essentially similar to the idea of target *undershoot*.

The variable realisation of vowels serves as a simple illustration of Johnson et al.'s model. The phonology-phonetics interface is assumed to assign the same 'peripheral' auditory  $F_1/F_2$  formant values to the 4 vowels /i/ ([+high, -low, -back, -round]), /æ/ ([-high, +low, -back, -round]), /ɑ/ ([-high, +low, +back, -round]), and /u/ ([+high, -low, +back, +round]) regardless of the phonetic context, the degree of stress or of speech rate and register. These hypertargets are indicated by the black dots in figure 1.1. More centralised vowel realisations, which are found in unstressed syllables for example, are derived by the subsequent application of phonetic rules, and so are effects of segmental context, such as the fronting of back vowels before coronal consonants (cf. Flemming 2001).

The hypothesis that auditory dispersion or contrast optimisation more generally plays a role in the structuring of (phonetic) sound inventories is not new to the model of Johnson et al. (1993). But they are among the first to present direct evidence for the idea that hyperarticulated targets play a role in speech processing. They report 3 *Method Of Adjustment* (MOA) experiments in which test subjects were asked to adjust the settings of a vowel synthesiser until the output matched what they perceived as the vowels in a list of visually presented stimulus words. The same set of subjects were asked to read the stimulus words in normal 'citation' and hyperarticulated forms (the latter elicited by way of feedback from the experimenter). The responses to the MOA task show that the subjects systematically selected sounds with more extreme formant values than the values they produced in the citation form reading task, even when a number of potential confounds (e.g., phonetic training of the test subjects) were eliminated: see figure 1.1. The vowel space of the hyperarticulated readings corresponds more closely to the boundaries found in the MOA experiments. Johnson et al.'s tentative conclusion from the observation that the test subjects treated the hyperarticulated vowels as representative for the stimulus words is that hypertargets are primary to reduced forms in speech production.

Johnson et al. (1993) do not offer a formal implementation of the mapping between parametric phonetic representations: such implementations are pro-

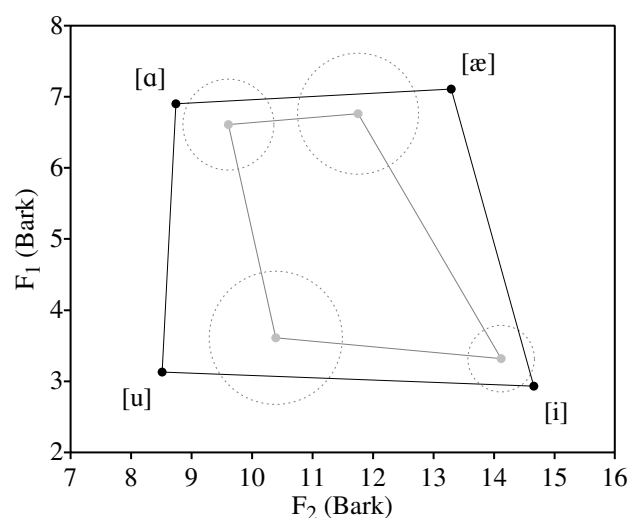


Figure 1.1: The *hyperspace effect* according to Johnson et al. (1993). The hyper-targets in black correspond to the MOA results for /i, æ, a, u/ of Johnson et al.'s experiment 1, converted to Bark; the grey dots represent the average values for the citation form readings of the same vowels in the same experiment (Johnson et al. 1993:520).

vided by Articulatory Phonology (Browman & Goldstein 1986 et seq.; Byrd 1996a) and the *Window Model* of coarticulation (Keating 1990a; see also e.g., Huffman 1993). What matters however is their notion of hypoarticulation as increased variability. For example, in contexts which allow little or no hyperarticulation, the physical articulatory system is instructed to produce the English vowels /i, æ, a, u/ with articulatory gestures that match the black dots in figure 1.1 very precisely. But in hypoarticulation contexts, the  $F_1$ - $F_2$  (and corresponding articulatory) values vary across wider ranges of values, e.g., those indicated by the dotted grey circles in figure 1.1 in a way that is determined by the effort put in by the speaker and the phonetic context.

This study assumes that hypoarticulation, conceived as a relaxation of auditory specificity, and implemented as a reduction of articulatory effort, is the driving force behind many phonetic rules. This auditory-articulatory take on hypoarticulation phenomena is borrowed from the *Hyperarticulation and Hypoarticulation* (H & H) theory of Lindblom (1990) and similar hybrid accounts (Boersma, 1998; Flemming, 2001) and contrasts with the purely articulation-driven views espoused by Articulatory Phonology and Kirchner (1998).<sup>5</sup>

<sup>5</sup>For an overview of coarticulation phenomena and models, see Farnetani (1997), and for def-

Two broad classes of phonetic rule that can be construed in terms of hypoarticulation are *coarticulation* and (phonetic) *reduction/lenition*. The former term will be used to refer to situations in which the realisation of a particular sound is influenced by that of a temporally close second sound with which it shares one or more (mechanically linked) articulators. For example, the precise constriction location of intervocalic /k/ in English and other languages depends on the quality of the flanking vowels: it is slightly fronter between front vowels but somewhat backed between back vowels. Conversely, the place of articulation ( $F_2$  in rough acoustic/perceptual terms) of vowels is often influenced by the neighbouring consonants: back vowels /u/ tends to be somewhat fronter between coronal than between velar consonants (Lindblom, 1963; Flemming, 2001). Although interactions between vowel place and consonant place of articulation have been phonologised by numerous languages, e.g., in terms of velar fronting (palatalisation) or the alternation between front and back velars that is common in the Turkic languages, they also occur as gradient processes.

These gradient processes can be understood in hypoarticulatory terms as follows. Vowels and non-labial consonants share an active articulator: the tongue. This means that if e.g., a front vowel and a back consonant are produced in sequence, the tongue has to be retracted a certain amount within a relatively short time span. The amount of retraction that is needed depends on the hypertargets for the vowel and consonant and the degree in which the realised targets are allowed to deviate from these hypertargets, i.e. the degree of hypoarticulation. All else being equal, if there is a low degree of hypoarticulation, both the vowel and consonant have to be realised with targets that are relatively ‘faithful’ to their hypertargets, which results in a relatively great amount of tongue retraction and hence a limited amount of observable place ( $F_2$ ) coarticulation. If the realised targets are allowed to deviate from the hypertargets to a greater extent, the front vowel *can* be realised with a less front articulation and the back consonant with more fronting, which means a smaller amount of tongue retraction. The reason that the vowel and consonant *are* realised with relatively close constriction locations is, according to many, that the smaller tongue displacement saves articulatory energy (e.g., Lindblom 1990; Boersma 1998; Kirchner 1998; Flemming 2001). Phrased in more general terms, relaxing constraints on the realisation of targets in auditory space allows for smaller transitions in articulator movements, and hence for a lower articulatory energy expenditure.

---

initions of articulatory effort, Boersma (1998) and Kirchner (1998). Note that the theory of coarticulation summarised in these paragraphs does not entail that speakers compute the articulatory energy involved in the realisation of an utterance (and a range of alternatives) before they produce it, as is suggested by the work of Boersma, Kirchner, and also Flemming (2001). Effort considerations could equally well enter the phonetic grammar if speakers receive some form of feedback about the energy consumed by the production of utterances with a given phonetic makeup, and simply learned from this experience to avoid overly difficult forms. See further below.

I will use the terms *reduction* and *lenition* for realisations of segments that deviate from their hypertargets in a way that can be described as a decrease in the overall magnitude (and speed) of the relevant articulator movements. A hypoarticulation-based account attributes lenition in this sense to exactly the same mechanism as coarticulation, although it involves the somewhat problematic notion of a *neutral vocal tract configuration*. The basic idea is that relaxation of auditory constraints on the realisation of a given target not only allows for deviations to accommodate the implementation of neighbouring sounds, but also a reduction in the magnitude of articulatory gestures with regard to some equilibrium point. This equilibrium point is often defined as the vocal tract configuration for schwa. For vowels, a gradient reduction in gesture size results in gradient centralisation whilst for stops it leads to shortening (to the extent that the duration of a stop is due to the magnitude of the closing gesture), affrication, spirantisation, or gliding, depending on the amount of gestural weakening. Because both phenomena are seen as reflexes of the same mechanism, the prediction of a hyperarticulation-based theory of coarticulation and reduction/lenition is that reduced sounds should always show an increased amount of coarticulation with neighbouring sounds, and vice versa.<sup>6</sup>

### 1.3.4 Hypoarticulation and prosody

Judging by the behaviour of reduction and coarticulation phenomena the degree of hypoarticulation varies at both global and local levels. Globally it varies with speech rate and register. For example, [Moon & Lindblom \(1994\)](#) show how vowel reduction and consonant-vowel coarticulation increase with decreasing clarity of speech, where clarity is defined in terms of the instructions given to the test subjects. Fast speech is generally considered to be conducive to hypoarticulation, and the evidence in the literature broadly supports this view. Studies such as [Lindblom \(1963\)](#), [Engstrand \(1988\)](#), [Byrd & Tan \(1996\)](#), [Kessinger & Blumstein \(1997\)](#), record increased undershoot and coarticulation of targets for pitch, VOT, and place for vowels and consonants. On the other hand, speech in noisy environments has often been claimed to be hyperarticulated: this phenomenon is also known as the *Lombard reflex*. The review by [Junqua \(1996\)](#) of work on speech in noisy environments notes several features also found in clear speech elicited by different methods, though there is also evidence for speaker-dependent and more fine-tuned adaptation of speech to specific types of noise.

A number of factors seem to condition more local fluctuations in hypoartic-

<sup>6</sup>Note that the view of lenition/reduction described here is similar to the conception of (phonological) lenition as the loss of phonologically marked structure that is developed by [Harris \(1994\)](#), especially if the resulting unmarked configurations are interpreted in terms of *phonetic underspecification* (see [1.3.5](#) below). An interpretation in these terms appears to be suggested by [Harris & Lindsey \(1995\)](#).

ulation. Since a lot of work on local hypoarticulation has focused on its articulatory reflexes, such fluctuations are now commonly referred to as *articulatory strengthening* and *weakening* (e.g., Pierrehumbert & Talkin 1992; De Jong 1995; Jun 1995; Gordon 1996; Byrd & Saltzman 1998; Hsu & Jun 1998; Keating et al. 1998; Fougeron 1999). The factors involved include (lexical) stress, morphosyntax, and information structure. The effects of the latter two variables are often assumed to be mediated by a *prosodic phrase structure* (Halliday, 1960; Selkirk, 1986; Nespor & Vogel, 1986; Pierrehumbert & Beckman, 1988; Ladd, 1996) and since lexical stress is part of prosody structure by virtually all definitions of the term, I will refer to their collective effects on phonetic realisation as *prosodic*.

Prosody introduces two major hypoarticulation asymmetries: one between (lexically) stressed and unstressed contexts, and a second one between constituent-initial and constituent-final contexts. Stressed syllables and constituent-initial positions are relatively resistant to reduction and coarticulation, and under the theory sketched in the previous section these environments should therefore be considered local hypoarticulation minima. Unstressed medial, and final contexts on the other hand, often exhibit consonant lenition, vowel reduction, and increased levels of coarticulation, and might therefore be regarded as local hypoarticulation maxima.

Observations about the relation between prosody and segmental realisation have been made both before, and outside the context of, recent experimental work explicitly couched in terms of articulatory strengthening. For example, Jones (1956) as well as Kahn (1976) highlight the role of lexical stress in the realisation of English fortis stops, which have more aspiration in the onsets of stressed syllables than elsewhere. The various lenition processes that affect English /t/ outside strengthening contexts is documented and analysed by Harris (1994). However, instrumental studies on articulatory strengthening have both quantified these and other phenomena, and demonstrated that they are much more general than might be gleaned from impressionistic descriptions of vowel reduction and consonant lenition. For example, Turk (1992) shows that, like alveolar stops, English labial and velar stops are subject to shortening in intervocalic contexts, even if the consequences are less perceptible than those of flapping.

Instrumental studies have also uncovered evidence indicating that the asymmetry between initial and final contexts is not restricted to the (prosodic) word level, but holds across higher levels of prosodic phrasing as well, in a way that is sensitive to juncture strength. For example, in a survey of 4 languages Keating et al. (1998) find that the amount of peak linguopalatal contact and seal duration (the duration of full oral tract constriction) in constituent-initial /t, n/ increases with the strength of the preceding juncture, and thus that within a given

constituent, peak contact and seal duration are greater in initial than in medial contexts. For example their EPG data for two French speakers show a mean maximal contact of > 60% of the measurement area for Intonation Phrase (*IP*)-initial /t/, which drops to just above 50% for *IP*-medial word-initial /t/. There is some evidence that the strengthening effects of stress and position are mutually reinforcing (i.e. initial stressed syllables are less hypoarticulated than stressed noninitial ones) but the effect is not simply additive (Lavoie, 2001). In addition, instrumental studies have established a correlation between the amount of segment-to-segment coarticulation and prosody. Work by De Jong et al. (1992) and De Jong (1995) shows that segments in syllables bearing lexical stress are less coarticulated than similar sequences in unstressed syllables.

### 1.3.5 Absent targets: phonetic underspecification

There is a considerable amount of data to suggest that where a given phonological contrast is neutralised, the resulting sounds sometimes lack targets for the phonetic parameters that signal that contrast in other, non-neutralisation environments. For example, chapter 3 discusses evidence adduced by Ernestus (2000) that the word-final laryngeal neutralisation ('final devoicing') of Dutch obstruents produces stops and fricatives without targets for phonetic voicing, segmental duration and other cues to [ $\pm$ tense]. Of course final obstruents in Dutch have voiced and voiceless intervals of definite lengths, but Ernestus claims that this voicing is completely derived from coarticulation. The oral tract configurations for stops and fricatives militate against the continuation of voicing after the offset of the preceding vowel or sonorant beyond certain (aerodynamically determined) points (see section 2.1 below), and utterance finally, this 'segment-internal' coarticulation results in the eponymous final devoicing. However, utterance medially, coarticulation with flanking (voiced) sonorant sounds and especially actively voiced lenis obstruents ([b, d]) is predicted to result in a greater amount of voicing for neutralised obstruents, and, as highlighted by experimental data in chapter 7, this is exactly what is observed.<sup>7</sup>

The analysis of final obstruent neutralisation in Dutch defended by Ernestus (2000) is an instance of a more general descriptive tool that gained popularity in the early years of laboratory phonology and is commonly known as *surface*, or *phonetic underspecification* (Pierrehumbert & Beckman, 1988; Keating, 1988). Note that phonetic underspecification is effectively the limiting case of hypoarticulation in the sense defined above: it describes sounds that allow the maximal amount of variability (that is physically possible) with regard to the underspecified phonetic dimension. So whilst the [+tense] obstruents of Dutch are specified

<sup>7</sup>Ernestus's analysis of Dutch final obstruent neutralisation is discussed in more detail in chapter 3 below.

as mostly voiceless and at least its [-tense] plosives as voiced for the larger part of their durations, the voicing of neutralised final obstruents is allowed to range across the whole continuum from fully voiceless to fully voiced. Phonetic underspecification of the complex of phonetic cues that signal [tense] therefore defines a third category of [0tense] (neutralised) obstruents in addition to the [ $\pm$ tense] stops and fricatives that occur in non-neutralisation contexts. This contradicts standard analyses of final laryngeal neutralisation in Dutch, which hold that neutralised obstruents are [+tense] and therefore phonetically indistinguishable from [+tense] obstruents in environments where the [tense] contrast is not suspended.

The account of Japanese tonal phonology in [Pierrehumbert & Beckman \(1988\)](#) is one of the original studies that developed phonetic underspecification in an area where full specification had been the explicit norm, and thus serves as a good illustration of the mechanics of the device. Japanese is a pitch accent language in which the presence and place of a tonal accent in a word is lexically contrastive, but not the shape of the tonal melodies of accented and unaccented words. Nevertheless, many earlier accounts claim that all syllables in Japanese are phonologically and hence phonetically specified for tone and so they represent the phonological melody of /moriya-no mawari-no o mawarisan/, *the Forests-neighbourhood policeman*, where the italicised segments indicate the sole accented syllable, approximately as in (1a). The most natural interpretation of this melody assigns high pitch targets to H and low targets to L and therefore derives a rise from /mo/ to /ri/ followed by a high plateau and a relatively abrupt fall between H-toned /no/ and /o/:

- (1) Specification of Japanese pitch contours (after [Pierrehumbert & Beckman 1988](#))

a. Full specification

L H H H H H H H L H L L L L  
mo ri ya no ma wa ri no o ma wa ri sa n

b. Phonetic underspecification

L H L H L L L L  
mo ri ya no ma wa ri no o ma wa ri sa n

However, Pierrehumbert and Beckman find that the pitch contours represented by this melody show a gradual fall from a peak corresponding to the H tone on /ri/ to the L on /o/. Moreover, systematic manipulation of the number of moras between the initial LH sequence and second L and varying the phonological length of the syllable carrying the second L shows that the slope of this contour is an approximately linear function interpolating between the pitch values of the first H and the second L. They conclude that the syllables in the /ya...no/ interval cannot be assigned pitch targets like the Hs on /ri/ and /ma/

or e.g., the L on /wa/, which correspond to clear local highs and lows in the pitch contour, but form a third distinct category of syllables with regard to phonetic interpretation in not bearing a pitch target. There are no phonological grounds for retaining the Hs on these syllables, because they do not mark lexical contrast directly or indirectly by conditioning the distribution of other features, and therefore [Pierrehumbert & Beckman \(1988\)](#) represent them as underspecified for tone targets, as in (1b).<sup>8</sup>

## 1.4 Formalism vs. functionalism

Since language is not, in its essence, a means for transmitting such [cognitive] information – though no one denies that we constantly use language for this very purpose – then it is hardly surprising to find in languages much ambiguity and redundancy, as well as other properties that are obviously undesirable in a good communication code. In sum, the theme of language as a game opens up perspectives that are by no means unattractive, so that others might wish to explore them further. ([Halle 1975:528](#))

We may say that a living body or organ is well designed if it has attributes that an intelligent and knowledgeable engineer might have built into it in order to achieve some sensible purpose, such as flying, swimming, seeing, eating, reproducing, or more generally promoting the survival and replication of the organism's genes. It is not necessary to suppose that the design of a body or organ is the best that an engineer could conceive of. Often the best that one engineer can do is, in any case, exceeded by the best that another engineer can do, especially another who lives later in the history of technology. But any engineer can recognise an object that has been designed, even poorly designed, for a purpose, and he can usually work out what that purpose is just by looking at the structure of the object. ([Dawkins 1988:21](#))

*Formalism* and *functionalism* are labels for hypotheses about the origins of the rules in the phonetic grammar. Formalism, which is normally only concerned with phonological processes, claims that such rules are motivated by a small number of grammar-internal principles that are essentially arbitrary with regard to the use of speech as a communication tool. This arbitrariness is highlighted

---

<sup>8</sup>See ([Pierrehumbert & Beckman 1988:chapter 2](#)) for the arguments against the idea that the contour between phrase-initial Highs and following Lows is a result of full tonal specification interacting with independent pitch range modification (i.e., declination).

by Halle's analogy between phonology and a mathematical game: the rules of the latter only exist for the sake of the game itself. Moreover, they can take any conceivable shape, as long as a limited number of basic constraints on the system as a whole (e.g., consistency) are respected. Functionalism, on the other hand, hypothesises that phonetic grammars are organised in ways that benefit speech perception, grammatical segmentation, lexical access, as well as speech production. In other words, functionalism claims that phonological and phonetic rules are designed to be communication tools.

As '-isms', formalism and functionalism represent claims about the phonetic (or more specifically the phonological) grammar as a whole. But testable formalist and functionalist hypotheses can be formulated for specific phenomena, and at least in specific cases, the controversy between the two paradigms can be resolved on empirical grounds. This section explores the types of prediction that are derived from formalist and functionalist theories, and goes on to argue for a 'diachronic' version of functionalism which holds that functional considerations enter the grammar in a stepwise fashion during language acquisition and change. One of the main advantages of this theory over 'synchronic' functionalism is that it can account for so-called *crazy rules*, as long as such rules can be decomposed into a diachronic series of small changes, each of which is functionally motivated.

### 1.4.1 Radical formalism

Taken to its logical conclusion, formalism predicts that the relation between phonological categories and their phonetic exponents is completely arbitrary. Foley (1977) and latterly Hale & Reiss (2000a,b) provide perhaps the closest approximations of this position. It entails that a set *a* of phonetic segments [p, t, k, f, s, x] should be equally likely to form a *phonological* natural class as a set *b* consisting of [p, ɭ, ð, ʉ, d, Ø]. In other words both sets are predicted to be equally probable phonetic interpretations of the phonological categories /p, t, k, f, s, x/. It follows that there should be languages in which the sounds in *b* exhibit what might be regarded as normal obstruent phonology: the ability to precede sonorants in syllable onsets, to trigger place assimilation in a preceding nasal, or to form [±tense] pairs (say, with [b, d, g, v, z, ɣ]) that are subject to neutralisation in word-final contexts.

The predictions of this radical formalism are falsified by the simple observation that phonological natural classes generally (although not always precisely) correspond to phonetic natural classes. Consequently, most models that would be counted in the formalist camp in the context of recent debates about the issue, are in fact hybrid frameworks, which incorporate notions such as articulatory/auditory *enhancement* (Stevens & Keyser, 1989). With the exception of Archangeli & Pulleyblank (1994), few of these models adhere to a well-defined

policy concerning the range of phonological phenomena that should be regarded as phonetically *grounded*, and it often seems to be a matter of common sense or the scope of formalist machinery. Nevertheless, even the ostensibly anti-functional [Kaye \(1989\)](#) maintains that phonological rules have an ultimate purpose as aids to grammatical segmentation and lexical access.<sup>9</sup>

The only feasible way of rescuing radical formalism is to claim that, as mental objects, phonological grammars operate on substance-free structures, but that language acquisition filters out those grammar-phonetic interpretation pairs that are impossible to use. This position, which borrows heavily from the functionalist theory of phonological change developed by [Ohala \(1981, 1993\)](#), is taken by [Hale & Reiss \(2000a\)](#). It implies that grammars treating sets *a* and *b* above as /p, t, k, f, s, x/ are identical at the level of phonological representation, and, as far as that representation is concerned, equally likely to occur. Because language learners have problems in acquiring this system of obstruents when it is paired with the sounds in *b*, it will only ever be interpreted in terms of *a* or a phonetically similar set of sounds such as [p<sup>h</sup>, t<sup>h</sup>, q<sup>h</sup>, φ, s̄, χ].

[Hale & Reiss \(2000a\)](#) then define the discipline *phonology* as the study of phonetically arbitrary systems that can be mentally represented, rather than as the study of phonetic systems that are selected by language learners. The advantage of this position is that it exempts their version of radical formalism from the all-too-obvious objections sketched above. But as they relinquish most of the (usual) predictions about the gross phonetic shapes of spoken language, it is unclear how models constructed according to this logic can be tested. [Hale & Reiss \(2000a,b\)](#) may be interested in ‘I-phonology’ (an abstract level of mental representation), but ‘E-Phonology’ (observations about speech production and perception) is the only available data.<sup>10</sup> Conversely, any descriptive generalisation concerning an E-phonological phenomenon can only be attributed to an I-phonological mechanism with some confidence if an external (acquisition-driven) explanation can be categorically ruled out. It would therefore appear that Hale and Reiss’s research program invests rather heavily in the potential (or perceived) limitations of acquisition-driven (functionalist) explanations of phonetic inventories and rules.

Worse, in theory it is possible that there are ‘latent’ principles of I-phonology that will never emerge in spoken language, because they are impossible to acquire and use for human speakers, whatever their phonetic exponence. In a sense therefore, the conception of phonology adopted by [Hale & Reiss \(2000a,b\)](#) is

<sup>9</sup>Perhaps the position of Kaye and other proponents of Government Phonology is better summed up as ‘opposed to *articulatory* phonetic explanations of sound patterns’. See [Harris & Lindsey \(1995, 2000\)](#).

<sup>10</sup>Hale & Reiss’s use of the term (phonological) *computation* in these two papers does not seem to refer to online language processing, and should probably be understood at the more abstract level of *computational theory* in the sense of [Marr \(1982\)](#).

comparable to a form of theoretical genetics investigating the space of ‘possible species’ as constrained by hypothetical ‘syntactic’ restrictions on nucleotide sequences, but without access to the chemistry that would enable it to test its claims.

### 1.4.2 Synchronic functionalism

Perhaps the most radical form of functionalism is represented by the ‘synchronically functional’ models of Boersma (1998), Kirchner (1998), Flemming (2001) and, in a slightly different way, Steriade (1997). These models imply that all rules of the phonetic grammar (both phonological and phonetic) are motivated on grounds of speech perception, ease of articulation, and other usage-based considerations. Moreover, they imply that the relative utility of a given utterance with respect to these functional considerations is computed online during speech production. For example, Steriade (1997) notes that the contrast between alveolar and retroflex consonants is less stable in word-initial and postconsonantal than in postvocalic contexts: if a language allows it in the former context it also maintains it in the latter, but the reverse does not hold. Steriade explains this contextual asymmetry in terms of the relative *perceptibility* of the contrast in question, i.e., the perceptual distance between corresponding alveolar and retroflex consonants. There is a marked difference in the  $F_3$  and  $F_4$  transitions for alveolar and retroflex consonants at the V-C boundary but not at the C-V boundary, and consequently it seems safe to assume that the perceptual distance between alveolars is greater after a vowel than after a consonant, where there are no  $F_3$  and  $F_4$  transitions. The *licensing-by-cue* model of Steriade (1997) suggests that information about the context-dependent relative perceptibility of the alveolar-retroflex contrast is encoded as such in speakers’ phonetic knowledge, and forms the basis for a cascade of phonological constraints on the distribution of retroflexes.

The claim that speakers make online judgments about the perceptual and articulatory disadvantages of phonetic forms (given the background noise in their immediate environment) is more explicit in Boersma (1998), Kirchner (1998), and Flemming (2001). The last of these presents an elegant model of consonant-to-vowel coarticulation that calculates the realised  $F_2$  (locus) targets for sequences of consonants and vowels as a function of their faithfulness to the relevant hypertargets, the effort involved in realising an  $F_2$  transition of a given size, and the importance speakers attach to these factors at a given time. The fact that this model can compute the relative amount of effort involved in any possible  $F_2$  transition strongly effectively entails that speakers are able to the same during online speech production.

Models that propose to model all of the phonetic grammar in these terms invariably founder on the observation that a number of well-documented phono-

logical rules, and even some *productive* phonological rules, lack synchronic motivation in terms of perceptibility, ease of articulation, or other usage-based considerations (e.g., [Bach & Harms 1969](#); [Anderson 1981](#); [Gussenhoven 1996](#)). Interest in such *unnatural* or *crazy* rules and their implications for phonological and phonetic models seems to be tied to (perceived) paradigm shifts in the field, and thus seems to emerge cyclically.

As a first example of a crazy rule, consider the dialectology of velar softening in Faroese. Velar softening, a change of a velar stop [k, g] to a palatoalveolar affricate [tʃ, dʒ] before nonlow front vowels is in itself a fully motivated process. As [Flemming \(1995\)](#) points out, velar obstruents can become fronted to palatals by coarticulation with vowels involving a coronal gesture. The resulting palatals are then likely to be reanalysed as palatoalveolar affricates because releasing a dorso-palatal occlusion tends to create a relatively high amount of friction. However, [Hellberg \(1980\)](#) demonstrates how the morphonology of Faroese tends to retain the reflexes of velar softening, even if vowel change has removed the original conditioning environment. Consequently, it is impossible to describe this phenomenon in a synchronic functional grammar of the type proposed by [Boersma \(1998\)](#), [Kirchner \(1998\)](#) and others, unless it is encoded directly into lexical forms and treated as inert debris of language change that is somehow left untouched by usage-based mechanisms.

(2) Faroese velar softening (data from [Hellberg 1980](#))

Orthography	Phonology	Gloss
<koma>	/koma/	come-INF.
<kemur>	/tʃemur/	come-2/3.SING.PRES.
<gav>	/gav/	give-PRET.
<geva>	/tʃeva/	give-INF.
<bøkur>	/bøkur/	book-NOM./ACC.PL.INDEF.
<bókin>	/boutʃin/	book-NOM.SING.INDEF.
<egg>	/εg:/	egg-NOM./ACC.SING.INDEF.
<eggið>	/εdʒ:ið/	egg-NOM./ACC.SING.DEF.

The examples in (2) represent a relatively abstract (orthography-driven) analysis of Faroese velar softening, illustrating how palatoalveolar fricatives before nonlow front vowels alternate with velar stops elsewhere (Hellberg's [č] and [j] have been replaced by [tʃ] and [dʒ]). These examples suggest that the process is synchronically motivated along the lines described by [Flemming \(1995\)](#). However, the transparency of the velar stop/ palatoalveolar affricate alternation in (2) is deceptive, because relatively recent sound changes in many modern dialects of Faroese have distorted the mapping between vowel quality and the place of articulation of dorsal stops. Thus, in a northern dialect described by [Hellberg \(1980\)](#), <tóku> *took*-PL. is realised with a nonlow and front suffix

vowel but nevertheless retains the velar stop that was motivated by the original high back suffix vowel: [touke]. In other words, in spite of the presence of the triggering environment in the surface form, the velar softening rule does not apply. Conversely, in the dialect of the island of Suðuroy, <fisikin> *fish-ACC.SING.DEF.* is realised with a back or centralised rounded suffix vowel but velar softening nevertheless applies: [fistʃən]. Note that the same dialect realises <fiskum> as [fiskøn].

In contrast to the northern and Suðuroy (southern) dialects of Faroese, a number of varieties retain the distinction between /i/ and /u/ in suffixes but have redistributed them. This redistribution process again obscures the relation between velar softening and surface vowel quality. For instance, velar stops are preserved before [-ir]-NOM.PL: cf. <røkur>, [rø:kir] *rock ledges*, <vikur> vi[vi:kir], *weeks*, <lungur> [luŋjir], *lungs*. On the other hand, palatoalveolar affricates appear before [-ør]-2/3.PRES.SING. (historical [-ir]): <vakir> [ve:tʃør], *is awake*, <tekir> [te:tʃør], *covers*. Note that where the present day quality of the suffixal vowel corresponds to its original value, velar softening is transparent: <sangir> [saŋdʒir], *songs*, and <leggur> [lɛgʝør], *puts*.

(3) Limburg Dutch diminutive formation (data from [Gussenhoven 1996](#))

UR	Phonetic form	Gloss
/du:m/ +/kə/	[dy:mkə]	thumb
/vo:t/ +/kə/	[vø:cə]	foot
/kra:x/ +/kə/	[kre:çskə]	collar
/snɔ:r/ +/kə/	[snɪrkə]	moustache
/bəl/ +/kə/	[bɛlkə]	ball

[Gussenhoven \(1996\)](#) provides a crazy rule from Limburg Dutch that proves even more problematic for theories of synchronic functionalism. In this group of dialects the suffixation of diminutive /kə/ causes the last stressed back vowel of a stem to front, and, if it is a low vowel, to raise (cf. the examples in 3). At one stage, this umlaut rule was a regular palatal harmony process triggered by a high front /i/ in the diminutive suffix. Although the phonetic grounding of vowel harmony is not fully understood, it seems likely that the process is rooted in vowel-to-vowel coarticulation and related (compensatory) perceptual processes ([Fowler, 1981](#); [Busá & Ohala, 1999](#)). At a later stage, Limburg Dutch reduced the suffix vowel to /ə/. Although (centralising) vowel reduction is itself an uncontroversially natural process, in this instance it removed the trigger for the umlaut rule, rendering it phonetically opaque in synchronic terms. Nevertheless the Limburg dialects retained the process as part of their morphology, and according to [Gussenhoven](#) it is synchronically productive. It is this productivity that is especially problematic for synchronic functional models since it indicates that synchronically unmotivated patterns are not necessarily inert, but

are at some level recognised and applied as rules by speakers.

### 1.4.3 Diachronic functionalism

The existence of crazy rules is sometimes touted as proof that phonological grammars are built around a non-functional core and are, to an extent, a mathematical game after all. However, the sorts of crazy rules that are documented in the literature merely seem to falsify synchronic versions of functionalism, but not an alternative theory, which I will label *diachronic* or *evolutionary* functionalism. This form of functionalism is central to the theories of language change pursued by Ohala (1981, 1993) and (Blevins, to appear), underpins several recent attempts to simulate language evolution (de Boer, 1999, 2001; Kirby, 1999; Briscoe, 2000; Kochetov, 2003), and is endorsed by Hale & Reiss (2000a), albeit not as part of what they consider the study of phonology to be about.

Rather than claiming that speakers are able to make online judgments about the effort involved in the production of an utterance and its precise perceptual consequences, diachronic functionalism views most phonetic behaviour as simply learned. Language learners are assumed to be fundamentally conservative in striving to copy the patterns they encounter in their speech community as faithfully as possible.<sup>11</sup> However, speech transmission is an inherently noisy process, both in the literal sense of ‘affected by background noise’ and because speech perception and production are not perfect, error-free, processes. The noise in the speech transmission chain is likely to introduce copying errors of various sorts. Although Ohala (1981, 1993) seems to assume that these copying errors are necessarily discrete at the level of lexical phonological contrast, given that the peripheral auditory and articulatory systems process continuously-valued representations, this assumption is unfounded. For example, on encountering a certain number of (partially) devoiced word-final [-tense] obstruents, a learner of English might conclude that voicing distinctions do not cue [ $\pm$ tense] in this environment. But if additional phonetic distinctions between [ $\pm$ tense] obstruents in terms of segmental duration,  $F_0/F_1$  perturbation, and release characteristics are sufficiently salient, there is little ground for this learner to decide that there is no phonological contrast at all, and to include a rule of word-final laryngeal neutralisation in his/her developing phonetic grammar.

The central claim of diachronic functionalism is that various forms of *feedback* received by language learners create a form of selectional pressure that determines whether the copying errors survive in their mature grammars as innovations with a chance of being passed on to the next generation of learners. One form of feedback is supplied by the learners’ own perceptual systems and

<sup>11</sup>Evolutionary functionalism does not require that the language acquisition process be fully inductive. In fact, both Kirby (1999) and Briscoe (2000) investigate scenarios in which language and an emergent UG co-evolve.

provides information e.g., about the relative amount of effort spent in producing an utterance (proprioceptive feedback) with a given phonetic make-up. The second form of feedback consists of the responses of the speech community to forms produced by the learner, which provides a measure of the communicative utility of an utterance with a particular phonetic make-up. This second type of feedback comes in a variety of linguistic and non-linguistic forms, and includes information both about the efficacy of a form in conveying the intended message, and its social status.

The probability of survival of a given phonetic form or pattern depends on the net amount of positive feedback received by the learner. Forms that incur a low amount of positive feedback are likely to be discarded whilst in favour of alternative phonetic encodings of the same message that receive a higher amount of positive feedback. On the assumption that effective communication (construed in the broadest sense possible) is the main goal of speaking and hence that feedback from the speech community receives considerable more weight than proprioceptive feedback, this selection process creates a bias towards forms that are easy to parse by listeners and are easy to produce by speakers to the extent that this does not interfere with parsing.<sup>12</sup> Thus, usage-based constraints can enter the phonetic grammar without speakers being able to assess their utility for various purposes in explicit terms, and diachronic functionalism removes all undesirable ‘teleology’ (Ohala, 1993) from the phonetic grammar.

The idea that functional considerations enter the phonetic grammar during acquisition has a number of important ramifications. First of all, because function-driven change is cumulative (successive generations each add their own innovations), diachronic functionalism predicts the existence of crazy rules, as long as they can be decomposed into a sequence of changes that are in themselves motivated by parsing or production considerations. Judging by the literature on the topic, this is at least typically the case: in fact, authors such as Bach & Harms (1969), Anderson (1981), and Gussenhoven (1996) make a point of demonstrating how crazy rules emerge from the aggregation of phonetically motivated changes. Note that this observation contradicts radical formalism (barring the version espoused by Hale & Reiss 2000a,b), which predicts that individual changes need not be functionally motivated and hence that crazy rules do not necessarily decompose in terms of such motivated changes. The latter position implies that a pattern along the lines of the Limburg Dutch diminutive illustrated in (3) could arise without an intermediate stage in which the suffix contains a high front vowel.

Second, as hinted above, evolutionary functionalism derives the presence of language usage-based constraints in the phonetic grammar as an epiphenomenon

---

<sup>12</sup>In this context, the term *parsing* should be understood as the totality of sound processing operations performed by a listener to decode a message.

of the language learning process. This entails, for instance, that retroflexes are not avoided in initial and postconsonantal contexts because speakers know that they are hard to distinguish from alveolars there, but simply because (a) learners fail to perceive a contrast between alveolars and retroflexes in these contexts and reanalyse all stops as alveolar; or (b) learners ‘inventing’ a contrast in these contexts (e.g., by reanalysing coarticulation differences between following rhotic and non-rhotic sounds) do not get sufficient positive feedback from their speech community (i.e., because there are no advantages from a parsing point of view).

Similarly, feedback-driven selection of innovations arising out of copying errors is able to account for the instability of, or gaps corresponding to, phonetically voiced [g] in [p, t, k, b, d, (g)] systems (offered by Boersma 1998 as an example of true teleology in language change). It seems probable that learners trying to produce voiced [g] occasionally stumble on nearby sounds in articulatory space such as voiced [ɣ, ŋ], voiceless fortis [x], or voiceless lenis [χ̥], all of which are somewhat easier to produce because they do not involve trying to maintain voicing behind a back constriction that allows for only limited oral cavity expansion (cf. chapter 2). All these sounds retain an important property of [g], i.e., its place of articulation, and compared to e.g., [c, q, tʃ, ʃ, ɟ, dʒ, ʝ, ɸ, ɹ, ɻ, ɲ] (ignoring the effects of flanking vowels), they are therefore relatively likely to be tolerated as substitutions by the speech community. Consequently, it seems safe to assume that they receive a relatively high amount of positive feedback, and the (correct) prediction follows that they are the most likely alternative candidates beside [g] to take on its structural role in a /p, t, k, b, d, g/ inventory.

Two further candidates that retain the place cues of [g] whilst being easier to produce in terms of voicing are voiceless fortis [k] and voiceless lenis [k̥]. Substitution of the former leads to neutralisation in production and perception of the [tense] contrast for the velar place of articulation, which may be tolerated by the speech community under certain circumstances. Substitution of [k̥] on the other hand, does not lead to full neutralisation, but depending on the other cues involved in the phonetic expression of [tense], may reduce the amount of contrast with /k/, which is in turn predicted to raise the chance of misperception and neutralisation by the next generation of learners. Thus evolutionary functionalism is able to handle both cases of apparent goal-driven behaviour by speakers (pace Boersma 1998) as well as gradient sound change (see above: pace Ohala 1993).

#### 1.4.4 The emergence of structure

From the point of view of the nonmodular phonetic grammar model described in section 1.3.2 above, a very important consequence of evolutionary functionalism is that it derives phonetic (and hence phonological) categories in continuous articulatory and perceptual space. This point is perhaps best illustrated by a

brief summary of the simulations carried out by [de Boer \(1999, 2001\)](#).

The architecture of de Boer's model consists of a population of 20 *agents* representing human language users. Every agent is endowed with an (initially empty) inventory of paired articulatory and auditory vowel targets, a vowel synthesiser (articulation model) and a vowel recogniser (perception model). Both articulatory and auditory space are modelled in continuous terms: there is no level of discrete representations that would hardwire category formation into the model. Articulatory targets are represented in terms of height, position and rounding whilst auditory targets are represented as a set of co-ordinates in  $F_1$ - $F_2$  space expressed on the Bark scale. The second formant is calculated as the perceptual  $F_2$  or  $F_2'$  ( $F_2$ -prime), which takes on board the contribution of higher formants in the acoustic spectrum to the perceived frequency of the second resonance peak (cf. [Chistovich & Lublinskaya 1979](#)).

Simulations consist of a series of *imitation games* between pairs of agents. Each game starts with an initiator transmitting a vowel sound generated from a randomly selected articulatory target in its inventory. The receiver, or *imitator* classifies this signal in terms of the perceptually nearest vowel in its own system, synthesises the corresponding articulatory target and sends it back to the initiator. An imitation game is labelled as successful if the response signal is classified as identical to the stimulus by the initiator, and the success or failure is relayed to the imitator in terms of a 'non-verbal' feedback signal. This feedback signal and the longer term communicative effectiveness of a vowel category (defined as the ratio between the number of times a vowel is used and the number of successful uses) determine how the vowel inventory of the agents is updated after every game. Vowel targets can be shifted in articulatory and auditory space, and vowel categories can be introduced, merged, or discarded. The mapping between feedback (history) and specific update operations introduce a bias in the model towards a vowel system that is shared by all members of the population: it favours high communicative effectiveness indices for all individual vowel targets in the inventories of all individual agents and the sum of these indices is maximal if all agents share the same inventory.

Two further properties of the model developed by [de Boer \(1999, 2001\)](#) are crucial to cumulative effect of the imitation games on the vowel inventories of the agents. The first is the addition of noise to the vowel signals transmitted between the agents. Technically speaking, this noise consists of transforming the signals in the  $F_1$  and  $F_2'$  domains randomly, but within fixed bounds that represent the 'noise level'. This means that vowel targets with overlapping noise ranges run the risk of being confused during imitation games. Given the communicative pressure described in the previous paragraph, noise addition therefore creates a bias towards auditory dispersion. Secondly, and essentially to keep lexical pressure on the model, a random vowel is added to the inventory of an

agent with a probability of .01 per game. This pushes the model away from a shared inventory with a highly effective single vowel.

After a certain number of imitation games, the model starts to converge on a relatively steady state in which the agents have highly similar inventories, with the spacing of vowels (and consequently their number) roughly inversely proportional to the level of noise in the transmission process. Every individual agent has a finite number of vowel targets with more or less stable co-ordinates that approximate the configurations of vowel targets in the rest of the artificial speech community, and can therefore be said to have developed a set of vowel *categories*. As a collective, the agents converge on clusters of targets in articulatory and auditory space that are similar to the phonetic clusters that realise lexical contrasts in human speech production.

Elsewhere (Jansen, 2001b) I have criticised some aspects of de Boer's methodology and the details of his interpretation of the simulation results. But these criticisms by no means undermine his basic conclusion that it is possible to generate vowel categories in continuous phonetic space on the basis of a noisy speech transmission chain and selection on the basis of feedback from a speech community, the two most important ingredients of diachronic functionalism. Intuitively speaking, the logic of this approach is perhaps easiest to apply to the development of vowel categories, but in principle, it is capable of generating categories in any sort of multidimensional space without the intervention of a discretely-valued level of representation, i.e. a separate phonological module.

For example, an extended version of de Boer's model should be capable of accounting for the phonetic properties associated with the [tense] contrast. As pointed out in chapter 2, there are good grounds to believe that the multiple cues many languages associate with the lexical contrast between /p, t, c, k, q/ and /b, d, ʒ, g, ɠ/ are organised in a mutually enhancing fashion. Under a diachronic functional theory this organisation would arise without the need for an explicit categorical [±tense] feature. Speakers would simply 'discover' the observed configurations of phonetic features by trial and error during the acquisition process: the combination of voicelessness with a short segmental duration for example would incur less positive feedback than the (commonly observed) combinations of (active) phonetic voicing with a short obstruent duration and (active) devoicing with long segmental duration. The same line of reasoning can be applied to the emergence of prosodic hierarchies in the vein of Nespor & Vogel (1986) or Pierrehumbert & Beckman (1988), which can of course not be explicitly encoded in a nonmodular phonetic grammar.

#### 1.4.5 Perceptibility

The theory of phonological change defended by Ohala (1981, 1993) revolves around the effects of perception errors during language learning. In one of

two possible scenarios, a learner fails to detect a phonological contrast in the speech of the surrounding speech community and therefore neutralises it in his/her developing grammar. In the second scenario, the learner interprets gradient context-dependent variation, due to e.g., coarticulation, as a reflex of a lexical phonological contrast and grammaticalises it as such, even it is a gradient phonetic rule in the speech of older speakers. Feedback-driven selection then determines whether these types of innovative neutralisation and phonologisation survive in the adult grammar of the learners and subsequent generalisations.

Thus, Ohala's model, and diachronic functionalism more generally relies relatively heavily on the notion of *relative perceptibility* or *salience*: the assumption seems warranted that contrasts are more likely to escape detection by learners when they are relatively imperceptible, and conversely, that relatively salient forms of gradient variation are more likely to be phonologised. Consequently, this notion deserves to be made a little more precise.

First, the relative perceptibility of a phonetic (hence phonological) contrast between two sounds can be defined in terms of the likelihood that two sounds are confused with each other by listeners. Studies of perceptual confusion such as [Miller & Nicely \(1955\)](#) show that this likelihood is far from the same for every possible pairing of sounds. For example, the voiced lenis fricatives of English are more likely to be confused with each other and lenis stops, than the corresponding voiceless fortis fricatives. Likewise, the relative perceptibility of a given phonetic category in a given context can be defined in terms of the frequency with which it is identified correctly by listeners. [Mielke \(2001\)](#), for instance, demonstrates how [h] is less perceptible at the end of an utterance (i.e., is identified correctly in a lower number of instances) than before a vowel.

Second, it appears that the relative perceptibility of a given contrast or a given sound in a particular context depends on a number of factors, including the number of available cues and their interaction with (e.g., masking by) the phonetic context they appear in, and the native language of a listener. The roles of both of these factors is demonstrated by the experiments reported in [Mielke \(2001\)](#), which show both language-specific effects in the perceptibility of [h], and crosslinguistic effects based on the availability of specific cues. Mielke's data shows how native speakers of Turkish and Arabic, languages in which [h] and similar sounds have a relatively wide distribution, are better at perceiving this sound across phonetic contexts than native speakers of English, in which [h] only occurs before stressed vowels, and French, which lacks contrastive [h] altogether. Despite these differences in overall identification levels, the effects of phonetic context are remarkably similar across languages. Thus, for all 4 languages, the lowest proportions of correct [h] (and non-[h]) identifications occur before voiceless obstruents and utterance finally. The most likely cause of this effect is the absence in this set of environments of the voicing/F<sub>0</sub> onset

that signals (the end of) [h] before voiced sounds. This mechanism is probably reinforced by the low salience of consonantal onset cues vis-à-vis offset cues, which has been demonstrated independently by Raphael (1981).

## 1.5 Conclusion: the phonetics-phonology interface revisited

Figure 1.2 depicts the model of speech production and perception described in the previous sections. The 'underlying' representations of this model do not consist of abstract phonological features, but of hyperarticulated articulatory and auditory targets represented in terms of continuously-valued features. These parametric representations have the same structure as the interface representations supplied/used by the peripheral perceptual and articulation systems. Articulatory representations can be conceived of as gestural scores in the Articulatory Phonology sense (Browman & Goldstein, 1986), *articulator windows* in the fashion of Keating (1990b), or the speech motor goals of Perkell et al. (1995). Irrespective of the choice of framework however, interface level articulatory representations specify all aspects of articulation that cannot be attributed to coarticulation, the anatomy of the vocal tract, or low-level reflexes.

Auditory forms encode the linguistic aspects of the acoustic form of speech sounds which is initially delivered by the peripheral auditory system. There is ample evidence that linguistic auditory processing imposes various forms of normalisation on the raw input signal and integrates individual acoustic cues into more abstract objects: as reviewed in chapter 2 for example, voicing,  $F_0$ , and  $F_1$  cues to [tense] may all be integrated into a single 'low frequency' feature. On the other hand, since native speakers of different languages (i.e., voicing and aspirating languages) respond differently to the presence vs. absence of the voicing component of this higher level perceptual feature, it must be assumed that some or more of the individual acoustic cues are differentiated at some stages of linguistic auditory processing.<sup>13</sup>

In the production of an utterance, articulatory hyperforms are filtered through a set of categorical and gradient rules. The former change (clusters of) hypertarget values in discrete steps, or 'remove' targets altogether, that is, phonetically underspecify sounds for one or more phonetic features. The latter set of rules acts in a continuous rather than discrete fashion, but since they operate on the same parametric representations, gradient rules may occasionally have the same effects as discrete, phonological rules. These rule blocks can be

---

<sup>13</sup>For expository reasons I have omitted the role of other sensory modalities, notably vision, in speech perception. Nothing crucial hinges on this. For a detailed discussion of the role of visual information in speech perception, and its integration with auditory information, see Massaro (1998).

interpreted in procedural or declarative terms, and (under the latter interpretation) may be regarded as generalisations over ‘clouds’ of stored exemplars (with hyperforms assuming some special status) or as devices used to construct parts of linguistic phonetic forms on the fly during speech production. Which of these interpretations is the most suitable for which (sub)sets of rules depends on data this study is not specifically concerned with (see e.g., [Levelt 1989](#)).

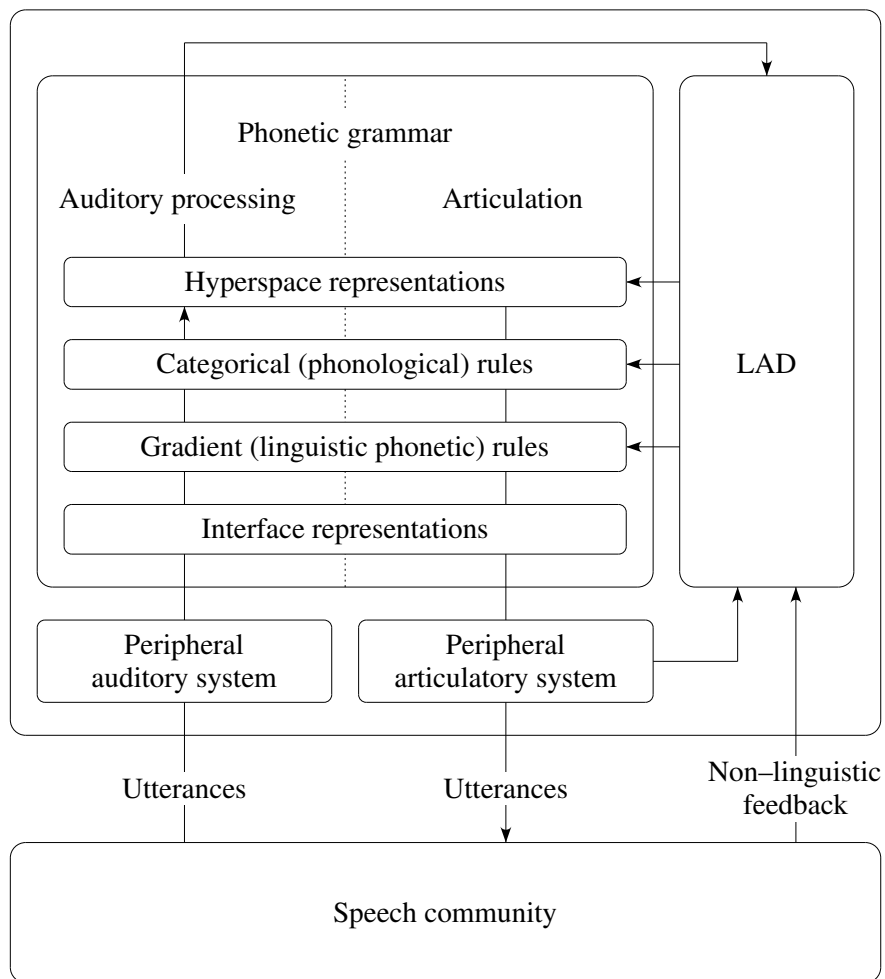


Figure 1.2: The model of speech production and perception adopted in this study.

The peripheral articulatory system responds to the instructions provided by the phonetic grammar by producing utterances, often to some sort of human

audience. The label *speech community* in figure 1.2 generalises over all possible forms of audience that are capable of providing some sort of feedback to the producer of the utterance. Any form of spoken feedback is processed by the original speaker's peripheral auditory system and delivered to the linguistic system, which maps it onto a grammatical form and ultimately some sort of meaning. No one would (still) claim that this mapping proceeds in a strict bottom-up fashion, reconstructing the hypothesised stages in the production process in a step-by-step fashion, and there is a reasonable amount of evidence to suppose that knowledge of phonetic and phonological rules aids this process. For instance, several researchers have found that reflexes of coarticulation or phonological assimilation do little or nothing to impede lexical access, whilst some have even suggested that the presence of context effects improves the sound-to-meaning mapping (Elman & McClelland, 1986; Gaskell & Marslen-Wilson, 1996, 1998; Quené & Krull, 1999). Similarly, Aylett (2000) reports psycholinguistic data which indicates that listeners benefit from the 'hypoarticulation contour' imposed on utterances by prosodic strengthening and weakening at constituent edges. It is for these reasons that the rule blocks straddle the auditory-articulatory divide in figure 1.2.

As argued in section 1.4.3, phonetic and phonological rules are not constructed on the basis of grammar internal formal templates or functional principles such as effort minimisation, but on the basis of learning, error, and feedback. Using traditional terminology, I have labelled the module responsible for (re)structuring the phonetic grammar on the basis of incoming information *Language Acquisition Device* (LAD). The use of this term highlights the role of the acquisition process in generating linguistic change and the incorporation of functional mechanisms, but does not imply that (re)structuring of the grammar ceases completely after the offset of the famous 'critical period' for language acquisition. The LAD receives data from a variety of sources, some of which are indicated in figure 1.2.

Because errors in perception and production, selectively incorporated into the phonetic grammar by the LAD, drive the form of phonological and phonetic rules, the formal statement of those rules becomes arbitrary. Phonological rules might be stated using the formalism adopted by Chomsky & Halle (1968) or in autosegmental terms, with (distinctive) features serving as notational shorthands for clusters of phonetic features, but as long as both frameworks are able to capture the relevant generalisations there are no empirical grounds for deciding between them. Phrased in more general terms, the framework adopted in this dissertation renders all empirical arguments for or against particular formalisms void, whether they concern, e.g., the advantages of autosegmental feature lattices over feature bundles, monovalent over bivalent feature representation, or declarative over procedural grammars.

Despite its differences with models of the phonology and the phonology interface typically encountered in the theoretical phonology literature (at least until recently), the model illustrated in figure 1.2 reconstructs a number of properties found in more traditional frameworks. By way of conclusion to this section and this chapter it is perhaps useful to point out some of the more important parallels.

First and foremost, as pointed out by Johnson et al. (1993) there is an important parallel between the hyperform-interface mapping and non-monotonic lexical-to-surface mappings in traditional phonological grammars: both lead to the loss or distortion of (lexical) information. For example, many generative models, including most current versions of Optimality Theory, in principle allow a lexical contrast between /i, e, a, o, u/ to be neutralised to phonological and phonetic [ə] on the surface by removing and/or replacing the relevant features. This mapping involves a loss of information in the sense that it is impossible to reconstruct the underlying vowel contrast on the basis of the forms exhibiting a reduction schwa. Similarly, phonetic vowel reduction can reduce a [i, e, a, o, u] distinction in hyperspace to, say, [ə, ə, ə], and whilst this process is incompletely neutralising, it does not allow for the original phonetic values to be reconstructed. For example, [ə] might correspond to hyperspace [i, e] or even [ɪ], but without additional (e.g., paradigmatic) information it is impossible to determine the underlying phonetic category.

Second, whilst the framework adopted here abolishes phonology as a separate, representationally distinct level of representation, it does not dispense with the notion of phonological contrast as a discontinuity in phonetic space. Although some, such as Port (1996) have implied that rules operating in phonetically discrete fashions do not exist, there is clear experimental evidence to the contrary (see Zsiga 1997 and chapter 2 below). Therefore, the diachronic functional model in figure 1.2 retains a set of phonological rules as opposed to a set of gradient phonetic rules, even if both types of rule operate on the same parametric phonetic representations. Which of the rules described in descriptive grammars or the theoretical literature as categorical indeed belong to this class, is simply an empirical matter.

Third, it is precisely the absence of a phonology-phonetics interface in the sense of e.g., Keating (1990a) that renders the framework in figure 1.2 similar in some ways to the monostratal models of Pierrehumbert & Beckman (1988) and Harris & Lindsey (1995, 2000). For example, the latter state that individual phonological elements, and consequently the lexical, intermediate and ‘surface’ forms composed of them are always phonetically interpretable. Phonological rules manipulate elements but do not transform them into (approximations of) interface representations. Thus, occurring on its own, the element A, is interpreted as a vowel with a low first and high second resonance, i.e., an unrounded low vowel. This view contradicts the position of Chomsky & Halle (1968),

restated more recently by Bromberger & Halle (1989), which holds that the purpose of phonological and phonetic rules is to progressively convert abstract underlying forms into structures that are understood at the interface levels. The model adopted here sides with Harris & Lindsey (1995, 2000) in the sense that hyperforms can be understood by the peripheral systems, in spite of the fact that the auditory/articulatory values encoded in hyperforms are not typical of interface forms encountered in speech production.

Fourth and finally, the LAD as conceived here corresponds to (certain versions of) *H-EVAL* in OT, albeit in a fairly abstract sense. The LAD evaluates forms produced by the speaker with respect to several forms of feedback ('constraints'), preferring forms that receive a certain amount of positive feedback ('a certain number of violation marks') over those that incur less positive feedback ('more violation marks'). The crucial difference is that the LAD processes feedback to forms that have been produced at a particular place and in the presence of a particular audience whereas *H-EVAL* is normally viewed as a device that determines which forms can (and will) be produced in the first place. Nevertheless, the basic idea that phonetic grammars are shaped by competing factors selecting optimal candidates from an array of alternatives (generated by *GEN* or errors in production and perception) is central to both standard OT models and the framework adopted here.