THE ELEMENTS OF FUNCTIONAL PHONOLOGY[1]

Paul Boersma
University of Amsterdam, The Netherlands

January 1997

(Please send any comments, criticisms, and challenges to boersma@fon.let.uva.nl)

Abstract

Phonological structures and processes are determined by the functional principles of minimization of articulatory effort and maximization of perceptual contrast. We can solve many hitherto controversial issues if we are aware of the different roles of articulation and perception in phonology. Traditionally separate devices like the segment, spreading, licensing, underspecification, feature geometry, and OCP effects, are surface phenomena created by the interaction of more fundamental principles.

[1] I thank Louis Pols and Mirjam Ernestus for their comments on earlier versions of this paper.

Contents

Introduction: from speaker to listener

The functional hypothesis for linguistics maintains that the primary function of a language is communication, and that languages are organized in a way that reflects this.

Consider the English utterance *tense*. Its underlying phonological form is

/tɛns/ (0.1)

I will take this to be the *perceptual specification* of the utterance: if the speaker produces the specified perceptual features in the specified order with the specified time alignment, the listener will recognize the utterance as /tɛns/, and a substantial part of the communication has succeeded. This basic insight should be reflected in our theory of grammar.

Several articulatory strategies can be followed to implement the utterance (0.1). In some varieties of English, a part of the dominant *articulatory implementation* is (time runs from left to right):

<table>
<tr><td>tongue tip</td><td>closed</td><td colspan="3">open</td><td colspan="3">closed</td><td>critical</td></tr>
<tr><td>velum</td><td colspan="3">closed</td><td colspan="2">open</td><td colspan="3">closed</td></tr>
<tr><td>glottis</td><td colspan="2">wide</td><td colspan="4">narrow</td><td colspan="2">wide</td></tr>
<tr><td>lips</td><td colspan="8">spread</td></tr>
</table>

(0.2)

This will give rise to an acoustic output that we can translate into the following table of perceptual phonetic events, time-aligned with the articulatory score (0.2) (*tr* = transition):

<table>
<tr><td>silence</td><td>+</td><td colspan="4"></td><td colspan="2">+</td><td></td></tr>
<tr><td>coronal</td><td></td><td colspan="3">burst tr.</td><td>side</td><td colspan="2"></td><td>bu. cont</td></tr>
<tr><td>voiced</td><td colspan="2"></td><td colspan="3">sonorant</td><td colspan="3"></td></tr>
<tr><td>noise</td><td></td><td>asp</td><td colspan="5"></td><td>sibilant</td></tr>
<tr><td>F1</td><td></td><td colspan="3">open mid</td><td colspan="4"></td></tr>
<tr><td>F2</td><td></td><td colspan="3">max</td><td colspan="4"></td></tr>
<tr><td>nasal</td><td colspan="3"></td><td colspan="2">+</td><td colspan="3"></td></tr>
</table>

(0.3)

In a microscopic transcription (§3.3), this *perceptual result* can be written as [[thɛɛ̃n_ts]] (“_” = silence). With the help of the processes of categorization and recognition, the listener may reconstruct /tɛns/.

The theory of *Functional Phonology*, introduced in this paper, claims that the principle of *mimization of articulatory effort* evaluates the articulatory implementation (0.2) and its competitors, and that the principle of *maximization of perceptual contrast* evaluates the differences between the perceptual specification (0.1) and the perceptual result (0.3). Together, these principles will determine which candidate articulatory implementation will actually be chosen to surface.

In the present paper, I will defend the hypothesis that the distinction between articulation and perception is an integral part of the grammar:

- Functional principles control both speech production and speech perception (§1).
- Phonology controls both the articulatory and perceptual specifications of speech production (§2).
- The traditional hybrid feature system should be replaced with separate systems of articulatory gestures and perceptual features (§2).
- The traditional hybrid phonological representations should be replaced with perceptual specifications and outputs, and articulatory implementations (§3).
- Both articulatory and perceptual principles can only be brought into the grammar if that grammar allows constraint violation (§4).
- Constraints against articulatory effort branch into many families that can be ranked individually in each language (§5).
- The finiteness of the number of feature values in every language is a result of general properties of motor learning and perceptual categorization (§6).
- Constraints against perceptual confusion (§7) branch into many families of input-output faithfulness, which can be ranked individually in each language (§8).
- An adequate account of phonological structures and processes needs a comprehensive approach to the interaction between faithfulness and articulatory constraints (§9).
- As an example, §10 describes how the realization of vowel height in phonetic implementation is determined by the interaction of two continuous constraint families, and how phonetic and pragmatic circumstances influence the result by shifting the rankings of the constraints.
- The local-ranking principle, rooted in general properties of motor behaviour and perception, determines which constraints can be ranked universally, and which must be ranked on a language-specific basis (§11). The examples of nasal place assimilation and obstruent voicing will illustrate the typological adequacy of this approach. It leads to a straightforward strategy for the phonologization of phonetic principles.
- Both segmental and autosegmental faithfulness are visible in the grammar (§12); they refer to “vertical” and “horizontal” perceptual connections, respectively.
- The degree of specification in (0.1) should actually be quite high. All the arguments for a theory of underspecification vanish if we distinguish between articulatory and perceptual features, and between high- and low-ranked specifications (§13).
- Many recalcitrant issues in the study of segmental inventories, sound change, and synchronic autosegmental phenomena like spreading and the OCP, can be solved with the help of the distinction between articulation and perception (§14; Boersma fc. a-e).

1 Functional principles

Functional principles were first expressed in explanations for sound change. According to Passy (1890), sound changes have the same cause that motivates the existence of language itself: “language is meant to convey information from one person to another as quickly and clearly as possible”.

1.1 Functional principles of speech production

Passy states the *principle of economy*: “languages tend to get rid of anything that is superfluous”, and the *principle of emphasis*: “languages tend to stress or exaggerate anything that is necessary”. His use of the terms *superfluous* and *necessary* expresses the idea that articulatorily motivated constraints may be honoured unless stronger perceptually motivated constraints are violated. Passy’s two composite principles easily let themselves be disentangled into the speaker-oriented principle of the *minimization of articulatory effort* and the listener-oriented principle of the *maximization of perceptual contrast*.

1.2 Functional principle of the communication channel

Passy’s “quickly” translates into the principle of the *maximization of information flow*: “put as many bits of information in every second of speech as you can”.

1.3 Functional principles of speech perception

On the part of the listener, we have the functional principles of *maximization of recognition* and *minimization of categorization*.

The listener will try to make maximum use of the available acoustic information, because that will help her recognize the meaning of the utterance.

On the other hand, in a world of large variations between and within speakers, the disambiguation of an utterance is facilitated by having large perceptual classes into which the acoustic input can be analysed: it is easier to divide a perceptual continuum into two categories than it is to divide it into five. Moreover, if a contrast between two perceptual classes is not reliable, i.e., if an acoustic feature is sometimes classified into an adjacent category, successful recognition is actually helped by not trying to use this contrast for disambiguating utterances: if the listener accepts the phonological ambiguity of an utterance, she will take recourse to alternative (semantic, pragmatic) disambiguation strategies, which might otherwise not have been invoked. Labov (1994) showed that this principle can be responsible for segment merger in cases of dialect mixture.

1.4 Functional hypothesis for phonology

Thus, I maintain that historical sound changes, synchronic phonological processes, and the structure of sound inventories are built in such a way that the following natural drives will be honoured:

(a) The speaker will minimize her articulatory and organizational effort, i.e., she will try to get by with a small number of simple gestures and coordinations.
(b) The speaker will maximize the perceptual contrast between utterances with different meanings.
(c) The listener will minimize the effort needed for classification, i.e., she will use as few perceptual categories as possible.
(d) The listener will minimize the number of mistakes in recognition, i.e., she will try to use the maximum amount of acoustic information.
(e) The speaker and the listener will maximize the information flow.

These principles are inherently conflicting:

- Minimization of effort conflicts with maximization of contrast.
- Minimization of categorization conflicts with maximization of recognition.
- Maximization of information flow conflicts with both minimization of effort and minimization of categorization (§8.6).
- Conflicts also arise *within* the various principles, e.g., the minimization of the number of gestures conflicts with the minimization of energy.

Making typologically adequate predictions about what is a possible language under this hypothesis, involves formalizing the various aspects of the functional principles (§4). We can achieve this by translating each of the principles (a) to (d) directly into several families of *constraints*, which will be identified in §5, §6, and §8. Since the principles are inherently conflicting, the constraints, if stated in their naked, most general forms, must be *violable*. We can expect, therefore, much from formalizing their interactions within a framework of constraint-*ranking* grammars, which, fortunately, is now available to the phonological community in the form of Optimality Theory. First, however, we must determine the nature of the phonological spaces (§2) and representations (§3) on which the constraints will be defined. This will lead to a replacement of the traditional hybrid features and representations with systems based on general properties of human motor behaviour and perception.

2 Articulatory, perceptual, and hybrid features

A thread of this work is the idea that features of speech sounds, language-dependent though they may be, can be divided into two large classes: articulatory and perceptual features. These two groups play different roles in phonology, and an awareness of the difference between them will solve many hitherto unsettled problems in several realms of phonological debate.

The difference between the two groups of features can be traced to their different roles in speech production and perception.

2.1 Articulation versus perception in speech production

Figure 2.1 shows a simplified view of how the articulatory and perceptual aspects of phonology are integrated into speech production. The point labelled “start” marks the interface of the rest of the grammar to the phonological/phonetic component. In the following paragraphs, I will explain this figure. The main point that I am trying to establish, is that phonology controls both the articulatory and the perceptual specifications of the utterance, i.e., both the representations that we saw in (0.1) and (0.2).

– Top right: length control. The speaker can control the tension of a muscle. For this, a direct *muscle command* (every term set in italics can be found in figure 2.1) is conducted by the α neuron fibers from the spinal cord or the brain stem to the muscle fibers, whose contraction then results in a change in the shape of the human body, e.g., a change in *vocal tract shape*. The length and length change of a muscle are measured by the *muscle spindles* (and the tension by the *tendon organs*), which send this information back (through the afferent fibers marked *1A*) to the *spinal cord* or the brain stem. If the muscle is stretched by an external cause, a direct excitatory synapse of the afferent with the α motor neuron then causes the *stretch reflex*: a compensatory contraction of the muscle.

With the help of the γ efferent fibers, the muscle spindles can be actively stretched, so that the afferents fool the spinal cord into thinking that the muscle itself is stretched by an external cause. Consequently, the reflex mechanism described above will cause the muscle to contract. Thus, while direct α activity would cause an uncontrolled contraction, this γ-loop system, which does not go further up than the spinal cord, can be used to control *muscle length* (Hardcastle 1976; Gentil 1990). The learning of a fast, shape-oriented gesture probably involves the learning of an efficient mix of α and γ activity, innervating the muscle spindles simultaneously with the other fibres.

Conclusion: the speaker can set her muscles to a specified length.

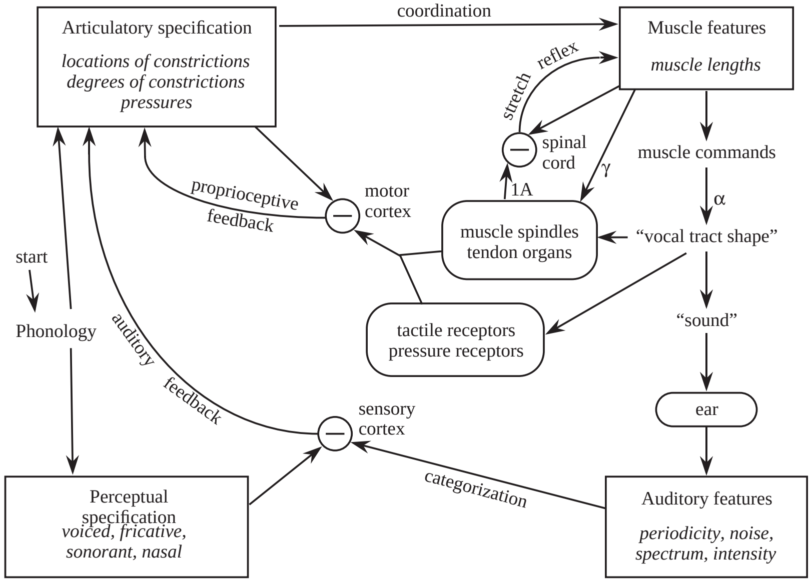


Fig. 2.1 Integration of phonology into speech production.
Rectangles = representations. Rounded rectangles = sensors.
Encircled minus signs = comparison centres. Arrows = causation.
α, γ, 1A = nerve fibers.

– Top left: control of position. For most gestures, the control of muscle length is not sufficient. Rather, the motor cortex specifies the actual position of the body structures. For the vocal tract, this means that the *locations* and *degrees of constrictions* are specified. That the muscle lengths are not the target positions specified in speech production, can be seen from bite-block experiments (Lindblom, Lubker & Gay 1979): speakers immediately compensate for the constraints on the jaw, even before phonating, in such a way that the tongue muscles bring about approximately the same area function in the vocal tract as in normally articulated vowels, while having very different shapes.

The proprioceptive sensory system, consisting of muscle spindles, tendon organs, *tactile receptors*, and *pressure receptors*, sends the information about the realized shapes back to the motor cortex, where it is compared to the intended shapes, i.e., the *articulatory specification*, and appropriate action is taken if there are any differences. This system is called *proprioceptive feedback*.

Conclusion: the speaker can directly control muscle tensions, muscle lengths, and the locations and degrees of the constrictions in the vocal tract. Hypothesis: the articulatory part of phonology specifies al these variables.

– Bottom right: auditory perception. The human *ear* will analyse any *sound*, perhaps one arising from a speech utterance, into *auditory features* like *periodicity* (pitch and noisiness), *spectrum* (timbre), and *intensity* (loudness), all of them functions of time. I will illustrate the perceptual part of speech production with the development of phonology in young children.

The infant is born with an innate control of some of the gestures that are also used in speech: breathing, vocal-fold adduction (crying), and repetitive jaw movements (drinking). Other gestures, like the movements of the limbs, are still largely uncoordinated. After a few months, the infant learns that she can control her environment (i.e., her perceptual impressions), by pulling some muscles. Like the use of one of her deltoid muscles gives her the visually pleasing result of a swinging object (her arm), a certain combination of expiration and vocal-fold adduction gives her the auditorily pleasing result of a periodic sound (voicing). A little later, when she has a command of some agonist/antagonist pairs, she will start exploring the benefits of repetitive movements; like hitting the mills and bells that are within her reach, she will superponate opening and closure gestures of the jaw on a background of phonation, thus getting nice alternations of silence and sound (babbling).

Conclusion: speakers learn the forward relationship between articulatory coordinations (top left) and perceptual results (bottom right).

– Bottom left: speech perception. At the time she starts to imitate the speech she hears, the little language learner will have to compare her own utterance with the model (*auditory feedback*). At first, the *perceptual specification* (initially, the adult utterance), is an unsegmented gestalt. The articulatory specifications, which she is now constructing for the sake of faithful imitation and the reproduction of her own speech, are not very sophisticated yet either, because the orosensory (proprioceptive) feedback mechanism is still under development.

But the child learns to group perceptual events into categories. For speech, this ultimately leads to a language-dependent *categorization* of perceptual features. The skilled speaker will also have highly organized articulatory specifications in terms of degrees of constrictions and air pressures, with a language-dependent degree of underspecification, determined by economical considerations, i.e., the balance between perceptual invariance and articulatory ease. She will use the auditory feedback only as a check and for maintenance.

Conclusion: the speaker can compare the realized perceptual categories with the perceptual specification of the utterance. Hypothesis: this is integrated into phonology.

2.2 The two targets of speech production: levels of phonological specification

For a skilled speaker, the perceptual specifications must be the *ultimate* (distal) targets of speech production. They cannot be the *immediate* (proximal) targets, because the auditory

feedback loop is much too slow for that. The immediate targets are the locations and degrees of constriction and the air pressures in the vocal tract. These proprioceptive targets can be monitored by the collective effort of tactile and pressure receptors, muscle spindles, tendon organs, and joint receptors.

The *task-dynamic* approach advocated by Kelso, Saltzman, & Tuller (1986) and Browman & Goldstein (1986, 1990), maintains that the input to an articulation model should consist of specifications of *tract variables*, such as locations and degrees of constrictions, as functions of time. This approach explicitly focuses on describing the coordination of the muscles of speech production: specification of these tract variables refers to *learned* motor behaviour. Kelso et al. notice, for example, that an experimentally induced perturbation of the movement of the jaw does not prevent the completion of the bilabial closure in [aba] or the achievement of an appropriate alveolar near-closure in [aza]. Thus, if the upper and lower teeth are externally constrained to be more than 1 cm apart, the required alveolar closure will still be attained. Crucially, however, the smallest bilabial closure will then be much larger than in the case of an unconstrained [aza]. Apparently (Kelso et al. argue), the immediate task for producing [b] is: “make a complete closure with the lips”, and for [z] it is: “make a near closure at the alveoli”. Crucially, the task for [z] does not specify bilabial closure at all; this is why there can be a large variation in the degree of bilabial closure during [z]. Therefore, there is some underspecification in the immediate targets of speech production.

However, as will be apparent from our separation of perceptual and articulatory specifications, a part of the ultimate *perceptual* specification of /z/ (in some languages) should be in these terms: “make a periodic sound that will produce strong high-frequency noise”. Speakers will learn that the only articulatory implementation (“task”) that achieves this, is: “make a near closure at the alveoli; meanwhile, the bilabial and dorsal constrictions should be wider than this alveolar constriction, the naso-pharyngeal port should be closed, the lungs should exert pressure, and the vocal cords should be in a position that enables voicing”. We see that the perceptual specification does require a constraint on bilabial closure after all (the lips must not be completely or nearly closed), and that the articulatory specification *follows* from the perceptual specification for /z/.

That the perceptual features, not the proprioceptive features, form the distal targets of speech production, can be seen in a simple experiment that embroiders on the bite-block experiments. If you ask someone to pronounce a central (e.g. Dutch) [a] with her teeth clenched, she will make compensating tongue and lip movements; however, because [a] is not specified for horizontal lip spreading, she will not draw the corners of her mouth apart, though this would yield a much more [a]-like sound; she will only learn this trick after some practice, using auditory feedback.

Conclusion: the articulatory specifications are the proximal targets of speech production, the perceptual specifications are the distal targets. Hypothesis: phonology controls both.

2.3 Perceptual specifications

The functional principle of maximization of perceptual contrast is evaluated in the perceptual space. Perceptual features include periodicity (voicing and tone), noise (frication, aspiration), silence, burst, continuancy, and frequency spectrum (place, nasality).

All these features are measured along continuous scales, but languages discretize these scales into a language-dependent number of *categories*. An example of the perceptual specification of labial sounds for a language that has two categories along the voicing, friction, sonorancy, and nasality scales, can be read from the following table, where '+' means 'present', '–' is 'absent' (suggesting a privative feature), and '|' is a perceptual contour, i.e., a temporal change in the value of a perceptual feature:

	p	f	v	b	m	w	p^h	ʋ	h^w	u	b̤	ũ	ṽ
voiced	–	–	+	+	+	+	–	+	–	+	+	+	+
noise	–	+	+	–	–	–	–|+	–	+	–	–|+	–	+
sonorant	–	–	–	–	+	+	–	+	–	+	–	–	+
nasal	–	–	–	–	+	–	–	–	–	–	–	+	+

(2.1)

– ***No universal feature values***. The language-dependency of perceptual feature values can be most clearly seen from the different divisions of the height continuum for languages with three and four vowel heights (§6): if the lowest vowel is [a] and the highest vowel is [i], a language with three vowel heights will have an "e" whose height is approximately midway between [a] and [i], and a language with four vowel heights will have two vowels close to canonical [ɛ] and [e]; this shows that the height continuum is divided on a basis of equal perceptual distance rather than on a basis of maximum use of universal binary features.

2.4 Articulatory specifications

The functional principle of minimization of articulatory effort is evaluated in the articulatory space, which consists of all the possible positions, shapes, movements, and tensions of the lips, cheeks, tongue tip, tongue body, velum, tongue root, pharynx walls, epiglottis, laryngeal structures, vocal folds, and lungs. The trajectory of the implementation of the utterance through this space is a voyage along many positions, each of which is characterized as a vector measured along scales of degree of closure or tension. Though these scales are continuous, languages discretize most of them. For instance, supralaryngeal degrees of closure can be: *complete* (usually brought about by a ballistic movement: plosives and nasals); *critical* (usually brought about by a controlled movement, which makes it precise enough to maintain friction noise or vibration: fricatives); *approximant* (strong secondary articulation, pharyngealization); *narrow* (0.3 - 1 cm^2; high vowels, glides, liquids, retracted tongue root); *open* (1 - 4 cm^2; neutral vocalic); or *wide* (4 - 15 cm^2; spread lips, advanced tongue root).

I classified these degrees of closure according to perceptual differences, i.e., every pair of successive labels is found somewhere in the world to contrast two phonemes on the same articulator. Still, there is nothing canonical, preferred, or universal about this subdivision. Besides the obvious articulatory implementation of the language-dependent subdivision of vowel height, here is an example with non-vocalic closures: Dutch contrasts a noisy voiced labiodental fricative ([viɫ] 'fell') and a noiseless approximant ([ʋiɫ] 'wheel'); in between those two, as far as noisiness and, therefore, degree of constriction are concerned, are the [v]-like sounds of German ([vaen] 'wine'), English ([vain] 'vine'), Afrikaans ([vət] 'white'), and French ([vil] 'city').

The labial, coronal and dorsal articulators can be used independently to a large extent in doubly articulated sounds (labial-velars, clicks) or even triply articulated sounds (Swedish [ɧ], Holland Dutch syllable-final <l> [$ɫ^w$]), but there are no sounds that use the same articulator twice (e.g. no clicks with dorso-palatal front closure). The articulatory space is organized in tiers, with one tier for every degree of opening and tension. The independence of these tiers represents the independence of the articulators, and reflects the independence of articulatory features in phonology.

An example of the articulatory specifications of some labial sounds in a language that would faithfully implement the perceptual features of (2.1), is given in (2.2) (0 = closed, 1 = critical, 2 = approximant, 3 = narrow, 4 = open, 5 = wide, | = time contour, 2-5 = from 2 to 5):

	p	f	v	b	m	w	p^h	ʋ	w̃	b̥	b̤	ɓ	h^w	u	ɔ
lip opening	0	1	1	0	0	3	0|2-5	2	3	0	0|2-5	0	3	3	4
tongue tip opening	2-5	2-5	2-5	2-5	2-5	3-5	2-5	2-5	3-5	2-5	2-5	2-5	3-5	4-5	5
tongue body opening	2-5	2-5	2-5	2-5	2-5	3	2-5	2-5	3	2-5	2-5	2-5	3	3	4
velum opening	0	0	0	0	4	0-1	0	0-1	4	0	0	0	0-1	0-1	0-2
pharynx opening	2-5	2-5	2-5	2-5	2-5	3-5	2-5	2-5	3-5	2-5	2-5	2-5	3-5	4-5	3
glottis opening	2-3	2-3	1	1	1	1	3	1	1	1	2	1	3	1	1
supralar. tension	+			–						–	–	–			

(2.2)

– ***Articulatory underspecification.*** There is a lot of underspecification in (2.2). For instance, if the lips are completely or almost closed, the coronal and dorsal constrictions have a lot of freedom: they can be anywhere between the approximant closure and a wide vocalic opening without affecting the perceptual features too much. As an example, consider the articulatory and perceptual features and specifications of [b] in the utterance

[aba]. During the pronunciation of [a], the tongue will be low in the throat, pulled down by the hyoglossus muscle. This state will last during the whole of the utterance [aba]. The jaw will travel a long distance in going from the [a] position to the [b] position and back again. The muscles of the lips will also make a closing-opening movement. If, however, the lips are less closed, as in [u], the coronal constriction should be quite wide so that it will not sound like a front vowel, and the pharyngeal constriction should also be quite wide so that the vowel does not sound more open or centralized. Thus, as already argued in §2.2, the articulatory specifications follow from the perceptual specifications.

Conclusion: articulatory underspecification is constrained by faithfulness to perceptual invariance.

2.5 Perceptual versus articulatory features

Though it is often the case that similar articulations produce similar perceptual results, as with most place features, there are several sources of asymmetry between perceptual and articulatory features. In the following, I will disentangle the hybrid features used in generative phonology.

– *Voicing*. If we define voicing as the vibration of the vocal cords, we are talking about the perceptual feature [voice], which refers to a high degree of periodicity in the sound. There is no single articulatory gesture that can be associated with voicing: for the vocal folds to vibrate, they must be close enough and air has to flow through the glottis with a sufficient velocity. The articulatory settings needed to implement the voicing feature, vary depending on the degree of constriction above the larynx. If the air is allowed to exit freely, as in sonorants, there is spontaneous voicing if the vocal folds have been adducted by the interarytenoid muscles; sufficient airflow is then guaranteed.

If the passage is obstructed, as in [b], active laryngeal or supralaryngeal gestures are often needed to maintain voicing, especially in initial position: the larynx may be lowered, the width of the glottis or the tension of the vocal folds may be adjusted, the walls of the pharynx, the cheeks, or the velum may be expanded passively or actively, or the stop may be pre-nasalized. The effects of all of these tricks have been confirmed in simulations with a simple model of the vocal tract (Westbury & Keating 1986) as well as with a more comprehensive model (Boersma 1993, 1995, in progress). Since it is not always easy to find out which trick (other than implosion or prenasalization) is used by a specific language, we can supply plain voiced obstruents with the implementationally formulated articulatory feature [obstruent voicing] (or Steriade’s (1995) suggestion [pharyngeally expanded], though the term “expanding” might be more correct).

Likewise, active gestures are sometimes needed for voiceless obstruents, especially in intervocalic position: widening or constriction of the glottis, raising of the larynx, stiffening of supralaryngeal walls, or active narrowing of the supralaryngeal tract. For this, we can similarly imagine a goal-oriented articulatory feature [obstruent devoicing].

Since assimilation processes are normally associated with changes of articulatory timing, we expect that obstruents can trigger voice assimilation, and that sonorants cannot. Acceptance of the distinction between articulatory and perceptual voicing features, will lead to a rejection of the main argument for underspecification in phonological processes (§13). Thus, an early decision to posit a single feature [voice] for underlying and surface representations resulted in the underspecification of sonorants for this feature: the fact that many languages do contrast voiced and voiceless obstruents but do not contrast voiced and voiceless sonorants, combined with the phonological inertness (with respect to spreading) of voicing in sonorants, was considered evidence for the analysis that sonorants were not voiced at all underlyingly; a late rule would insert the voicing feature for sonorants. A distinction between an articulatory voicing feature, which only applies to obstruents because sonorants are spontaneously voiced, and a perceptual voicing feature common to sonorants and voiced obstruents, would quite simply solve the mysteries associated with the voicing problem. However, this will not go without a struggle: the one phenomenon that seems immune to a simple functional approach, NC voicing (i.e., the phenomenon that plosives tend to be voiced after nasals), tempted Itô, Mester & Padgett (1995) into the following remarks:

> “the trouble lies not with [voice], (...) the challenge is to resolve the paradox without destroying the unity and integrity of the distinctive feature [voice].” (Itô, Mester & Padgett 1995, p. 581)

Their resolution of the paradox entails that nasals, *because* they are redundantly voiced, like to share a non-redundant voicing feature with their neighbours. No explanation is given for the absence of CN voicing. An articulatory explanation was advanced by Hayes (1995): the velum goes on raising even after the moment of closure, so that the enlarging pharyngeal cavity facilitates the maintenance of voicing; the exactly reverse situation from the CN case. The question how such details are phonologized, is answered in §11.

– *Noise*. In the phonological literature, fricatives are economically divided into *non-strident* (/ɸ/, /θ/, /x/) and *strident* (/f/, /s/, /ʃ/, /χ/). In contrast with what the label suggests, this division is based on distributional grounds: the strident fricatives are louder (make more noise) than their non-strident counterparts *on the same articulator* (Chomsky & Halle 1968, p. 327), and are, therefore, on the average more suitable for human communication in a world with distances and background noise; the non-strident fricatives, on the other hand, often alternate, or are historically related to, plosives at the same place of articulation; as so happens, plosives tend to occur at locations where perfect closures are easy to make (bilabial, corono-postdental, dorso-velar), and fricatives prefer locations with small holes (labio-dental, corono-interdental) or unstable structures (dorso-uvular). From the perceptual standpoint, however, we could divide the continuous noise scale into four levels of a combined loudness/roughness nature (which is rather arbitrary, especially for the non-peripherals):

- [aspirated]: as in [h], [p^h], and so-called “voiceless sonorants”.
- [mellow friction]: resulting from airflow through a smooth slit ([ɸ], [x]).
- [strident friction]: airflow along sharp edges ([f], [θ]) or loose structures ([χ]).
- [sibilant]: a jet of air generated in one place (alveolar) and colliding at a rough structure at another place (teeth): [s], [ʃ]; this causes a 15 dB intensity increase with respect to the normal strident [θ][2]. According to Ladefoged (1990a), the distance between the lower and upper teeth is critical[3], and sibilants are the only English sounds with a precise specification for jaw height (see the discussion below for vowel height).

The epenthesis of a vowel in English *fishes* versus *myths* is due to the equal specifications for [sibilant] in base and affix (§14.2, Boersma fc. b), not to a missing stridency contrast on the labial articulator as proposed by Yip (1988).

– Sonorant. Chomsky & Halle’s (1968) definition of sonorants is that they are “sounds produced with a vocal tract configuration in which spontaneous voicing is possible” (p. 302). This is neither an articulatory nor a perceptual definition, and, as such, not likely to play a role in phonology. Since, as Ladefoged (1971) states, “the rules of languages are often based on auditory properties of sounds”, I will simply take [sonorant] to refer to a high degree of loudness and periodicity that allows us to hear a formant structure[4]. Thus, [sonorant] implies [voice]. Its implementation is as follows. From the openings associated with each articulator, we can derive the following abstract openings:

- Oral opening. This equals the minimum of the labial, coronal, and dorsal openings.
- Suprapharyngeal opening. The maximum of the oral opening and the nasal opening.
- Supralaryngeal opening. Minimum of suprapharyngeal and pharyngeal openings.

These derivative features can help as intermediaries in formulating the mapping from articulatory to perceptual features. For instance, the supralaryngeal articulatory setting needed for spontaneous voicing is:

$$\textit{supralaryngeal opening} \geq \text{“approximant”} \qquad (2.3)$$

This condition is not sufficient, of course. Vocal-fold adduction and lung pressure have to be added.

[2] Which the reader may verify by saying [sθsθsθsθsθ].

[3] The reader may verify that she cannot produce a faithfully sibilant [s] with a finger between her teeth.

[4] This raises the question whether [sonorant] can be considered a primitive feature at all: it can be seen as a *value* of a loudness feature, or as a derived feature based on the presence of formant structure.

– Fricatives versus approximants. So-called voiceless sonorants are just very mellow fricatives (aspirates). The binarily categorizing language of table (2.1) shows a perceptual contrast between fricatives and approximants, but only if these are voiced ([v] and [ʋ]), not if they are voiceless ([f] and [h^w]). This is because a voiced approximant will not produce friction, but a voiceless (aspirated) articulation with the same degree of closure, will. So, voiced fricatives and approximants can easily occur together in such a language (e.g., Dutch [v] and [ʋ]), because voiced fricatives are noisy and voiced approximants are not; their voiceless counterparts cannot occur together in such a language, because voiceless fricatives and voiceless approximants only differ in their *degree* of noisiness, which would force the listener to distinguish between the categories [aspirated] and [fricative].

– Nasality. The perceptual feature [nasal] more or less coincides with the articulatory feature [lowered velum]. But not precisely. Table (2.2) shows a less restricted nasal specification for [ɔ] than for [u]. A slightly open nasopharyngeal port is allowed in lower vowels, because it can hardly be heard if the oral opening is large (Van Reenen 1981). Thus, the same small amount of velum lowering may give rise to a perception of nasality in high vowels, and of no nasality in low vowels.

– Continuant. This feature has been used to distinguish plosives from fricatives, and to be able to treat nasal and “oral” stops as a natural class. As a perceptual feature for audible oral airflow, I will replace it with [oral]; thus, [f], [h], and [a] are oral, and [p] and [m] are not, while [ã] is both oral and nasal. This move reflects the articulatory symmetry between the nasal and oral pathways. However, because most speech sounds are oral but not nasal, commonness considerations (§8.5) lead us to expect that the values [–oral] and [+nasal] play more visible roles in phonological processes than their counterparts [+oral] and [–nasal].

In another respect, oral stricture works just like velar stricture: the degree of perceived oral airflow does not necessarily reflect the degree of closure. A sound made with the articulatory setting for a labial fricative will normally lose its friction when the velum is lowered: the air will follow the path of lowest resistance[5]. This is why nasalized fricatives like [ṽ][6] in table (2.1) are so rare in the languages of the world; to make one, you’ll have to come up with a very precise setting of your tongue blade, with different muscle tensions and positions from normal fricatives. Again, the perceptual specification determines the articulatory gestures.

[5] You can check this by pinching your nose, making a “nasal” [z], and then suddenly releasing your nose.

[6] If we take a perceptual definition for [ṽ]. The IPA is a hybrid notation system, and often ambiguous: if [i] and [u] are vowels with minimal F_1, what does the IPA symbol [y] mean? Is it a front rounded vowel with minimal F_1, or a vowel with the tongue shape of [i] and the lip shape of [u]?

If two articulations produce the same sound, the easier one is more likely to be used. At most places of articulation, a complete closure is easier to make than a critical closure, because it involves a ballistic instead of a controlled movement (Hardcastle 1976). For labiodentals, even a ballistic movement often results in an incomplete closure; so, labiodental plosives are very rare, but labiodental nasals quite common. Every non-labiodental nasal forms a natural class with its corresponding plosive because both are implemented with the same ballistic articulatory gesture, e.g., [complete labial closure].

– ***Plosives***. The intervocalic plosive in [ata] is perceptually marked by a sequence of formant transition [[t˺]] + silence [[_]] + release burst [[t]] + formant transition. Their has been a gigantic literature about the importance of all these cues in the perception of speech. While the formant transitions are shared with most other consonants at the same place of articulation, the silence and the burst together signal the presence of a voiceless plosive. In [[thɛɛ̃n_ts]], both release bursts are heard, but silence associated with the first [t] merges with the ambient stillness, thus giving up its identity. A cluster of plosives, like /atpa/, is pronounced with overlapping gestures in most languages (with French as a notable exception), so that the result [[at˺_ːpa]] shows the demise of the main place cue for the recognition of [coronal]. In English, this may lead to place assimilation ([ap˺_ːpa]]), because the articulatory gain of not having to perform a blade gesture outweighs the perceptual loss of losing the remaining place cue. We will see (§11, Boersma fc. a) that this kind of phonetic detail can be expressed directly in the grammar of spreading phenomena.

– ***Duration***. Duration could be called a *derived* perceptual feature, because the perception of duration presupposes the recognition of another feature (the presence of sound, timbre) as being constant. In the above example of place assimilation, the duration of the silence was preserved, which is a sign of the independence of the silence cue for plosives.

– ***Vowel height.*** According to Kenstowicz (1994, p. 20), "we may interpret [+high] as the instruction the brain sends to the vocal apparatus to raise the tongue body above the neutral point". However, since different tongue muscles are involved in [i] and [u], such a standpoint testifies to a view that speech is organized very differently from other motor activities: no proprioceptors for non-low tongue height are known; the correlation of vowel height with jaw height is weak, regarding the highly varying strategies that speakers adopt to implement this feature (Ladefoged 1990). Therefore, with Ladefoged (1971, 1990a) and Lindau (1975), I will assume that vowel height inversely corresponds to the first formant (F_1), i.e., that the phonological effects of vowel height correspond to the perception of the first peak in the excitation pattern of the basilar membrane in the inner ear (the higher the vowel, the lower its F_1). Simplistically, the muscles used in implementing vowel height are roughly: genioglossus (higher front vowels), styloglossus (higher back vowels), and hyoglossus (low vowels).

Vowel height does define natural classes in inventories and rule targets (as a result of perceptual categorization, see §6), but vowel harmonies and assimilations are largely confined to the more articulatorily tractable features of rounding, backness, and *advanced tongue root*; the rule ɔ → o / _ i is relatively rare (as compared with ɔ → ø / _ i), and assimilation of vowel height is expected to occur only if all the vowels involved use the same articulator, as in ɛ → e / _ i. Apparent exceptions are treated in Boersma (fc. a).

– ***Tensions.*** A direct relation between articulation and perception is found in the tension of the vocal cords, which is the main determiner of the pitch of voiced sounds. The tension of the lung walls determines the subglottal pressure, which influences the loudness (spectral slope and intensity) and pitch of the perceived sound. A rather indirect relation between articulation and perception is found with the tension of the walls of the pharynx and the cheeks, which can play a role in the voicing of obstruents.

– ***Place***. The perceptual distinction between the various places of articulation is primarily made on the basis of the associated auditory spectra. For vowels, the first formant, which is in the lower part of the spectrum and represents the degree of closure, seems to be an independent perceptual feature; it disappears in the transitions to neighbouring obstruents. Thus, place information for vowels is restricted to the upper part of the spectrum, and we can imagine that it is a multi-valued perceptual feature, encompassing [front], [back], and [round]; all these colour features assume [sonorant]. In the auditory spectrum, the front-back distinction is represented by the *second formant* (F_2); I will take it to specify the strongest spectral peak above the first formant[7]. Specifying the value "max" for F_2 means that F_2 should be at a maximum given F_1; this is most faithfully rendered by producing a front vowel with lip spreading. The value "min" specifies a minimum value of F_2 given F_1; this is most faithfully implemented as a rounded back vowel. No "enhancement" of an allegedly distinctive feature [back] by an allegedly redundant feature [round], as proposed by Stevens, Keyser & Kawasaki (1986) for reasons of lexical minimality, is implied here: the two gestures just implement the same perceptual feature symmetrically.

For consonants, place cues can be found in the formant transitions from and to neighbouring sounds. Other cues must be found in noises (fricatives and release bursts). The perceptual place feature is a rather continuous path through a multidimensional space, ranging from [bilabial] to [glottal], and does not respect the discrete articulatory

[7] Known in the phonetic literature as F_2', the usual definition of F_2 being: the *second* spectral peak, measured from 0 Hz upwards. This peak is commonly determined by a computer program that is forced to find five peaks between 0 and 5000 Hz. For [i], this second peak (at 2500 Hz or so) usually incurs a much weaker impression on the inner ear than the third and fourth peaks, which tend to conspire to build a very strong perceptual peak near 4000 Hz.

distinctions between the articulators: labiodental and corono-dental fricatives sound quite similar, and so do corono-postalveolars and dorso-palatals; perceptually, [glottal] must be included in the set of values of the [place] feature (adjacent to [epiglottal]), though it shows no formant transitions to surrounding vowels because these have glottal constrictions, too. For nasals, the place information contained in the various oral side branches is very weak: an isolated nasal stop produced with simultaneous lip and blade closures will sound as [n] in the dark, and as [m] if the listener sees the speaker: the visual cue overrides the auditory cue. Release cues without noise occur for nasal stops and laterals[8].

Vocalic place cues can be used with stops and fricatives to a certain extent: in many languages, lip rounding contributes to the perceptual contrast between [s] and [ʃ]. By contrast, lip rounding does not influence at all the stationary part of the sound of [n][9].

2.6 The speech-neutral position and privative features

Some features must be considered *privative* (mono-valued, unary), because only a single value can be phonologically active (Anderson & Ewen 1987, Ewen & Van der Hulst 1987, Van der Hulst 1988, 1989, Avery & Rice 1989). For instance, only [+nasal] is thought to be able to spread.

Steriade (1995) provides an articulatory explanation for the existence of privative features. The presence of an articulatory gesture like [lowered velum], she argues, is qualitatively different from its absence, because it constitutes a deviation from the speech-neutral position (Chomsky & Halle 1968, p. 300).

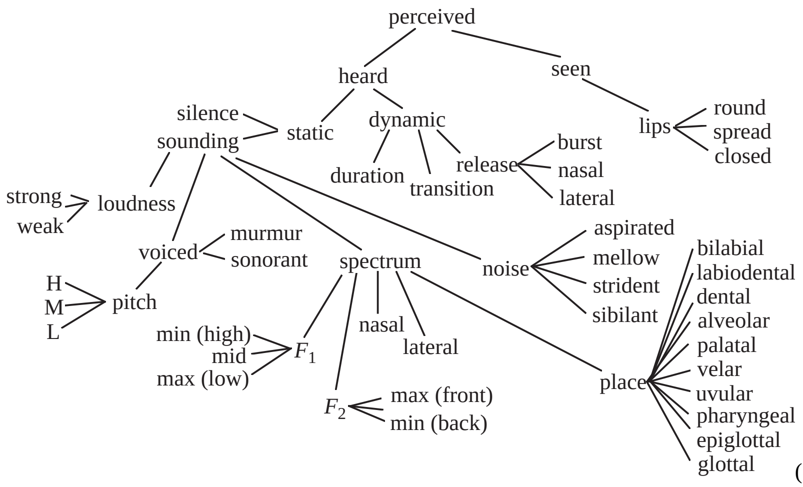
The only real neutral position is the one in which most muscles are relaxed, namely, the neutral position for breathing, which involves a wide glottis and a lowered velum. The alleged speech-neutral position would have glottal adduction and a raised velum, which involve active muscular effort (interarytenoid and levator palatini).

This speech-neutral position can only be explained with reference to requirements of perceptual contrast: we can produce better spectral contrasts for non-nasals than for nasals, and voicing allows us to produce tone contrasts, better formant structures, and louder sounds. Thus, nasal sounds will occur less often in an utterance than non-nasal sounds, and voiceless sounds will occur less often than voiced sounds. Instead of a *neutral* position, we now have the *most common* position.

So, instead of invoking a mysterious speech-neutral position, it seems more appropriate to explain privativity directly by arguments that start from the frequency of occurrence of the feature values in the average utterance: the presence of a perceptual feature like [nasal] is quantitatively different from its absence, because the latter would not signal any deviation from the more common non-nasality. In §8.5, I will show that differences in the phonological activities of various articulatory gestures can be related directly to the listener's adaptation of recognition strategies to frequency differences in the corresponding perceptual features. I will argue there and in §13 that the common values like [–nasal] are not absent, but only relatively invisible because of their weak specifications.

2.7 Feature geometries

The above story gives rise to the following partial geometry of implications for the presence of perceptual features (conjunctions are shown by vertical branches, disjunctions by horizontal branches):



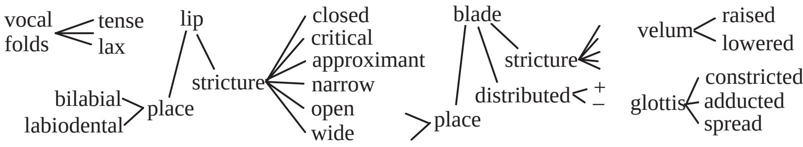
(2.4)

This figure only shows perceptual dependencies, so it does not show which features cannot co-occur because of articulatory constraints; for instance, an aspirated sonorant is easy ([ɦ]), but a sibilant sonorant would be much harder to produce. Some of the implications have to be taken with a grain of salt, as it is not unthinkable that pitch is perceived on voiceless syllables (as in Japanese), etc.

The implicational geometry for articulatory gestures is extremely flat, because of the near independence of the articulators:

[8] You can hear them if you record [ana] or [ala], create a backward copy of this sound, and compare the two CV transitions.

[9] Try saying [nːː] and superpose the lip movements of [wiwiwi]. The colour does not change. An analogous experiment with [ŋːː] and [wawawa] shows velar excitation of a closed front cavity.



(2.5)

The picture that arises from these geometries is rather different from the hybrid feature geometries that have been proposed by Clements (1985), Sagey (1986), McCarthy (1988), and Keyser & Stevens (1994). Those geometries will be seen to result from a confusion of the roles of articulatory and perceptual features (§14.3).

2.8 Conclusion

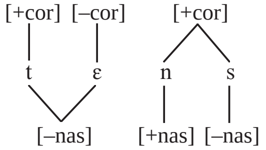
As the examples show, the relations of the traditional hybrid features with their supposed articulatory and acoustic correlates are rather vague. Every instance of asymmetry between articulatory and perceptual features causes problems to theories that do not distinguish them. Therefore, now that phonological theories have gotten rid of the early generative segmentality, binarity, representations, grammar organization, and rule ordering, the time has come to replace the content of the features with concepts rooted in general properties of human motor behaviour and perception.

3 Hybrid, articulatory, and perceptual representations

The purpose of linguistic proposals for phonological representations is the efficient description of phonological structures and processes. Derived from the evidence of language data, the usual phonological representation of an utterance is a hybrid of articulatory and perceptual specifications.

3.1 Hybrid representations

If we return to the English word *tense*, we see that linear phonology (Chomsky & Halle 1968) described it as a sequence of four bundles of binary features, called *segments*: /t+ɛ+n+s/. The *autosegmental* approach (Leben 1973, Goldsmith 1976) stressed the autonomy of the various features:

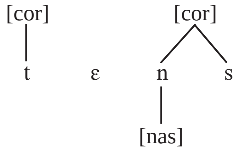


(3.1)

This would seem phonetically more satisfying, as it reflects the independence of the articulators and heeds two other principles that can be seen as consistent with articulatory phonetics: the *Obligatory Contour Principle* (OCP: “adjacent identical autosegments are forbidden”) ensures that the single coronal gesture of /ns/ is represented as a single feature value, and the *No-Crossing Constraint* (NCC: “association lines do not cross on the same plane”) ensures that the two successive coronal gestures of /t/ and /ns/ are represented as two separate feature values.

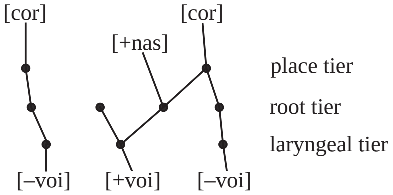
Important predictions of these representational constraints are that phonological processes cannot change two non-adjacent identical elements at a time, and that they cannot change only a single element out of a sequence of two adjacent identical elements. Thus, they allow only a limited range of primitive phonological processes, like *delinking* and *spreading*. From the functional point of view, these processes are advantageous if delinking is seen as the deletion of an articulatory gesture, and spreading as the change in the timing of an articulatory gesture, often in order to compensate for the loss of another gesture; for instance, in the common process of place-assimilation of nasals (/n+b/ → [mb]), the coronal gesture is deleted, and the labial gesture is extended in such a way that the nasal still has consonantal perceptual properties. However, this interplay between articulatory and perceptual needs could not be expressed in autosegmental phonology, because articulatory features like [closed tongue blade] could not by distinguished from perceptual features like [consonantal].

The advent of theories of privative features (§2.6), whose presence is qualitatively different from its absence, brought phonology again somewhat closer to function. In the interpretation of Archangeli & Pulleyblank (1994), the representation of /tɛns/ is[10]



(3.2)

Theories of Feature Geometry (Clements 1985, Sagey 1986, McCarthy 1988) subsumed the features [labial], [coronal], and [dorsal] under the [place] node, the features [voiced], [spread glottis], and [constricted glottis] under the [laryngeal] node, and all features together under the *root node*. For instance, a partial representation of /tɛns/ along the lines of Archangeli & Pulleyblank (1994) would be



(3.3)

Articulatory detail was put under the relevant articulator node: the [coronal] node dominates the feature [±anterior], and the [labial] node dominates [±labiodental]. The idea of this implicational interpretation of feature geometry is that if a node spreads, the dependent features also spread; for instance, place assimilation of /n+f/ can only give /ɱf/, never /mf/, because [labial] cannot spread without its dependent [labiodental].

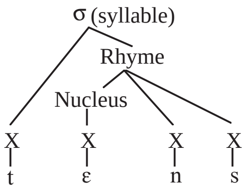
Directly under the root node are those features that we would associate with independent articulatory tiers, for instance, [nasal]. The features that do not spread, except if the whole segment spreads, can be seen as part of the root node. These *major class features*, it will come as no surprise, are exactly the perceptual features [sonorant] and [consonantal].

The remaining traditional feature [continuant] causes the greatest problems. If it is associated with the stop/fricative distinction, it should be dependent on each articulator tier, and, indeed, we see that clicks in Nama (Beach 1938) can have separate specifications for continuancy on their coronal and dorsal articulators. A reason *not* to put the feature [continuant] there is the fact that continuancy does not necessarily spread if the articulator spreads.

[10] The interpretation of the NCC and OCP implicit in (3.2) is the only alternative that stays compatible with the gestural analogy. It makes it hard to describe long-distance anti-repetition phenomena as OCP effects, but this is actually an advantage, as shown in §14.2 and Boersma (fc. b).

In §14.3, I will show that only implicational hierarchies as in (2.4) and (2.5) can be maintained, and that the place node and the problems with [continuant] are illusions caused by the interaction of more fundamental perceptual and articulatory phenomena.

Finally, theories of metrical phonology (Clements & Keyser 1983, Hyman 1985, McCarthy & Prince 1986, Hayes 1989) would propose hierarchical structures like (after Blevins 1995):

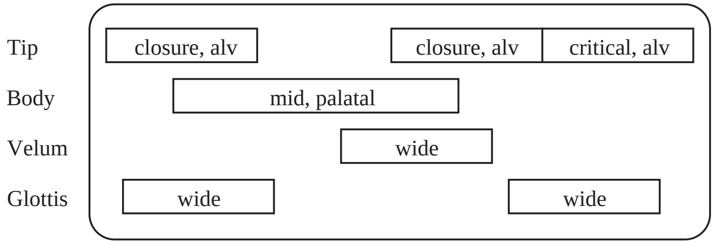


(3.4)

In this work on Functional Phonology, I will not touch metrical phenomena like accent, stress, and rhythm, because these have no obvious functional correlates in the speech-production and perception systems other than purely organizational principles: if we want to know what those principles are, we can only look at how languages handle them, and the current bottom-up approach, which starts from physiological principles, seems impossible.

3.2 Articulatory phonology

An interesting attempt to get at least one of the representations right, is Articulatory Phonology (Browman & Goldstein 1984, 1986, 1989, 1990a, 1990b, 1992, 1993): each articulator has its own tier, and the *gestural score* is a representation of the values on all relevant tiers. For instance, Bird & Klein (1990) give the following gestural score for the English word /tɛns/:



(3.5)

This representation shows the three overlaps between the four consecutive segments: the glottal widening, needed to make the stop voiceless, is continued after the release of the stop, giving the result of aspiration or a voiceless vowel; the lowering of the velum before the closing of the tongue tip causes nasalization of the preceding vowel; and the raising of the velum before the lowering of the tongue tip, which is needed to create the conditions for sibilant noise, causes an intrusive stop (silence + burst) to appear between /n/ and /s/ (Fourakis & Port 1986, Clements 1987).

In Articulatory Phonology, the values on the tiers represent immediate articulatory specifications only: these are the proximal targets of speech production and implement the forward path that we saw in the top left of figure 2.1, typical of skilled motor behaviour. But the auditory system will monitor the acoustic result, and the speaker/listener will assess the faithfulness of the perceptual result to the original perceptual specification: between the stretches of gestural specification in (3.5), for instance, the articulators return to their neutral positions, but the freedom of the articulators to go anywhere depends on the local perceptual specification of this utterance.

As a theory of phonology, therefore, Articulatory Phonology neglects the organizing power of perceptual invariance and segmental linearization. The solution to this problem involves a radical discrimination between the underlying perceptual specification, candidate articulatory implementations, and perceptual surface representations.

3.3 Functional phonology: the specification – articulation – perception triad

All the representations that we saw in §3.1 were proposed on the basis of studies of phonological structures and processes: the top-down approach. In this paper, I will use the bottom-up approach: to derive what languages could look like, starting from the capabilities of the human speech-production and perception system.

When turning a set of functional explanations into a theory of phonology, the first step is to posit the existence of *underlying forms*. In perceptuomotor terms: the intended effects of my movements on the environment. In speech terms: specifications of how my utterances should sound. We can see in figure 2.1 why phonology is different from other parts of the grammar: as a control mechanism for motoric events, it contains a feedback loop, which compares the perceptual result of the utterance with its specification. My hypothesis is that all strata of our phonological system mirror this loop, although it can only actually be proven to apply to phonetic implementation. This approach allows various degrees of abstractness in underlying specifications at each stratum, and the output of each stratum will generally be different from its input (about the number of strata that we need, see §14.6 and Boersma (fc. e)).

Thus, I propose the following three representations within each stratum:

1. ***Specification:***
   The underlying form (input), specified in perceptual features.

2. ***Articulation:***
   A candidate implementation, expressed on articulatory tiers.
3. ***Perception:***
   The surface form (output), expressed in perceptual features.

As an example, we show a fairly complete (“phonetic”) specification for /tɛns/ (the symbols /t/ etc. are nothing more than mnemonic symbols for bundles of feature specifications, reminding us of the predominant segmentality of English phonology):

Specify:	/t/	/ɛ/	/n/	/s/
timing	C or X	V, X, or µ	C, X, or µ	C, X, or µ
coronal	burst		+	
voice		sonorant	sonorant	
noise	aspirated			sibilant
F1		open mid		
F2		max		
round				
nasal			+	

(3.6)

This specification contains exclusively perceptual features, whose content was discussed in §2.5. The criterion for entering a specification in this table is the answer to the question whether the value of that feature matters for the recognition of the utterance as more or less representing the English word /tɛns/. The formalization of the verb *matter* and the adverbial phrase *more or less* will be presented in §8.

Besides the values of perceptual features, the table also specifies relations of simultaneity and precedence between the features. Thus: there is an “open mid” specification somewhere; the *first* segment is specified as voiceless (simultaneity relation between C and [voiceless]); there is a link between voicelessness and sibilancy; aspiration precedes voicing; a V precedes [nasal]. The specification also implicitly tells us what should *not* be there: no labial burst (because there is no labial specification), no voiced sibilancy (because these features are not simultaneous); no nasality during the vowel (because the privative feature [nasal] is not specified for the vowel).

The usual articulatory implementation of /tɛns/ in English and its perceptual result are as follows:

Articulate:

<table>
<tr><td>tip</td><td>closed</td><td colspan="3">open</td><td colspan="3">closed</td><td>critical</td></tr>
<tr><td>body</td><td colspan="8">open</td></tr>
<tr><td>velum</td><td colspan="3">closed</td><td colspan="2">open</td><td colspan="3">closed</td></tr>
<tr><td>glottis</td><td colspan="2">wide</td><td colspan="4">narrow</td><td colspan="2">wide</td></tr>
<tr><td>lips</td><td colspan="8">spread</td></tr>
</table>

Perceive:

<table>
<tr><td>silence</td><td>+</td><td colspan="4"></td><td colspan="2">+</td><td></td></tr>
<tr><td>coronal</td><td></td><td colspan="3">bu. tr.</td><td>side</td><td colspan="2"></td><td>bu. cont</td></tr>
<tr><td>voice</td><td colspan="2"></td><td colspan="3">sonorant</td><td colspan="3"></td></tr>
<tr><td>noise</td><td></td><td>asp</td><td colspan="4"></td><td>sibilant</td></tr>
<tr><td>F1</td><td></td><td colspan="3">open mid</td><td colspan="4"></td></tr>
<tr><td>F2</td><td></td><td colspan="3">max</td><td colspan="4"></td></tr>
<tr><td>rounding</td><td colspan="8"></td></tr>
<tr><td>nasal</td><td colspan="3"></td><td colspan="2">+</td><td colspan="3"></td></tr>
<tr><td></td><td>_</td><td>t h</td><td>ɛ</td><td>ɛ̃</td><td>n</td><td colspan="2">_</td><td>t s</td></tr>
</table>

(3.7)

– *Articulation.* In the articulatory representation, time runs from left to right on each tier, and the tiers are time-aligned; thus, there are no simultaneous articulatory contours in this example. The specification on each tier is complete, for consonants as well as for vowels.

From all possible articulations that implement /tɛns/, table (3.7) shows the one that involves the fewest contours. The openness of the tongue body and the spreading of the lips are only needed for giving the correct vowel height during /ɛ/. During the other parts of the utterance, these shapes may remain the same, since they would not interfere with the perceptual invariants of /t/, /n/, and /s/; here, a less spread lip shape would give almost the same perceived utterance, though a *complete* labial closure must be forbidden. In reality, lip spreading is achieved during the closure of /t/, and undone during /n/ or /s/; this is related to the fact that the active maintenance of lip spreading costs more energy than keeping the lips in a neutral position. Thus, there is a conflict between two aspects of laziness: minimization of number of contours and minimization of energy (for the fomalization of this conflict, see §5.2).

– *Perception.* In the representation of the uncategorized (“acoustic”) perceptual result, time runs from left to right on each tier, and the tiers are time-aligned with each other and with the articulatory tiers above. If a feature has no value, no value is shown (see the noise tier); for some binary features, only positive values are shown, suggesting privativity (§8.9). In the perceptual score, many features are specific to either the consonantal or the vocalic class of sounds, in line with the implications shown in (2.4).

A complete (i.e., intervocalic) plosive is represented as a sequence of (pre-consonantal) transition (tr), silence, and release burst (bu). On the coronal tier, [side] means the acoustical correlate of the oral side branch with a coronal closure (barely distinguishable from other oral closures), and [cont] means a continuant coronal sound.

– *Microscopic transcription*. Though the set of perceptual tiers is the ultimate surface representation of the utterance, a linear transcription would be more readable. Because all phonetic details will be involved in assessing the faithfulness relations between specification and output, such a transcription should be very narrow. Instead of a traditional narrow transcription like [t^{h}ɛ̃n^{t}s], we shall use a transcription that introduces a new symbol in the string every time that any perceptual feature changes its value. For instance, the coronal gesture in /ata/ will normally be heard as transition + silence + burst; this will give [[at˺_ta]] in a *microscopic* transcription:

- A transition is denoted in microscopic phonetic notation as an unreleased stop: [t˺].
- Silence is denoted by an underscore: [_].
- A release burst is denoted by the symbol for the stop itself: [t].

Thus, a readable shorthand for the perceptual result is [[thɛɛ̃n_ts]]. The [h] part could equally well be transcribed as a voiceless vowel [ɛ̥].

4 Formalization of functional principles

We see that the specification /tɛns/ (3.6) and the perceptual result [[thɛɛ̃n_ts]] (3.7) are different: there are several aspects of *unfaithfulness* of the perceptual result to the specification. These differences arise through properties of the speech-production system, and their interactions with properties of the speech-perception system. The properties and their interactions will be formalized in the following sections.

Functional principles can be expressed explicitly as output-oriented *constraints* on articulations and on specification-perception correspondences. In order to state these constraints in an unconditional way, without reference to exceptions, the constraints should be considered *violable*; within the theory of Functional Grammar (from which I devised the name of Functional Phonology) this relation between generality and violability was formulated by Dik (1989, p. 337) in a theory of constituent ordering.

For the resolution of the conflicts between violable constraints, I will use the strict-ranking strategy of Optimality Theory (Prince & Smolensky 1993). Though this theory originated in the generativist tradition (its original version explicitly denied any role for function in the grammar), it is a very promising framework for expressing the interactions of functional principles.

The principle of the minimization of articulatory effort thus translates into families of articulatorily motivated constraints, formulated within a space of articulatory gestures (§5), and the principle of the maximization of perceptual contrast translates into families of perceptually motivated faithfulness constraints, formulated within a space of perceptual features (§8)[11]; the faithfulness constraints of speech perception are also formulated within a space of perceptual features (§6).

In §5 to §8, we will formulate the functional constraints and their universal rankings. The remaining part of this paper will centre on their interactions.

[11] The idea of articulatorily versus perceptually motivated constraints was conceived independently by Jun (1995) and Hayes (1995, 1996a,b).

5 Articulatory effort

In his *Dictionary of Phonetics and Phonology*, Trask (1996) calls the principle of *maximum ease of articulation* “A somewhat ill-defined principle sometimes invoked to account for phonological change”. In this section, we will formalize effort in this section, and turn it into a well-defined principle that will be seen to work for phonetic implementation (§10), segment inventories (§14.4, §14.5, Boersma forthcoming c, d), and autosegmental processes (§14.1, §14.2, §14.3, Boersma forthcoming a, b).

As we will see below, previous attempts to formalize articulatory effort run short of several generalizations, because they try to express articulatory effort into one variable. The relevant constraint in such an approach would be (the asterisk can be read as “no”):

Def. *EFFORT (*effort*)

“We are too lazy to spend any positive amount of *effort*.” (5.1)

The constraint-ranking version of minimization of effort would then be stated as:

Minimization of effort:

“An articulation which requires more effort is disfavoured.” (5.2)

This would be formalized into a universally expected constraint ranking:

$$\text{*E}{\small\text{FFORT}}\ (x) >> \text{*E}{\small\text{FFORT}}\ (y) \Leftrightarrow x > y \qquad (5.3)$$

However, articulatory effort depends on at least six primitives: energy, the presence of articulatory gestures, synchronization of gestures, precision, systemic effort, and coordination, and languages seem to be able to rank these separate measures individually to a certain extent. All of these will prove to be crucial in phonology.

5.1 Energy

A formula for the physiological effort needed by a muscle is at least as involved as

$$\int \left(ma + F_{el}\right) v\, dt + \int F_{el} v_0\, dt \qquad (5.4)$$

where

t = time. Ceteris paribus, the longer the utterance, the more energy.
x = displacement of the muscle.
$v = dx/dt$ = the velocity of the moving muscle. For a constant force, the power spent is higher for higher velocity.
m = mass to move.

$a = d^2x / dt^2$ = the acceleration. The heavier the moving structures, the more energy is spent in accelerating them.

F_{el} = elastic forces and forces exerted by other muscles (gravitational forces can be included here). Stretching other muscles costs energy.

v_0 = some constant expressing the energy needed for an isometric contraction. Applying a force costs energy, even in the absence of motion.

Negative integrands should be ignored in (5.4), because no energy can be regained by the muscle.

The energy constraint against a positon change, i.e., a slow movement of an articulator from one position to the other, is associated with the *work* done by the muscle, i.e., the term $\int F_{el} v \, dt$ in (5.4). It can be expressed as:

Def. *DISTANCE (*articulator*: *a* || *b*)

"An *articulator* does not move from location *a* to *b*, away from the neutral position." (5.5)

The universal ranking of these constraints is given by the following principle:

Minimization of covered distance:

"An articulator moving away from the neutral position prefers to travel a short distance." (5.6)

This is expressed in a constraint-ranking formula as:

$$\text{*DISTANCE}\ (\textit{articulator}\text{: } x_1 \parallel x_2) \gg \text{*DISTANCE}\ (\textit{articulator}\text{: } y_1 \parallel y_2) \Leftrightarrow \left| x_1 - x_2 \right| > \left| y_1 - y_2 \right| \qquad (5.7)$$

This is expected to hold within each articulator in every language.

The energy constraint against maintaining a non-neutral position of an articulator is associated with the energy spent in holding an isometric contraction, i.e., the term $\int F_{el} v_0 \, dt$ in (5.4). It can be expressed as:

Def. *HOLD (*articulator*: *position*, *duration*)

"An *articulator* stays at its neutral position, i.e., it is not held in any non-neutral *position* for any positive *duration*." (5.8)

The universal ranking of these constraints are given by the following principles:

Minimization of extension:

"An articulator likes to stay near the neutral position." (5.9)

Minimization of duration:

"A non-neutral position should be maintained as short as possible." (5.10)

In formulas, where the position *x* is measured relative to the neutral position:

$$\text{*HOLD}\ (\textit{articulator}\text{: } x, \Delta t) \gg \text{*HOLD}\ (\textit{articulator}\text{: } y, \Delta t) \Leftrightarrow |x| > |y| \qquad (5.11)$$

$$\text{*HOLD}\ (\textit{articulator}\text{: } x, \Delta t) \gg \text{*HOLD}\ (\textit{articulator}\text{: } x, \Delta u) \Leftrightarrow \Delta t > \Delta u \qquad (5.12)$$

In a model for vowel inventories, Ten Bosch (1991) constrained the articulatory space with a boundary of equal effort, which he defined as the distance to the neutral (straight-tube, [ə]-like) position. In terms of the ranking (5.11), this would mean having all *HOLD constraints undominated above a certain displacement *x*, and all constraints maximally low for smaller displacements.

Finally, equation (5.4) contains the term $\int mav \, dt$, which expresses the fact that a displacement costs more energy if it has to be completed in a short time, at least if no energy is regained in the slowing down of the movement. The related constraint is:

Def. *FAST (*articulator*: *a* || *b*, *duration*)

"An *articulator* does not complete its displacement from *a* to *b* in any finite *duration*." (5.13)

The universal ranking within this family is given by:

Minimization of speed:

"Faster gestures are disfavoured." (5.14)

This can be formalized as

$$\text{*FAST}\ (\textit{articulator}\text{: } a \mid b, \Delta t) \gg \text{*FAST}\ (\textit{articulator}\text{: } a \mid b, \Delta u) \Leftrightarrow \Delta t < \Delta u \qquad (5.15)$$

The *DISTANCE, *HOLD, and *FAST constraint families associated with a certain articulator, can probably not be freely ranked with respect to one another, because there are no signs that the production system, let alone phonology, treats them individually. Rather, we could regard them as aspects of a general articulator-specific *ENERGY (*articulator*: *x*(*t*)) constraint, to whose ranking they contribute additively. This *ENERGY constraint is ranked by its energy value (5.4). The *ENERGY constraint clan is active in the case of phonetic implementation (§10), but will be seen to show surprisingly little *organizational* power, especially seen in the light of the extensive use it has been made of in the literature on the phonetic simulation of sound inventories (for a discussion on this subject, see §14.4, Boersma fc. c).

5.2 Number of gestures

The *number of articulatory contours* on the gestural tiers is a first rough measure of the organizational effort of an utterance. The constraints that favour a reduction of the number of articulatory contours, express the qualitative difference between making and not making a gesture: the loss of a gesture implies a discrete organizational articulatory gain.

In this coarse measure, therefore, the *amount* of movement does not matter (by definition). Compare the simplest implementations of /apa/ and /awa/:

	a	p	a
lips	wide	closed	wide
pharynx	narrow		

	a	w	a
lips	wide	narrow	wide
pharynx	narrow		

(5.16)

Both contain two contours, so they are equally difficult in that respect.

The *number* of movements does matter. Compare /tɛnt/ with /tɛns/:

	ɛ	n	t
velum	closed	open	closed
blade	wide	closed	

	ɛ	n	s
velum	closed	open	closed
blade	wide	closed	crit

(5.17)

The utterance /tɛns/ ends with two contours, and is therefore more difficult organizationally than /tɛnt/.

The constraint family associated with the minimization of the number of contours can be called *GESTURE:

Def. *GESTURE (*gesture*)
“A *gesture* is not made.” (5.18)

For instance, the constraint *GESTURE (blade: closure) can be held responsible for the deletion of the coronal gesture in Dutch /n+p/ sequences. Since *GESTURE has no continuous parameters, there is no universal ranking within this family. A universal *tendency* within the *GESTURE family, however, is expected to be

*GESTURE (*gesture*$_1$) >> *GESTURE (*gesture*$_2$) ⇔
⇔ effort (*gesture*$_1$) > effort (*gesture*$_2$) (5.19)

Such a ranking expresses an articulatory markedness relation across articulators. As with implicational markedness statements, these rankings can probably only be determined or predicted for “neigbouring” gestures. For instance, the larger rate of occurrence of coronal plosives with respect to labial plosives in most languages, may be attributed to the universal ranking *GESTURE (lips) >> *GESTURE (blade). However, the ranking of these constraints with respect to, say, *GESTURE (lowered velum) is not only difficult to determine; it is plausible that languages have a free choice in this ranking. For instance, there are a few languages without labial plosives, and a few other languages without nasal stops; this can be interpreted as the typology expected from a free ranking of *GESTURE (lips) with respect to *GESTURE (lowered velum).

Although (5.19) may express cross-linguistic and intralinguistic markedness relations, it is not valid in the realm of articulatory detail within a language. Rather, the finiteness of available articulatory tricks in every language forces us to admit that

*GESTURE (*gesture*) is undominated with probability 1 (5.20)

where *gesture* spans the infinite number of thinkable articulations in the human speech apparatus. This effect is due to *motor learning*: only those few gestures that the child has managed to master during the acquisition of her speech, are associated with a violable *GESTURE constraint. For instance, speakers of English apparently have a low *GESTURE (corono-alveolar closure) constraint, because they obviously know how to make alveolar plosives; the *GESTURE (corono-dental closure) constraint, on the other hand, is ranked high. Speakers of French have the reverse ranking. Considerations of minimization of energy, therefore, seem not to be involved.

The emergence of motor skills is reflected in the reranking that takes place during speech development. Children start out with very few usably low-ranked *GESTURE constraints. While learning, the acquisition of coordinative skills causes the emergence of more low *GESTURE constraints, giving the *ENERGY constraints a chance to play a role.

Now that we have two constraint families, we can study an interaction. Below (3.7), I discussed the conflict between an active maintenance of lip spreading and the organizational problem of issuing a command to move the lips back to their rest position. In terms of tension control, the conflict is between *HOLD (risorius: 20% active, 100 ms) and *GESTURE (risorius: relax from 20% active); in terms of length control, the conflict is between *HOLD (risorius: 40% spread, 100 ms) and *GESTURE (risorius: from 40% spread to neutral); and in terms of the control of articulator position, the conflict is between *HOLD (lips: 40% spread, 100ms) and *GESTURE (lips: from 40% spread to neutral). The un-English implementation (3.7) would be the result of the ranking *GESTURE (relax lips) >> *HOLD (lips: spread, 100ms):

/tɛns/	*GESTURE (relax lips)	*HOLD (lips: spread)
☞ thɛɛ̃n̜t̜s̜		*
thɛɛ̃nts	*!	

(5.21)

It should be noted that a candidate without any lip spreading (i.e., satisfying *GESTURE (lips: spread)) is ruled out by the specification of maximum F_2 (§8).

Now that we have constraint interaction, we can predict a typology. Languages that have the ranking *GESTURE (relax lips) >> *HOLD (lips) are expected to maintain any non-neutral lip shape as long as possible, because that would minimize the number of articulatory contours, since there is always a chance that a following strong perceptual rounding specification requires the same lip shape. A typical phonologization of this effect would be the restriction of its domain to the morphological word: this would give a rightward rounding harmony, spreading from the strongly specified root onto weakly specified suffixes, like in many Turkic languages. Languages that have the ranking *HOLD (lips) >> *GESTURE (relax lips) will return to a neutral lip shape as soon as possible; their weakly specified suffixes typically contain central vowels, as in many Germanic languages.

5.3 Synchronization

It is difficult to synchronize two articulatory contours exactly. If /tɛns/ is produced maximally faithfully as [[t^{h}ɛns]][12], we have a perfect synchronization of the nasal opening gesture with the dorsal closing gesture, and a synchronization of the nasal closing gesture with the dorsal opening gesture. This is depicted in the gestural score as the synchronization of the relevant contours:

Articulate:

velum	clos	open	clos
blade	wide	clos	crit

(5.22)

The resulting perceptual features and microscopic transcription are:

Perceive:

nasal		+	
coronal		+	
voiced	sonorant		
friction			sib
	ɛ	n	s

(5.23)

This output [[ɛns]] is perfectly faithful to the input. However, the required articulatory implementation apparently involves the violation of two contour-synchronization constraints (the “|” stands for an articulatory contour, i.e., a change in position or tension of the articulator):

[12] The aspiration is considered part of the specification.

Def. *SYNC (*articulator*$_1$: *from*$_1$ | *to*$_1$; *articulator*$_2$: *from*$_2$ | *to*$_2$[; Δ*t*])

> “The movement of *articulator*$_1$ from *from*$_1$ to *to*$_1$ is not synchronous with the movement of *articulator*$_2$ from *from*$_2$ to *to*$_2$ [within any finite time span Δ*t*].” (5.24)

For a discrete version of *SYNC, the temporal distance parameter Δ*t* can be left out; it is then assumed to be “zero” for practical (perhaps perceptual) purposes. The universal ranking within the *SYNC family is given by:

Minimization of synchronization:

> “Two articulatory contours on different gestural tiers like to be far apart.”(5.25)

This can be formalized as

$$\begin{gathered} {}^*\text{SYNC}\ (articulator_1\text{: } from_1 \mid to_1;\ articulator_2\text{: } from_2 \mid to_2;\ \Delta t) >> \\ >> {}^*\text{SYNC}\ (articulator_1\text{: } from_1 \mid to_1;\ articulator_2\text{: } from_2 \mid to_2;\ \Delta u) \Leftrightarrow \\ \Leftrightarrow\ |\Delta t| < |\Delta u| \end{gathered} \tag{5.26}$$

The two *SYNC constraints violated in [[ɛns]] would be:

*SYNC (velum: closed | open; apex: open | closed)
*SYNC (velum: open | closed; apex: closed | critical)

Both of these constraints can be satisfied by a different timing:

Articulate:

velum	closed	open		clos ed	
blade	wide		closed		crit

Perceive:

nasal		+			
coronal			side		cont
voiced	son				
noise					sib
	ɛ	ɛ̃	n	_ t	s

(5.27)

The resulting sound in that case is [[ɛɛ̃n_ts]]. Of course, this is different from the input /ɛns/ (it violates some FILL constraints, §8.9), but this is no reason to feel uncomfortable, because we have Optimality Theory to handle constraint interactions.

5.4 Precision

In his “quantal theory of speech production”, Stevens (1989) states that languages prefer those articulations whose acoustic result is not very sensitive to the accuracy of the articulation. For instance, an [i] is characterized by the proximity of its third and fourth formants; this closeness is preserved for a large range of tongue positions around the optimal palatal position. Thus, Stevens' account can be translated into the principle of the minimization of the articulatory precision needed to reach a reproducible percept, as he stated in a comment on Keating's (1990) window model of coarticulation (Stevens 1990).

Another working of precision is the cross-linguistic predominance of plosives over fricatives. After all, it is easier to run into a wall than to stop one inch in front of it. Thus, *controlled* movements, as found in fricatives and trills, involve more precision than *ballistic* movements, as found in stops (Hardcastle 1976).

The relevant constraint family can be written as

Def. *PRECISION (*articulator*: *position* / *environment*)
"In a certain *environment*, a certain *articulator* does not work up the precision to put itself in a certain *position*." (5.28)

The environment will often be something like *left* _ *right*, which stands for "between *left* and *right*", where *left* and *right* are preceding and following articulatory specifications, often on the same tier. For instance, the constraint acting against the precision (constant equilibrium position of the lungs) needed to hold your breath between the inhalatory and exhalatory phase is expressed as (when your upper respiratory pathways are open):

*PRECISION (lungs: hold / in _ out)

Quite probably, it is much more difficult to temporarily hold your breath during the course of an exhalation. This means that the constraint just mentioned is universally ranked below

*PRECISION (lungs: hold / out _ out)

5.5 Coordination

There is no principled difference between assuming that the number of vowels in a language is finite, and assuming that vowel systems are structured within themselves, i.e., that they can be expressed in smaller units. Having a finite number of vowels means having a finite number of tricks, and there is no principled reason why these tricks could not be perceptual features and articulatory gestures, instead of segments as a whole. So: [e] and [o] form a natural class because of equal F_1 (perceptual feature), while [e] and [i] may form a natural class because of an equal place of articulation (articulatory gesture).

A first rough measure of the systemic effort of a language would be the number of articulatory and perceptual tricks needed to speak and understand that language, plus the number of combinations of these tricks that the language uses. For instance, if we find the sound change /k/ > /k^h/ in a language, there is a good chance that *all* voiceless plosives get aspirated *at the same time*, as that would keep the number of trick combinations at a manageable level: the trick combination "plosive + voiceless" is replaced by the trick combination "plosive + aspiration", whereas if the other voiceless plosives would not become aspirated, the language ends up with having the two trick combinations "plosive + voiceless" and "plosive + aspiration" at the same time. Alternatively, if the sound change /k/ > /k^h/ renders the sound system asymmetric, this principle may work later on in simplifying the now unbalanced system by causing the aspiration of /p/ and /t/, too.

The principle examined here is very important in building sound systems, and is usually called *maximum use of available features*, though, as we saw in our example, this term should be extended with: *and their combinations*.

Because every combination of articulatory tricks has to be learned, we have the following constraints:

Def. *COORD (*gesture*$_1$, *gesture*$_2$)
"The two gestures *gesture*$_1$ and *gesture*$_2$ are not coordinated." (5.29)

As with *GESTURE, most of these constraints are undominated.

These negativces relations between gestures are the common situation in speech development. Skilled speakers, on the other hand, have many *positive* relations between gestures, resulting from the acquired coordinations that implement the perceptual specifications of the utterances of the language.

For instance, Dutch has two perceptually contrasting degrees of voicing for plosives: fully voiced and fully voiceless. Both require an active laryngeal adjustment in their articulatory implementations. Now, a lax voiceless stop, as the English or South-German word-initial "b", which requires no actions of the laryngeal muscles, can hardly be pronounced consciously by native speakers of Dutch; instead, it must be elicited by an extralinguistic experiment, for instance, the simulation of a repetitive mandibular gesture like the one found with babbling infants.

Another example is the extreme difficulty displayed by Dutch students when learning to produce unrounded back vowels: it seems to have to be either an unrounded front vowel modified with a backing gesture of the tongue body, or a rounded back vowel modified with a spreading gesture of the lips. No-one, on the other hand, has any trouble in producing the extralinguistic sound expressing disgust, which combines voicing, lip spreading, and dorsal approximation. That sound, again, can hardly be produced without pulling the facial muscles that are associated with disgust but are superfluous for producing unrounded back vowels.

Thus, while plosives and rounded back vowels require complex coordinations not mastered by beginners, adults have several constraints that are the results of the plasticity of the human motor system:

Def. IMPLY (*gesture*$_1$, *gesture*$_2$) $\equiv \exists$ *gesture*$_1$ $\Rightarrow \exists$ *gesture*$_2$
"The presence of *gesture*$_1$ implies the presence of *gesture*$_2$." (5.30)

This is an example of language-specific effort. Several muscles can only be pulled as a group (at least when speaking). These coordinations are language-specific and reflect the organizational shortcuts that are typical of experienced speakers. The cross-linguistic pervasiveness of some of them have led some phonologists to ascribe to them the status of universal principles. For instance, numerous underspecificationists want us to believe that the implication [+back] → [+round] is a universal (innate) default rule, whereas, of course, the tendency that back vowels are round is related to their maximal perceptual contrast with front vowels. If we stay by the functions of language, we can unequivocally assign the roles of cause and consequence.

Still, we have to ask in how far (5.30) plays a role in the phonology of the language. It is quite probable that we have to invoke it for explaining the phenomena found in second-language acquisition: the trouble for speakers of English in producing unaspirated French plosives is not due to a perceptual failure or low PARSE constraint, but must be attributed directly to the need to bypass a soft-wired (i.e., built-in but not innate) coordinative structure. Thus, the language-specific constraint (5.30) must play a role in articulatory implementation, i.e., the speaker uses it to her advantage in minimizing the number of higher neural commands, delegating some of the more automatic work to the more peripheral levels; in this way, [+back], with its automatic implication of [+round], is a simpler command than [+back; –round]. On the other hand, in explaining sound inventories, the combination [+back; +round] must be considered more complex than [+back; –round], because it involves one more active gesture; the requirements of perceptual contrast then force the implementation of the more complex combination. From the functional standpoint, we would like to postpone the assumption of innate implementation rules to the arrival of positive evidence.

5.6 Global or local rankings of effort?

It is probable that the first steps of learning to move or speak are chiefly controlled by the principle of the minimzation of the number of gestures, and that later on, the development of coordination makes the minimization of energy a more important criterion. In general, however, it is hard to determine how to rank the various effort principles with respect to one another; not only for the linguist, but also, I would like to propose, for the speaker.

In discussing the relation between motor activity and effort in sports, it is impossible, for instance, to give a universal answer to the question whether skating or skiing is the more difficult of the two: it depends on the learning history of the person who performs these activities; but it is a universal fact that skiing becomes more difficult for very steep slopes, and that skating requires more effort on poor ice or if the rider is making a contest out of it.

Likewise, a speaker cannot assign numerical values to the various principles of effort, but she can locally rank different kinds of efforts within the separate families, along the lines of (5.7, 5.11, 5.12, 5.15, 5.26). The rankings across the families are determined by the learning history, i.e., by the language environment in which the speaker has grown up.

If languages differ as to what kinds of effort they consider important, a global measure of effort is not feasible. So I hypothesize that the holistic ranking (5.3) is not valid, and that only the rankings within the separate families are universal:

Local-ranking hypothesis for articulatory constraints:

> "A constraint cannot be ranked universally with respect to a constraint in a different family; and constraints within a family can only be ranked universally if only a single parameter is varied." (5.31)

Apart from being a negative condition on possible rankings, this is also a positive condition on the freedom assigned to every language: all ranking of constraints across families or of constraints with two different parameters, is free. An example of the single-parameter condition in (5.31) is: a language can freely rank its *HOLD constraints as long as the rankings (5.11) and (5.12) are honoured.

If this hypothesis is true, speech researchers will not have to try to assign numerical values to articulatory effort: we can get along with simple local rankings, and these can be predicted from known relations of monotonicity between effort on one side, and extension, duration, speed, number of contours, synchronization, precision, and coordination on the other.

5.7 Ranking by specificity

Another intrinsic ranking applies to the articulatory constraints. The gesture [bilabial closure] is, on the average, more difficult to make than the gesture [labial closure], because the underspecification of the latter would allow a labiodental implementation if the phonotactics of the situation favoured that:

Minimization of specificity of articulatory constraints:

> "For articulatory constraints, more specific constraints are ranked above less specific constraints." (5.32)

It can be formalized as

$$(A \Rightarrow B) \Rightarrow \text{*G{\small ESTURE}}\ (A) >> \text{*G{\small ESTURE}}\ (B) \qquad (5.33)$$

Ranking (5.33) can be used as a universal ranking condition for *PRECISION constraints: the larger the window, the lower its *PRECISION constraint.

Ranking (5.33) is the reverse of an analogous ranking for perceptual constraints (see §8.10).

5.8 A restriction on functional rankings of articulatory constraints

Articulatory constraints cannot be ranked by considerations of perceptual importance. For instance, an alleged ranking *GESTURE (labial / stem) >> *GESTURE (labial / affix) or *GESTURE (labial / –stress) >> *GESTURE (labial / +stress), where the "/" means "in the domain of", would confuse articulatory constraints with faithfulness constraints: the ranking of *GESTURE (labial) can only depend on its articulatory environment. In §10 and §11 I will show that asymmetries between the surfacing of gestures in environments of varying degrees of perceptual importance, arise from dependencies in the rankings of faithfulness constraints.

5.9 Conclusion

Gestural constraints like *GESTURE and *COORD and phonotactic constraints like *SYNC can be thought of as motivated by the principle of minimization of articulatory effort. These constraints are violable and can therefore be stated in general terms, so that they can be thought to be language-independent and phonetically motivated. Their rankings with respect to heterogenous constraints must be language-specific.

6 The emergence of finiteness

The most salient aspect of sound inventories is their finite size: each language uses a finite number of underlying lexical phonological segments or feature values. The functional explanation for this fact contains two sides: the finiteness of the number of articulatory features, and the finiteness of the number of perceptual features.

Prince & Smolensky (1993) maintain that any theory of phonology can only be called 'serious' if it is "committed to Universal Grammar" (p. 1). The learning algorithm of Tesar & Smolensky (1995) explicitly assumes "innate knowledge of the universal constraints" (p. 1). They also have to assume that there are a finite number of constraints. However, we have seen for articulatory constraints (§5), as we will see for perceptually motivated constraints (§8), that there are an infinite number of them. In this section, I will show that, though the constraints themselves are universal, separate languages warp the continuous articulatory and perceptual spaces in such a way that each language ends up with a unique set of allowed gestures and specificational elements (features): the articulatory space is warped by motor learning, which lowers a few articulatory constraints, and the perceptual space is warped by categorization, which lowers some constraints of speech perception.

6.1 Feature values are not innate

If we talk about certain linguistic phenomena as being 'universal', we can mean either of two things: first, in the sense of Universal Grammar, that these phenomena exemplify *innate* properties of the human language faculty; secondly, that languages tend to have these phenomena because the functions of communication are similar in most languages, and because our speech-production organs and our ears are built in similar ways. Though these two views need not be conflicting as they stand, I will take the stronger functional position: that humans are capable of learning to speak without the necessity of innate phonological feature values, i.e., that languages can make their own choices from the perceptual and articulatory possibilities identified in §2.

As we see from the success of sign languages for the deaf (Brentari 1995), a phonology can be based on the capabilities of any motor system (talking, signing) and any sensory system (audition, vision) considered suitable for expressing intentions, wishes, and thoughts. We must conclude that nature did not force any specific motor system upon us for communication. This supports the view that we are not confined to using a universally fixed set of features if we choose to use the speech apparatus for our communication.

As an example, consider the division of the vowel height continuum. All too often, vowels are put into categories on the basis of a dogmatic "principle" that states that all languages use the same feature set (Kenstowicz 1994, Clements & Hume 1995). The International Phonetic Alphabet, for instance, seems to have been developed for languages with four vowel heights, having [ɛ] and [e] to represent front unrounded mid vowels. However, in most languages with three vowel heights (e.g., Spanish, Russian, Japanese), the height of this vowel is in between [ɛ] and [e]. This means that vowels are distributed along the height dimension in a way that enhances the perceptual contrast between them, and not according to a universal set of binary features, not even, I would like to conjecture, “underlyingly”.

The illusion of a universal set of features probably originated in the fact that the speech systems of most humans are very much alike, so that many languages do use the same features. Generalizing this to assuming a universal innate set of features is unwarranted.

Though there is no such thing as cross-linguistic sameness, much work in contemporary phonology is done to find the allegedly universal features, and put them into larger classes and hierarchies (manner versus place features, or major class features versus the rest). For instance (emphasis added):

> “*since* features are universal, feature theory *explains* the fact that all languages draw on a similar, small set of speech properties in constructing their phonological systems. *Since* features are typicaly binary or one-valued, it also *explains* the fact that speech sounds are perceived and stored in memory in a predominantly categorial fashion.” (Clements & Hume 1995, p. 245)

My position on this subject is that the causal relationships in these assertions should be reversed: because of the content of the constraints on human speech production and perception, different languages may sometimes show up with similar feature sets, and the functional interpretation of categorization predicts into how many values a perceptual feature space can be divided. An analysis of the emergence of language-specific features from an infinite universal pool of possible articulations and perceptual categories, is advanced in the remaining part of this section.

6.2 Constraints in speech production

Most articulatory gestures have to be learned. Before this is accomplished, all *GESTURE constraints are ranked quite high, but once a gesture has been learned because it occurs in a mastered word, the relevant *GESTURE constraint must have descended below the relevant faithfulness constraint. But this will facilitate the surfacing of the gesture in other words, too. For instance, a language with a click consonant will probably have more than one click consonant, because some of the coordinations required for those other clicks have been mastered already for the first consonant. Likewise, speakers of a language with corono-dentals stops will have trouble with the corono-alveolar stops of other languages, and vice versa; there is no universal preference for either of these implementations of coronal stops.

Thus, in the end, though most *GESTURE constraints are still undominated (see (5.20), some of them are so low as to allow the gestures to be made. This means that gestures and coordinations are the articulatory building blocks of sound inventories:

Articulatory inventory constraints:

“Low-ranked *GESTURE and *COORD constraints determine the finite set of allowed articulatory features and feature combinations.” (6.1)

This explains not only the finiteness of the segment inventory, but also (partly) the symmetries that we find inside inventories.

6.3 Functional constraints in speech perception: categorization

Because of the overwhelming variation in the world they live in, human beings organize their view of the world with the help of *categories*. Besides reducing cognitive load, categorization leads to fewer mistakes in identifying groups of things that we had better treat in the same way.

Like the production, the perception of speech has to be learned, too. The process of speech recognition entails that an acoustic representation is ultimately mapped to an underlying lexical form. A part of this process is the categorization of the acoustic input (figure 2.1). This section will describe the relation between the acoustic input and the perceptual result in terms of the *faithfulness and categorizarion constraints of speech perception*.

First, it is desirable that an acoustic feature is recognized at all by the listener. The following constraint requires a corresponding perceived feature value for every acoustic feature value (the subscript *i* denotes correspondence):

Def. PERCEIVE (*f*) $\equiv \exists x_i \in f_{ac} \Rightarrow \exists y_i \in f_{perc}$

“A value *x* on a tier *f* in the acoustic input is recognized as any corresponding value *y* on the same tier.” (6.2)

As always in Optimality Theory, the constraint has to be interpreted as gradiently violable: each unrecognized feature incurs one violation mark; this differs from the purely logical interpretation of “$\exists x_i \Rightarrow \exists y_i$” or its alternative “$\forall x_i \exists y_i$”, which means the same.

An analogous constraint DONTPERCEIVE requires that a recognized feature should have a correspondent in the acoustic input.

Secondly, it is undesirable that an acoustic feature value is recognized as something which is normally associated with a very *different* acoustic feature value. For instance, a vowel with a F_1 of 600 Hz is most properly perceived as a lower mid vowel, and a recognition as a high vowel is disfavoured. The following faithfulness constraint militates against distortions in perception (the asterisk can be read as “don’t”):

Def. *WARP (f: d) $\equiv \exists x_i \in f_{ac} \wedge \exists y_i \in f_{perc} \Rightarrow |x_i - y_i| < d$

> "The perceived value y of a feature f is not different from the acoustic value x of that feature by any positive amount of distortion d." (6.3)

Note that if a feature is not perceived, *WARP is not violated because the acoustic input feature has no correspondent: it is then *vacuously satisfied*. In other words, this constraint can be subject to *satisfaction by deletion*, the suggestion of which is enahanced by its negative formulation.

Because it is worse to perceive [ɛ] as /i/ than it is to perceive [ɛ] as /e/ (as will be proved in §8.2), *WARP has the following universal internal ranking:

Minimization of distortion:

> "A less distorted recognition is preferred over a more distorted recognition." (6.4)

This can be formalized as

$$*\text{WARP}\ (\textit{feature}: d_1) >> *\text{WARP}\ (\textit{feature}: d_2) \Leftrightarrow d_1 > d_2 \qquad (6.5)$$

Together, (6.3) and (6.5) assert that if a higher *WARP constraint is violated, all lower *WARP constraints are also violated.

Besides the above faithfulness constraints, and analogously to the *GESTURE family (5.18), which is an inviolable constraint for most of the universally possible gestures, we have a family of constraints that express the learnability of categorization:

Def. *CATEG (f: v) $\equiv \exists x_i \in f_{perc} \Rightarrow x_i \neq v$

> "The value v is not a category of feature f, i.e., a perceptual feature f cannot be recognized as the value v." (6.6)

Analogously to the situation with *GESTURE, as stated in (5.20), we have

*CATEG (*feature*: *value*) is undominated with probability 1 (6.7)

where *value* spans the whole range of values that *feature* can attain along its continuous auditory dimension. This expresses the finiteness of available perceptual categories within a language: *CATEG is high-ranked for almost all values, and low-ranked only for a small number of discrete values.

The interaction of the *CATEG, PERCEIVE, and *WARP constraints in recognition is the subject of the following section.

6.4 Categorization along a single perceptual dimension

As an example, we will take a look at the interaction of the constraints for the recognition of an auditory feature f that can have any value between 0 and 1000 along a continuous scale: the first formant, with a scale in Hz[13] . If PERCEIVE is undominated (i.e., every acoustic input will be categorized), and *WARP is ranked internally in the universal way, and *CATEG is ranked high except for the values $f = 260$, $f = 470$, and $f = 740$, then a partial hierarchy may look like (the dependence on f is suppressed from now on):

PERCEIVE
*WARP (400)
*WARP (300)
*CATEG (280), *CATEG (510), *CATEG (590) etc. etc. etc.
*WARP (240)
*WARP (140)
*WARP (100)
*CATEG (260)
*CATEG (740), *CATEG (470)
*WARP (50)
*WARP (20) (6.8)

Note that all the *WARP constraints not mentioned here do belong somewhere in this ranking, according to (6.5), and that all the *CATEG constraints not mentioned in (6.8) take fourth place in ranking, together with *CATEG (280). We will now see how this constraint system controls the recognition of any input value f_{ac} between 0 and 1000.

First, consider the input [260], which is a phonetic realization of f with a value of 260 (e.g., a vowel pronounced with a first formant of 260 Hz). We see that this auditory input is recognized as /260/ (in this tableau, the constraints have been abbreviated):

[260]	PERC	*W(400)	*C(280)
			*C(510)
			*C(590)
				*W(240)	*W(100)	*C(260)	*C(470)
							*C(740)
								*W(30)

☞ /260/						*		
/470/					*!		*	*
/740/		*!		*	*		*	*
nothing	*!							

(6.9)

The candidates /470/ and /740/, though chosen in (6.8) to be stronger categories than /260/, lose because honouring them would violate some stronger *WARP constraints.

[13] For (6.5) to be valid, we should use the perceptually calibrated Bark scale instead, but since the current case is meant as an example only, we use the more familiar physical frequency scale.

The winning candidate violates only the *CATEG(260) constraint, which cannot be helped: satisfying all *CATEG and *WARP constraints would require violating PERCEIVE.

The case of an input that is quite close to one of the preferred categories, yields an analogous result, as shown in the following tableau for the realization [510], which will be recognized as /470/:

[510]	PERC	*W(400)	*C(280)
			*C(510)
			*C(590)
				*W(240)	*W(100)	*C(260)	*C(470)
							*C(740)
								*W(30)

/260/				*!	*	*		*
☞ /470/							*	*
/510/			*!					
/740/					*!		*	*
nothing	*!							

(6.10)

In this case, we must consider the candidate /510/, which satisfies all *WARP constraints, but violates the strong *CATEG(510) constraint. Thus, because it is worse to map the input into the non-existing category /510/ than to distort the input by 40 Hz, the input [510] maps to the output /470/.

Another case is the recognition of an input that is not close to any of the good categories. The following tableau shows the recognition of [590]:

Input: [590]	PERC	*W(400)	*C(280)
			*C(510)
			*C(590)
				*W(140)	*W(100)	*C(260)	*C(470)
							*C(740)
								*W(30)

/260/				*!	*	*		*
☞ /470/					*		*	*
/590/			*!					
/740/				*!	*		*	*
nothing	*!							

(6.11)

The output candidate /470/, being 120 Hz off from the input, violates *WARP (119) but not *WARP (120). Thus, it is slightly better than the candidate /740/, which violates *WARP (149). So we see that stray inputs like [590] are put into the "nearest" category.

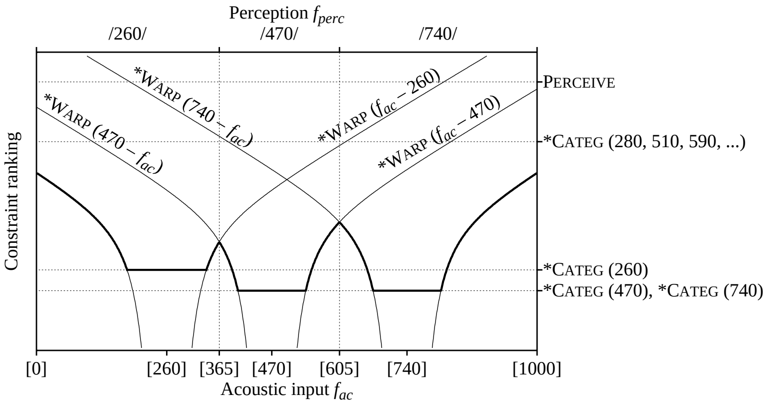


Fig. 6.1 Categorization of the input along a continuous auditory parameter. The curves represent the heights of the *WARP constraints in the cases that the auditory input is recognized as /260/, /470/, or /740/. The thick curve represents the height of the highest violated constraint if the categorization divides the domain into the three parts shown at the top.

Generalizing from these three examples, we can draw a picture of the recognition of all possible inputs between [0] and [1000]. Figure 6.1 shows the relevant PERCEIVE and *CATEG constraints as horizontal dotted lines, and the three *WARP constraints *WARP $\left(\left|f_{ac}-260\right|\right)$, *WARP $\left(\left|f_{ac}-470\right|\right)$, and *WARP $\left(\left|f_{ac}-740\right|\right)$ as functions of the auditory input parameter f_{ac}.

The picture shows that PERCEIVE is ranked as high as *WARP (550): the curve *WARP (740 – f_{ac}) crosses the PERCEIVE line at $f_{ac} = 190$; also, *CATEG (280 etc.) are as high as *WARP (350): the same curve crosses that *CATEG line at $f_{ac} = 390$. Two *criteria* (category boundaries) emerge exactly half-way between the categories, at 365 and 605. Note that though /260/ is a weaker category than /470/ (its *CATEG constraint is higher), the location of the boundary between the /260/ and /470/ equivalence classes is not influenced by this height difference: the height of the horizontal thick line above '260' in the figure does not influence the location of the cutting point of the two *WARP curves at [365], unless this line would actually be higher than the cutting point. This is an example of *strict ranking*: the two struggling *WARP constraints determine the outcome, without being influenced by any lower-ranked third constraint (§8.5 will show that the height of *CATEG correlates with the width of the *WARP curve, so that the criterion does shift).

In a more realistic model of speech recognition, the thick curve in figure 6.1 does not represent the ultimately recognized category. In the phase of recognition proper (seen here as occurring "after" categorization), which involves lexical access and information on context and syntax, we must assign a probabilistic interpretation to the curve (§8.2, §8.5): it only shows the *best* candidate, i.e., the candidate with highest probability of

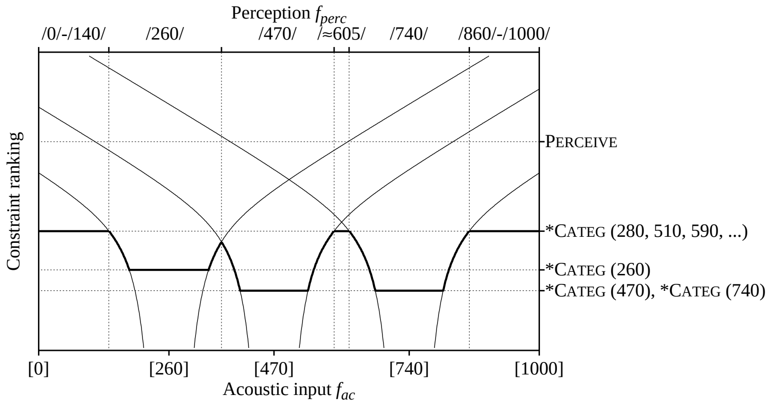


Fig. 6.2 Categorization along a one-dimensional continuum, if the *CATEG constraints for the poor categories are ranked rather low.

being correct; other, lower-ranked, candidates have lower probabilities, and a global optimization algorithm will find the best time path through the candidates.

6.5 Special case: weak categories

If the *CATEG constraints of the poor categories are ranked low enough, they can interact with *WARP constraints. In this case, highly distorted categorizations will not take place. Instead, inputs that are far away from the centre of the equivalence class of a strong category, will be recognized into one of the poor categories:

[590]	PERC	*W(400)	*W(110)	*C(280)
				*C(510)
				*C(590)
					*W(100)	*C(260)	*C(470)
							*C(740)
								*W(30)

/260/			*!		*	*		*
/470/			*!		*		*	*
☞ /590/				*				
/740/			*!		*		*	*
nothing	*!							

(6.12)

Figure 6.2 shows the classification of any input between [0] and [1000] in the case of low poor-category constraints.

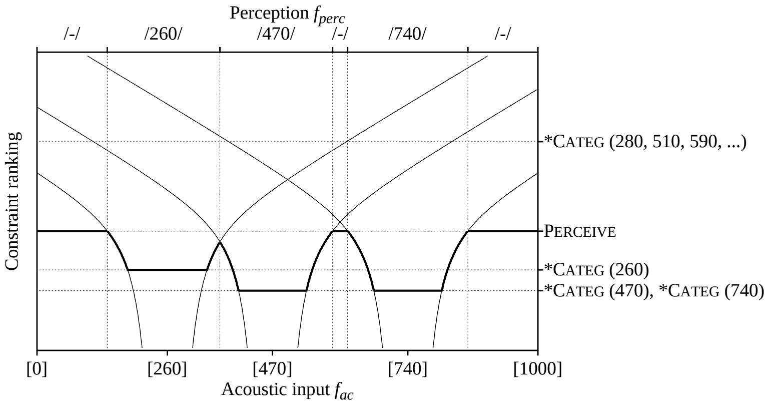


Fig. 6.3 Categorization along a one-dimensional continuum, if the PERCEIVE constraint is ranked low. Non-recognition is denoted as “/-/”.

6.6 Special case: unparsed features

If the PERCEIVE constraint is ranked low, it is allowed to interact with the *WARP constraints. In this case, highly distorted categorizations will not take place; instead, inputs that are far away from the centre of the equivalence class will not be recognized (“/-/” stands for “not recognized”):

[590]	*W(400)	*C(280)
		*C(510)
		*C(590)
			*W(110)	PERC	*W(100)	*C(260)	*C(470)
							*C(740)
								*W(30)

/260/			*!		*	*		*
/470/			*!		*		*	*
/590/		*!						
/740/			*!		*		*	*
☞ /-/				*				

(6.13)

Figure 6.3 shows the classification of any input between [0] and [1000] in the case of a low PERCEIVE constraint.

6.7 Dependence on environment

The ranking of the constraints of speech perception depends on several external and internal phenomena:

- A higher frequency of occurrence of a certain category in the vocabulary of a language means that that category is recognized more often, and, therefore, that categorization into this category is easier. Thus, frequently visited categories have low *CATEG constraints. This is formalized and proved in §8.5.
- A higher frequency of occurrence also lowers the distinctive power of a feature value and, with it, the height of the PERCEIVE constraint for this feature.
- The presence of background noise, too, reduces the importance of the classification of the individual features; thus, it lowers the ranking of PERCEIVE.
- More variation in the acoustics of a feature value gives more latitude in the admittance to the corresponding category, and this leads to relatively low *WARP constraints for high distortions (“wide” *WARP functions).

6.8 Merger

We can now predict what happens when two categories come to overlap. The source of the overlap is usually an increase in the variation in the production, often caused by the merger of a migrating group of people with another population that speaks a related but slighly different dialect.

Because of the large variation, the *WARP functions will be wider, as shown in figure 6.4. The more common (stronger) category (550) will have the lower *CATEG constraint; figure 6.4 shows us that this will lead to a shift of the criterion in the direction of the weaker category (to “442”). As every input greater than 442 will be classified as belonging to the stronger category, this criterion shift will again increase the rate of recognition into the stronger category, and decrease the rate of recognition into the weaker category. As a result of this, the *CATEG constraint of the stronger category will become lower, and that of the weaker category will become higher. This will cause a further criterion shift. Apparently, the larger class is eating away at its peer, and this positive-feedback mechanism will ultimately send the weaker class into oblivion (unless checked by the requirements of information content, see §8.6): an irreversible process of lexical diffusion ends up as a blind law of sound change. The resulting merger of the categories may well result at first in an asymmetry between production and perception: the speaker may still know that she produces a contrast, but the listener may be indifferent to it, because not considering the information found in a poorly reproducible contrast may decrease the error rate of the recognition.

The problem in figure 6.4 can also be solved by the weaker category moving away from the encroaching stronger one (push chain).

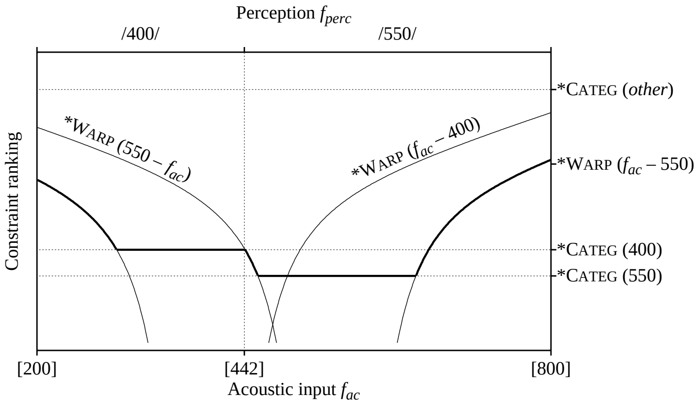


Fig. 6.4 The recognition into two overlapping categories of unequal strength.

6.9 Conclusion

The finiteness of sound inventories is explained by the articulatory inventory constraints (§6.2) and their perceptual counterpart:

Perceptual inventory constraints:

> “Low-ranked *CATEG constraints determine the finite set of allowed perceptual features.” (6.14)

The term “features” here is used in a broad sense: it may refer to values on a continuous auditory scale (e.g., F_1 or F_2), or to combinations of those (e.g., a location in a vowel triangle). Functionally, there is no reason why features should be one-dimensional; some latitude in the dimensionality of primitive perceptual spaces would explain why besides languages with highly symmetric vowel systems, we also find languages with asymmetric vowel systems; in the former case the language has several vowel-height and vowel-place categories, in the latter case it has vowel-quality categories. It is only natural that the languages with two distinct dimensions of categorization have more vowels than those with direct categorization of the two-dimensional quality space.

We can also draw an important conclusion from our functional standpoint: though all constraints may be universal, the features that build inventories are language-specific. For instance, all languages have the same constraints against the categorization of all vowel heights, i.e., they all have *CATEG (F_1: x) for all possible values of x. In every language, almost all of these constraints are undominated (see (6.7)). But though all languages have the *CATEG (F_1: 320 Hz) and *CATEG (F_1: 620 Hz) constraints, only a language with two categorizable vowel heights has them at a low rank, so that this language shows vowel heights at 320 Hz (“i”) and 620 Hz (“a”). Its sister language, with three vowel heights, has the same constraints, but has three different *CATEG constriants at a low rank, giving recognizable heights at 260 Hz (“i”), 470 Hz (“e”), and 740 Hz (“a”). Finally, a typical language with four vowel heights will have them around 240 Hz (“i”), 380 Hz (“e”), 560 Hz (“ɛ”), and 780 Hz (“a”). The interaction of *ENERGY and faithfulness constraints dictates the dependence of the peripheral heights (“a” and “i”) on the number of vowel heights (see §10.8), and the interaction of *WARP constraints determines the positions of the categories. The use of the label “a” with all three languages should not mean that we pose a universal category /a/, and the label “e” (which is especially arbitrary for the three-height language) does not mean the same for languages with three and four vowel heights: there is no universal vowel /e/. Thus, from the universal *CATEG family emerges a language-specific division of the vowel-height dimension, which is, moreover, partially determined by the functional principle of maximal minimal contrast. This leads to an important conclusion:

The functional view: there are no universal phonological feature values

> “The continuous articulatory and perceptual phonetic spaces are universal, and so are the constraints that are defined on them; the discrete phonological feature values, however, are language-specific, and follow from the selective constraint lowering that is characteristic of the acquisition of coordination and categorization.” (6.15)

7 Perceptual distinctivity

As with the maximization of articulatory ease, Trask (1996) calls the principle of *maximum perceptual separation* “a somewhat ill-defined principle sometimes invoked to account for phonological change”. But, again, we will see that it can be expressed in a linguistically meaningful way.

A global interpretation of maximization of contrast would express it in one measure, for instance, the probability of confusion. A utilitarian optimization strategy would then minimize the total number of confusions that would occur in a long series of utterances. An egalitarian optimization strategy, by contrast, would minimize the maximum confusion probability. The latter option is more in line with the idea behind Optimality Theory, where the highest-ranked constraint, i.e., the constraint against the largest problem, outranks all others. Interestingly, Ten Bosch (1991) showed that in a model of vowel inventories, the optimization strategy of maximizing the minimum distance between pairs of vowels, performed better than maximizing a global contrast measure along the lines of Liljencrants & Lindblom (1972) or Vallée (1994). An output-oriented contrast constraint would be

Def. *CONFUSION (*confusion*)

> “We are too petty to allow any positive amount of *confusion*.” (7.1)

The constraint-ranking version of minimization of confusion would then be stated as:

Minimization of confusion:

> “Within the set of all pairs of utterances with distinctive meanings, the pairs with higher confusion probabilities are disfavoured.” (7.2)

This rote functionalism is obviously not supported by the facts. It would predict, for instance, that sound changes would change only those words that are most easily confused with others, or that otherwise homogeneous sound changes would have exceptions where they would create homonyms. Such phenomena are very unusual, especially for gradual processes such as vowel shifts. This is explained by the facts of categorization: if categories are important, they move as a whole, dragging along all the words in which they occur. If the movement is gradual, there is no way for isolated lexical items to stay behind; only for sound changes that involve category jumps, like processes of lexical diffusion, it could be functionally advantageous not to jump if that would increase homonymy.

I will now review some possible ways of measuring contrast or confusion.

7.1 Discrete measures

A rough measure of the contrast between two utterances is the number of differing features. For instance, the difference between [v] and [p] is larger than the distance between [b] and [p]: two features (voicing and frication) versus one feature (voicing).

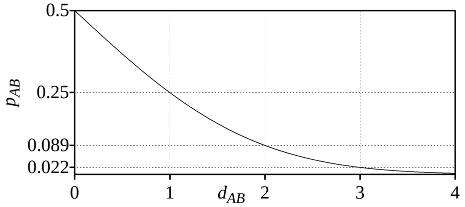
More precision can be achieved if we recognize the fact that the existence of a salient feature may partially obscure another contrast. Thus, the voicing contrast between [b] and [p] will probably be larger than the contrast between [f] and [v], because the presence of frication noise distracts the attention from other features. This statement has its roots in intuitive knowledge about the workings of the human ear. If not, we could equally well have brought forward that “the voicing contrast between [b] and [p] will probably be *smaller* than the contrast between [f] and [v], because the *absence* of frication noise distracts the attention from other features”. We know, however, of two properties of the auditory mechanism: firstly, the presence of noise may mask spectral information from other sources; secondly, periodic noise bursts (as in [z]) have a lower degree of periodicity than a truly periodic signal (as in [b]), thus giving a smaller periodicity contrast for the fricatives than for the plosives. A large say in the matter comes from perception experiments (though these are heavily influenced by language-specific categorization), which agree that [b] and [p] are perceptually farther apart than [f] and [v] (for Dutch: Pols 1983). The unmarkedness of plosives as compared to fricatives, as can be induced from the data of the languages of the world, can partly be traced back to this asymmetry.

A little more precision yet can be achieved if we take into account some asymmetries of the speech organs. Thus the voicing contrast between [k] and [g] will be smaller than the voicing contrast between [p] and [b], because of the different volumes of expandable air involved in helping to maintain the contrast.

7.2 Combining various perceptual dimensions to a global contrast measure

There exists a universal measure for the perceptual contrast between any two events (e.g., sounds) A and B. This measure is the *confusion probability* of A and B, and is defined as the probability that event A will be perceived as event B, which need not be equal to the probability that event B will be perceived as A. If this confusion probability *is* symmetric with respect to A and B (i.e., if there is no *bias* for either A or B), and A and B differ along only one acoustic/perceptual dimension, the confusion probability often bears a monotonic relationship with the *distance* between A and B along that dimension. This distance can then be expressed as a number of *difference limens* (units of just noticeable differences), and, if the variation along the scale is small in comparison with the total length of the scale, this number of difference limens may well exhibit an almost universal relationship with the confusion probability. Thus, if the distance between A and B is one difference limen, the confusion probability is 25% (this is one definition of a difference limen); if the perceptual measurements are drawn from a Gaussian distribution, and the distance is two difference limens, the confusion probability is 10%; for three difference limens, it is 2.4%; for four, 0.47%. The confusion probability is given by the formula

$$p_{AB} = \tfrac{1}{2}\left(1 - \operatorname{erf}\left(d_{AB} \cdot \operatorname{inverf}\left(\tfrac{1}{2}\right)\right)\right) \tag{7.3}$$



where d_{AB} is the difference between A and B, expressed in difference limens, and erf is related to the primitive of the Gaussian distribution function. If there are three events A, B, and C, there are two special cases. The first special case is if all three events differ along the same dimension, and B is perceptually somewhere between A and C. The distance between A and C can then be expressed as

$$d_{AC} = d_{AB} + d_{BC} \tag{7.4}$$

The second special case is if B and C differ along a dimension that is perceptually independent of the dimension along which A and B differ. The confusion probability between B and C can then be expressed as

$$p_{AC} = p_{AB} \cdot p_{BC} \tag{7.5}$$

Now, in order to derive an equation for the distance between A and C, we approximate (7.3) by

$$p_{AB} \approx e^{-\left(\frac{d_{AB}}{\alpha}\right)^2} \tag{7.6}$$

or

$$d_{AB}^2 \approx -\alpha^2 \log p_{AB} \tag{7.7}$$

We can now rewrite (7.4) as

$$\begin{aligned} d_{AC}^2 &\approx -\alpha^2 \log p_{AC} = -\alpha^2 \log\left(p_{AB} \cdot p_{BC}\right) = \\ &= -\alpha^2 \log p_{AB} - \alpha^2 \log p_{BC} \approx d_{AB}^2 + d_{BC}^2 \end{aligned} \tag{7.8}$$

which is the perceptual counterpart of the global articulatory equation (5.4).

If we realize that both equations (7.4) and (7.8) are Euclidean distance measures (for one dimension and two independent dimensions, respectively), we can conclude that the distance in the perceptual space can be measured as if this were a Euclidean space, provided that it is calibrated in units of one difference limen along every independent dimension. For instance, if the intensities of two sounds differ by 3 difference limens, and their pitches differ by 4 difference limens, the perceptual distance between these sounds can be expressed as “5 difference limens”.

To sum up, measuring every perceptual dimension with a dimensionless difference-limen scale allows us to compare distances along very different kinds of dimensions, and to compute in a natural way the total distance between any pair of events, provided that the Gaussian hypothesis and the strong hypothesis of separability (7.5) holds. And, of course, they do not normally hold. For instance, the total confusion probability may depend only on the maximum constituent confusion probability (a case of strict ranking):

$$p_{AC} = \max\left(p_{AB}, p_{BC}\right) \tag{7.9}$$

or, in the other direction, (7.4) might hold even if the pairs AB and BC differ along perceptually independent dimensions (city-block distance), so that the two sounds of our example differ by 7, instead of 5, difference limens.

7.3 Perceptual salience versus dissimilarity

Kawasaki (1982) draws our attention to the acoustic correlates of two aspects of the maximization of contrast. First, she points out that languages tend to disfavour contrasting, but acoustically very similar, sounds: poorly distinguishable sequences such as [gla] and [dla] tend not to co-occur in languages; Kawasaki calls this *maximization of dissimilarity*. Secondly, sequences of acoustically similar sounds such as [wu] or [ji] are avoided in the world's languages in favour of sequences with a greater acoustical dynamic variation like [wi] or [ju]. Kawasaki calls this *maximization of perceptual salience*.

Kawasaki defines perceptual salience as the amount of change of the perceptual features within an utterance. Her formula is

$$\sum_i \int \left(\frac{dP_i(t)}{dt} \right)^2 dt \tag{7.10}$$

where P_i are perceptual features (in Kawasaki’s case, formants in mel). The combination of the various dimensions seems to follow (7.8); the use of the squares cause (7.10) to be sensitive to the rate of change of the parameter, interpreting rapid changes as more salient than slow ones.

An analogous formula for the perceptual contrast between the utterances *a* and *b* as

$$\sum_i \int \left(P_{a,i}(t) - P_{b,i}(t) \right)^2 dt \tag{7.11}$$

In §7.2, we saw how perceptual features of different origin (e.g., voicing, tone, spectrum, and loudness) can be combined in such a formula if we know the difference limens of all of them.

7.4 Global or local contrast measures?

In §5.6, I argued for restricting the measurability of the ranking of articulatory effort to minimally different pairs of situations. The same holds for perceptual contrast.

In discussing similarity, it is impossible to give a universal answer to the question which pair is more alike: a horse and a cow, or an apple and a peach. But most people would agree that a horse is more similar to a cow than it is to a duck, and that an apple is closer to a pear than to a peach. Likewise, the listener cannot assign numerical values to the various degrees of contrast, but she can rank locally different contrasts. Thus, the main thing we will have to know about contrasts is the monotonicity of the relation between distance and contrast: the higher the distance between two sounds along a single acoustic/perceptual scale, the lower their probability of confusion.

8 Specificational and faithfulness constraints

The functional principle that surface forms with different meanings should be sufficiently different, can be implemented by a pair of requirements: the underlying forms should be sufficiently different, and every underlying form (*specification*) is close to the corresponding surface form (*perceptual result*).

Each candidate articulation in the specification-articulation-perception triad (§3.3) may produce a different perceptual result. The differences between the input specification and the perceptual output are caused by articulatory constraints, which tend to decrease the perceptual contrast between utterances. For instance, if the constraint against the laryngeal gestures that implement the voicing contrast for obstruents is ranked high, underlying /ba/ and /pa/ will fall together; and honouring the constraint against the synchronization of the velar and coronal gestures in /tɛns/ 'tense' will make it sound like the output of /tɛnts/ 'tents'. Thus, the principle of maximization of perceptual contrast can be translated:

- indirectly: into families of faithfulness constraints that state that aspects of the specification should appear unaltered in the output;
- directly: into the contrast-dependent rankings of these constraints.

A global formulation would be:

Def. FAITH (*d*)
"The perceptual output should not be different from the specification by any positive difference *d*." (8.1)

The constraint-ranking version of maximization of contrast would then be stated as:

Maximization of faithfulness:
"A less faithful perceptual output is disfavoured." (8.2)

This would be formalized into a universally expected constraint ranking:

$$\text{FAITH}\ (d_1) \gg \text{FAITH}\ (d_2) \Leftrightarrow d_1 > d_2 \qquad (8.3)$$

Just as with the constraints of articulatory effort, the faithfulness constraints branch into several families, which cannot be universally ranked with respect to each other along the lines of (8.3), which uses a global measure of contrast like equation (7.8). The various aspects of the underlying specification will be identified in the following sections.

8.1 Faithfulness in phonetic implementation

The first thing that is apparent from the specification (3.6) is the presence of features. For instance, the morpheme /tɛns/ contains specifications for [coronal], [+nasal], and [lower mid]. Because the speaker will try to accomodate the listener, it is desirable that the acoustic output contains something (anything) that corresponds to them. Analogously to the PERCEIVE constraint of perception, the speaker would adhere to the following imperative of correspondence:

Def. PRODUCE $(f) \equiv \exists x_i \in f_{spec} \Rightarrow \exists y_i \in f_{ac}$
"A value *x* on a tier *f* in the specification has any corresponding value *y* on the same tier in the acoustic output." (8.4)

An analogous constraint DONTPRODUCE, which can be formalized by reversing the implication in the definition of PRODUCE, requires that anything in the acoustic output has a correspondent in the specification (cf. DONTPERCEIVE in §6.3).

Mostly, the speaker is also intent on maximizing the probability of correct recognition of her utterance. So, analogously to *WARP, we would have a constraint that penalizes the variation of production, as far as this leads to deviant acoustic results:

Def. *VARY $(f\text{: } d) \equiv \exists x_i \in f_{spec} \wedge \exists y_i \in f_{ac} \Rightarrow \left| x_i - y_i \right| \le d$
"The produced value *y* of a perceptual feature *f* is not different from the specified value *x* by any positive amount of variation *d*." (8.5)

The wording of this constraint is deliberately symmetric between input and output. Like *WARP, *VARY is satisfied vacuously if the underlying feature has no correspondent in the acoustic signal: this may occur in situations where it is better not to produce a feature than to produce the wrong value. The universal ranking within the *VARY family is

Minimization of variation:
"A less deviant production is preferred over a more deviant production".(8.6)

This can be formalized as

$$\text{*VARY}\ (\textit{feature}\text{: } d_1) \gg \text{*VARY}\ (\textit{feature}\text{: } d_2) \Leftrightarrow d_1 > d_2 \qquad (8.7)$$

The picture presented here of the listener is that she will hear every acoustic output as it is. As we have seen, however, the effects of categorization discretize the perceptual output and, therefore, the perceptual specification. Discretized versions of PRODUCE and *VARY will be presented below.

8.2 Faithfulness in phonology

The listener will not rank the acoustically realized feature values directly along continuous scales. Rather, she will categorize the acoustic input into perceptual feature values along one-dimensional scales (“before” recognition of the utterance). The standpoint of Functional Phonology, inspired by the presence of an auditory-feedback loop (§2.1, figure 2.1), is that the version of faithfulness that plays a role in the organization of spoken language, evaluates the difference between the perceptual specification and the perceptual features *as categorized* by the listener.

We can view the medium of information transfer between speaker and listener as a system of parallel communication channels, each of which represents one perceptual tier. Each tier tries to transmit serially events associated with a particular perceptual feature. The presence of a message on each tier is transmitted successfully if the PRODUCE and PERCEIVE constraints are both satisfied (also in the unlikely case that PRODUCE and DONTPERCEIVE are both violated):

Def. TRANSMIT $(f \,/\, x) \equiv \exists x_i \in f_{spec} \Rightarrow \exists y_i \in f_{perc}$

> “The value (category) x on a tier f in the specification corresponds to any category y on the same tier in the perceptual output.” (8.8)

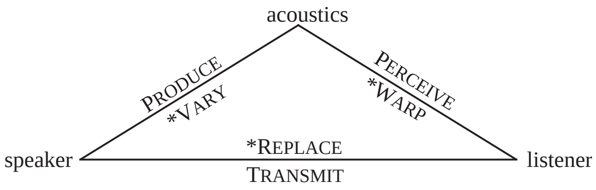
And, again, we have DONTTRANSMIT, which is satisfied if both DONTPRODUCE and DONTPERCEIVE are satisfied (or if DONTPRODUCE and PERCEIVE are both violated).

Analogously to *WARP and *VARY, we have a constraint that penalizes the difference between the specified and the perceived feature value:

Def. *REPLACE $(f: x, y) \equiv \exists x_i \in f_{spec} \wedge \exists y_i \in f_{perc} \Rightarrow \left| x_i - y_i \right| \le d$

> “The perceived category y on a tier f is not different from the specified value x by any positive distance d.” (8.9)

Thus, the effect of TRANSMIT is the product of the effects of PRODUCE and PERCEIVE, and the effect of *REPLACE constraint is the convolution of the effects of *VARY and *WARP. The communication process can thus be summarized as



(8.10)

The phonology handles TRANSMIT and *REPLACE constraints, because language users are speakers and listeners at the same time, and do not know about the acoustic medium.

In contrast with *VARY, which worked, by definition, along a perceptually homogeneous scale, *REPLACE has to be parametrized with the feature values x and y, because its ranking depends on the distances and strengths of the categories, as will be seen below and in §8.5. The universal ranking within the *REPLACE family, based on the principle that the listener will compensate for near categorical errors more easily than for distant errors (by adapting the recognition probabilities, see §8.5), is:

Minimization of categorization error:

> “A production that gives rise to a less distant categorization error is preferred over one that leads to a more distant error”. (8.11)

This can be formalized as (if y_1 and y_2 are on the same side of x):

$$\text{*REPLACE}\ (\mathit{feature}: x, y_1) \gg \text{*REPLACE}\ (\mathit{feature}: x, y_2) \Leftrightarrow \left| y_1 - x \right| > \left| y_2 - x \right| \qquad (8.12)$$

Because of the discreteness of categorization (if only a finite number of *CATEG constraints are dominated), it now becomes sensible to talk about a homogeneous version like *REPLACE (*feature*: x): “do not replace the feature value x by *any* different value”. With *VARY, this would have made no sense because *VARY (f: 0) would always be at the extreme lower end of the constraint system: it is utterly unimportant to have a F_1 which is within 0.01 Hz from the desired value, whereas recognizing, say, /tɛns/ as the neighbouring /tæns/ could already constitute a noticeable problem. The constraint family associated with generalizing this over all values of x, could be called *REPLACE (*feature*); if featural correspondence is forced by segmental correspondence (§12), such a family can be identified with the homogeneous segment-based IDENTIO (*feature*) constraint proposed for hybrid features by McCarthy & Prince (1995). However, we will see in §8.5 that the ranking of *REPLACE generally depends on its arguments x and y.

The ranking effects of (8.12) will be seen only for features that have been divided into many categories, like vowel height. Thus, for English /tɛns/, the outputs [tæns] and [tɛns] will be less offensive than the output [tins]. We can see how this works if we assign numeric values to the variation. For instance, figure 8.1 shows the distributions of the acoustic and perceptual results of a large number of replications of four vowel heights with specifications of 260, 430, 580, and 810 Hz, assuming a Gaussian model with equal standard deviations of 100 Hz (which could be caused by variations within and between speakers and by background noise).

With the help of figure 8.1, we can make a probabilistic version of what was presented in figure 6.1 as strict ranking. The shaded area in figure 8.1 represents the events in which /ɛ/ was intended, but /e/ was recognized. Its area is 0.218 (relative to the area under the Gaussian curve). The following table shows the conditional probabilities $P\left(f_{perc} = y_j \,\middle|\, f_{prod} = x_i \right)$ (the “|” reads as “given that”) of classifying the four intended categories x_i into each of the four categories y_j available for perception:

f_{prod}↓ f_{perc}→	/i/	/e/	/ɛ/	/æ/	$P(f_{prod} = x)$
/i/	0.802	0.191	0.007	$8\cdot10^{-6}$	0.25
/e/	0.198	0.575	0.223	0.004	0.25
/ɛ/	0.009	0.218	0.648	0.125	0.25
/æ/	$2\cdot10^{-6}$	0.001	0.124	0.875	0.25
$P(f_{perc} = y)$	0.252	0.246	0.251	0.251	

(8.13)

The right column contains the marginal probabilities $P\left(f_{prod} = x_i\right)$ of the four intended classes x_i, and the bottom row contains the total probabilities of finding each of the four initial recognitions y_j: $P\left(f_{perc} = y_j\right) = \sum_i P\left(f_{perc} = y_j \mid f_{prod} = x_i\right) P\left(f_{prod} = x_i\right)$.

Under the assumption of complete categorical perception, the best global strategy for the recognition of the categories is for the listener to assume the following Bayesian probabilities for the intended sounds, as functions of the initial categorization:

$$P\left(f_{prod} = x \mid f_{perc} = y\right) = \frac{P\left(f_{perc} = y \mid f_{prod} = x\right) P\left(f_{prod} = x\right)}{P\left(f_{perc} = y\right)} \qquad (8.14)$$

This results in the following table for these probabilities (the sum of each row is 1):

f_{perc}↓ f_{prod}→	/i/	/e/	/ɛ/	/æ/
/i/	0.795	0.196	0.009	$2\cdot10^{-6}$
/e/	0.194	0.584	0.221	0.001
/ɛ/	0.007	0.223	0.646	0.124
/æ/	$8\cdot10^{-6}$	0.004	0.125	0.871

(8.15)

We can now see that a more distant *REPLACE violation is worse than an “adjacent” *REPLACE violation: if the speaker produces [tʰens], the listener hears /tʰens/ but the candidate /tʰɛns/ has still a probability of 22.1% of being the correct candidate; if the speakers produces [tʰins], the listener hears /tʰins/ and the candidate /tʰɛns/ has only a probability of 0.9% of being the correct candidate. Thus, during the process of recognition, which, apart from the initial phonological classification, involves the lexicon, the syntax, and the semantics, the candidate /tʰɛns/ has a much larger chance of emerging on top if the production was [tʰens] than if the production was [tʰins].

The conclusion of this is that even in the idealized case of complete categorical perception before recognition, the *REPLACE constraints can be universally ranked. The reader would probably have believed this without all the above machinery, but we will need it again for a more complicated case below.

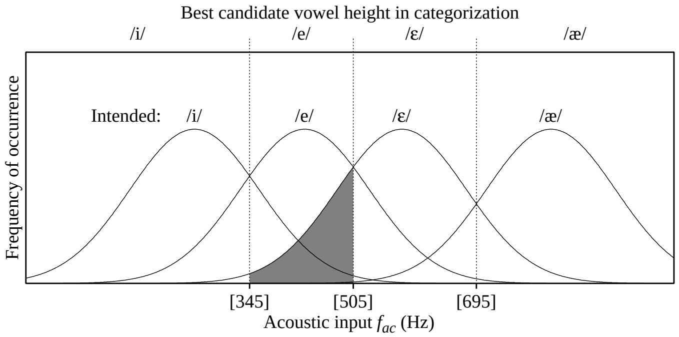


Fig. 8.1 The curves represent the variation in the production of four equally strong categories. The horizontal axis is the realized acoustic result. Along the top, the speaker’s optimal classification is shown.

8.3 The emergence of equally spaced categories

In figure 8.1, we see that the centres of the production distributions are not necessarily equal to the centres of the perceptual categories. For /ɛ/, the centre of the production was 580 Hz, whereas the centre of the perceptual category came out as 600 Hz, which is the midpoint between the two criteria that separate /ɛ/ from /e/ and /æ/. This seems an unstable situation. The speaker will cause fewer confusions in the listener if she produces an /ɛ/ right into the middle of the perceptual category, namely at 600 Hz. Thus, the slight asymmetry that arises in figure 8.1 as a result of the different distances from /ɛ/ to /e/ and /æ/, may cause a category shift from 580 to 600 Hz. This shift causes the criteria to move to the right, which induces a new shift. The equilibrium will be reached when the centre of the /ɛ/ category will be in the middle between the centres of /e/ and /æ/, i.e., at 620 Hz. Thus, the drive to equalize the category centres of production and perception favours the emergence of equal spacings between the categories, if they are equally strong.

Another prediction of this model is that languages tend to have their back vowels at the same heights as their front vowels, because they use the same F_1 categorization. If the number of back vowels is different from the number of front vowels, there is a tension between minimization of the number of height categories that have to be recognized, and equalization of the height distinctions among the front and back vowels separately.

8.4 Extreme feature values

In figure 8.1, the extreme categories /i/ and /æ/ behave differently from /ɛ/. If we assume an undominated PERCEIVE constraint, all feature values above 695 Hz will be perceived as /æ/. There is no centre, then, of the perceptual category /æ/; rather, its value is specified as "max" (maximal). The associated production constraint is

Def. MAXIMUM $(f\!: v) \equiv \exists x_i \in f_{spec} \wedge \exists y_i \in f_{ac} \Rightarrow \left(x_i = \text{“max”} \Rightarrow y_i > v \right)$

"If the value *x* on a tier *f* in the input is specified as "max", its acoustic correspondent *y*, if any, should be greater than any finite value *v*." (8.16)

For the non-categorizing listener, this constraint ensures the lowest probabilities of recognition into the adjacent category /ɛ/. The universal ranking is:

Maximization of the maximum:

"For "max" specifications, lower produced values are worse than higher values." (8.17)

This can be formalized as

$$\text{MAXIMUM } (\textit{feature}\!: v_1) >> \text{MAXIMUM } (\textit{feature}\!: v_2) \Leftrightarrow v_1 < v_2 \qquad (8.18)$$

Of course, analogous MINIMUM constraints should also be assumed.

The name of MAXIMUM is deliberately ambiguous. On the one hand, it can be seen as a universal constraint, because its logical formulation asserts that it only actively applies to features specified as "max". On the other hand, it can be seen as a language-specific output-oriented constraint (see §14.6): "the value of *feature* is maximal".

Since it is impossible for the produced value to reach infinity, the actually realized value will depend on the interaction of the MAXIMUM constraints with the articulatory constraints, which tend to disfavour extreme perceptual results (see §10.4).

*8.5 Weak and strong categories: the ranking of *REPLACE as a result of markedness*

This section describes a strategy for determining universal rankings of *REPLACE constraints.

Of the labial and coronal gestures, the coronal seems to be the 'easiest', since it is this articulator that is used most in many languages (the three stops most common in Dutch utterances are /n/, /d/, and /t/), and it can often occur in places where the labial gesture cannot. If this argument is correct (the asymmetry could also be due to coronals making better voicing contrasts etc.), we hereby identify the universal tendency *GESTURE (lip) >> *GESTURE (blade). But if there are more coronal than labial gestures in an average utterance, the distinctivity of the acoustic correlate of the labial gesture is larger than that of the coronal gesture. In this section, we will see how the listener reacts to this bias.

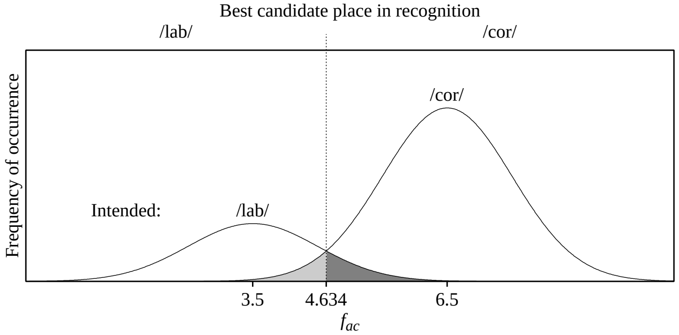


Fig. 8.2 Variation in production and acoustics causes an overlap of acoustic regions, leading to probabilistic recognition strategies in the listener.

Imagine that we have two gestures, [lip] and [blade], and that the lip gesture is more difficult (or slower) than the blade gesture. Thus, *GESTURE (lip) >> *GESTURE (blade). The result of this is that in a certain language, the blade gesture is used three times as much for plosive consonants than the lip gesture. Imagine further that the perceptual categories that correspond with these gestures are [labial] and [coronal], both measured along a perceptual dimension of place. What is the best categorization strategy for the listener, i.e., where along the place dimension does she have to put her criterion for distinguishing the two feature values in order to make the fewest mistakes?

Suppose that the auditory inputs from both gestures show variations (perhaps from imperfections in the production or from background noise) whose distributions can be described by Gaussian curves with equal standard deviations σ. Figure 8.2 shows, then, the distributions of the auditory input of a large number of replications of lip and tip gestures, produced with a ratio of 1 to 3, where the distance between the averages μ_1 and μ_2 is 3σ. The curve for [coronal] is three times as high as the curve for [labial].

The best criterion for discriminating the two categories is the point along the place dimension where the two curves cross, which is to the left of the mid-point between the averages, or, to be precise, at

$$\frac{\mu_1 + \mu_2}{2} - \frac{\sigma^2 \ln 3}{\mu_2 - \mu_1} \qquad (8.19)$$

With this criterion, the total number of confusions (the shaded area) is minimal: if you shift the criterion to the left or to the right, the shaded area will still contain everything that is shaded in figure 8.2, and a little more.

We can now derive a bias for confusion probabilities. We see from the figure that the shapes of the shaded areas to the left and to the right of the criterion are very similar, which tells us that the expected absolute number of incorrect [labial] categorizations is about equal to the number of incorrect [coronal] categorizations. However, the *probability* that a lip gesture is recognized as [coronal] equals the shaded area to the right of the criterion, *divided by* the total area under the [labial] curve, and the probability that a blade gesture is recognized as [labial] equals the shaded area to the left of the criterion, divided by the total area under the [coronal] curve. So we must expect from the ratio of the areas of the Gaussians that the probability that a lip gesture is recognized as [coronal] is approximately three times as high as the probability that a blade gesture is recognized as [labial]. The exact ratio, as a function of the distance between the averages, is

$$\left(\tfrac{1}{2}-\tfrac{1}{2}erf\left(\tfrac{1}{2}\sqrt{2}\left(\frac{d}{2}-\frac{\ln 3}{d}\right)\right)\right)\Bigg/\left(\tfrac{1}{2}-\tfrac{1}{2}erf\left(\tfrac{1}{2}\sqrt{2}\left(\frac{d}{2}+\frac{\ln 3}{d}\right)\right)\right) \qquad (8.20)$$

where d is the distance between the averages, expressed in standard deviations (in figure 8.2, d is 6.5 – 3.5 = 3). For strongly overlapping distributions, which can occur if the background noise is very strong, the ratio increases dramatically. Thus, we predict that relatively uncommon feature values will be mistaken for their relatively common neighbours, more often than the reverse, and that this bias is stronger for higher levels of background noise. This prediction is corroborated by some data:

- Pols (1983) for Dutch: inital /m/ is recognized as /n/ 26.1% of the time, the reverse confusion occurs 10.4% of the time; the plosives show a slight reverse bias: 5.4% versus 7.1%.
- Gupta, Agrawal & Ahmed (1968) for Hindi: initial /m/ becomes /n/ 67 times, the reverse occurs 27 times; /p/ → /t/ 66 times, the reverse 7 times (all sounds were offered 360 times).
- English /θ/ is more often taken for /f/ than the reverse.

This asymmetry will inform us about the ranking of *REPLACE (place: lab, cor) versus *REPLACE (place: cor, lab). The example of figure 8.2 gives the following confusion probabilities, obtained by dividing the shaded areas by the areas of the Gaussians:
$P\left(place_{perc}=\text{cor} \mid place_{prod}=\text{lab}\right)=12.8\%$, $P\left(place_{perc}=\text{lab} \mid place_{prod}=\text{cor}\right)=3.1\%$.
Thus, from every 100 replications of a [place] specification, we expect the following numbers of occurrences of produced and perceived values:

prod↓ perc→	lab	cor	total produced
lab	21.8	3.2	25
cor	2.3	72.7	75
total perceived	24.1	75.9	100

(8.21)

Doing the Bayesian inversion (8.14) (for our pre-categorizing listener) from the columns in this table, we can see that the probability that a perceived [labial] should be recognized as a produced [coronal], is 2.3 / 24.1 = 9.6%. In figure 8.2, this is the ratio of the lightly shaded area and the sum of the two areas at the left of the criterion. Likewise, the probability that a perceived [coronal] should be recognized as [labial], is 3.2 / 75.9 = 4.2%. In other words, perceived labials are far less reliable than perceived coronals.

Now consider a language in which underlying NC clusters arise from the concatenation of two morphemes. If coronals are three times as common as labials, 9/16 of those clusters will be /anta/, 1/16 will be /ampa/, and both /amta/ and /anpa/ will occur 3/16 of the time. We will now determine which of the two, /amta/ or /anpa/, will be more likely to show place assimilation.

If /amta/ is produced as [anta] (because the speaker deletes the labial gesture), the listener assigns the candidate /amta/ a probability of 4.2% · 95.8% = 4.1% (at least if she makes the [coronal] feature of [n] correspond to the [labial] feature of /m/; see §12 for a discussion of this *segmental hypothesis*). If, on the other hand, /anpa/ is produced as [ampa], the candidate /anpa/ still has a probability of 9.6% · 90.4% = 8.7%. Comparing these figures, we see that for a successful recognition of NC clusters, it is much more detrimental to replace a [labial] specification with a [coronal] output than the reverse. This means that a faithful surfacing of the labial place feature is more important than a faithful surfacing of the coronal place feature. Thus, because the speaker is also a listener, the constraint *REPLACE (place: lab, cor) must be ranked higher than *REPLACE (place: cor, lab). This gives the following partial universal ranking tendency of *REPLACE, written as segmental filters:

*REPLACE

$$\begin{tikzcd}
{*/\mathrm{p}/ \to \text{cor}} \arrow[d, dash] & {*/\mathrm{m}/ \to \text{cor}} \arrow[d, dash] \\
{*/\mathrm{t}/ \to \text{lab}} & {*/\mathrm{n}/ \to \text{lab}}
\end{tikzcd}$$

(8.22)

We thus see that the weaker specification (which may, at the surface, look like *underspecification*, see §13) of coronals is the ultimate result of an asymmetry in articulatory ease (or any other cause that leads to a frequency bias). This unmarkedness conspiracy can be summarized as follows:

*GESTURE (lower lip) >> *GESTURE (tongue tip)
→
frequency (tongue tip) > frequency (lower lip)
→

$$\text{frequency (place = coronal)} > \text{frequency (place = labial)}$$
$$\rightarrow$$
$$P\left(place_{perc} = \text{cor} \,\middle|\, place_{prod} = \text{lab}\right) > P\left(place_{perc} = \text{lab} \,\middle|\, place_{prod} = \text{cor}\right)$$
$$\rightarrow$$
$$P\left(place_{prod} = \text{lab} \,\middle|\, place_{perc} = \text{cor}\right) < P\left(place_{prod} = \text{cor} \,\middle|\, place_{perc} = \text{lab}\right)$$
$$\rightarrow$$
$$P\ (prod = /\text{amta}/ \mid perc = [\text{anta}]) < P\ (prod = /\text{anpa}/ \mid perc = [\text{ampa}])$$
$$\rightarrow$$
$$\text{*R}{\small\text{EPLACE}}\text{ (place: labial, coronal)} \gg \text{*R}{\small\text{EPLACE}}\text{ (place: coronal, labial)} \qquad (8.23)$$

Since, like labials, *dorsal* stops are also less common than coronals in most languages, the same ranking is expected for dorsals versus coronals (see §11.7 for Tagalog). Ranking (8.23) predicts that there are languages that show assimilation of coronals but not of labials and dorsals, namely, those languages where an articulatory constraint like *GESTURE is ranked between the two *REPLACE constraints (§11):

/anpa/	*REPLACE (place / _V)	*REPLACE (place: lab, cor / _C)	*GESTURE	*REPLACE (place: cor, lab / _C)
[anpa]			*!	
☞ [ampa]				*
/amta/				
☞ [amta]			*	
[anta]		*!		

(8.24)

Note that the ranking difference between *GESTURE (lips) and *GESTURE (blade) must be small for this to work; they are represented here as a single homogeneous constraint. The deletion of the coronal gesture in [ampa] is accompanied by a lengthening of the labial gesture; thus, the candidate [aãpa] must lose because a constraint for the preservation of the link between nasality and non-orality outranks *HOLD (labial). We fix the direction of assimilation by noting that perceptual place contrasts are larger before a vowel than in other positions because of the presence of an audible release, so that the environmentally conditioned universal ranking *REPLACE (place: *x*, *y* / _V) >> *REPLACE (place: *x*, *y* / _C) appears to be valid. The environments “_V” and “_C” refer to material present in the *output*, because that is the place where perceptual contrast between utterances must be evaluated (§14.6).

We thus derived a picture that is radically different from Prince & Smolensky (1993), who confound articulatory and perceptual principles by stating that “the constraint hierarchy [*PL/Lab >> *PL/Cor] literally says that it is a more serious violation to parse labial than to parse coronal” (p. 181). Moreover, they attribute this to “Coronal Unmarkedness”, an alleged principle of Universal Grammar. We can replace the claim of built-in references to phonetic content with a functional explanation: the ranking (8.23) follows from a general principle of perception: the adaptation of the listener’s expectations to variations in the environment.

8.6 Information

Following the reasonings from §6.8 and §8.5, you could think that the [coronal] category would eat away at the [labial] category until there were no labials left. In general, there are no classification errors if there is only a single category. However, this process is checked by another principle of communication: “maximize the information content of the average utterance” (§1). The information (measured in bits) that can be stored in every instance of a feature is

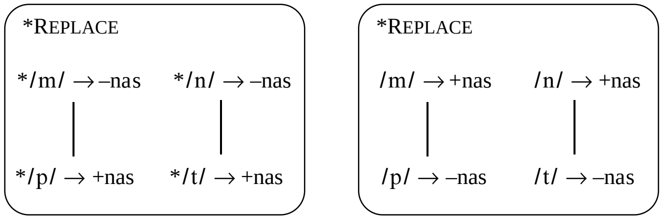
$$-\sum_i P\left(f = x_i\right) \log P\left(f = x_i\right) \qquad (8.25)$$

where the sum is over all categories. For instance, if a binary feature has two equally common categories, the information content is 1 bit per instance. If a binary feature has a category that occurs three times as much as the other category, the information content is $-0.75 \cdot \log_2 0.75 - 0.25 \cdot \log_2 0.25 \approx 0.8$ bits per instance. This means that for transferring 4 bits of information, an utterance should have a length of five instead of four instances of such a feature, which is not a world-shattering problem. However, if the frequency ratio of the two categories is 1000, a 100 times greater length would be required. Somewhere, an optimum exists, and it may be found by a technique analogous to the one that will be developed for the interaction between articulatory effort and perceptual contrast in §10.4.

8.7 Binary features

Several features are categorized with only two values in most languages. A typical example is [nasal], which can be seen as having the possible values “max” and “min”, which we can write as [+nasal] and [–nasal] because our notation does not have to heed any other values. Somewhat more symmetrically, we have [H] and [L] on the tone tier in some languages.

For binary features, the *REPLACE constraints are simplified to having a single argument: *REPLACE (nasal: +, –) is not any different from *REPLACE (nasal: +), because [+nasal] cannot be replaced with anything but [–nasal]. So we write *REPLACE (+nasal), *REPLACE (H) etcetera. Analogously to the argument of §8.5, we can posit universal rankings for binary features as functions of the commonness of their values. For the feature [nasal] (§2.6), this would give the following universal ranking:



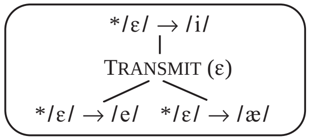
(8.26)

Next to the usual filter notation on the left, we see an equivalent positive notation on the right: *REPLACE constraints expressed directly as specifications. This is possible only for binary features. A word of caution is appropriate here: the positive formulation of the specification /m/ → [+nas] obscures the fact that the constraint is vacuously satisfied if correspondence fails, e.g., if no segment corresponding to /m/ appears in the output; the correct interpretation is more straightforward with a negative formulation.

8.8 Correspondence strategy for binary features

Correspondence is a part of the input-output relationship, and as such it is evaluated by the faithfulness constraints; no separate theory of correspondence is needed.

We will consider an interesting interaction between the correspondence constraint TRANSMIT and the identity constraint *REPLACE for features with few values. In the case of the four-valued height feature discussed in §8.2, the listener could have followed the strategy of finding out the vowel height by guessing. On the average, this would give a result that is 1.25 categories away from the intended category (1.25 is the average of 0, 1, 2, 3; 1, 0, 1, 2; 2, 1, 0, 1; 3, 2, 1, 0). Such a strategy would, therefore, be only slightly less advantageous than recognizing an intended category into an adjacent category, but more advantageous than a recognition that is off by two categories. This gives the following ranking:

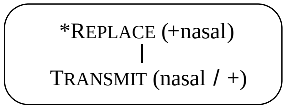


(8.28)

A similar ranking would be derived for PERCEIVE and *WARP.

A probabilistic argument will also work. In the case of the vowel in /tɛns/, it would not have been worse for the speaker not to produce any value for F_1 at all, than to produce an /e/. If the listener has to find out the vowel height by guessing, the /ɛ/ will have a probability of 25%, which is not worse than the probability that a perceived /e/ should be recognized as /ɛ/, which was 22.1% in our example. The probability that a perceived /æ/ should be recognized as /ɛ/ is even smaller: 12.5%. So, with the locations and widths of §8.2, all *REPLACE constraints would be ranked higher than TRANSMIT.

This situation is even stronger for features with two categories: it will always be better not to produce any value (50% correct from guessing), than to produce the wrong value (always less than 50%); or, by guessing, you will be half a category off, on the average, and by choosing an adjacent category you will be one category off, which is worse. Thus, binary features will always have a stronger *REPLACE than TRANSMIT constraint:



(8.29)

Now, because a violation of TRANSMIT will automaticaly cause the satisfaction of the higher *REPLACE, the best strategy for the listener will be not to make the output feature correspond to the specification at all, if no other constraints interact with TRANSMIT:

tɛns | +nas$_i$	*REPLACE (+nas)	TRANSMIT (nas / +)
tɛts | –nas$_i$	*!	
☞ tɛts | –nas$_j$		*

(8.30)

If the listener follows the strategy described here, the *REPLACE constraint will be invisible in her grammar, and a single combined TRANSMIT-*REPLACE constraint, equally highly ranked as the original TRANSMIT, will do the job. It combines a negative with a positive attitude:

Def. *DELETE $(f{:}\ x) \equiv \exists x_i \in f_{spec} \Rightarrow \left(\exists y_i \in f_{perc}{:}\ y_i = x_i\right)$

“An underlyingly specified value x of a perceptual feature f appears (is heard) in the surface form.” (8.31)

For instance, we have *DELETE (tone: H) and *DELETE (nasal: +), which can easily be abbreviated as *DELETE (H) and *DELETE (+nasal). Note that *DELETE (*feature*) cannot be satisfied by deletion of its bearing segment, in other words: *DELETE (*feature*) can actually force the parsing of whole segments, if ranked above *DELETE (timing: X).

Because of the impossibility of vacuous satisfaction of *DELETE, a positive name would be appropriate. In line with current usage, which refers to the surfacing of underlying material with the term "parsing", we will sometimes use the name PARSE, which originally comes from Prince & Smolensky (1993), who restricted it to the parsing of a prosodic constituent, like a segment, into a higher constituent, like a syllable. McCarthy & Prince (1995) coined a similar constraint MAX-IO, as an analogy with MAX-BR, which stated that a Reduplicant should take the maximum number of segments from the Base. For the faithfulness of hybrid features, some names based on the slightly inappropriate PARSE and MAX are: PARSE$^{\text{FEAT}}$ (Prince & Smolensky 1993), PARSEFEAT (Itô, Mester & Padgett 1995), MAXF (Lombardi 1995), MAX(FEATURE) (Zoll 1996). Also, in a declarative wave, we may decide to give this constraint no name at all, taking the specification "/+nasal/" or "∃[+nasal]" to mean: "there should be a [+nasal] in the output". In any case, a universal ranking for [nasal] is given by

*DELETE (+nasal)
|
*DELETE (–nasal)

PARSE (+nasal)
|
PARSE (–nasal)

(8.32)

which expresses the cross-linguistic preference for the assimilation of [+nasal] as in /akma/ → [aŋma], over the assimilation of [–nasal] as in /aŋpa/ → [akpa]. Besides promoting the presence of specified material in the output, a specification also implicitly states that unspecified material does *not* surface. If *REPLACE dominates DONTTRANSMIT, we have

Def. $*\text{INSERT}\,(f{:}\,y) \equiv \exists y_i \in f_{perc} \Rightarrow \left(\exists x_i \in f_{spec}{:}\; x_i = y_i\right)$
"A value *y* of a perceptual feature *f*, that is heard in the surface form, corresponds to the same underlying feature value." (8.33)

For instance, a replacement of /H/ by /L/ now violates both *DELETE (H) and *INSERT (L), if the listener manages not to make the two values correspond:

$/\text{H}_i/$	*REPLACE (H)	TRANSMIT (tone / H) *DELETE (tone: H)	DONTTRANSMIT (tone / L) *INSERT (tone: L)
$/\text{L}_i/$	*!		
☞ $/\text{L}_j/$		*	*

(8.34)

Again, because of its combined negative/positive interpretation, a positive name like FILL (Prince & Smolensky 1993) or DEPF (McCarthy & Prince 1995)) could be used instead of *INSERT. For the feature [nasal], we could superficially translate (8.32) into the fixed ranking (still restricted to assimilation):

*INSERT (–nasal)
|
*INSERT (+nasal)

FILL (–nasal)
|
FILL (+nasal)

(8.35)

but this would only be valid under a linear OCP-less interpretation of perceptual correspondence (§12).

The overlapping functions of *DELETE and *INSERT for binary features will be collapsed in §8.9 for those features which can be considered monovalent.

As an example of how a feature reversal as in (8.34) may come about, consider the floating H prefix found in Mixteco, as analysed by Zoll (1996: ch. 2). An underlying LM sequence as in /kìku/ 'child', enriched with the H affix, gives a HM sequence (/kíku/):

/kìku/ + H	*DELETE (tone: H / H-affix)	*DELETE (tone: L / base)
kìku	*!	
☞ kíku		*

(8.36)

Zoll notes that the constraint MAX(FEATURE) (i.e., a homogeneous *DELETE (tone)) does not do the job, not even if helped by IDENT(F), which is roughly a segment-based homogeneous *REPLACE (tone). This situation is reason for Zoll to propose a constraint MAX(SUBSEG), which we could translate as a homogeneous *DELETE (tone / floating). However, I can think of no functional explanation as to why the ranking of a constraint should depend on whether a feature is linked or not. Rather, two alternative approaches (also touched upon by Zoll), combined in the formulation of (8.36), follow from the theory of Functional Phonology developed so far.

First, we note that §8.5 proved that *DELETE constraints should be parametrized with feature values, because their ranking depends on the commonness of the feature values. For Mixteco, we could have *DELETE (tone: H) >> *DELETE (tone: L), or Zoll's MAX(H) >> MAX(L). With such a ranking, a floating L-affix would only be able to affect one of the eight possible tone sequences of Mixteco (namely, MM), whereas the floating H-affix affects four of them (MM, LH, LM, and ML); this would explain why Mixteco does not have any L-affixes.

The second possibility is conditioning the ranking by the base/affix opposition: *DELETE (tone / H-affix) >> *DELETE (tone / base), or Zoll's MAX (affix) >>

MAX (base). This would be the approach when H and L values are equally common in the language, so that neither of them can be considered unmarked. Morphological conditioning of faithfulness is quite common: the cross-linguistic tendency *DELETE (*feature* / base) >> *DELETE (*feature* / affix) has an obvious functional explanation (it is more important to keep all the information in content morphemes than to keep all the information in function morphemes), and manifests itself in the preference for base-to-affix spreading above affix-to-base spreading in vowel-harmony systems. The reversal of this ranking in the Mixteco case, where a failure to parse the H tone would obscure the entire affix, can be attributed to the idea that it is more important to keep *some* information about the affix than to keep *all* the information about the base. I would like to contend that functional arguments like these are the real explanations for facts of ranking (this is note the sole role of function in the grammar: even if most of the rankings are given, function is needed in describing the competence of the speaker, at least at the postlexical level, as shown in §10 and §11.5).

8.9 Privative features

Unary features are a special kind of binary features (§2.6).

For nasality, the probability of correct categorization depends on the quality of the nasality cues (heights of spectral peaks and depths of valleys) in the acoustic signal. It is probable that the categorization of this feature for almost all existing languages has resulted in two perceptually distinct values for the feature [nasal]: *present* and *absent*. With many aspects of perception, there is an asymmetry, a qualitative difference, between presence and absence. Also in this case, non-nasality is the default: perceptually, nasality is associated with some extra peaks and valleys in the auditory frequency spectrum, as compared to the more common spectra of vowels. Thus, we can posit the existence of a single-valued perceptual feature of nasality, and (3.6) contains the specification [nasal]. The following constraint ensures that it is present in the output:

Def. PARSE (f) $\equiv \exists x_i \in f_{spec} \Rightarrow \exists y_i \in f_{perc}$
“A specified feature *f* appears (is heard) in the surface form.” (8.37)

This constraints plays the parts of both TRANSMIT and *REPLACE, because you cannot replace a value of a unary feature with any other value, and it is equivalent to *DELETE. Thus, if PARSE (nasal) is violated, /tɛns/ will surface as [tʰɛts].

Not all features are privative. The feature [sibilant], for instance, is not a clear candidate for a privative feature: failure to satisfy an alleged PARSE (sibilant) would result in the output [tɛnt] or [tɛnθ]; but the latter is better because TRANSMIT (noise) probably dominates *REPLACE (noise: sibilant, strident) (§2.5), in contrast with requirement (8.29) for the existence of PARSE.

Also, we may have PARSE (coronal) and PARSE (labial), if the separate place features have their own tiers instead of being values of a perceptual feature [place]. But this is doubtful. For instance, the fact that it is less offensive to replace [θ] with [f] than to replace it with [χ], suggests a single perceptual feature [place], with *REPLACE constraints ranked by the perceptual contrasts of their argument pairs.

The global ranking of PARSE for unary features could be thought to depend on:

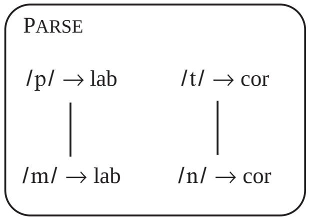
Maximization of conservation of salience:

> “The greater the distinctive power of a feature (value), the higher the ranking of its specification.” (8.38)

Thie parenthesized “value” in (8.38) suggests that multivalued features may also show a presence/absence asymmetry. On the noise scale, for instance, we have [aspirated], [fricative], and [sibilant], next to the absence of noise. For instance, if [sibilant] is a salient feature value, the contrast between [asa] and [ata] is large, so that the probability of the candidate /asa/ if the listener hears [ata], is low; if [aspiration] is a less salient feature, the contrast between [aka] and [akʰa] is small, so that the probability of the candidate /akʰa/ is reasonably high, even if the listener hears [aka]. This would imply that TRANSMIT (noise / sibilant) >> TRANSMIT (noise / aspiration): it is less bad for the speaker to leave out underlying aspiration than it is for her to leave out sibilancy.

However, it will be strongly language-dependent what features are considered salient and what are not. After all, it is a common property of human perception that it is difficult to compare unlike entities along scales like “conspicuity”, “salience”, or “notability”. For instance, people would disagree about whether a *duck* or a *lamp post* were the more conspicuous of the two. Thus, the conjecture (8.38), which, by the way, expresses the same idea as the *production hypothesis* of Jun (1995) (though that referred to acoustic cues, not perceptual features, see §11.8), is subject to language-specific variation and can, at best, be used to explain cross-linguistic *tendencies*, or the workings of very large salience/conspicuity contrasts, such like that between an *airplane* and a *tulip* (though even that depends on the environment).

For practical purposes, the ranking (8.38) is valid only for comparisons of a feature *with itself* in different environments. A clear example (for non-unary features) is the confusion probability of [m] and [n], as compared with the confusion probability of [p] and [t]. Measurements of the spectra of these sounds agree with confusion experiments (for Dutch: Pols 1983), and with everyday experience, on the fact that [m] and [n] are acoustically very similar, and [p] and [t] are farther apart. Thus, place information is less distinctive for nasals than it is for plosives. This means that for the understanding of the utterance, the emergence of the underlying place information in the actual phonetic output is less important for nasals than for plosives. In constraint terminology, we can express this as a general ranking of two parsing constraints, namely that PARSE (*place* / plosive) dominates PARSE (*place* / nasal). An alternative terminology would represent these constraints directly as *specifications*, e.g., /m/ → [labial]. A partial representation of the PARSE family will then look like (cf. 8.26):



(8.39)

A more accurate account would use *REPLACE instead of PARSE, as in §11.

The unary-feature version of both DONTTRANSMIT and *INSERT is:

Def. FILL (f:) ≡ $\exists y_i \in f_{perc} \Rightarrow \exists x_i \in f_{spec}$

"A feature f that is heard in the surface form, also occurs in the specification." (8.40)

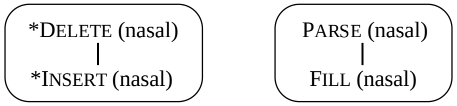
The implementation of /ɛns/ as [[ɛɛ̃n_ts]], chosen in order to satisfy two synchronization constraints, involved the epenthesis of a silence plus stop burst, in other words, a violation of FILL (plosive) (or FILL (silence) and FILL (burst) if we are talking autonomous-cue faithfulness, but according to §11.8, we should not). The alternative ordering of the two contours between [n] and [s] would give [[ɛɛ̃nəs]] or so, epenthesizing a syllable and thus violating FILL (timing: σ). Depending on the language, one or the other is worse. However, the epenthesis of [t] in this environment of a coronal nasal and a coronal obstruent is not as bad as the epenthesis of [t] between the elements of the sequence [ia]; a [j] would be more appropriate there. This means that the ranking of the FILL constraints depends strongly on the environment, and that the ranking is especially low if the syntagmatic perceptual salience of the utterance is hardly increased, as is the case in [[ɛɛ̃n_ts]] (continuous place information) and in [ija]:

Minimization of intrusive salience:

"The greater the distinctive power of a feature, the higher the problem of its insertion into the output." (8.41)

We can note that the implementation of /ɛn/ as [[ɛɛ̃n]] does not violate any FILL constraint, because all output features are already present in the input. Reversing the order of the two contours involved would give [ɛt^{n}], with epenthesis of a nasal plosive, which would be worse than the other candidate no matter how the language ranks the FILL constraints. This explains the universal preference for the implementation of the vowel-nasal transition with the nasal contour first; the few languages that do implement /ɛn/ as [ɛt^{n}], may heed a "path" (§8.11) constraint against the insertion of simultaneous nasality and vocalicness.

An interesting property of faithfulness constraints is that they do not distinguish between unary and binary features. If we conjecture, analogously to the universal ranking for place features (8.22), that the relative uncommonness of nasals in the average utterance causes the universal rankings (8.26), (8.32), and (8.35), we could equally well phrase this in privative terminology as the following near-universal ranking for the unary perceptual feature [nasal]:



(8.42)

So, whether or not we specify a value for [–nasal] in (3.6), makes little difference, if any.

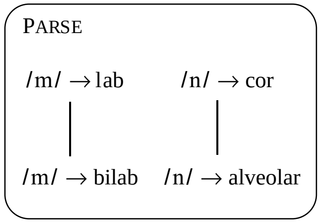
8.10 Ranking by specificity

Besides considerations of contrast, PARSE can be universally ranked by the specificity (perceptual precision) of its arguments. Like in the case of MAX, where a less specific constraint like F1 > 500 Hz was ranked higher than a more specific constraint like F1 > 700 Hz, we have analogous constraints for place of articulation. For instance, an /m/ is specified for [bilabial], but its [labial] specification must be stronger, because all bilabial consonants must necessarily be labial. For instance, Dutch /m/ may assimilate to a following labiodental consonant, but not to anything else; its labiality, therefore, seems more important than its bilabiality. Likewise, an /n/ is specified for [alveolar], but its [coronal] specification is stronger. These are instances of a more general principle:

Minimization of specificity:

"More specific perceptual features are less urgent than less specific features." (8.43)

This is in line with the functional principle "if you cannot have it all, settle for something less", and completely in accord with that maxim of Optimality Theory, *minimal violation*. After (8.22) and (8.39), we have our third partial universal hierarchy for place faithfulness:



(8.44)

The general principle (8.43) can be formalized as

$$(A \Rightarrow B) \Rightarrow \text{PARSE}\ (B) >> \text{PARSE}\ (A) \qquad (8.45)$$

or, as a generalization of PARSE (bilabial ∨ labiodental) >> PARSE (bilabial) ("∨" = "or), which, like the universal ranking of MAXIMUM, expresses the lower importance of narrow perceptual windows:

$$\text{PARSE}\ (A \vee B) >> \text{PARSE}\ (A) \qquad (8.46)$$

Note the asymmetry between articulation and perception, and between markedness and specificity:

*GESTURE (lab) >> *GESTURE (cor) ; PARSE (lab) >> PARSE (cor)
*GESTURE (bilab) >> *GESTURE (lab) ; PARSE (lab) >> PARSE (bilab) (8.47)

Because of this asymmetry, the PARSE and *GESTURE hierarchies generally interact in such a way that there is a working-point where the perceptual problems arising from imperfect contrastivity equal the problems associated with articulatory effort and precision; an example of this will be shown in §10.

There can be counterexamples to hypothesis (8.43), forced by other constraints. In §12.7, we will see an example of the somewhat perverse principle "if I cannot have it all, I'd rather have nothing".

The above example was somewhat unrealistic, because it hinges on a hierarchical place feature, divided into several (perceptual!) articulator features. If we accept the continuity of the perceptual place feature, so that the cross-articulator contrast between [θ] and [f] is smaller than the within-articulator contrast between [θ] and [ʃ], the ranking in (8.44) reduces to the less spectacular rankings of *REPLACE (bilabial, alveolar) >> *REPLACE (bilabial, labiodental) etc., which can be immediately related to confusion probabilities.

The asymmetry in (8.47) can be formulated in terms of precision: precise articulations are disfavoured, and precise productions are not needed.

8.11 Simultaneity constraints

Besides separate feature values, the specification (3.6) contains information about simultaneity of features. For instance, the /n/ of /tɛns/ is specified as simultaneously nasal and coronal. Simultaneous feature values on the perceptual tiers *f* and *g* can combine to new feature values on a combined tier *f* × *g*. For instance, the combination [coronal nasal] may be a member of a higher-level perceptual feature, say, [spectrum], and have its own correspondence and faithfulness constraints, which I will call *path constraints* as a tribute to Archangeli & Pulleyblank (1994), who use the term "path" to refer to simultaneously occurring features or nodes[14]:

Def. TRANSMITPATH $(f \times g) \equiv \exists x_i \in f_{spec} \times g_{spec} \Rightarrow \exists y_i \in f_{perc} \times g_{perc}$

"Every value *x* on the tiers *f* and *g* in the specification corresponds to any category *y* on the same tiers in the perceptual output." (8.48)

Def. *REPLACEPATH $(f \times g\!: x, y) \equiv \exists x_i \in f_{spec} \times g_{spec} \wedge \exists y_i \in f_{perc} \times g_{perc} \Rightarrow \left| x_i - y_i \right| \le d$

"The perceived category *y* on the tiers *f* and *g* is not different from the specified value *x* by any positive distance *d*." (8.49)

Def. *DELETEPATH $(f \times g) \equiv \exists x_i \in f_{spec} \times g_{spec} \Rightarrow \exists y_i \in f_{perc} \times g_{perc}$

"A specified combined unary feature on the tiers *f* and *g* appears (is heard) in the surface form." (8.50)

Def. *INSERTPATH $(f\!: \times g) \equiv \exists y_i \in f_{perc} \times g_{perc} \Rightarrow \exists x_i \in f_{spec} \times g_{spec}$

"A combined unary feature on the tiers *f* and *g* that is heard in the surface form, also occurs in the specification." (8.51)

For our example /tɛns/, the output [tʰɛs] would violate TRANSMITPATH (place × nasal), and the output [tʰɛms] would violate *REPLACEPATH (place × nasal: +nas cor, +nas lab), which is a more precise formulation than *REPLACE (place: cor, lab / +nas), because the latter wording is not explicit about whether the environment "+nas" should refer to a feature in the input or in the output or in both (but it must be the output, because that is where contrast is evaluated), and whether the input and output [+nas] should have to stand in correspondence; according to (8.49), they do not have to (and often, they do not, see §12), because links are autonomous.

Normally, we will write the constraint PARSEPATH (nas & cor) simply as PARSE (nas & cor) or PARSE (coronal nasal), expressing the unity of composite features. This constraint might be expected to be ranked below the less specific PARSE (nas) and PARSE (cor) (§8.10), so that it would be redundantly violated in [tʰɛms], [tʰɛts], and [tʰɛs], and visibly violated in [tʰɛ̃ts], which satisfies both PARSE (cor) and PARSE (nas). A recalcitrant ranking of PARSEPATH (nas & cor) above PARSE (cor) and PARSE (nas) may yield an all-or-none behaviour of the surfacing of /n/; a possible case of this is shown in §12.7.

The inevitable companion of a complex PARSE is a complex FILL. For instance, [tʰɛms] would violate FILLPATH (nas & lab) (which can simply be written as FILL (labial nasal)) as well as FILL (lab). Possible cases of crucial high rankings of this constraint are presented throughout §12.7. The usual output of /tɛns/, [[tʰɛɛ̃n_ts]], violates FILL (nasal mid vowel) and FILL (coronal plosive).

[14] Within a Containment version of OT with hybrid features, Itô, Mester & Padgett 1995 suggest PARSELINK and FILLLINK as constraints for faithfulness of association lines. They use it as part of a homogeneous FAITH constraint.

8.12 Precedence constraints

In /tɛns/, the feature value [sibilant] should occur after the vowel (this is satisfied) and after [nasal] (also satisfied), and [nasal] should occur after the vowel (partly violated). The candidate [snɛt] would violate both of these ordering relations, except the basic CVC ordering. For segments, McCarthy & Prince (1995) proposed a constraint LINEARITY to handle this. The featural version is:

Def. PRECEDENCE $(f\text{: }t;\ g\text{: }u) \equiv \exists t_i, u_j \in f_{spec} \wedge \exists v_i, w_j \in f_{perc} \Rightarrow \left(t_i < u_j \Rightarrow v_i < w_j\right)$

“A pair of contours at times *t* and *u*, defined on two perceptual tiers *f* and *g* and ordered in the specification, have the same ordering in the output, *if they occur there*.” (8.52)

This constraint can be satisfied by deletion, because the relevant TRANSMIT constraints independently control the presence of perceptual features in the output. This constraint expresses the difficulty of making reversely ordered feature values correspond to each other. For instance, does the underlying sequence /H_iL_j/, if surfacing as LH, correspond to L_iH_j or to L_jH_i? The answer depends on the relative ranking of PRECEDENCE (tone) and *REPLACE (tone).

To clarify this, consider the relation between the input /bɛrk/ and the output /brɛk/ on the root tier (in a language that disallows branching codas, for instance). If we subscript the input as /bɛ$_i$r$_j$k/, the output candidates /br$_i$ɛ$_j$k/ and /br$_j$ɛ$_i$k/ must be evaluated separately. Because an output /r/ is made to correspond with an input /ɛ/, the first of these candidates violates *REPLACE (ɛ, r). The second candidate violates a precedence constraint on the root tier. If we call the process metathesis, the second analysis must win:

/bɛ$_i$r$_j$k/	*CC]$_\sigma$	*REPLACE (ɛ, r)	*REPLACE (r, ɛ)	PRECEDENCE (root: ɛ, r)
bɛ$_i$r$_j$k	*!			
br$_i$ɛ$_j$k		*!	*!	
☞ br$_j$ɛ$_i$k				*

(8.53)

A violation of PRECEDENCE brings about metathesis. While this phenomenon can be seen as advocating segmental integrity, this typically segmental behaviour can also arise as a consequence of the dominance of combinatory feature constraints, not necessarily at the root level. For instance, PARSE (lower mid front vowel) and PARSE (vibrant sonorant), expressing the perceptual unity of some feature paths, would have sufficed in this case, but would, admittedly, have been less simple and generalizing. On the other hand, metathesis also exists on the featural level. Consider, for instance, the complicated correspondence relations in /hˈufnit/ → [snˈuftit] ‘I don’t want to eat that’, spoken by Jildou (aged 1;10): it involves hopping of the feature [nasal] to a position where it is better licensed (in her speech), leaving behind a coronal stop.

8.13 Alignment constraints

Coincidence relations exist between the beginnings and ends of the feature values in the specification. These often occur at equal times in a simple representation like (3.6): in /tɛns/, the nasal should start where the coronal starts, the vowel should end where the nasal starts, and [sibilant] should start where [nasal] ends. We can formulate a constraint that requires approximate simultaneity of the contour pairs in the output:

Def. *SHIFT $(f\text{: }t;\ g\text{: }u;\ d) \equiv \exists t_i, u_j \in f_{spec} \wedge \exists v_i, w_j \in f_{perc} \Rightarrow \left(t_i = u_j \Rightarrow v_i - w_j < d\right)$

“A pair of contours (edges) at times *t* and *u*, defined on two perceptual tiers *f* and *g* and simultaneous in the specification, are not further apart in the output (if they occur there) than by any positive distance *d*.” (8.54)

*SHIFT expresses the difficulty for the listener to reconstruct the simultaneity of contours, and the triple implication can be logically reversed:

Correspondence-strategy interpretation of *SHIFT:

“If two contours in the output do not coincide by anything less than *d*, they do not correspond to simultaneous contours in the input.” (8.55)

A universal ranking of *SHIFT is

Minimization of shift:

“A less shifted contour pair is preferred over a more shifted pair.” (8.56)

This can be formalized as

$$\text{*SHIFT } (f\text{: }t;\ g\text{: }u;\ d_1) \gg \text{*SHIFT } (f\text{: }t;\ g\text{: }u;\ d_2) \Leftrightarrow d_1 > d_2 \tag{8.57}$$

The formulation (8.54) is sensitive to the direction of the shift, and, therefore, to the order of the two arguments: we do not take the absolute value of $v_i - w_j$. Thus, [[t^{h}ɛɛ̃n_ts]] violates *SHIFT (coronal: –|+; nasal: –|+; 50 ms), because the coronal closure lags the lowering of the velum by 50 ms; likewise, it violates *SHIFT (vowel: +|–; nasal: –|+; 50 ms), and *SHIFT (sibilant: –|+; nasal: +|–; 30 ms). In a phonologized situation, time will be measured in moras (or so), instead of seconds. With unary features, we cannot refer to minus values, so we will have to refer to edges: we have *SHIFT (cor: Left; nas: Left), which does some of the work of PARSE (nas cor); *SHIFT (voc: Right, nas: Left), which does some of the work of FILL (voc nas); and *SHIFT (sib: Left; nas: Right), which expresses adjacency. In general, *SHIFT (*a*: Left; *b*: Left) expresses left alignment of *a* and *b*, *SHIFT (*a*: Right; *b*: Right) expresses right alignment, *SHIFT (*a*: Left; *b*: Right) militates against material intervening between *a* and *b*, and *SHIFT (*a*: Right; *b*: Left) militates against overlap.

If we get rid of the confusing edges, we can rephrase the four *SHIFT constraints as LEFT (*a*, *b*, *d*), RIGHT (*a*, *b*, *d*), *INTERVENE (*b*, *a*, *d*) (note the order of the arguments), and *OVERLAP (*a*, *b*, *d*).

Other general alignment constraints have been proposed. The best known is ALIGN (McCarthy & Prince 1993b):

Def. ALIGN (cat_1, $edge_1$, cat_2, $edge_2$)
"for every morphological, prosodic, or syntactic category cat_1, there is a category cat_2 so that $edge_1$ of cat_1 and $edge_2$ of cat_2 coincide." (8.57)

There are several differences between *SHIFT and ALIGN:

(a) ALIGN is homogeneous, i.e., it is not ranked by the amount of misalignment or intervening or overlapping material. It does incur a number of marks which is proportional to the extent of the violation, but this only allows ALIGN to interact with itself in the grammar. If this is realistic behaviour, the more restricted ALIGN should be preferred over *SHIFT in this respect.

(b) ALIGN is asymmetric with respect to its arguments: it is vacuously satisfied if cat_1 is missing, but not if cat_2 is missing (except under the assumption of Containment). No motivation for this asymmetry has ever been given. The alternative constraint ANCHOR, proposed by McCarthy & Prince (1995), does not show this asymmetry.

(c) ALIGN is symmetric with respect to overlap versus intervention, whereas *SHIFT allows to be ranked differently for these functionally very different situations.

(d) ALIGN is partly a positive constraint: deletion of cat_2 typically causes it to be violated. However, surfacing of cat_2 is independently controlled by its transmission constraint, so vacuous satisfaction should be allowed.

(e) ALIGN is formulated as a binary constraint; it needs a separate clause for assessing the number of violation marks. *SHIFT solves this problem with its distance parameter.

(f) ALIGN is morpheme-specific: it states the preferred positions of its arguments as constraints, whereas other (e.g., featural) specifications are part of the underlying form. *SHIFT is more consistent: if morphology is taken care of representationally, i.e., by time-aligning two contours in the input specification, the *SHIFT constraints automatically evaluate the deviations from this representational alignment. Thus, *SHIFT is language-independent, though its ranking (not its arguments) can be morphologically conditioned.

(g) ALIGN is not a faithfulness constraint. Instead of relating input and output, it evaluates the output in a declarative manner. Its implicational formulation allows it to be used for controlling *licensing*, if that happens to involve the edge of a domain. As Zoll (1996) shows, licensing does not always refer to edges, so a separate licensing constraint is needed anyway, like Zoll's COINCIDE (*a*, *b*) "if (the marked structure) *a* occurs in the output, it must be within a domain (strong constituent) *b*".

The binarity problem was noted by Zoll (1996), and she proposes an alternative:

Def. NO-INTERVENING (ρ; E; D)
"there is no material intervening between ρ and edge E in domain D."(8.58)

For concatenative affixation, Zoll rewords this as "if there is an element *x* in the base, and an affix *y*, *x* does not intervene between any part of *y* and the edge of the word"; the usual interpretation of gradient violation incurs one mark for every *x* that violates this. Besides solving the binarity problem (e), the negative formulation of this constraint fixes the problems of asymmetry (b), and vacuous satisfaction (d). Despite the existence of a COINCIDE constraint, however, NO-INTERVENING can still be misused for licensing purposes, because it still evaluates the output only. Moreover, the *empirical* (rather than technical) differences between ALIGN and NO-INTERVENING are few (Zoll does not provide any).

The largest empirical difference between *SHIFT and ALIGN/NO-INTERVENING is the distance parameter. While both ALIGN and NO-INTERVENING must be considered gradient constraints (in their workings), *SHIFT is a family of binary constraints with fixed internal ranking based on the distance between the realized edges.

First, we will see that *SHIFT can do the work of ALIGN. I will take the cherished example of Tagalog um-infixation, but analyse it very differently from Prince & Smolensky 1993; McCarthy & Prince 1993a, 1993b et seq. The prefixation of the root /basa/ with the actor-trigger morpheme /um/ (Schachter & Otanes 1972) gives /bumasa/ 'read', and /um/ + /ʔaral/ gives /ʔumaral/ 'teach' (that's the difference: not /um/ + /aral/ → /umaral/, because prefixation of another actor trigger gives /mag/ + /ʔaral/ → /magʔaral/ 'study', not */magaral/, showing that the glottal stop can be considered underlyingly present). The undominated licensing constraint ONSET "every syllable has an onset" (rather than the very violable NOCODA, which we may only need for cluster-initial loans like /gr(um)adwet/) forces violation of the lowest possible *SHIFT constraint:

/u$_i$m$_j$ | ʔ$_k$aral/	*$_σ$[V	PRECEDENCE	FILL (ʔ)	*OVERLAP (um, base, σσ)	*OVERLAP (um, base, σ)
u$_i$m$_j$ʔ$_k$aral	*!				
ʔ$_l$u$_i$m$_i$ʔ$_k$aral			*!		
ʔ$_k$u$_i$m$_i$ʔ$_l$aral		*!	*		
☞ ʔ$_k$u$_i$m$_j$aral					*
ʔ$_k$aru$_i$m$_j$al				*!	*

(8.59)

Some differences with ALIGN and NO-INTERVENING appear. Because *OVERLAP refers to an alignment difference between the right side of /um/ and the left side of /ʔaral/, the amount by which it is violated in /ʔumaral/ is actually /ʔum/. The output-oriented left-alignment constraint ALIGN (um, Left, Stem, Left) measures the distance between the left edge of the substring /um/ and the left edge of the entire string (stem) /ʔumaral/, which is /ʔ/. The non-directional constraint NO-INTERVENING measures the distance between the substring /um/ and the left edge of the entire string /ʔumaral/, which is also /ʔ/ (the constraint is non-directional, i.e., able to flip between right and left according to which side is closest to the specified edge of the word).

Intuitively, describing the violation as /ʔ/ seems preferable, and we could get this result with a faithfulness constraint that honours the left-aligned specification of /um/ instead of its adjacency to the base: the idea is that the "stem" already occurs in the input specification: it is the entire string /um | ʔaral/ as specified in the input. The violated constraint would then be LEFT (um, "stem", C), and *OVERLAP (um, base, σσ) would be replaced with LEFT (um, "stem", CVC), giving a tableau completely analogous to (8.59).

However, there is some very scant (probably dubious) evidence that the *OVERLAP constraints as stated in (8.59) are appropriate for Tagalog: if *OVERLAP (um, base, σσ) dominates FILL (C) (the two are not crucially ranked for /ʔumaral/), we can explain the fact that Tagalog has no bisyllabic infixes. For instance, the instrument-trigger morpheme /ʔipaŋ/   which Prince et al. would analyse as /ipaŋ/, is a regular prefix (/ʔipaŋ-hiwa/ 'cut with', not */h̤-ipaŋ-iwa/), and Prince et al. provide no explanation for the fact that bisyllabic "vowel-initial" prefixes are exceptions to the generalization that all and only the vowel-initial consonant-final prefixes show infixation.

Positing *SHIFT as a family predicts that its members can interact with other constraints separately, i.e., that it shows *inhomogeneity* effects. Now, ALIGN has always been considered a homogeneous constraint, so it would be interesting to find inhomogeneous alignment effects. Such an effect can be found in Yowlumne[15] glottalization (Newman 1944; Archangeli 1984; Archangeli & Pulleyblank 1994; Zoll 1994, 1996), in its interaction with vowel shortening.

The Yowlumne durative morpheme can be represented as the suffix /^{ʔ}aː/, where /^{ʔ}/ represents a floating [glottal plosive][16] feature (Archangeli 1984). This feature prefers to dock on the rightmost post-vocalic sonorant, with which it combines to give a single glottalized segment: /t^{s}aːw-/ 'shout' + /^{ʔ}aː/ gives [t^{s}aːw^{ʔ}aː]. We see that [w^ʔ] (the glottal constriction is centred around the middle of [w]) acts as a single segment: an utterance like *[t^{s}aːwʔaː] would be ill-formed in Yowlumne, because this language only allows CV, CVC, CVV syllables, so that CVVCCVV is not syllabifiable, and CVVCVV is. These syllabification requirements often lead to shortening of vowels: /ʔiːlk-/ 'sing' + /^{ʔ}aː/ gives [ʔel^{ʔ}kaː], where we see the expected glottalization and shortening of an ill-formed VVCCV to VCCV. If there are no glottalizable sonorants, as in /maːx-/ 'procure' (the /m/ is not post-vocalic), the result is a full glottal stop: it appears in /maxʔaː/, with shortening of the long vowel, which proves that /xʔ/ must be analysed as a consonant cluster, not as a single glottalized obstruent. Finally, the glottal stop does not surface if there is no glottalizable sonorant and no licit syllabification: /hogn-/ 'float' + /^{ʔ}aː/ gives [hognaː], not *[hognʔaː]; syllabification requirements could be satisfied by an otherwise well-attested epenthesis procedure, which could give a well-syllabified *[hoginʔaː], but glottalization does not appear to be able to enforce this.

Zoll (1994) notes that the output [t^{s}aːw^{ʔ}aː] violates a base-affix alignment constraint by one segment, because the left edge of the suffix coincides with the left edge of the segment [w^ʔ], and the right edge of the base [t^{s}aːw] coincides with the right edge of that segment. In order to satisfy ALIGN, the result should have been [t^{s}awʔaː], with a separate glottal-stop segment; but this would force shortening of the long vowel /aː/ in the base to [a]. Apparently, the constraint TRANSMIT (timing), or, more precisely, PARSE (μ), dominates ALIGN. In the following tableau, I have translated this idea into the current framework (with some undominated syllable-structure licensing constraints):

/t^{s}aːw | ^{ʔ}aː/	*VVC]$_σ$	*$_σ$[CC	*DELETE (μ)	*OVERLAP (base, suffix, C)
☞ t^{s}aː.w^{ʔ}aː				*
t^{s}aːw.ʔaː	*!			
t^{s}aː.wʔaː		*!		
t^{s}aw.ʔaː			*!	

(8.60)

[15] Also known as Yawelmani, which is a plural form denoting members of the tribe (Newman 1944:19; Zoll 1996: ch. 1: fn. 13).

[16] I use the perceptual formulation of this feature instead of the usual hybrid [constricted glottis].

The above account works for all suffixes that start with a floating glottal stop. However, Yowlumne has more suffixes with latent segments, and Zoll (1994, 1996) argues that these should be treated in the same way: like /ˀaa/, the suffix /ʰnel/ ‘(passive adjunctive)’ does not trigger epenthesis: when suffixed to /hogon/ ‘xx’, it gives [hogonnel], not *[hogonihnel] or so. However, it does induce vowel shortening, suggesting the ranking of ALIGN above *DELETE (μ):

/maxaː | ʰnel/	*VVC]$_\sigma$	*$_\sigma$[CC	*OVERLAP (base, suffix, σ)	*DELETE (μ)
ma.xaːh.nel	*!			
ma.xaː.hnel		*!		
☞ ma.xah.nel				*
mah.xaː.nel			*!	

(8.61)

Thus, Yowlumne would be a case for *OVERLAP (base, suffix, σ) >> *DELETE (μ) >> *OVERLAP (base, suffix, C), showing that alignment can work as an intrinsically ranked family of independently interacting constraints.

For the Yowlumne facts, other analyses may be possible. Zoll (1994) did not notice the discrepancy described above, but still, her 1996 version takes care of it. The non-directionality of NO-INTERVENING solves the problem of [tˢaːwˀaː]: the *right* edge of the [glottal stop] feature perfectly aligns with the right edge of the base, so NO-INTERVENING is not violated. Therefore, the homogeneity of Zoll’s alignment constraint is preserved.

Some of these problems relate to the idea that the real problem with infixation is not its lack of alignment, but its violation of the integrity of the base or the affix or both. This cannot be handled by general contiguity constraints, like those proposed by McCarthy & Prince (1995), because these also militate against epenthesis of new material. Rather, a constraint like *MIX (base, affix) could rule out outputs that correspond to the underlying morphemes in affix-base-affix or base-affix-base order (or, as in [ʔelˀkaː], base-affix-base-affix). That would be a morphological constraint, whereas *SHIFT only refers to phonological material, though its ranking could be morphologically conditioned.

8.15 Global or local ranking of faithfulness constraints?

As was the case with effort constraints, and follows from the argument in §7.4, the perceptually motivated constraints of speech production cannot be ranked in a universal way, except for local variations. Phonology translates system-wide contrast into a system of local, manageable universal rankings and language-specific rankings of non-neighbouring constraints. In §11, we will see the role of this principle in the phonologization of phonetic principles.

8.16 Conclusion

The *faithfulness* constraints favour the correspondence and similarity between the perceptual specification of the input to the speech-production mechanism and the perceptual result of each candidate articulatory implementation. Functionally, these constraints can be attributed to the principle of maximizing perceptual contrast: they try to bring all (often contrasting) feature specifications to the surface of the utterance. These constraints are thus perceptually based, although some of them are cast in terms that look deceptively articulatory in nature.

If underlying autosegments are freely floating objects, PARSE and FILL would be the only faithfulness constraints we need, but in reality we will also have to deal with constraints that favour the surfacing of any underlying simultaneity, precedence, and alignment.

The question of the correspondence between input and output features and their combinations is deferred to §12.

9 Interaction between articulation and perception

In §8, we met with some interactions between various kinds of faithfulness constraints. In the following sections, we will see how faithfulness constraints interact with the articulatory constraints identified in §5.

In §3.3, I stated that the perceptual output should look *more or less* like the specification. Constraint ranking determines what is more and what is less. In the /tɛns/ example, the following interactions between articulatory and faithfulness constraints occur:

- In the output [[t^{h}εɛ̃n_ts]], all forward faithfulness constraints (TRANSMIT and *REPLACE) are satisfied, i.e., all specified feature values emerge in the output: /t/ → [aspirated], /ɛ/ → [voiced], /ɛ/ → [max F2], /s/ → [sibilant], etc.
- The articulatory implementation shows the minimum number of *GESTURE violations given complete forward faithfulness. The constantly spread lips involve an appreciable violation of *HOLD.
- There are no simultaneous articulatory contours, so there are no violations of *SYNC.
- The complete satisfaction of *SYNC must sometimes lead to epenthesis. The chosen order of the nasal opening gesture and coronal closing gesture gives no epenthesis, because the resulting [ɛ̃] contains no perceptual features that are not present in [ɛ] or [n] as well. The chosen order of the nasal closing gesture and the coronal medial release gesture, however, leads to epenthesis of silence and a coronal release burst. Thus, *INSERT (plosive) is violated.
- The path constraints *INSERT (nasal vowel) and *INSERT (aspirated mid front vowel) are violated.

The following constraint tableau evaluates some candidate implementations for /tɛns/. The candidates are shown with a microscopic transcription, which should suggest the articulatory as well as the acoustic result, and with the probable early-categorized perceptual results, which determine the faithfulness:

/tɛns/	PARSE	*SYNC	*GESTURE	*INSERT (plosive)
(a) [[thɛns]] /tɛns/		*!*	*******	
(b) [[thɛs]] /tɛs/	*!*******		****	
☞ (c) [[thɛɛ̃n_ts]] /tɛnts/			*******	*
(d) [[thɛɛ̃ns]] /tɛns/		*!	*******	
(e) [[thɛ̃ĩs]] /tɛ̃ĩs/	*!		******	

(9.1)

The candidate /tɛ̃ĩs/, which violates PARSE (consonantal) and FILL (oral / nasal), is not a well-formed utterance in English, but it is the result of feature-level categorization, as assumed in §8. This is midway between gestalt recognition of the utterance (or segments) and grammaticization of separate acoustic cues (§11.8).

A concise justification of the specification (3.6) can now be given:

- Perceptual vowel features (as opposed to articulatory gestures) do not matter for the non-vowels (though, of course, the perception of nasality requires its own spectral features), so vowel features are not shown for /t/, /n/, and /s/. In constraint language: the perceptual distinctivity between rounded and unrounded /s/ is so small, that the relevant PARSE constraints are very low, so low that we cannot determine the underlying value, because it will always be overridden by an articulatory constraint[17]. The only way to construe a rounding value for /s/ is by noticing that an isolated /s/ is pronounced without rounding; so there may be a specification after all, but a very weak one. However, the *GESTURE (lips) constraint may be strong enough to override any rounding specification for /s/; suddenly, we cannot determine the underlying rounding value of /s/ any longer, because it would always be overridden (but see §13.8 for a strategy for determining an underlying value even in cases like this).
- In the same way, no coronal specification is needed for /ɛ/.
- Some values can be predicted from the values of other features. For instance, the coronal burst of /t/ forces the minus value for the [nasal] feature. But this is only true if the burst is parsed. For instance, if the specified /t/ is pronounced (and heard) as [n] (probably forced by an articulatory constraint), we may not only have a violation of PARSE (plosive), but also a violation of FILL (nasal & coronal).

[17] This does not mean that the rounding of [s] cannot play a role in the recognition of /su/ versus /si/.

- The vowel /ɛ/ is specified for [+voiced], because a voiceless vowel would be unacceptable in English. This specification is redundant in the sense that all English vowels are voiced. To capture this generalization, the lexicon might just contain the specification “vowel”, and some rules filling in the values of [sonorant] and [voiced]. However, for the determination of constraint satisfaction, we need the [+voiced] value, because a voiceless implementation of /ɛ/ is obviously unfaithful to the specification, and must, therefore, violate PARSE (voice). Our specification, therefore, is more phonetic than the minimal lexical specification. See also §13.2.
- We included a specification of [–nasal] for /ɛ/, because English vowels show up as oral, especially in isolation.

Whether /n/ and /s/ share a single coronal specification can be doubted, because of the different cues involved in /n/ and /s/, but I represented them that way in (3.6) so as not to give the impression that specifications are ‘linear’ rather than autosegmental. The question is whether there is anything against adjacent identical autosegments on specificational tiers, for instance, whether we should collapse the two [voiced] specifications for /ɛ/ and /n/. See §12.

In /tɛns/, the output correspondents of the [coronal] specifications of /t/ and /n/ must be separate: although the output [ɛns] satisfies one [coronal] specification, it does violate PARSE (coronal), because the listener will not be able to link the single recognized /coronal/ to the corresponding features of both /n/ and /t/ (because of the precedence constraints of §8.12, she will probably link it with /n/). Whether the [coronal] specifications of /n/ and /s/ should also have separate correspondents is another matter: they may be part of a homorganic NC cluster singly specified for [coronal] (§12).

9.1 Inherent conflicts

If we think of combining the articulatory and perceptual drives that build sound systems and determine phonological processes, we must conclude that not all functional principles can be honoured simultaneously.

For instance, the principles of maximizing perceptual salience and minimizing articulatory effort seem to be on especially bad terms. However, there are some utterances that combine a minimal number of articulatory contours with a maximal number of perceptual contours: the utterance [b̥ab̥a] only involves one opening and one closing gesture for each syllable without any laryngeal activity; the very young even manage to produce this utterance without lip gestures, only using their jaw muscles. Thus, this utterance involves only a minimal *GESTURE violation and no *SYNC violation at all; moreover, the labial closing gesture is ballistic, so no precision constraints are violated. The perceptual contours of [b̥ab̥a], on the other hand, are many: silence vs. loudness, voiceless vs. voiced, low vs. high first formant. This explains the preference of languages for the alternation of consonants and vowels.

As another example, the combination of maximization of salience and minimization of physical effort favours small movements that yield swift variations in perceptual parameters. This interaction predicts exactly the reverse effects from Stevens’ holistic precision criterion (§5.4). A comprehensive theory of functional phonology will show more interesting conflicts between the various articulatory and perceptual needs.

9.2 No interaction constraints

Our standpoint assumes a rigorous division of labour between articulatory and faithfulness constraints, so does not allow the use of surface-true constraints that can be reanalysed as an interaction between articulatory and perceptual needs. For instance, a constraint like “nasals assimilate in place to any following consonant” (§11.6) is not allowed in the grammar, because it should be seen as the interaction of a constraint that minimizes articulatory gestures, and a constraint that tries to preserve place contrasts. The relative weakness of the latter constraint for nasals as compared with plosives, causes the surface-truth of the hybrid constraint in some languages.

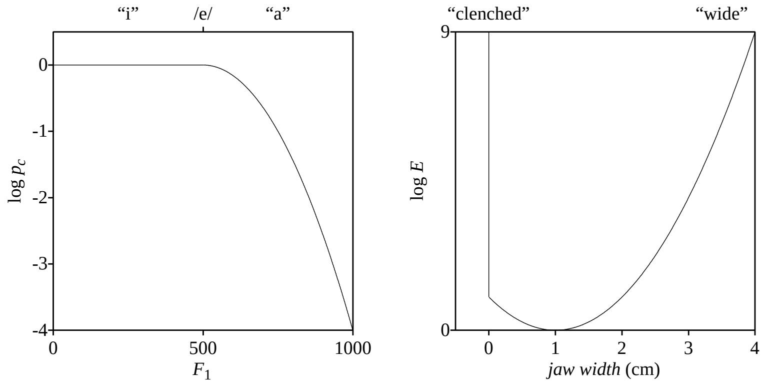


Fig. 10.1 Confusion probability as a function of first formant (left), and energy expenditure as a function of jaw width (right).

10 An example of acoustic faithfulness: vowel reduction

We will show the interaction between specification, articulation, and perception in phonetic implementation, using as an example the phenomenon of the reduction of the vowel /a/ in various contexts, in a language with the front vowels /a/, /e/, and /i/.

10.1 Specification: perceptual constraints

The perceptual specification of the vowel /a/ includes directions to make its height contrastive with that of its neighbours.

In our example, its nearest neighbour will be an /e/ with an F_1 of 500 Hz. The probability of confusing /a/ with /e/ as a function of the first formant of the realization of /a/, is roughly as shown in figure 10.1 (on the left): if the realized F_1 is 500 Hz, confusion with /e/ is complete, and confusion is much less for larger distances. Ideally, we should use a frequency scale calibrated in difference-limen units (§7.3), but if we crudely assume that we can use a linear Hz scale and that formula (7.3) for the relation between distance and confusion probability holds, the logarithm of the confusion probability is a parabolic function of the distance in Hz between the F_1 of the realization of /a/ and the F_1 of the neighbouring /e/. We then specify the vowel /e/ on the F_1 tier as [500 Hz], the vowel /i/ as [“min”], and the vowel /a/ as [“max”] (i.e., minimum vowel height).

In the phonetic implementation, actual values will have to be assigned to the first formant of /a/. Because of the continuous range of F_1, the [“max”] specification will branch into an infinite number of constraints, ranked logically according to the principle that a less restrictive specification is ranked lower than a more restrictive specification (8.17): thus, for the maintainance of the height contrast it is more important for /a/ to have its F_1 at least 100 Hz away from that of its neighbour, than it is to have its F_1 at least 200 Hz away. The constraint “F_1 is maximal” will therefore be divided up into a continuously parametrized constraint family MAXIMUM (F_1, *f*), or just ($F_1 > x$), where *f* is a frequency, and a partial ranking within this family is:

$$(F_1 > 600 \text{ Hz}) \gg (F_1 > 700 \text{ Hz}) \gg (F_1 > 800 \text{ Hz}) \qquad (10.1)$$

Instead of ranking these three arbitrary members only, we can express the logical ranking of the complete family as

$$(F_1 > x_1 \,/\, env) \gg (F_1 > x_2 \,/\, env) \Leftrightarrow x_1 < x_2 \qquad (10.2)$$

where *env* is any environment (the everything else that is kept equal). Hence, the falling slope between 500 and 1000 Hz in figure 10.1 (left-hand side) can be interpreted as the rankings of these specificational constraints along an arbitrary scale of importance.

10.2 Articulatory constraints

To find the actual resulting F_1 value, the MAXIMUM constraints have to be matched by articulatory constraints. A very high F_1 is difficult to produce, because of the strong jaw and tongue depression needed.

Consider first the /a/ spoken in isolation. The jaw opening, which is much wider for a typical /a/ than if all the muscles are relaxed, must be maintained by an isometric contraction of the mylohyoid and other jaw depressors (for simplicity, the tongue is ignored). According to formula (5.4), this involves more energy as the opening gets wider, because the elastic restoration forces increase. Figure 10.1 (right-hand side) shows the effort as a function of the jaw width, measured at the teeth: the resting width is 1 cm and all other widths take some amount of continued muscle activity, shown by the parabolic curve; widths below 0 cm are impossible to achieve, so the curve shoots off into space there. According to (5.11), we can translate this energy hierarchy into a *HOLD constraint hierarchy, analogously to the MAXIMUM constraint family of the previous section. This is reflected in the following formula (for openings wider than neutral), where I use the more general term *ENERGY (§5.1):

$$\text{*ENERGY (jaw opening} = x_1) \gg \text{*ENERGY (jaw opening} = x_2) \Leftrightarrow x_1 > x_2 \qquad (10.3)$$

Hence, the curve in the right-hand side of figure 10.1 can be interpreted as the rankings of these articulatory constraints along an arbitrary scale of importance.

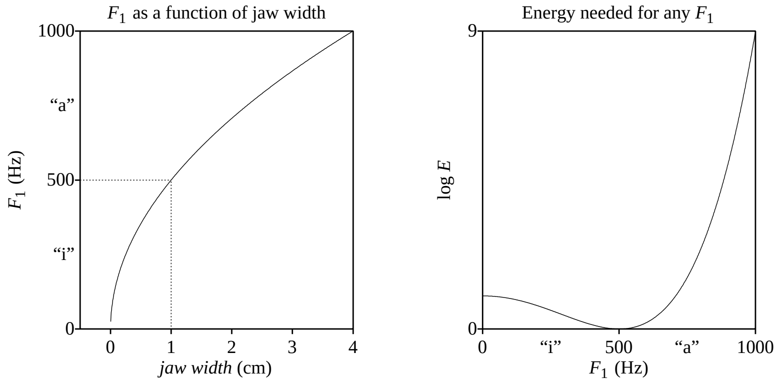


Fig. 10.2 The realized first formant as a function of the jaw width (left), and the energy needed to realize any F_1 (right).

10.3 Articulation-to-perception transformation

If we know the relative heights of all the MAXIMUM and *ENERGY constraints, we can compute the resulting F_1 value if we know the relation between jaw opening and F_1. Let's assume that this relation is (figure 10.2, left-hand side):

$$F_1 = 500 \text{ Hz} \cdot \sqrt{\frac{\textit{jaw width}}{1 \text{ cm}}} \qquad (10.4)$$

Thus, with a neutral jaw width of 1 cm, the first formant is 500 Hz, and a width of 4 cm is needed to increase it to 1000 Hz. Of course, this is a gross simplification of all the factors contributing to F_1, but it expresses the idea that the more peripheral a vowel must be, the more energy must be spent to achieve the necessary vocal-tract shape.

10.4 Interaction between articulatory and perceptual constraints

The right-hand side of figure 10.2 now shows the energy needed to reach a given F_1. It is computed from

$$\log E = \left(\frac{\textit{width}}{1 \text{ cm}} - 1\right)^2 = \left(\left(\frac{F_1}{500 \text{ Hz}}\right)^2 - 1\right)^2 \qquad (10.5)$$

Now that we know both the confusion probability and the needed energy as functions of F_1, we are in the position to compare the rankings of the two constraint families.

The following tableau shows four candidates for the expression of the underlying feature value [max F_1], for a certain choice for the interleaving of the two constraint families ("*ENERGY (jaw opening = *x*)" is abbreviated to "*E(*x*)"):

[max F_1]	*E(4cm)	F_1>600	*E(3cm)	F_1>700	*E(2cm)	F_1>800	*E(1cm)
550 Hz		*!		*		*	*
650 Hz				*!		*	*
☞ 750 Hz					*	*	*
850 Hz			*!		*		*

(10.6)

From these four candidates, the winner is 750 Hz. The first two candidates have a too low F_1, and the fourth candidate involves a too difficult gesture (jaw more than 3 cm wide).

We can represent the same intertwining of the constraint families with the two curves in figure 10.3a. As a measure of the "importance" of the specificational constraints, we take $10 + 5 \log p_c$; as a measure of the importance of the articulatory constraints we take $3 + \log E$. The two curves cross at about 750 Hz. To the left of this point, the perceptual constraint is the stronger, so that it forbids candidates with low F_1; to the right of the crossing, the articulatory constraint forbids candidates with a large jaw opening; at the crossing, both constraints are equally strong, and there must be a stable equilibrium here because we cannot optimize two interdependent quantities at a time. Thus, the OT optimization criterion is:

***Minimize the maximum problematic phenomenon*:**

> "The working point of a system of continous constraints is located where the two strongest optimization principles pose equal problems." (10.7)

We should compare this strategy with the strategy most commonly found in the literature: that of minimizing a weighted sum over the various factors. Figure 10.3b shows the resulting curves of adding $\log E$ to $\frac{1}{2} \log p_c$, $\log p_c$, $2 \log p_c$, and $5 \log p_c$. The gross features of these functions vary wildly, and only the third function has a minimum between 500 Hz and 1000 Hz. This should be compared with figure 10.3c, where $\log E$ is *subtracted* from the four functions $5 + \log p_c$, $1 + \log p_c$, $5 + 5 \log p_c$, $1 + 5 \log p_c$, after which the absolute value is taken. Though the four zeroes appear at somewhat varying locations, they all lie well within the region of interest.

The cause of the trouble is the fact that it is a poor optimization strategy to add a monotonically increasing function to a monotonically decreasing function; the result strongly depends on the precise shape of these functions, as well as on the weighting factor. By contrast, the presence of a cutting point in figure 10.3a does not depend on the

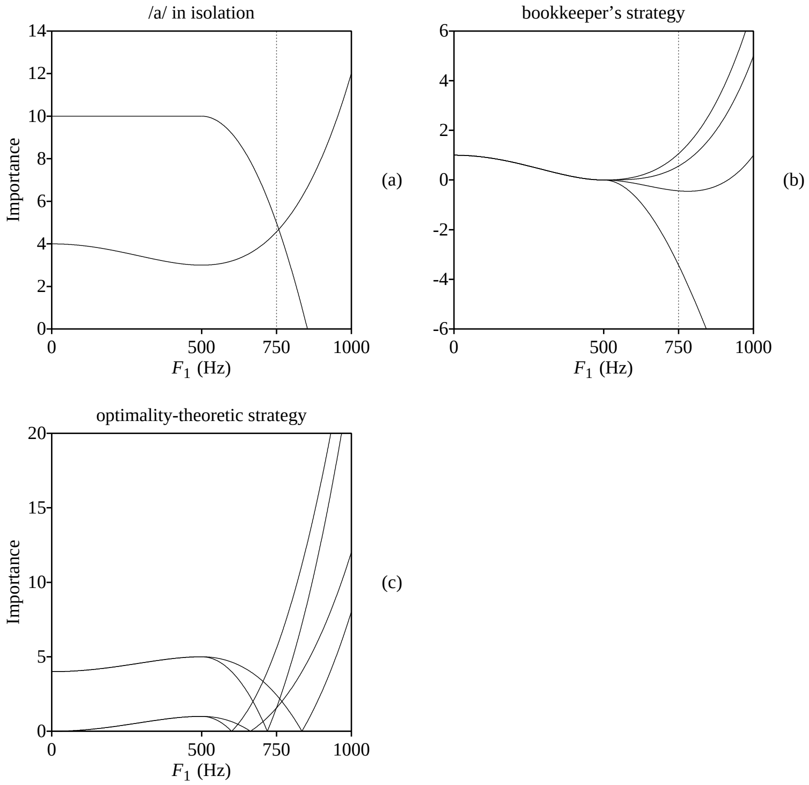


Fig. 10.3 Construction of the working point (the realized F_1) for the interacting perceptual and articulatory constraints in the phonetic implementation of /a/.

exact shapes of the functions, as long as these are monotonic. A comparable strategy of minimizing the maximum problem (in his case, vowel contrast) was shown by Ten Bosch (1991) to outrank Liljencrants & Lindblom's (1972) global optimization criterion for simulating vowel systems with phonetic principles; yet, Vallée (1994), in the same kind of simulations, returns to additive global optimization criteria, meticulously adapting her distance functions to the needs of stability. We must conclude, however, that the OT-compatible strategy of minimizing the largest problem is a more robust way of showing the presence of an equilibrium point.

We shall now turn to the environmental conditioning of the interleaving of the perceptual and articulatory constraint family, and prove that phonetic explanations can be adapted very well to an optimality-theoretic framework.

10.5 Dependence on stress

As usual, the ranking of the MAXIMUM constraints depends on the environment if the environment influences the distinctivity. Now, all distinctions are fainter in an unstressed than in a stressed environment (the average background noise masks more of the spectrum). This gives the functional ranking

$$(F_1 > x \;/\; +\text{stress}) \gg (F_1 > x \;/\; -\text{stress}) \qquad (10.8)$$

Thus, in unstressed position, the MAXIMUM constraints are ranked lower, and if the stressed position has its constraints ranked as in the previous tableau, the ranking in unstressed position may be as in the following tableau:

[max F_1]	*E(4cm)	*E(3cm)	F_1>600	*E(2cm)	F_1>700	*E(1cm)	F_1>800
550 Hz			*!		*	*	*
☞ 650 Hz					*	*	*
750 Hz				*!		*	*
850 Hz		*!		*		*	

(10.9)

Suddenly, the optimal candidate is only 650 Hz. The previous winner (750 Hz) now involves a jaw width (more than 2 cm) that costs too much in relation to the importance of very high F_1.

Figures 10.4a and 10.4b show curves of the constraint families in stressed and unstressed positions. Figure 10.4a is the same as 10.3a, i.e., the isolated /a/ is thought of as stressed. In the unstressed situation of figure 10.4b, the lowering of the PARSE family with respect to the stressed environment causes the working point to move down to 650 Hz. In the ultimate unstressed case, the MAXIMUM curve falls entirely below the *ENERGY curve, so that the *ENERGY constraints determine the working-point all by themselves: the resulting working-point is the minimum of the *ENERGY curve, i.e., the neutral position of the jaw, and the only vowel left in the system is a vowel with an F_1 of 500 Hz. Here we see an example of how the weakness of a faithfulness constraint can cause a change in the language's inventory of sounds; in Boersma (fc. c) we will defend the hypothesis that the interaction between articulatory and perceptual constraints indeed

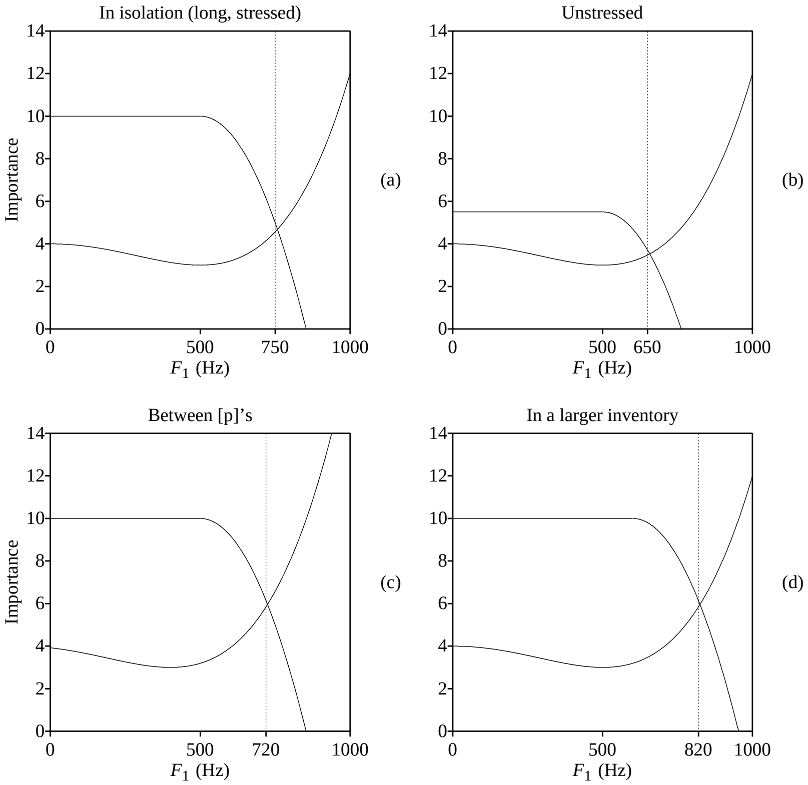


Fig. 10.4 The influence of various environments on the working-point in the interaction between a perceptual and an articulatory constraint.

determines the exact shape of every sound inventory, *including its size* (which is different from what all other phonetically-based models have done so far).

10.6 Dependence on surrounding consonants

Very probably, the energy, and thereby the ranking of the separate constraints of this family, does not depend on stress. The energy does depend, however, on the position of the articulators before and after the vowel. A given jaw opening is easier to achieve before the isolated [a] than in the utterance [pap], which involves two lip closures that can only be brought about with the help of a closing jaw. According to equation (5.4), the movement costs more energy as the distance to travel is larger, either because of the extra duration of the gesture, or because of the higher velocity and acceleration.

*ENERGY (jaw=*x* / [pap]) ≫ *ENERGY (jaw=*x* / [pa]) ≫ *ENERGY (jaw=*x* / [a])(10.10)

The constraint *ENERGY (jaw=*x* / [ap]) also belongs between the highest and lowest constraints in this formula, but can be ranked a priori with *ENERGY (jaw=*x* / [pa]) only if we can find a way of locally comparing [ap] and [pa] (i.e., seeing them as differing in one respect only), presumably by an argument involving time asymmetry.

If we want to know the resulting F_1, we can make a tableau like the previous one. Instead of weakening PARSE constraints, we now see strengthening *ENERGY constraints, but this produces the same kind of shift of these families with respect to each other. Again, therefore, the resulting F_1 will be lower in [pap] than in the ideal isolated [a]. This can also be seen in figure 10.4c: the zero-energy position of the jaw is more closed than in the isolated "environment", so the *ENERGY constraint curve moves to the left with respect to figure 10.4a, which results in a lower working-point.

10.7 Dependence on duration

A fast movement takes more energy than a slow movement. According to equation (5.4), if a given trajectory in space must be walked twice as fast, the double velocity combines with the double acceleration to give a fourfold increased power expenditure. Because the gesture is finished in half time, this leaves us with a doubled energy cost:

*ENERGY (jaw opening = *x* / –long) ≫ *ENERGY (jaw opening = *x* / +long) (10.11)

Along the lines of the previous sections, this will mean that the resulting F_1 is lower for short vowels than for long vowels. If we assume that the isolated /a/ was long, figure 10.4c shows the displacement of the *ENERGY curve with respect to the curve of figure 10.4a, which again results in a lower working-point.

10.8 Dependence on inventory size

Above, we considered a front-unrounded vowel system consisting of /a/, /i/, and only one mid vowel with an F_1 of about 500 Hz. Now imagine that we have two mid vowels instead of one. Their F_1 values are likely to be around 400 and 600 Hz. The nearest neighbour to /a/ is now the higher mid vowel with an F_1 of 600 Hz. This means that the MAXIMUM curve of figure 10.4a should now by centred around 600 Hz. This is shown in figure 10.4d. The 100-Hz change in the formant of the nearest neighbour causes the working point to move up by 70 Hz. The working-point does not move by 100 Hz, because the *ENERGY curve is not horizontal; thus, though the preferred F_1 of /a/ rises, the distance to its nearest neighbour decreases by 30 Hz.

10.9 Comparison to other models

Because of its comprehensive nature, the account of vowel reduction presented here is in accordance with almost every theory about it. From the presentation above, we can conclude that the shorter an open vowel is, the lower its F_1 will be; this is in line with Lindblom's (1963, 1990b) target undershoot model. Note that what happens here is not *centralization*, but *coarticulation*: the vowel triangle gets smaller because the low vowels *rise* in the direction of their neighbouring consonants; for low vowels, this is the same as centralization, but there is no articulatory or perceptual gain in centralizing *high* vowels in unstressed or shortened environments. This is in accord with the findings of Van Bergem (1995), who showed that high vowels did not centralize in these positions.

But we must also conclude that vowel reduction in unstressed syllables is caused by two phenomena: first, because of the lower intensity the contrasts are smaller, so that it becomes less important to maintain them; secondly, because of their unimportance, unstressed syllables will be shorter than stressed syllables, and this will reduce the vowels further because of the extra energy that would be needed to bring them to their 'long' position. As usual, a comprehensive optimality-theoretic account proves capable of reconciling the articulatory and perceptual viewpoints.

Two other vowel-reduction ideas should be noted here. Van Son (1993) showed that in rapid speech, a professional radio announcer was able to compensate for the shorter vowel durations by raising the velocity of the articulators in such a way that the same formant values were reached as in the slow-speech setting. Well, there are not many situations where faitfhulness is ranked higher than in the case of a speaker whose living depends on being clearly understood by a million people at the same time.

The other idea is that not the isolated long stressed vowel, but a short vowel in an unstressed environment might be the 'target' defined in our lexicon for that vowel, and that the clear stressed varieties are actually perceptual and/or articulatory enhancements over these moderately contrastive vowel targets (Koopmans-van Beinum 1980). Now, in the account presented in this chapter, the question is whether we need 'targets' at all: none of the four situations depicted in figure 10.4 was granted the status of a 'target', and the concept is meaningless in the context of interacting continuous constraint families. It all depends on your frame of reference.

Finally, we must note that listeners can compensate for the variation that results from the constraint interactions in various environments. For instance, so-called 'target undershoot' can be compensated for by a mechanism of 'perceptual overshoot' (Lindblom & Studdert-Kennedy 1967). For understanding the structure of sound systems, the existence of these mechanisms helps explain why listeners are so resilient that speakers may let their faithfulness constraints be dominated by so many articulatory constraints that phonology stays such an interesting subject.

10.10 Conclusion

The argument can be extended for peripheral vowels other than /a/. Peripheral front vowels are specified for maximum F_2, given their values of F_1. A high F_2 (with constant F_1) is achieved by a combination of wide pharynx (so that the tongue body does not have to be constricted too much), strongly bulging tongue, and strong lip spreading; relaxing these conditions in any way will lower the F_2. Peripheral back vowels are specified for minimum F_2, which, with constant F_1, is achieved by strong lip rounding and a back closure, the location of which depends on F_1. Any more neutral vocal-tract shape would give a higher F_2. So we see that the peripherality of front unrounded and back rounded vowels is subject to the same mechanisms as the lowness of /a/.

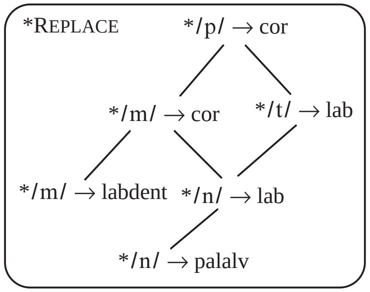
We have thus found formal functional explanations for the following well-attested facts of language:

- Vowels are less peripheral in unstressed than in stressed position.
- Vowels are more peripheral when spoken in isolation than when embedded in an utterance.
- Long vowels are more peripheral than short vowels.
- The vowel triangle is larger for large inventories than for small ones.
- In a large inventory, vowels are closer together than in a small inventory.

The model can be extended to other cases, most notably the interaction between *PRECISION and *SYNC. Like the acquisition of coordination facilitates recurrent use of combinations of articulatory gestures, the acquisition of categorization facilitates recognition of the discrete elements that make up the utterance, and is translated into a reranking of the *PRECISION and *SYNC constraints by changing the boundaries between which the articulations are constrained in order to produce a reproducible percept. Another field where the balancing model will lead to an optimum is the interaction between the informational constraint (maximum entropy, §8.6) on the one hand, and minimization of effort and categorization on the other.

11 Typology and phonologization: the local-ranking hypothesis

We can combine (8.22), (8.39), and (8.44) into the following partial grammar of (near-)universal rankings:



(11.1)

The lines in this figure connect pairs that vary along a single perceptual dimension (place), or that vary minimally in their environment (plosive/nasal), or that vary minimally in their degree of specificity. These minimally different pairs could be locally ranked according to universal principles of commonness (§8.6), environment-dependent contrast (§8.9), or the distinction between essentials and side-issues (§8.10).

As already touched upon in §5.6, §7.4, and §8.14, the remaining pairs in (11.1) cannot be ranked locally in this way, and we will propose that speakers and listeners cannot rank them in this way either. This leads to the hypothesis that phonology can rank but not count, or, more accurately:

Local-ranking principle (LRP):

> “Universal rankings are possible only within a single constraint family, for the same feature or gesture, for the same sign of the articulatory or perceptual deviation”, e.g. (5.7), (5.11), (5.12), (5.15), (5.26), (5.33), (6.5), (8.7), (8.12), (8.18), (8.45), (8.57), (10.2). (11.2a)
>
> “A near-universal ranking is possible only between a pair of constraints whose arguments or environments differ minimally”, e.g., (5.19) (for lip vs. blade), (8.23), (8.29), (8.32), (8.42), (10.8). (11.2b)
>
> “Cross-linguistic variation is expected for other pairs, though *tendencies* are expected for rankings based on global measures of effort or contrast, and the strength of the tendency depends on the difference between the two global measures”, e.g. (5.3), (5.19) (for blade vs. velum), (7.2), (8.3), (8.38), (8.41). (11.2c)

Of course, the transitivity of the strict-ranking scheme causes such rankings as *REPLACE (bilabial, coronal / plosive) >> *REPLACE (bilabial, labiodental / nasal) to be near-universal, too.

The LRP is a special case of Ladefoged’s (1990) statement: “there is no linguistically useful notion of auditory distinctiveness or articulatory economy in absolute terms”. Instead of going along with Ladefoged’s pessimistic view of the possibility of doing anything interesting with these principles in phonology, we can be glad that the LRP allows the linguist in her quest for universals to restrict herself to local, more manageable, variations, instead of tediously trying to measure the ingredients of equations (5.4) and (7.8).

11.1 Freedom of ranking

By itself, nearly every constraint can be ranked very low in one language, and very high in the other.

After the speakers of a language have learned a gesture, the corresponding *GESTURE (*gesture*) constraint is often very low; for other languages, however, it may still be undominated. For instance, a language typically has no apico-palatal closures at all, or it has a more or less complete set like /ɭ/, /ɳ/, /ɖ/, and /ʈ/.

The same is true of the *COORD families. Consider, for instance, the “anterior”-dorsal coordination found in oral suction consonants: a language typically has no click consonants at all, or it has a complete set with three, four, or five anterior releases, one or two dorsal closures, and several *manners* chosen from voiceless, voiced, aspirated, nasal, prenasalized, and glottalized.

The same, again, is true of *CATEG constraints: every language makes its own choice of subdividing the continuous parameter of vowel height or the continuous parameter of the voice-onset time of plosives.

Thus, the height of many *GESTURE, *COORD, and *CATEG constraints varies cross-linguistically from maximally high to maximally low. Universal notions of “easy” and “difficult” gestures and coordinations do not play any role in the description of any particular language. At best, these notions could explain statistical tendencies such as the relatively modest rate of occurrence of apico-palatal gestures and velaric ingressive coordinations when compared with, say, apico-alveolar gestures and labial-velar approximants.

We have seen that the possibility of universal ranking within a family is subject to the condition of *ceteris paribus* (“if everything else stays equal”): we can only impose an a-priori ranking on constraint pairs that differ minimally. There is no simple way in which we could predict the universal ranking of the labiality of /m/ and the coronality of /t/. The local-ranking principle proposes that there *is* no such universal ranking; this would mean that we expect that some languages rank the labial parsing constraints as a group above the coronal parsing constraints, and others rank the parsing constraints for plosives as a group above those for nasals:

Typological prediction of the local-ranking principle:

> "Languages can freely rank any pair of constraints that cannot be ranked by the LRP (11.2ab) directly or by transitivity." (11.3)

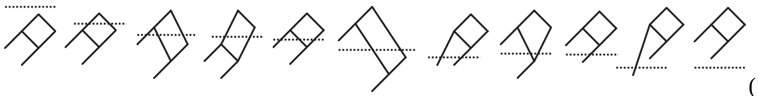
Stated as bluntly as this, (11.3) is too strong; after all, most people would agree that competitive skating is more difficult than riding a bike slowly, and that a horse is more different from a duck than an apple is from a pear. Thus, very large differences of effort and contrast will still be visible in the typology of languages (11.2c). We predict that *only* very large differences of effort and contrast will be visible in the ranking of non-minimally different pairs of constraints.

So it seems that we need only look at the local (one-feature) variation to predict universal or near-universal ranking, and that many of the more distant constraint pairs must be ranked in the grammar of each language. Restricting ourselves to these relative terms dismisses us of the task of finding global measures of effort or distinctivity: if languages do not care, why should the linguist?

11.2 Combinatorial typology

Prince & Smolensky's (1993) view of the freedom of ranking goes by the name of *factorial typology*: if there are four constraints, these can be ranked in 4! (*four-factorial*) = 24 ways. The local-ranking principle, however, restricts the freedom of ranking. If we have two families of three constraints, and the constraints within these families can be ranked according to universal principles, the rankings of each set of three constraints is fixed. The number of possible rankings should then be divided by 2!·2! = 4, leaving six ways in which languages are allowed to rank them. In general, with two families of *m* and *n* constraints, we have $\binom{m+n}{m}$ possible rankings: the number of *combinations* of *m* elements within a set of *m* + *n*.

The typical way to test the ranking of *REPLACE constraints is to split up the family by using a homogeneous *GESTURE constraint: all faithfulness constraints ranked above it will be satisfied; those below may be violated. Random variation in the ranking of this *GESTURE constraint determines the number of possible languages. For our ranking (11.1), we get 11 possibilities (the homogeneous *GESTURE is shown as a dotted line):



(11.4)

For consonants in onset position, the rightmost of this figure usually holds: all place contrasts surface. For consonants in coda position before another consonant, the PARSE constraints are ranked lower, and place assimilation may result. The leftmost of these figures depicts the situation in which all coda consonants assimilate to any following consonant.

11.3 Implicational universals

The connections in (11.1) allow us to state the following implicational universals for the assimilation of place (also put forward by Mohanan 1993):

- If plosives assimilate, so do nasals (at the same place). (11.5)
- If labials assimilate, so do coronals (with the same manner). (11.6)

The fact that there is no connection in (11.1) between */m/ → [coronal] and */t/ → [labial], means that (11.5) and (11.6) are independent from each other: there will be languages where nasals assimilate, but plosives do not, and there will be languages where coronals assimilate, and labials do not, and the inclusion of any language in the first group is independent from its inclusion in the second group, as we proved in §11.2. Thus, we have the following corollary:

Independence of implicational universals:

> "The local-ranking principle ensures that two implicational universals, if not transitively related, are independent from each other." (11.7)

The reverse is also true. Independence of the two implicational universals (11.5) and (11.6) gives the diamond-shaped part of (11.1), not two independent pairs of constraints. Thus, the hypothesis that (11.5) and (11.6) are translatable into the two independent rankings PARSEPLACE (plosive) >> PARSEPLACE (nasal) and PARSEPLACE (labial) >> PARSEPLACE (coronal), would predict that there are no languages where only coronal nasals assimilate, in contrast with the prediction of (11.5) and (11.6).

11.4 Case: place assimilation of nasal stops

We expect that *REPLACE (coronal, labial / plosive) and *REPLACE (labial, coronal / nasal), shown in (11.1), can be ranked in either way, depending on the language. That this accurately represents the situation in the languages of the world, will be illustrated with data on place assimilation of nasals in Dutch and Catalan.

In Dutch, nasal consonants at the end of a word have the tendency to change their place of articulation to that of an immediately following consonant. However, this tendency is not the same for all three nasal consonants (/n/, /m/, /ŋ/). The velar nasal /ŋ/ is always realized as a velar, irrespective of the place of the following consonant:

/dɪŋ/ 'thing' + /pɑkə/ 'take' → /dɪŋpɑkə/ 'take thing'
/dɪŋ/ 'thing' + /trɛkə/ 'pull' → /dɪŋtrɛkə/ 'pull thing'

/dɪŋ/ ‘thing’ + /kɛɪkə/ ‘watch’ → /dɪŋkɛɪkə/ ‘watch thing’ (11.8)

The alveolar nasal /n/ takes on the place of any following consonant, which can be velar, uvular, bilabial, labiodental, or palatalized alveolar:

/aːn/ ‘on, at’ + /pɑkə/ ‘take’ → /aːmpɑkə/ ‘take on’
/aːn/ ‘on, at’ + /vɑlə/ ‘fall’ → /aːɱvɑlə/ ‘attack’
/aːn/ ‘on, at’ + /trɛkə/ ‘pull’ → /aːntrɛkə/ ‘attract’
/aːn/ ‘on, at’ + /kɛɪkə/ ‘watch’ → /aːŋkɛɪkə/ ‘look at’
/aːn/ ‘on, at’ + /ʀaːdə/ ‘guess’ → /aːɴʀaːdə/ ‘advise’ (11.9)

The bilabial nasal /m/ is always realized as a labial, but may surface as labiodental before labiodental consonants:

/ʊm/ ‘about’ + /poːtə/ ‘plant’ → /ʊmpoːtə/ ‘transplant’
/ʊm/ ‘about’ + /vɑlə/ ‘fall’ → /ʊɱvɑlə/ ‘fall over’
/ʊm/ ‘about’ + /trɛkə/ ‘pull’ → /ʊmtrɛkə/ ‘pull down’
/ʊm/ ‘about’ + /kɛɪkə/ ‘watch’ → /ʊmkɛɪkə/ ‘look round’
/ʊm/ ‘about’ + /ʀɛɪə/ ‘drive’ → /ʊmʀɛɪə/ ‘make a detour’ (11.10)

This situation could be captured by the following naive superficial constraint system (from high to low):

(a) PARSE (dorsal), PARSE (labial), PARSE (nasal)
(b) NC-HOMORGANIC: “A sequence of nasal plus consonant is homorganic”
(c) PARSE (bilabial)
(d) PARSE (coronal) (11.11)

For instance, we see that the sequence /m + k/ must surface as [mk], because that only violates constraint (b), whereas [ŋk] would violate the higher-ranked constraint (a):

/m+k/	PARSE (labial)	NC-HOMORGANIC	PARSE (bilabial)
☞ mk		*	
ŋk	*!		*
nk	*!		*
ɱk		*	*!

(11.12)

On the other hand, /m + f/ must surface as [ɱf], as the highest violated constraint in this case is (d), whereas [ɱf] would violate constraint (b)[18]:

/m+f/	PARSE (labial)	NC-HOMORGANIC	PARSE (bilabial)
mf		*!	
☞ ɱf			*

(11.13)

11.5 Optionality

Language variation can simply be viewed as a variation in the ranking of constraints. For instance, for those speakers whose /m/ is always bilabial, constraint (c) ranks higher than constraint (b). But reranking is possible within a single grammar, too. Native speakers of Dutch often object to the reality of the constraint hierarchy that I showed above for the place assimilation of nasal consonants. Beside the fact that many people maintain that they always pronounce /m/ as a bilabial (and some of them actually do), people express considerable disbelief about the whole theory because “all those assimilation rules are optional”; they state that if they want to speak clearly, there need not be any place assimilation at all. Some opponents restrict their objections to the assimilation of /m/.

They are right of course. If your utterance is not understood at the first try, the importance of perceptual contrast rises with respect to the importance of articulatory effort, and you may repeat your utterance with fewer assimilations and more “parsed” features. In terms of constraint ordering, this means that perceptual constraints rise with respect to articulatory constraints. From the Dutch data, for instance, it seems warranted to state that “homorganic nasal plus consonant” first falls prey to “parse bilabial” (people start out saying [ɔmvɑlə] for /ɔm + vɑlə/), and that “parse coronal” only wins in situations where separate syllables are “spelled out” ([ɪnvɑlə] instead of [ɪɱvɑlə] for /ɪn + vɑlə/). This stylistic variation is the reason why we can rank “PARSE bilabial” above “PARSE coronal”, although the two can never be in conflict. The strength of the objections to the assimilation of /m/, expressed by some people, can now be seen, not as an overreaction to a mild constraint reranking, but as a defence against the shattering of the illusion of the discrete inviolability of the “PARSE bilabial” constraint.

On the other hand, we could also imagine that there are situations (highly predictable words; singing without the need to be understood) where articulatory constraints may rise with respect to perceptual constraints. In our example, we could expect that the first thing

[18] Because of the hybrid formulation, which bypasses the OCP for PARSE constraints, /m/ → [labial] is not violated. See §12.

to happen is that the velar nasal assimilates to a following uvular consonant (*onraad* vs. *vangrail*).

11.6 Problems with surface constraints

Most languages do not exhibit the combination of assimilation of /n/ and faithful parsing of /m/. But Catalan (Recasens 1991) and Dutch do. Instead of a cross-linguistically optional assimilation rule, we have a structural constraint, whose ranking determines whether we see the phenomenon: in Dutch and Catalan, it is ranked higher than in the non-assimilating languages (Limburgian), but lower than in the fully assimilating languages, like Malayalam (Mohanan 1986). Cross-linguistic optionality is thus automatically caused by the ranking of the constraints, and not an isolated coincidence.

A problem arises when we extend our example to clusters of plosive plus consonant. In Dutch, these clusters are not subject to the same assimilations as clusters of nasal plus consonant. For instance, though /n + x/ combines to /ŋx/, not /nx/, its counterpart /t + x/ is rendered as /tx/, not /kx/. The only assimilation I can think of is the assimilation of an alveolar plosive to a following palatalized coronal consonant, but it is hard to find even one example.

We could encompass all stops (nasals and plosives) in a single superficial grammar:

(a) PARSE (dorsal), PARSE (labial), PARSE (nasal)
(b) NC-HOMORGANIC
(c) PARSE (bilabial)
(d) PARSE (coronal)
(e) “A sequence of plosive and consonant is homorganic”
(f) PARSE (alveolar) (11.14)

In terms of functional principles, this is clearly wrong. NC-HOMORGANIC is an ad-hoc constraint, the result of a confusion of articulatory and perceptual constraints (§9.2); as such, it is found in the generative literature. For instance, Lombardi (1995) states: “in a language like Diola the constraint causing nasals to assimilate is high ranked, but whatever could cause other consonants to assimilate is low ranked”. What the *whatever* is, makes a large difference in explanation. Making the wrong choice here will eventually have repercussions throughout our theory of grammar.

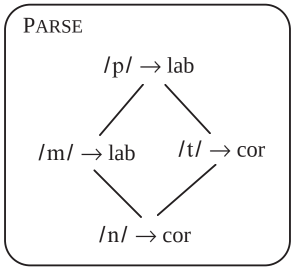
The articulatory gain of the homorganicity of plosive plus consonant must actually be *equal* to the gain of the homorganicity of nasal plus consonant, since it involves exactly the same articulatory phenomena: spreading of a place feature, and deletion of another. It is not the articulatory constraints, but the faithfulness constraints that are ranked differently. So, PARSE (coronal) is more important for plosives than for nasals, because its violation spans a larger contrast for plosives than for nasals. Therefore, the correct ranking is something like (assuming equal articulatory effort for the various oral closing gestures):

(a) /ŋ, k/ → [dorsal], /m, p/ → [labial], /t/ → [coronal]
(b) /p/ → [bilabial]
(c) *GESTURE (tongue tip), *GESTURE (upper lip), *GESTURE (back of tongue)
(d) /m/ → [bilabial], /t/ → [alveolar]
(e) /n/ → [coronal], /n/ → [alveolar] (11.15)

This ranking is not only in accordance with the data (it shares that with (11.14)), but it is also in agreement with the ranking (11.1), which was derived from functional principles.

11.7 Typology of place assimilation of nasals

The constraint ranking found in (11.15) contains some universal rankings, shown in this figure, which abstracts away from the second argument of *REPLACE:



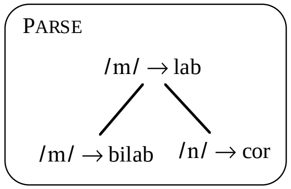
(11.16)

The solid lines in this figure reflect the universal ranking of place-parsing constraints for plosives above those for nasals, and the almost universal ranking of the parsing of labial features above coronal features. Depending on the ranking of the *GESTURE constraints, this predicts the following possible place-assimilation systems:

- Nothing assimilates (Limburgian).
- Only coronal nasals assimilate (Dutch).
- All coronals assimilate, but labials do not (English).
- All nasals assimilate, but plosives do not (Malayalam).
- All nasals and all coronals assimilate (language?).
- Everything assimilates. (11.17)

These are exactly the six that can be expected with a “combinatorial typology”. In those exceptional languages where the dorsal articulator is as commonly used for stops as the coronal articulator, we may find that PARSE (labial) >> PARSE (dorsal) also holds: in Tagalog, for instance, /ŋ/ will often assimilate (though not as often as /n/), and /m/ will not (Schachter & Otanes 1972); this seems to be a counterexample to Jun's (1995) cautious suggestion that "if velars are targets of place assimilation, so are labials". Note that with the separate rankings PARSE (lab) >> PARSE (cor) and PARSE (place / plosive) >> PARSE (place / nasal), as proposed by Jun (1995), the Dutch data cannot be explained (§11.3). Therefore, the dependence of contrast on the environment should generally be included in the environment clause of the constraint, and not be directly expressed as a constraint ranking, as Jun does. In other words, influences of the environment are *additive*, and not subject to strict ranking: they *add* to the ranking of the faithfulness constraint. Implicational universals respect this addivity.

For the finer place structure of nasals, we have the following universal ranking, simplified from (11.1):



(11.18)

Again, the two subordinate specifications are not neighbours, and can be ranked freely. This gives the following typology for assimilation of nasals to a following labiodental consonant:

- Nothing assimilates.
- Only /m/ assimilates: Central Catalan (Recasens 1991: 252, 256).
- Only /n/ assimilates: many speakers of Dutch.
- Both /m/ and /n/ assimilate: Mallorca Catalan and the other speakers of Dutch.
- Everything assimilates. (11.19)

Thus, we see that the only freely rankable pair of constraints (/m/ → [bilab] and /n/ → [cor]) can be shown to be actually ranked differently in a variety of Catalan and a variety of Dutch.

11.8 Perceptual versus acoustic faithfulness

As we will see in almost every example, the ranking of PARSE is usually determined by its environment. For the assimilation example /atpa/ → [[ap̚_ːpa]], there are two possibilities:

1. It violates PARSE (coronal / _C) or, in a loose declarative way, /t/ → coronal / _C. This is the approach found in the present work.

2. It violates PARSE (t̚) or *REPLACE (t̚, p̚). The first of these is analogous to Jun's (1995) account of place assimilation. There is an obvious problem in the autonomous ranking of separate place cues: because of the strict-ranking principle of OT, the cues do not additively contribute to the perception of place. I cannot tell whether this is Jun's intention; the use of a environment-conditioned constraint translatable as PARSE (place / onset) suggests that it is not.

The choice between the two approaches may be helped with the following argument: faithfulness is, in the end, a relation between specification and *perception*, not between specification and *acoustics*. Therefore, the effects of categorization should be taken into account. Now, if we accept that [coronal] is a perceptual category, and [t̚] is only an acoustic cue (see §12.3), and if we believe that strict ranking is the way that our grammar works, we must conclude that the grammar contains strictly rankable faithfulness constraints for [coronal], and that there is no evidence for such constraints for [t̚]. If we exclude constraints like PARSE (t̚) from the grammar, the possibility of additive contribution of acoustic cues to perceptual categorization is preserved (analogously to the aspects of *ENERGY, see §5.1). We already saw (in §11.4) that additivity of environmental information, which has a lot in common with additivity of acoustic cues, is needed to explain the data of Dutch place assimilation.

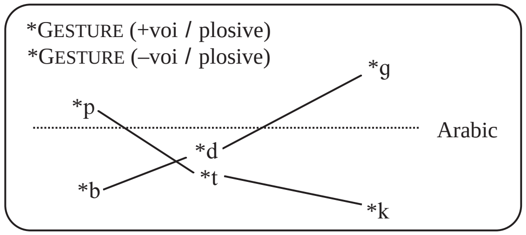
Thus, we opt for PARSE constraints for perceptual features, provided with environment clauses. The interpretation of the environment "C" that occurs in PARSE (coronal / _C), is that it refers to a consonant present in the *output*, not in the input, because the ranking of the faithfulness constraints should reflect the perceptual contrast between the output results [[at̚_ːpa]] and [[ap̚_ːpa]]. The relevant constraint is not *REPLACE (t̚, p̚), but *REPLACE (coronal, labial / C).

11.9 Constraint generalization

Depending on the relative heights of /m/ → [lab] and /t/ → [cor] and the homogeneous *GESTURE constraint, there must be languages where (11.16) can be simplified as PARSE (lab) >> PARSE (cor) (English) or as PARSE (place / plosive) >> PARSE (place / nasal) (Malayalam). This is a trivial case of generalization, empirically void because no difference can be detected with the full constraint system (11.16); it just means that there are no constraints in between the two, so that they appear as homogeneous. Only if the English and Malayalam-type languages occur much more often than the Dutch-type languages, could we conclude that languages like to use generalized constraints.

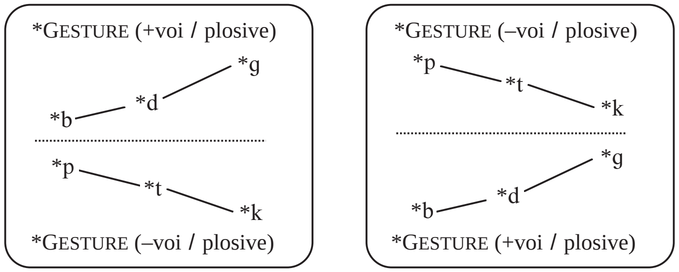
Another trivial case of generalization is the following. The near-universal hierarchy PARSE (place / onset) >> PARSE (place / coda) (which, accidentally, we need in order to derive the direction of place assimilation in §11.4), can be replaced with the ranking PARSE (place / onset) >> PARSE (place) without any empirical consequences, though the number of violation marks can be higher in the second case (if an onset place specification fails to surface). Note that we do not have to stipulate an Elsewhere principle to make this work. With this strategy, only PARSE (place / onset) and PARSE (place) need ever occur in grammars, and the constraint PARSE (place / coda) could be called *ungrounded* in the sense of Archangeli & Pulleyblank (1994). Here, the constraint PARSE (place / coda) can be considered superfluous because one of its simplifications would do as well.

As an example, we will now see how the articulatory problems of the voicing contrast in plosives can be generalized in the grammar. Because the amount of air that can expand above the glottis depends on the place of constriction, and some air expansion is necessary to keep the vocal folds vibrating for some time, a /g/ is more difficult to voice than a /d/ or a /b/ (Ohala 1976; see also Boersma 1989, 1990, fc. c, fc. d, to appear). For voiceless plosives, the situation is the reverse of this. Thus, we get the following global hierarchy of articulatory effort (a “phonetic difficulty map” in the words of Hayes 1996) for voicing contrast in plosives:



(11.20)

The lines in this figure connect universal rankings of voicing difficulty. Note that there are no lines connecting *p and *b, because languages, according to the LRP, cannot rank the effort of the two different gestures (say, pharynx widening and vocal-fold abduction) in a universal manner. Nevertheless, from the fact that more languages have a gap in their plosive system at /g/ than at /k/, and more languages have a gap at /p/ than at /b/, we may conclude that the phonetic difficulties are close to those portrayed in the figure. We can see, then, that Arabic actually respects the global hierarchy: it lacks /p/ and /g/ (and /ɢ/), as shown in the figure with a dotted line, which represents a homogeneous PARSE (±voi) constraint. In general, however, languages are free to rank the two families, so we expect to find lots of the following rankings:


*GESTURE (+voi / plosive)
*g
*d
*b
*p
*t
*k
*GESTURE (–voi / plosive)
*GESTURE (–voi / plosive)
*p
*t
*k
*g
*d
*b
*GESTURE (+voi / plosive)

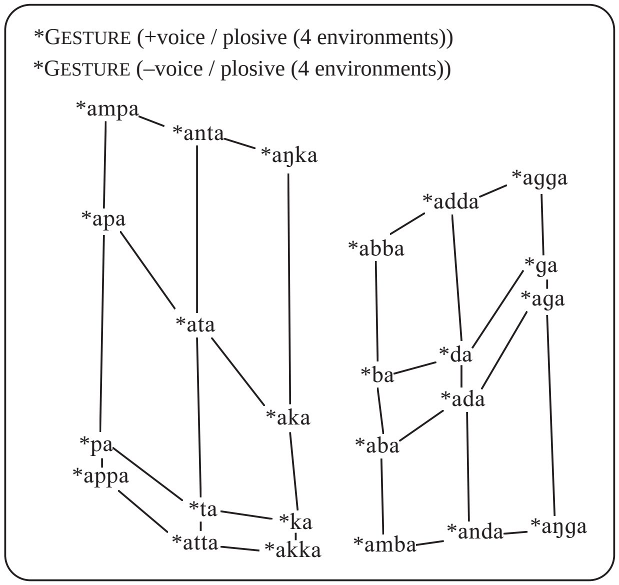
(11.21)

If the global map of (11.20) is correct, we expect to find a larger number of languages of the type pictured on the left of (11.21), than of the type on the right, i.e., we will find more languages with only voiceless stops than with only voiced stops: a tendency expected by the principle of (11.2c).

But there is a difference with the PARSE hierarchy seen before. Once that a gesture has been learned, its *GESTURE constraint falls to a low position in the overall constraint hiererchy. Because voicedness and voicelessness are implemented by very different gestures, the separations depicted in (11.21) are expected to be much more common than a grammar that allows a plosive inventory restricted to [b], [t], and [k]; this is different from faithfulness, because, say, learning of the perceptual feature value [+voice] automatically involves learning of the perceptual feature value [–voice].

11.10 Phonologization

The procedure of the previous section can be extended from single gestures to coordination. The phonetic hierarchy (11.20) would look differently for plosives in initial position (less easily voiced than elsewhere), geminate plosives (hard to voice), intervocalic plosives (easy to voice), and post-nasal plosives (hard to devoice). Hayes (1996) gives a tentative measure for the effort associated with all 24 cases, based on Westbury & Keating’s (1986) aerodynamic vocal tract model, which predicts the possibilities of voicing on the basis of transglottal pressure. Though Hayes uses a global effort measure, we should respect the fact that voicing and devoicing strategies use different gestures, so Hayes’ numbers can be pictured as follows, if we take into account the local-ranking principle:



(11.22)

(In Hayes' table, [pa] and [aba] tie, and the seven utterances at the bottom have zero effort.) With the algorithm of random reranking, subject to the local-ranking principle (which fixes the rankings that are expressed with lines in (11.22)), several universals follow automatically:

- There are languages with voiceless plosives in every position except post-nasally (see Pater 1996).
- There are languages which only allow voiceless geminates (Japanese).
- If voiced plosives are allowed initially, they are also allowed intervocalically and postnasally (if plosives are allowed there at all, of course).
- If voiced coronals are allowed, so are voiced labials (in the same position, and if labials are allowed at all).
- Et cetera.

Besides these near-universals, several tendencies can be predicted from the global height of the constraints in the phonetic map (11.22):

- The average *b is ranked lower than the average *p, so gaps at /p/ are more common than gaps at /b/.
- The average *g is ranked higher than the average *k, so gaps at /g/ are more common than gaps at /k/.

- The average *aNÇa is ranked higher than the average *aÇːa, so voiced geminates are more common in languages with geminates than post-nasal voiceless plosives in languages with post-nasal plosives.
- Et cetera.

The local-ranking principle may lead to a phonological constraint ranking that is very different from the global phonetic ranking in (11.21).

Dutch, for instance, allows /aŋka/ and not /aŋga/, although the map shows that the latter must be much less difficult (and Dutch has some voiced plosives). This is possible because the local-ranking principle allows the right half of the map to be turned counterclockwise by almost 90 degrees, without disturbing the fixed rankings, so that the quartet *agga >> *ga >> *aga >> *aŋga may dominate all other voiced-plosive constraints. These fixed rankings do predict that if a language does not allow /aŋga/ (but does allow post-nasal stops), it also disallows the near-universally worse /aga/, /ga/, and /agːa/. For Dutch, this prediction is borne out: the language simply lacks a /g/ phoneme. Thus, ranking (11.22) allows the generalization of four constraints to the simple *g.

The perfect mirror image of the Dutch example is found in Arabic and was dubbed “very striking” by Hayes (1996: 10). Arabic has the voiced geminate [bː] but not the voiceless geminate [pː], though the phonetic map shows that *abba is ranked much higher than *appa in a global effort space. Now, the left-hand side of the map (11.21) may be turned clockwise by almost 90 degrees, so that the quartet *ampa >> *apa >> *pa >> *appa may dominate all other voiceless-plosive constraints. These fixed rankings do predict that if a language does not allow /apːa/ (but does allow geminates), it also disallows the near-universally worse /pa/, /apa/, and /ampa/. For Arabic, this prediction is borne out: the language simply lacks a /p/ phoneme. Thus, ranking (11.22) allows the generalization of four constraints to the simple *[–voi / labial plosive].

A word must, then, be said about Hayes' solution for this phenomenon. To assess the “effectiveness” of the generalized constraint *p, he computes its average ranking number as the average of the ranking numbers of *appa (8), *pa (9.5), *apa (19), and *ampa (24)[19], as counted from the bottom in (11.22); the result is 15.1. The effectiveness of the generalized *b is 11.1, which is the average of the ranking numbers of *abba (18), *ba (13), *aba (9.5), and *amba (4). Now, Hayes' criterion of *inductive* (i.e., learnable) *grounding* identifies *p as grounded because its effectiveness is greater than that of all its simpler or equally simple “neighbours” *b, *t, *k, *[lab] and *[–voice]. In the same way, *b is not grounded because all of its neighbours *p, *d, *g, *[lab] and *[+voice] are more effective (a single one would have been enough to make it ungrounded). Hayes proposes that only grounded constraints make their way into the grammar.

There are several problems with Hayes' approach. First, it would mean that *[cor] and *[dors], which we can identify as *GESTURE (blade) and *GESTURE (body), do not

[19] This is an equivalent reformulation of Hayes' very different-looking algorithm.

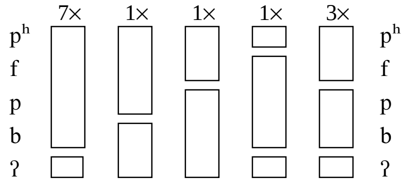
occur in the grammar because the more effective constraint *[lab] is a neighbour, an obviously undesirable result in the light of our example of place assimilation. Another serious problem with inductive grounding is that it is a procedure based on a global effort map, and, as such, only capable of deriving tendencies, not universals. For instance, the average ranking of the voicedness constraints in (11.21) is somewhat higher (12.7) than that of the voicelessness constraints (12.3), predicting that there are languages with exclusively voiceless plosives, and no languages with exclusively voiced plosives. Though this is a strong tendency with as few exceptions (Maddieson 1984: Alawa and Bandjalang) as the "near-universal" hierarchies *d >> *b (Proto-Indo-European; Maddieson 1984: Mixe, Cashinahua) and *g >> *d (Maddieson 1984: Acoma), the question is whether languages with a single series of plosives bother at all about making them voiceless or voiced; rather, they are likely not to show active devoicing at all, giving, on the average, a "lax voiceless" stop which does not violate any glottal *GESTURE constraint. The surprise of Westbury & Keating (1986) at finding that most languages with a single stop series have voiceless stops even in intervocalic position, whereas their model predicted that these should be more easily voiced than voiceless, may be due to an oversimplification in their model: even if the transglottal pressure is sufficiently high to allow voicing, a supraglottal closure should be accompanied by an active laryngeal closing gesture in order to withstand the voicing-adverse passive vocal-fold abduction caused by the rising intraglottal pressure, as seen in the comprehensive vocal-tract model of Boersma (1993, 1995, to appear). As an example (with unrealistic figures), consider the passive and active contributions to glottal widening in five obstruents (PCA = posterior cricoarytenoid, IA = interarytenoid)[20]:

sound	supra laryngeal closure	passive widening	active widening	muscle	total widening	acoustic result
p^h	closed	3 mm	3 mm	PCA	6 mm	aspirated
f	critical	2 mm	2 mm	PCA	4 mm	voiceless
p	closed	3 mm	1 mm	PCA	4 mm	voiceless
b	closed	3 mm	–3 mm	IA	0	voiced
ʔ	open	0 mm	–2 mm	IA	–2 mm	voiceless

(11.23)

In the column "total widening", we see the glottal strictures in the order of Ladefoged (1973). Gandour (1974), however, notes that the natural classes of initial obstruents in 13 tone-split rules in the histories of various Tai languages point to an order of [p^h, f, p, b, ʔ]. These tone splits are collected in the following table, where the natural classes are shown as rectangles:

[20] This simple example ignores supralaryngeal voicing gestures and the muscle-spindle reflex, which may bring the vocal folds together again after 20 ms of passive widening.

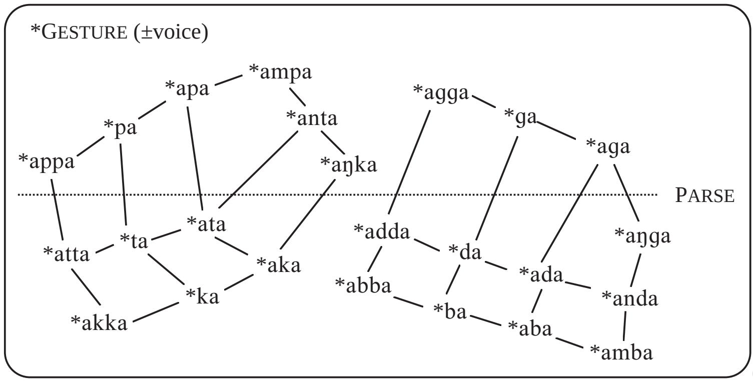


(11.24)

Gandour's solution to the disparity between Ladefoged's order and the Tai data involves a hierarchical ordering between the binary perceptual feature [±vibrating] and the multi-valued articulatory feature [glottal width]. Note, however, that sorting the five obstruents by their degree of *active* widening in (11.23) would also give the order [p^h, f, p, b, ʔ]. If there is some realism in my picture of passive glottal widening, this explains Westbury & Keating's surprise as well as the highly skewed distribution of homogeneously voiceless versus homogeneously voiced plosive systems: an active widening of 0 mm, as may be appropriate in a system without any voicing contrasts, leads to a total width of 3 mm for plosives, as can be seen in (11.23), and these may be considered "lenis voiceless". Thus, this skewed distribution cannot be taken as evidence of a universally ungrounded *[–voice] in systems that have to maintain a faithful voicing contrast in obstruents.

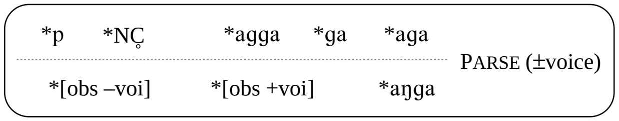
The conclusion must be that inductive grounding does not redeem Hayes' promise (1996: 5) that "we seek to go beyond mere explanation to achieve actual description". Rather, a much simpler strategy based on local ranking, which does not need a global effort map, correctly generalizes phonetic principles to phonological constraints. Just turn the symmetric diamond ◊ by almost 45 degrees in either direction.

As an example, consider a language which lacks /p/, /g/ (except post-nasally), and post-nasal voiceless plosives. Such a language should be able to exist according to the fixed rankings in (11.22). Without changing the ranking topology of this map, we can transform (11.22) into:



(11.25)

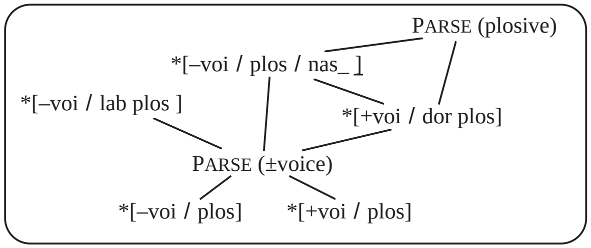
This can be simplified as



(11.26)

Thus, a simplification like (11.26) is allowed by the local-ranking principle. Cross-linguistically, languages seem to prefer these simplifications over drawing a dotted PARSE (±voice) line through the middle of (11.22). This effect cannot be explained by an asymmetry in the learning of voicing versus devoicing gestures, since the language of (11.25) obviously uses both of these gestures to a large extent. Rather, its success lies in the simplification itself: (11.26) needs fewer constraints than the average language that can be derived from (11.22) by restricted random reranking.

The remaining complexity with /g/ in (11.26) can be resolved by noting that if the language had a homogeneous *g constraint, there would be no way to parse a dorsal nasal-plosive sequence, as *[aŋka] is ruled out by *NC̥. Therefore, a strong PARSE (plosive) constraint may force the surfacing of [aŋga]. The following constraint system can handle this:



(11.27)

Note that if PARSE (plosive) is ranked at the top, *[–voi / plos / nas_] must dominate *[+voi / dor plos]; with the reverse ranking, underlying dorsal nasal-plosive sequences would show up as [aŋka] instead of [aŋga]: a minimal difference.

From the 24 articulatory constraints that we started with, only five remain, even in this relatively complex language. The reformulation of the *GESTURE constraints in (11.27) is explained below.

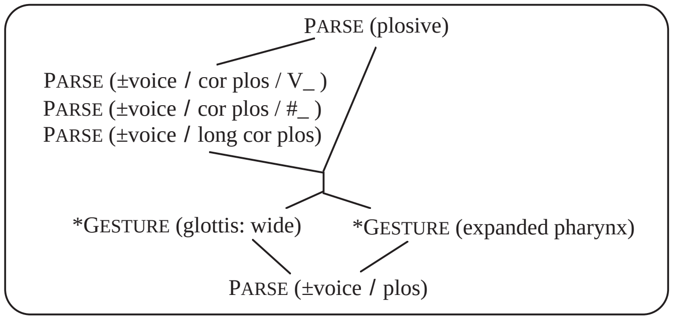
*11.11 Homogeneous *GESTURE or homogeneous PARSE constraints?*

The reader may have noticed that in §11.2 to §11.7, a homogeneous *GESTURE constraint was used to divide up interestingly ranked PARSE families, whereas in §11.10 a homogeneous PARSE constraint was used to divide up interestingly ranked *GESTURE families. Clearly, we cannot have both at the same time. In this section, I will solve this mystery and show that a phonetic map like (11.22) and a language like (11.25) can also be described with homogeneous *GESTURE constraints and varying PARSE constraints.

First, we can note that the articulatory constraints in (11.27) are explicitly shown in an "implementational" formulation: *[–voi / lab plos] >> *[–voi / cor plos] means that it is more difficult to make a labial plosive voiceless than to make a coronal plosive voiceless. Of course, this ranking can only be fixed if these formulations refer to the same degree of perceptual voicing for the labial and coronal cases. Thus, more effort is required for the implementation of the [aba] - [apa] contrast than for the [ada] - [ata] contrast, *given that the perceptual contrasts are the same in both cases.* Now, equal contrasts mean equal PARSE constraints (§8), so use of a homogeneous PARSE (±voice) constraint for all places is legitimate.

While the PARSE (voice) constraints are equally high for the various places, the *GESTURE constraints are not. The implementationally formulated constraint *[–voi / lab plos] is really something like *GESTURE (glottis width: 3 mm), and *[–voi / cor plos] is something like *GESTURE (glottis width: 2 mm), which is universally ranked lower, if the gesture is considered made from a state of phonation-friendly vocal-fold adduction.

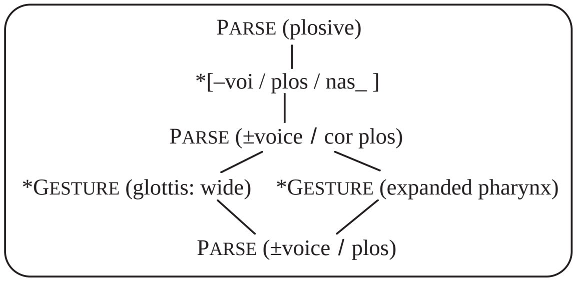
The voicing theory described above is perception-oriented. We can also devise an articulation-oriented theory, namely, one that says that only particular gestures are learned. For instance, if we learn to use the gesture "glottis width: 2mm" for the implementation of voiceless plosives, a /p/ will surface as less voiceless than a /t/. Likewise, with equal voicing gestures (pharynx expansion or so), a /g/ will come out as less voiced than /k/. Thus, PARSE (±voice) will be ranked less high for labials than for coronals, and PARSE (±voice) will be ranked less high for dorsals than for coronals. For post-nasal position, PARSE (±voice) will be ranked very low because post-nasal plosives with a 2mm glottis-width gesture will be voiced in such an environment, so that the perceptual contrast with the result of the expanded-pharynx gesture is very small. A working constraint hierarchy is:


PARSE (plosive)
PARSE (±voice / cor plos / V_)
PARSE (±voice / cor plos / #_)
PARSE (±voice / long cor plos)
*GESTURE (glottis: wide)
*GESTURE (expanded pharynx)
PARSE (±voice / plos)

(11.28)

This yields a language slightly different from (11.27): for labial and dorsal plosives and for post-nasal plosives, no voicing contrast exists, and neither of the gestures will be used for them. The automatic results may be something like [aba], [b̥a], [apːa], [ag̊a], [ka], [akːa], [amba], [anda], and [aŋga]; the minimal difference referred to below (11.27) does not exist. Note that it is no coincidence that both *GESTURE constraints in (11.28) seem to be on the same height: if PARSE (±voice) falls below one of them, the voicing contrast is neutralized, so that PARSE (±voice), whose ranking depends on contrast, falls further.

To sum up, the ranking in (11.27) expresses the articulatory problem of implementing the perceptual voicing feature faithfully, whereas (11.28) expresses the resistance against using articulatory gestures that do not result in good perceptual voicing contrasts. Real languages will allow both of these ideas to play a role in the grammar. For instance, the simplest constraint ranking for the languages in (11.27) and (11.28) would be



(11.29)

This uses only six constraints; both (11.27) and (11.28) needed one more. The ranking (11.29) expresses the following ideas: except for coronals, the voicing contrast in plosives, as implemented by a fixed pair of gestures, is so low that is too unimportant to maintain; for coronals, therefore, the contrast is maintained, except in post-nasal position, where the implementation of [–voice] is too difficult.

11.12 Licensing

In the previous section, we noted two different ways to phonologize articulatory constraints.

In the first interpretation, articulatory phonological constraints directly militate against certain fixed articulations. Typical examples are all the constraints propsed in §5, most notably *GESTURE.

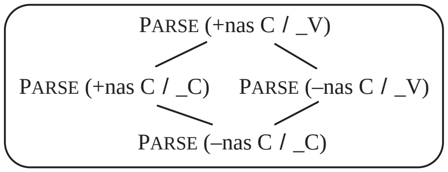
The second interpretation emerges from an interaction with perceptual requirements, and sees articulatory phonological constraints as constraints against the effort of implementing fixed perceptual results; their arguments, therefore, are perceptual features, not articulatory gestures. A typical example is *[–voi / plos / nas_]. Generally, we can call these constraints *licensing constraints*:

Def. licensing constraints: *[*f*: *v* / *env*]
> "The value *v* on a perceptual tier *f* is not implemented in the environment *env*." (11.30)

Licensing constraints seem the only way to reconcile a functional approach with a single system of features: articulatory gestures may be removed from the grammar. However, as seen in §11.11, these licensing constraints are *GESTURE constraints in disguise, and can be universally ranked with the procedures of §5. In §13.2, we will see that the more fundamental *GESTURE constraints are probably needed: the fact that most languages with voiceless nasals also have aspirated plosives; this can most easily be explained directly with the lowness of *GESTURE (spread glottis), and not with constraints like *[asp] and *[–voi / nasal]. Note that because of their grounding in more basic articulatory constraints, a functional ranking of licensing constraints such as *NÇ >> *VÇV is legitimate, but a similarly-looking ranking of assimilation constraints such as *[np] >> *[tp] is not: the former ranking may involve articulatory constraints only (as in 11.27), whereas the second ranking crucially involves an interaction with faithfulness constraints.

11.13 Nasal assimilation

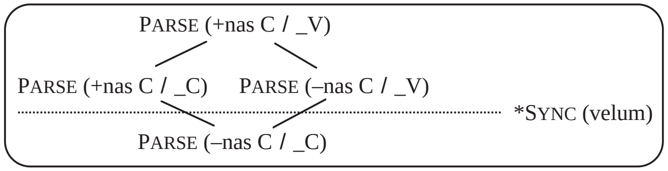
In Sanskrit, word-final plosives assimilate to following nasals: /ak+ma/ → [aŋma]. From the commonness considerations of §8.5 ([+nasal] is less common than [–nasal], because fewer constrasts can be made with [+nasal] sounds than with [–nasal] sounds), we can expect that this is a less offensive change than assimilation of [–nasal], as in /aŋ+pa/ → [akpa]. Also, we can expect that onset specification are stronger than coda specifications, as with our example of place assimilation. This leads to the following near-universal ranking:



(11.31)

The presence of “C” in the argument of PARSE makes this an explicitly segmental formulation, a shorthand for PARSEPATH (nasal & root) or PARSEPATH (nasal & timing), though it could be replaced with a formulation involving higher prosodic units (by replacing “C” with “μ” or “σ”, for instance).

According to the local-ranking principle, all rankings not shown with straight lines in (11.31) are free. Sanskrit makes the following choice:



(11.32)

The relevant articulatory constraint is not from the *GESTURE family, but from the *SYNC family, and militates against a velar movement inside a CC cluster.

We expect the following typology:

(a) Nothing assimilates (most languages).
(b) Plosives assimilate to a following nasal (Sanskrit).
(c) Coda consonants assimilate their nasality to the following [±nas] consonant (spreading of [–nas] is found in the North-Germanic sound change /ŋk/ → /kː/).
(d) Plosives assimilate to a nasal on either side. (11.33)

There are only four (not six) possibilities, because (c) and (d) both already satisfy *SYNC (velum). Note that none of the four violates FILL (+nas).

The typology (11.33) is equivalent to the following set of independent implicational universals for nasal spreading within consonant clusters:

(a) If [–nas] spreads, so does [+nas].
(b) If [+nas] spreads rightward, it also spreads leftward. (11.34)

11.14 Conclusion

Starting from a typological interpretation of the local-ranking principle, we derived a successful strategy for simplification of the grammar:

The functional view of the phonologization of functional constraints

> “From all the grammars allowed by the local-ranking principle, languages tend to choose a grammar in which many constraints can be generalized over their arguments or environments.” (11.35)

12 Correspondence: segmental integrity versus featural autonomy

In §8, I proposed a large number of faithfulness constraints. The workings of some of these are likely to overlap. Though all these constraints can be defended by invoking functional principles, phonology may be special in that it allows only a subset of them to play a role in language. In this section, we will compare two hypotheses for a reduction of the number of necessary faithfulness constraints:

a. Segmental integrity:

> "All featural faithfulness relations are transferred through the segment, which is the complete bundle of simultaneously present features." (12.1)

The typical representative of this approach is the "linear" *correspondence theory* of McCarthy & Prince (1995), who used the following constraints:

- MAX-IO: **if** the input contains a segment, **then** this segment should also be in the output (like our PARSE, but for segments, not features).
- IDENT-IO (*f*): **if** the input segment **and** the corresponding output segment both contain the feature *f*, **then** the two values of this feature should be equal (like our *REPLACE).

For instance, IDENT-IO (voice) is satisfied if the value for the feature [voice] in the input is equal to the value for [voice] in the *corresponding segment* in the output, and it is violated if these values are unequal. But if either the input or the output does *not* contain the bearing segment, the constraint is *not* violated.

b. Featural autonomy:

> "Every specified feature has its own faithfulness constraints, which try to bring it to the surface." (12.2)

Archangeli & Pulleyblank (1994) simply state that "the notion of segment is both inadequate and superfluous" and that phonology works with features, nodes, and links (though they do incorporate a root tier). Our account of §8 also brought up featural faithfulness as predominantly autonomous, disregarding correspondence through segments, controlling faithfulness with constraints like:

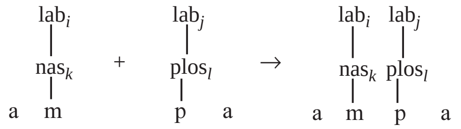
- PARSE (feature: *x*): **if** the input contains the feature value *x*, **then** *x* should also be in the output.

In the examples of §11, however, I tacitly handled the faithfulness of features by using segments as their domains. In the following, we will relieve this tension and consider the relative merits of the linear and the autosegmental approaches.

12.1 Perception is segmental

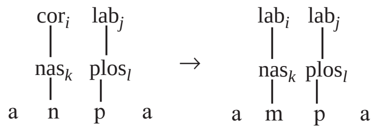
With a distinction between articulation and perception, there is a very simple solution to the everlasting problem of segmental versus autosegmental processes: the consonant cluster in [ampa] contains a single articulatory labial gesture, but is heard as containing two separate instances of a perceptual feature [labial]. Thus, we can evaluate our faithfulness constraints via linearly ordered segments, and still understand that assimilation is spreading of an articulatory gesture. In this way, we have the best of both worlds.

In §11, we assumed the segmental interpretation of faithfulness to our advantage. For instance, we did not mark the concatenation /ʊm/ + /poːtə/ → /ʊmpoːtə/ with a violation of PARSE (labial). Thus, the correspondence in this example is like this:



(12.3)

Another process discussed earlier, the assimilation /an+pa/ → [ampa], can be seen as a replacement of [coronal] with [labial] on the perceptual place tier, but only if we represent the two [labial] feature values of the output as separate:



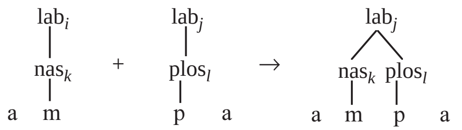
(12.4)

Now, it might just be the case that this is the correct rendering of the perceptual score, and that autosegmental representations respecting the OCP are limited to the articulatory score. Such a hypothesis would express a nice functional correlate of the tension between segmental and autosegmental phenomena: there is a single lip gesture, but separate labial sounds.

But in those cases where features are not neatly lined up, it is often difficult to even count the number of segments in an utterance. For instance, does the usual pronunciation of *tense* with an intrusive stop lead to four or five segments on the surface? And there are several other problems with the segmental approach.

12.2 OCP-driven epenthesis

I argued earlier (§11.8) that several place cues can collectively contribute to the perception of a single value of the perceptual place feature. For instance, transition and burst together will make us hear a single instance of [labial] in [apa], microscopically [[ap˺_pa]]. Such a single labial would surely also be perceived in a prenasalized stop as in [a$^{\text{m}}$pa]. But a homorganic cluster as in [ampa] is, in many languages, by far the most common nasal-plosive sequence, and it would be advantageous to the listener to hear them as a cluster with a single place, in accordance with what happens gesturally. Thus, we could re-represent (12.3) with autonomous features as



(12.5)

In a segmental approach, no constraint at all is violated: MAX-IO is satisfied because all underlying segments appear in the output, and the resulting [m] corresponds maximally faithfully with underlying /m/. In an autosegmental approach, by contrast, we have only *one* labial feature left in [mp], whereas the two underlying segments /m/ and /p/, coming from two different morphemes with separate lexical representations, contribute *two* labial specifications. Therefore, we have a violation of PARSE (labial), and the utterance is indistinguishable from an utterance with a single underlying dorsal gesture (i.e., a tautomorphemic homorganic nasal-plosive cluster). This violation of PARSE is a faithfulness problem, so we expect interactions with other constraints, such as FILL.

As an example, consider the following data from Geleen Limburgian, where the diminutive suffix /kə(n)/ shows epenthesis of [s] when attached to a stem that ends in a dorso-velar[21]:

pop (pl. popə) ‘doll’	pøpkə	
lɑm̄p (pl. lɑm̄pə) ‘lamp’	læm̄(p)kə	
kom̄p (pl. kǿm) ‘bowl’	kǿmkə	
bɔūm (pl. bœ́ym) ‘tree’	bœ́ymkə	
dû:f (pl. dū:və) ‘pigeon’	dŷ:fkə	
ʃtʀɔ̂:t (pl. ʃtʀɔ̂:tə) ‘street’	ʃtʀœ̂:cjə ([c, ɲ] = palatalized alveolar)	
bɛt (pl. bɛdə) ‘bed’	bɛcjə	
mɑn̄ (pl. mǽn or mɑ́nə) ‘man’	mǽnkə (place assimilation forbidden)	
bɑl̄ (pl. bǽl) ‘ball’	bǽlkə	
kɑʀ̄ (pl. kǽʀ) ‘cart’	kǽʀkə	
jɑs (pl. jæs) ‘coat’	jæskə	
ɦū:s (pl. ɦū:zəʀ) ‘house’	ɦøskə (irreg. vowel)	
kes (pl. kestə) ‘chest’	keskə	
kɑɲ̄c (pl. kǽɲ) ‘side	kæɲ̄cjə	
ɦoɲ̄c (pl. ɦǿɲ) ‘dog’	ɦǿɲcjə	
wóɲ (pl. wóɲə) ‘wound’	wǿɲcjə	
vœʃ (pl. vœʃə) ‘fish’	vœʃkə	
bǽlʃ (pl. bæl̄ʒə) ‘Belgian’	bǽlʃkə	
blɔk (pl. blœk) ‘block’	blœkskə	
ɦɛk (pl. ɦɛgə) ‘hedge’	ɦɛkskə	
plɑŋ̄k (pl. plæŋ̄k) ‘plank’	plæŋ̄kskə	
dɛŋ̄k (pl. dɛ́ŋəʀ) ‘thing’	dɛ́ŋskə	
ɔ́ux (pl. ɔūɣə) ‘eye’	œ́yxskə	
lé:x (pl. lé:xtəʀ) ‘light’	lé:xskə	(12.6)

In Correspondence Theory, this epenthesis cannot be represented, because a violation of DEP-IO (= FILL) is always worse than no violation at all, independently of the relative rankings of MAX-IO, DEP-IO, and IDENT-IO (place):

/ŋ+k/	MAX-IO	DEP-IO	IDENT-IO(place)
*☞ [ŋk]			
[ŋsk]		*!	

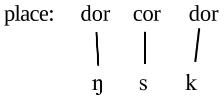
(12.7)

In the purely autosegmental approach, PARSE (dorsal) may be strong enough to force epenthesis:

/ŋ+k/	PARSE (dorsal)	FILL (sibilant)
[ŋk]	*!	
☞ [ŋsk]		*

(12.8)

With the epenthesis of [s], PARSE (dorsal) is no longer violated, because the two dorsal specifications of /ŋ/ and /k/ are now separated on the perceptual place tier:



(12.9)

[21] The highly productive diminutive morpheme is expressed as: umlaut (fronting of back vowels); softening (leaving only the sonorant from underlying sonorant + voiced plosive sequences); tone change (changing an underlying circumflex tone into an acute, but not before a voiceless consonant); and the suffix /-kə(n)/.

Though some theories (e.g., McCarthy 1988) might still consider the two unary [dor] features adjacent because there is no conflicting value on the same tier, we cannot represent them with one specification without getting the ‘gapped’ representation that Archangeli & Pulleyblank (1994) militate against. Going back to the fundamentals (i.e., function), we see that there is perceptual separation on the place tier: there are no separate perceptual coronal and dorsal tiers[22].

12.3 Horizontal and vertical correspondence

In §6, we handled the acoustics-to-perception faithfulness of an utterance consisting of a single feature; in such a case, the question what input feature values correspond to what output feature values, has a simple answer. But if an utterance contains multiple simultaneous feature values across tiers and multiple ordered feature values within tiers, the correspondence question becomes more complicated. The general idea is that it is favourable for the listener to perceive a set of acoustic cues or perceptual features that often occur together, as a single feature.

One aspect of this occurring together is the grouping of simultaneously occurring features, discussed in §8.11. If the “vertical” path constraints are strong, we can expect segment effects.

The other aspect of occurring together is the grouping of acoustic cues or feature values that occur after one another. If cue A is usually followed by cue B, they may be recognized as a single feature value; I used this idea in §11.8 to account for the additivity of environmental conditions that was necessary to explain the data of Dutch place assimilation. If the “horizontal” temporal identity constraints are strong, we can expect autosegmental effects.

In OT, every conflict is resolved by a constraint, so the conflict between the segmental representation (12.3) and the autosegmental representation (12.5) must be handled by a constraint as well. I propose the following pair of listener constraints for the temporal correspondence between the acoustic input and the perceptual result

Def. OBLIGATORYCONTOURPRINCIPLE-AC (*f*: *x*; *cue*$_1$, *m*, *cue*$_2$)
“A sequence of acoustic cues *cue*$_1$, *cue*$_2$ with little intervening material *m* is heard as a single value *x* on the perceptual tier *f*.” (12.10)

Def. NOCROSSINGCONSTRAINT-AC (*f*: *x*; *cue*$_1$, *m*, *cue*$_2$)
“A sequence of acoustic cues *cue*$_1$, *cue*$_2$ with much intervening material *m* is not heard as a single value *x* on the perceptual tier *f*.” (12.11)

[22] Long-distance “OCP effects” that forbid the use of the same articulator twice within a domain, are due to a *Repeat constraint that works exclusively with articulatory gestures (§14.2, Boersma fc. b).

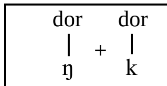
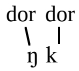
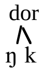
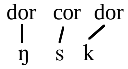
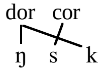
These constraints promote a maximally easy perceptual organization. The more often the cues occur together, the greater the chance that they are perceived as a single feature; this is true for simultaneity (segmentalism) as well as temporal ordering (autosegmentalism). They can be universally ranked by such things as temporal distance of the cues, rate of occurrence within morphemes versus across morphemes, etc. For instance, I believe that OCP-AC (place: labial; [[p˺_p]]) is ranked so high that the plosive in [apa] is represented almost universally with a single perceptual place specification. Not much lower would be the constraint that forces us to hear a geminate consonant as having a single place value: OCP-AC (place: labial; [[p˺__p]]). Lower still would be the constraint against hearing homorganic nasal-plosive clusters as having a single place value: OCP-AC (place: labial; [[m_p]]). The NCC-AC constraint would be ranked in the other direction: the more intervening material, the higher the ranking.

The phonological counterparts of the acoustics-to-perception constraints (12.10) and (12.11) have to refer to perceptual features, not acoustic cues. They can be stated as:

Def. OCP (*f*: *x*, *y* / *env*)
“A sequence of values *x* and *y* on the perceptual tier *f* are heard as a single value in the environment *env*.” (12.12)

Def. NCC (*f*: *x*, *y* / *env*)
“A sequence of values *x* and *y* on the perceptual tier *f* are not heard as a single value in the environment *env*.” (12.13)

The following tableau evaluates the Limburgian case again:

	OCP (place: dor / nas | plosive)	NCC (place: dor / nas | [s] | plosive)	PARSE (dor)	FILL (sib)
	*!			
			*!	
☞ 				*
		*!	*	*

(12.14)

The constraints OCP (place: dor; side | [s] | burst) and NCC (place: dor; side | _ | burst) are probably ranked quite low. We also see that the NCC constraint in this tableau is superfluous: the branching [dor] would be ruled out because it violates PARSE (dor). Note that association lines cross in the fourth candidate, for there is a single perceptual place tier.

Consider now the English past tenses /hɛd-ɪd/ ‘headed’ versus /kæn-d/ ‘canned’. Epenthesis is forced only between homorganic plosives:

cor + cor d d	OCP (place: dor; trans | _ | burst)	PARSE (cor)	FILL (σ)	OCP (place: dor; side | _ | burst)
cor cor d d	*!			
cor d d		*!		
☞ cor cor d ɪ d			*	
cor d ɪ d		*!	*	

(12.15)

Between a nasal and a plosive, no epenthesis occurs:

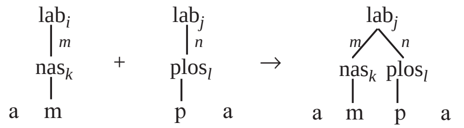
cor + cor n d	OCP (place: dor; trans | _ | burst)	PARSE (cor)	FILL (σ)	OCP (place: dor; side | _ | burst)
☞ cor cor n d				
cor n d		*!		*
cor cor n ɪ d			*!	
cor n ɪ d		*!	*	

(12.16)

The OCP-based account described here manages the data of Limburgian and English well and does the typological prediction that if heteromorphemic homorganic nasal-plosive clusters undergo epenthesis, then plosive-plosive clusters undergo epenthesis as well.

But there is still a problem. There seems to be a segmental intuition that the perceptual loss of identity of the first /d/ in /d+d/ → /dː/ is greater than the loss of identity of /n/ in /n+d/ → /nd/. It would be nice if we could express this intuition with a variation in the ranking of a faithfulness constraint, instead of burdening the listener with a dual-coronal representation of /nd/.

We can respect the perceptual OCP (place) in /nd/ if we notice that no identity is lost on the combined place and nasal tiers. We can rewrite (12.5) as



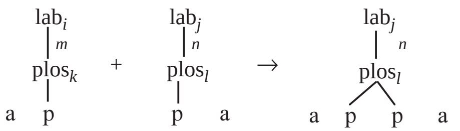
(12.17)

On the combined labial-nasal tiers, correspondence is between feature combinations, not between the separate features: it is (nas lab)$_m$, not (nas$_k$ lab$_i$), and the former is preserved in the output, though PARSE (lab) is still violated. With a homogeneous unviolated OCP, violations, (12.16) becomes:

cor + cor n d	OCP (place: dor)	PARSE (nas & cor)	FILL (σ)	PARSE (cor)
cor cor n d	*!			
☞ cor n d				*
cor cor n ɪ d			*!	
cor n ɪ d			*!	*

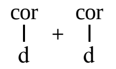
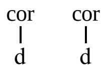
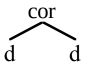
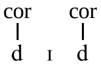
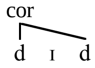
(12.18)

The analogue of (12.17) for plosive-plosive clusters is:



(12.19)

The same constraint system as in (12.18) will now have to evaluate /hɛd+d/:

	OCP (place: dor)	PARSE (plosive & cor)	FILL (σ)	PARSE (cor)
	*!			
		*!		*
☞ 			*	
		*!	*	*

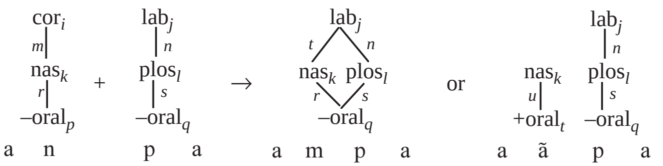
(12.20)

Technically, we could have done the job with the near-universal ranking PARSE (cor / plosive) >> PARSE (cor / nasal), derived earlier (8.39) from considerations of perceptual confusability, but this is a coincidence that we probably cannot use for all cases.

The somewhat unsettling problem with (12.18) is that even for the seemingly trivial case of /m+p/ → /mp/ we need a ranking like PARSE (nas & lab) >> PARSE (lab), against the functional ranking principle of §8.10.

For the assimilation case (12.4) of /an+pa/, we have to understand why the candidate [ampa] is better than [aãpa]. In the segmental account, this is because the non-orality (or consonantality) of /n/ is preserved in [m] but not in [ã][23]. In the autosegmental account, however, non-orality is not even preserved in [ampa], because this feature is shared with [p]:

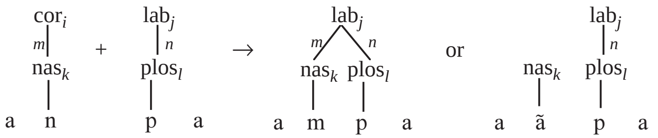
[23] We cannot yet say that consonantality is not subject to the OCP because it belongs in the root node. Such things have to be derived, not posited, in a functional phonology.



(12.21)

We see that both candidates violate PARSE (cor), PARSE (nas & cor), PARSE (–oral & cor) (though not shown, this path must be present), and PARSE (–oral), and that [ampa] also violates FILL (nas & lab), while [aãpa] violates FILL (+oral), PARSE (+nasal & –oral), and FILL (+nasal & +oral). Note that this example shows that PARSE (–oral) is not necessarily the same as FILL (+oral) in autosegmental perceptual phonology. Actually, however, FILL (+oral) is not violated in this case, since [ã] must share its [+oral] value with [a]. The real conflict, therefore, is between FILL (nas & lab) on the one hand, and PARSE (+nasal & –oral) and FILL (+nasal & +oral) on the other. Languages that highly estimate the preservation of nasal non-orality, will end up with [ampa]; those that do not like to hear a labial nasal where it is not specified, will end up with [aãpa]; in both cases, cross-tier faithfulness constraints decide the issue. If the /p/ in (12.21) were a fricative, there would only be one change: [ampa] would not violate PARSE (–oral).

The process /an+pa/ → [ampa] can be represented with less violation of correspondence than in (12.21). Though TRANSMIT (place) may be ranked lower than *REPLACE (cor, lab), this situation may well be reversed for the combined feature [place × nasal] (§8.8): TRANSMIT (place × nasal) may be ranked higher than *REPLACE (cor & nas, lab & nas) because within the combined [place × nasal] space, [cor & nas] and [lab & nas] are relatively close together. Instead of (12.21), we get



(12.22)

The input [cor & nas]$_m$ now corresponds to the output [lab & nas]$_m$. The main candidates are evaluated (without any constraints involving [–oral]) according to:

/an+pa/	PARSE (nasal)	*GESTURE (blade)	TRANSMIT (place × nasal / _C)	*REPLACE (nas cor, nas lab / _C)
anpa		*!		
ampa (12.21)			*!	
☞ ampa (12.22)				*
aãpa			*!	
apa	*!		*	

(12.23)

(The candidate [apa] loses as long as PARSE (nasal) dominates *GESTURE (velum).) This evaluation, involving the correspondence in (12.22), is the interpretation of the example of §8.5 and §11.4, which involved the less accurate constraint *REPLACE (cor, lab / _C). We see that strong "vertical" constraints like TRANSMITPATH can force segment-like behaviour: the faithfulness constraints in (12.23) look suspiciously like MAX-IO and IDENT-IO (place), but note that we can only use the latter pair of constraints if we do not consider the [ã] in [aãpa] to be a segment (it could be transcribed as [ãpa], with [ã] corresponding to both /a/ and /n/; see §12.5); with our more restricted path constraints, such a stipulation is unnecessary.

Conclusion: we need path constraints whose featural coherence is greater than that of autonomous features, but smaller than that of a segment.

An inherent problem in autosegmentalism. In the autosegmental approach, subtraction may sometimes be evaluated as addition.

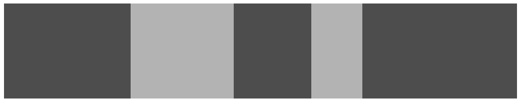
The process /dap/ → [dãp] violates FILL (nasal) and FILL (nasal & vowel), whereas /dam/ → [dãm] violates only FILL (nasal & vowel). Therefore, the former process is always worse: insertion of a marked value or privative feature is worse than spreading.

Likewise, the process /dãp/ → [dap] violates PARSE (nasal) and PARSE (nasal & vowel), /dãm/ → [dam] violates only PARSE (nasal & vowel). Therefore, the former process is always worse: deletion of a marked feature is worse than deletion of its association line only.

The symmetry seen in the /dap/ and /dam/ cases is related to the idea that the perceptual contrast between [dap] and [dãp] is larger than that between [dam] and [dãm], a difference that can be ascribed to the general process of lateral inhibition (a nasality contrast is less easy to hear next to a nasal). An asymmetry is due to the markedness of [+nasal] (§8.5): the process /dap/ → [dãp] must be less bad than /dãp/ → [dap], suggesting that for marked feature values, PARSE violations are worse than FILL violations.

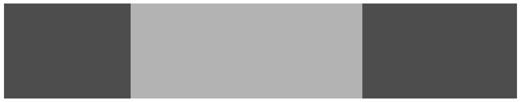
But problems arise with /pap/ and /mam/. Let us assume that the distinction between [pap] and [pãp] is larger than the distinction between [mam] and [mãm].

The process /pap/ → [pãp] violates FILL (nasal) and FILL (nasal & vowel), whereas /mam/ → [mãm] violates PARSE (nasal) and FILL (nasal & vowel). The violation of PARSE (nasal) can be illustrated with the following metaphor. Suppose we start with a sequence of dark-light-dark-light-dark rectangles:



(12.24)

If we paint the middle rectangles in a light shade of grey, one dark rectangle is lost:

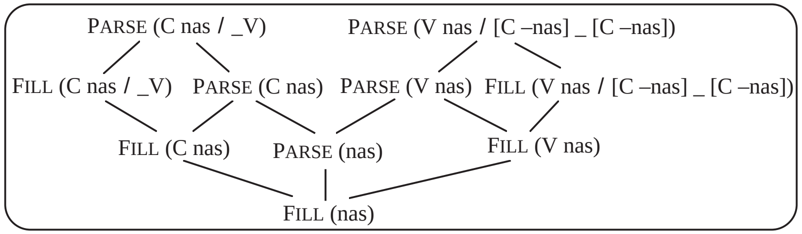


(12.25)

As we see, however, one light rectangle is also lost. Adding nasality to the vowel in [mam] thus violates PARSE (nasal). Now, if the [pap] - [pãp] distinction is larger than the [mam] - [mãm] distinction, the change /pap/ → [pãp] is more offensive than the change /mam/ → [mãm], so that FILL (nasal) must dominate PARSE (nasal).

The process /pãp/ → [pap] violates PARSE (nasal) and PARSE (nasal & vowel), whereas /mãm/ → [mam] violates FILL (nasal) (like going from (12.25) to (12.24)) and PARSE (nasal & vowel). If the latter process is less bad than the former, PARSE (nasal) must dominate FILL (nasal), so there is a contradiction with the previous pair.

We can get out of the predicament only by assuming such rankings as PARSE (+nas & vowel / [–nas & cons] _ [–nas & cons]) >> PARSE (+nas & vowel / [+nas & cons] _ [+nas & cons]), and the same ranking for FILL, together with low rankings of PARSE (nasal) and FILL (nasal). So, we can finally replace the naive nasality faithfulness rankings of §8 with an accurate grammar (cf. 11.31) that handles all cases of the spreading of nasality to adjacent plosives and vowels:



(12.26)

12.4 Floating features

The faithfulness of floating features cannot be represented at all within Correspondence Theory, because these features are by definition not underlyingly connected to a segment, which makes IDENT-IO insensitive to them. This was already recognized by McCarthy & Prince (1995); the solution they suggest is the generalization of the MAX and DEP constraints to autosegmental features. However, this would involve not just a simple generalization of Correspondence Theory to the featural domain, because some constraint families will show considerable overlap: the separate need for the IDENT-IO family (together with MAX-IO) will be severely reduced as a consequence of the existence of the MAX (*feature*) family (though in a *comprehensive* approach we may need all of them).

Zoll (1996) explicitly evaluates correspondence through output segments, even for floating features (if these dock onto segments). For instance, Zoll argues that in Inor, the verb /kəfəd/ plus the masculine floating affix [round], which together give [kəfʷəd] (because [round] will dock on the rightmost labial or dorsal segment), should be analysed as if both underlying /f/ and underlying [round] correspond to the output segment [fʷ]. This would lead to the following evaluation:

/kəfəd/ + [round]	MAX (SEG)	MAX (SUBSEG)	IDENT (F)
☞ kəfʷəd			
kəfəd		*!	
kəfʷəz			*!
kəfʷə	*!		

(12.27)

Several remarks are in order.

First, Zoll holds the underlying /f/ to correspond to surface [fʷ] without violating IDENT(F) (a constraint that requires that the featural make-up of corresponding segments should be the same), because Zoll “follow[s] the proposal of Orgun 1995 and 1996 in assessing violations of IDENT(F) only in cases of absent or differing specifications, but not when the output correspondent is more specified than the input”. As we have seen in §8.5, we can explain such an asymmetry between input and output without such stipulations: it follows directly from the markedness of the feature [round] in the average utterance and the listener’s optimal recognition strategy, which leads to the near-universal ranking PARSE (round) >> FILL (round). In other words, it is worse to replace /fʷ/ with [f] than to replace /f/ with [fʷ]. Thus, the segment-based constraint IDENT(F) is superfluous.

More important is the fact that both (sub)segmental MAX constraints can be replaced with featural correspondence constraints. In the winning candidate, PARSE (round) is satisfied. The only problem with [kəfʷəd] is that [round] has been linked with a labial consonant; but this is less bad than linking [round] with the coronal consonant (in Inor), although that is final. The complete constraint system gives:

/kəfəd/ | [round]	PARSE (cor)	FILL (noise)	FILL (rnd &cor)	PARSE (rnd)	FILL (rnd & lab) FILL (rnd & dor)	*SHIFT (σσ)	*SHIFT (σ)
☞ kəfʷəd					*		*
kʷəfəd					*	*!	*
kəfəd				*!			
kəfʷəz		*!					*
kəfədʷ			*!				
kəfʷə	*!						*

(12.28)

I misused the constraint PARSE (cor) for assessing the loss of the final segment in [kəfʷə]; the question of featural or temporal segmentality (i.e., whether we should have taken PARSE (root) or PARSE (timing) instead of PARSE (cor)) is independent from the question of featural correspondence discussed here. The *SHIFT family evaluates the suffixal specification of [round], as suggested by the “|” in the representation; note that this constraint is vacuously satisfied if the floating [round] does not surface, and that it rates [kʷəfəd] as worse than the winner (§8.13), it will be replaced with a continuous family.

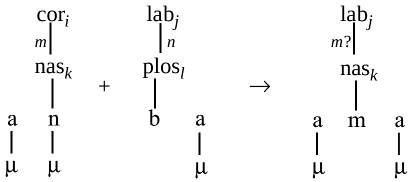
12.5 Fusion

In the simple fusion /n+b/ → [m] (e.g., Tagalog /maŋ+bili/ → [mamili] ‘buy’), one segment disappears. First, assume that the deleted segment is /b/. In Correspondence Theory, this means that there is one violation of MAX-IO. This must be brought about by a higher-ranked constraint, say the anti-cluster constraint *CC. However, because underlying /n/ now corresponds to surface [m], we also have a violation of IDENT-IO(place). In that case, as (12.29) shows, the candidate [n] would always be better, independently of the ranking of MAX-IO and IDENT-IO (place). The second strategy would be to assume that the deleted segment is /n/. In this case, the output candidate [m] must correspond to the input /b/, violating IDENT-IO (nasal). Correspondence Theory would then predict the output [b], independently of the ranking of MAX-IO and IDENT-IO (nasal). Thus, the output [m] cannot possibly win, unless it corresponds to both input segments:

/n_i+b_j/	*CC	MAX-IO	IDENT-IO (place)	IDENT-IO (nasal)
[$n_i b_j$]	*!			
[m_i]		*	*!	
☞ [n_i]		*		
[m_j]		*		*!
☞ [b_j]		*		
[m_{ij}]		?	?	?

(12.29)

To represent the fusion /n+b/ → [m] correctly, Correspondence Theory would have to be extended appreciably, because it is no trivial matter to decide whether MAX-IO or IDENT-IO are satisfied or not in (12.29). The autosegmental account, by contrast, views features as independent of segments. The fusion process is shown as follows:



(12.30)

PARSE (coronal) and PARSE (plosive) are violated, but the universal frequency-based rankings of PARSE (labial) above PARSE (coronal) and PARSE (nasal) above PARSE (plosive) guarantee the output [m]:

/n+b/	*CC	PARSE (μ)	PARSE (labial)	PARSE (nasal)	PARSE (coronal)	PARSE (plosive)
[nb]	*!					
☞ [m]		*				*
[n]		*	*!			
[b]		*		*!	*	
[d]		*	*!	*!		

(12.31)

So, fusion is most easily described with PARSE constraints for fully autonomous features.

12.6 Phonetic substance of epenthesis

An assessment of the featural content of epenthesized segments is impossible within Correspondence Theory: IDENT-IO is insensitive to any extra segment in the output, exactly because the epenthesized segment has no correspondent in the input. In the autosegmental account, the most faithful epenthesis candidate (i.e. the one that violates the least severe FILL constraint) will be the one that adds the fewest features or paths to the output (unless, of course, the epenthesis is meant to separate identical elements, as in §12.2).

12.7 Subsegmental satisfaction by segmental deletion

As we can see from its definition, IDENT-IO can be satisfied by means of the deletion of a segment. An example of this may be found in Limburgian, where the /n/ in the masculine singular ending of articles and adjectives is only licensed by following laryngeal consonants and coronal stops: /dən/ 'the' + /dāːx/ 'day' becomes /dəndāːx/ 'the day' (likewise: dən-tīːt 'the time'), but /dən/ + /bɛ̂ːʀ/ 'bear' becomes [dəbɛ̂ːʀ]: rather than deleting only the coronal gesture, which would give *[dəmbɛ̂ːʀ], the whole segment is deleted (likewise: də-ʃtɛīn 'the stone').

– ***Segmental account***. Apparently, IDENT-IO outranks MAX-IO (we use an ad-hoc nasal-consonant (NC) homorganicity constraint to make [nb] ill-formed):

/dən+dāːx/	IDENT-IO(place)	NC-HOMORGANIC	MAX-IO (ə<u>n</u>C)
☞ dəndāːx			
dədāːx			*!

(12.32)

/dən+bɛ̂ːʀ/	IDENT-IO(place)	NC-HOMORGANIC	MAX-IO (ə<u>n</u>C)
dənbɛ̂ːʀ		*!	
dəmbɛ̂ːʀ	*!		
☞ dəbɛ̂ːʀ			*

(12.33)

Thus, in this case, the segmental account seems appropriate. We will now see that all attempts to describe the phenomenon with the assumption of featural autonomy, are problematic.

– *Autosegmental account*. If autosegments were autonomous, constraint satisfaction by deletion could not occur: independently of the ranking of the two PARSE constraints involved, *[dəmbɛ̂ːʀ], which violates PARSE (coronal), would always be a better candidate than [dəbɛ̂ːʀ], which violates both PARSE (coronal) and PARSE (nasal). To solve this problem, we could put the constraint *GESTURE (velum) in between the two PARSE constraints:

Version 1 (covertly segmental)	PARSE (cor/_V)	*GESTURE (blade)	PARSE (cor/_C)	*GESTURE (velum)	PARSE (nas/ə_C)
☞ dəndāːx		*		*	
dədāːx		*	*!		*
dənbɛ̂ːʀ		*!		*	
dəmbɛ̂ːʀ			*	*!	
☞ dəbɛ̂ːʀ			*		*

(12.34)

All rankings in this tableau are crucial: any other ranking of the same constraints would give a different result. The idea is that the inviolable parsing of the place features of the onset consonant (/d/) forces the tongue-tip gesture and thereby *licenses* the surfacing of coronality in the nasal consonant (because the two segments share the same gesture). A nice result, and we can relate the rarity of this phenomenon to the critical ranking that is needed: even if we assume that PARSE (cor / _V) is universally undominated, there are 24 possible rankings of the four remaining constraints, and only one of those rankings produces the correct result. Too bad there’s a flaw. In a truly autosegmental framework, [dəndāːx] actually violates PARSE (coronal), according to the OCP; in §12.2, it was proved that Limburgian considers a homorganic nasal-plosive sequence to have a single [coronal] specification. But [dəndāːx] does not violate the segmental-integrity constraint PARSE (nasal & coronal), which is part of the specification and requires the co-occurrence of two perceptual features:

Version 2 (illogical)	PARSE (cor/_V)	*GESTURE (tongue tip)	PARSE (cor)	PARSE (nas & cor)	*GESTURE (velum)	PARSE (nas/ə_C)
☞ dəndāːx		*	*		*	
dədāːx		*	*	*!		*
dənbɛ̂ːʀ		*!			*	
dəmbɛ̂ːʀ			*	*	*!	
☞ dəbɛ̂ːʀ			*	*		*

(12.35)

Rather strange in this proposal, however, is the crucial ranking of the more general PARSE (nas) below the more specific PARSE (nas & cor), allowing a *GESTURE constraint to intervene, contrary to the universal logical ranking defended in §8.10. It seems we’ll have to use a constraint against [m]: not against [m] in this position in general ([əmb] is an otherwise licit sequence), but against [m] where there is no underlying labial nasal; in other words, FILL (nas & lab), which is unviolated:

Version 3	PARSE (cor/_V)	*GESTURE (tongue tip)	PARSE (cor)	FILL (nas & lab)	PARSE (nas/ə_C)	*GESTURE (velum)
☞ dəndāːx		*	*			*
dədāːx		*	*		*!	
dənbɛ̂ːʀ		*!				*
dəmbɛ̂ːʀ			*	*!		*
☞ dəbɛ̂ːʀ			*		*	

(12.36)

The two “nasal” constraints have been crucially reranked. An undominated FILL (nasal & dorsal) constraint is needed as well. This account takes care of the fact that Limburgian is adverse to nasal place assimilation in general. The obvious functional reason for ranking FILL (nas & lab) so high is that the result of violating it is the creation of an otherwise licit path (or the creation of an existing phoneme, so to say), thus crossing the border between two main categories.

The crucial ranking of PARSE (nas/ə_C) >> *GESTURE (velum) in (12.36) is needed to ensure the surfacing of the /n/ is [dəndāːx]. In (12.35), the reverse ranking was needed to get rid of the [m] in [dəmbɛ̂ːʀ]. There are three reasons to prefer (12.36):

1. With (12.36), we understand the general resistance of Limburgian against place assimilation of nasals. No association lines should be added.
2. In (12.35), PARSE (nas & cor) is crucially ranked with respect to *GESTURE (blade). In (12.36), FILL (nas & lab) is not crucially ranked with the constraints to its left. Therefore, (12.36) is the simpler grammar.
3. If we accept the ease of correspondence between /n/ and [m], we cannot use PARSEPATH or FILLPATH, but should use TRANSMITPATH and *REPLACEPATH instead. This gives the same ranking as with FILLPATH:

Version 4	PARSE (cor/_V)	*GESTURE (tongue tip)	PARSE (cor)	*REPLACE (nas cor, nas lab)	TRANSMIT (nas/ə_C)	*GESTURE (velum)
☞ dəndāːx		*	*			*
dədāːx		*	*		*!	
dənbɛ̂ːʀ		*!				*
dəmbɛ̂ːʀ			*	*!		*
☞ dəbɛ̂ːʀ			*		*	

(12.37)

Whether we represent this phenomenon with PARSE (nas & cor) “we are only interested in /n/ if it stays coronal”, or as FILL (nas & lab) “do not create an [m] where there is no /m/”, both the marked PARSE ranking and the combinatory FILL constraint express an attitude to the segment that is contrary to the idea of autonomous features.

Though the above example seems to make a case for the “segmental” approach, Lombardi (1996) notices that there are no languages that satisfy a final-devoicing constraint by deletion of underlying voiced segments only. Thus, a grammar that allows /at#/ to surface as [at], but forces /ad#/ to become [a], does not occur. Nevertheless, this is what a ranking of IDENT-IO (voice) above MAX-IO would have to give:

/at#/	CODAVOICELESS	IDENT-IO (voice)	MAX-IO
☞ at			
a			*!

(12.38)

/ad#/	CODAVOICELESS	IDENT-IO (voice)	MAX-IO
ad	*!		
at		*!	
☞ a			*

(12.39)

If the typological interpretation of Optimality Theory, namely that all thinkable rankings give possible grammars and that all possible grammars are given by a thinkable ranking, is correct, the non-existence of the above grammar must lead us to conclude that IDENT-IO (voice) is not a viable constraint. If we consider, instead, the feature [voice] as an autonomous autosegment, we can replace the offensive constraint with PARSE (voice); even if we rank this above PARSE (segment) (which is the same as MAX-IO), there is no deletion:

/at#/	CODAVOICELESS	PARSE (voice)	PARSE (segment)
☞ at			
a			*!

(12.40)

/ad#/	CODAVOICELESS	PARSE (voice)	PARSE (segment)
ad	*!		
☞ at		*	
a		*	*!

(12.41)

This gives the correct result (final devoicing), since deletion of the segment is no way to satisfy PARSE (voice).

12.8 Conclusion

To sum up: the overall rarity of featural constraint satisfaction by deletion of a segment, and typical autosegmental effects such as fusion, OCP-driven epenthesis, and floating features pose insuperable problems to a linear version of Correspondence Theory.

So we use PARSE (*feature*), and if we need control over the exact location of features in the output, which is the rationale behind any segmental approach, we can use path constraints like FILL (*feature*$_1$ & *feature*$_2$). The idea is that all aspects of segmentality are as violable as any other constraints.

The grammar of most languages apparently handles segmental effects, which are caused by “vertical” connections between perceptual tiers, as well as autosegmental effects, which are caused by “horizontal” connections between perceptual cues.

13 Degrees of specification

In current theories of *underspecification*, segments are either completely specified underlyingly for a given feature, or not specified at all for that feature. In this section, I shall defend the view that all specifications are violable PARSE constraints, and that underspecification is not a separate principle of phonology, but that it is, instead, an illusion created by normal interaction of faithfulness constraints.

The term *underspecification* is used for two not necessarily related phenomena: the fact that some features are redundant in underlying representations (e.g., the segment /m/, being a sonorant, does not have to be specified for [+voice]), and the fact that some features (like [coronal]) are more likely not to surface than some other features (like [labial]). In the rest of this section, I shall address both of these phenomena.

13.1 Different feature systems for inventories and rules

In a formal formulation of a phonological rule, a natural class is often represented by a bundle of features. Such a bundle specifies the features common to the segments that undergo the rule. Usual phonological practice uses the same features for rules as it does for describing the contrasts in sound inventories:

> “redundant phonological features are mostly inert, neither triggering phonological rules nor interfering with the workings of contrastive features.” (Itô, Mester & Padgett 1995, p. 571)

However, the number of features used for describing sound inventories is usually the minimum that is needed to catch all the possible contrasts. There is no *a priori* reason why these should be the same as those needed in rules. For instance, languages might never contrast more than two values for the feature [voice]; nevertheless, the involvement of segments bearing this feature in phonological processes like voicing assimilation is likely to depend on the actual implementation of the voicing feature in the language at hand. I will show that there are also *empirical* reasons for not assuming the identity of distinctive and inclusive features[24].

13.2 Redundant features

The segment /m/ is allegedly underspecified for the feature [voice]. From the functional viewpoint, however, it is completely specified as /labial, nasal, stop, voiced, sonorant, consonant, bilabial/: a complete set of perceptual features. Voicing is an inalienable facet

[24] cf. Archangeli & Pulleyblank (1994: 52): “both unpredictable, lexically specified F-elements as well as completely predictable F-elements may play either active or inert roles in the phonologies of different languages”.

of the listener's idea of how the segment /m/ should sound, i.e., if it is not voiced, it is less of an /m/. The non-redundant feature [sonorant] might be sufficient, because an unvoiced [m̥] is not sonorant any longer, but an /m/ made non-sonorant will be more /m/-like if it is voiced than if it is voiceless, so [+voiced] is independently needed. Consider the situation where the common cold obstructs your nasal tract. The following tableau shows the three relevant candidates, and the solution that you are likely to choose:

/m/ + cold	*Gesture (open nasal tract)	Parse (nasal)	Parse (voice)	*Gesture (lowered velum)
[m]	*!			*
☞ [b]		*		*
[p]		*	*!	*

(13.1)

Though the articulatory distances of both [b] and [p] to [m] are comparable, the perceptual distance of [b] to [m] is much smaller than the [p] – [m] distance. We see that the superiority of [b] over [p], can only be explained if the constraint Parse (voice) is allowed to compete, i.e., if the feature [voice] is present.

Of course, if you consider this strategy a part of phonetic implementation, which would be a stratum that is ordered *after* redundant feature values have been filled in, you would consider this example phonologically irrelevant. Therefore, I'll have to address the positive evidence that has been brought up for the underspecification of voicing for sonorants.

The idea that some features are redundant in underlying representations, is based on two, not necessarily related, reasons: redundancy for describing inventories, and inertness in phonological rules. I will tackle both.

– *The inventory argument:* "in many segment inventories, all sonorants are voiced but obstruents exhibit a voiced/voiceless contrast; therefore, sonorants are not underlyingly specified for voice".

To make a sonorant, like /m/, voiceless, you have to actively widen your glottis to a large extent; otherwise, because the airflow is not seriously obstructed above the larynx, the vocal folds will not cease to vibrate. In an obstruent, like [b] or [p], voicelessness is brought about more easily, because the labial closure decreases the glottal airflow, which disfavours vocal vibration; instead, sustaining the vibration now requires some extra effort. In other words, for a voiceless [m̥] we need aspiration, and for voiceless [p] only a condition that we'll vaguely call "obstruent-voiceless", and we can assume a fixed ranking of the directly articulatory constraint *Gesture (glottis: spread) above the implementationally formulated (licensing) constraint *[–voiced / obstruent] (see §11.12).

On the perceptual side, we have the Parse (voice) constraints. Now, voiceless nasals are barely audible in many situations, and their place distinctions are nothing to write home about either. By contrast, voiceless plosives have salient release bursts with strong place cues. So, according to the minimal-confusion hypothesis, we can rank Parse (voice / nasal) below Parse (voice / obstruent).

The common inventory described above is a result of the following ranking, where we assume that the categorization is so unrestrictive as to allow the recognition of /m/, /m̥/, /b/, /p/, and /p^h/, and that all three voicing features are subject to the same Parse (voice) constraint (the licensing constraint has been replaced with its appropriate articulatory constraint):

input	output	*Gesture (spread glottis)	Parse (voice/plos)	Parse (voice/nas)	*Gesture (obs –voi)
/m/	☞ [m]				
/m̥/	[m̥]	*!			
	☞ [m]			*	
/b/	☞ [b]				
/p/	[b]		*!		
	☞ [p]				*
/p^h/	☞ [p]		*		*
	[p^h]	*!			*

(13.2)

The resulting inventory is { m, b, p }, independent of the ranking of the rightmost two constraints. If we reverse the first two constraints, the inventory will be { m, b, p, p^h }. So four of the six possible rankings give an inventory that contains more voicing contrasts in obstruents than in sonorants, and even the inventory with the aspirated obstruent does not contain a voiceless sonorant. The two remaining possible rankings, however, will show us that nothing special is going on. First, if we rank both *Gesture constraints (in their fixed order) above both Parse constraints (in *their* fixed order), the inventory will be { m, p }. Finally, if we rank both Parse constraints above both *Gesture constraints, we get { m, m̥, b, p, p^h }. Apart from the richness of some of these inventories along the voicing dimension for obstruents, which is a result of the assumptions mentioned earlier[25], the four types of inventories predicted here are exactly the ones that are attested in actual languages. The typological predictions are:

- As an automatic result of the fixed ranking of the two PARSE constraints and the fixed ranking of the two *GESTURE constraints (and not of an inherent property of sonorants), /m̥/ is rare in inventories. Of the 317 languages considered in Maddieson (1984), only 3 have /m̥/.
- If a language has voiceless sonorants like /m̥/, it must also have aspirated plosives[26] like /p^{h}/.

This predicted implicational universal is borne out by the facts (all the languages mentioned also have a series of voiced nasals):

- Of the three languages with /m̥/ in Maddieson (1984), only Otomi and Klamath are presented with aspirated plosives, whereas Hopi is only reported to have plain voiceless stops. However, Voegelin (1956) explicitly states that exactly those Hopi dialects that have voiceless nasals, also have pre-aspirated plosives that contrast with plain plosives. In the description of Toreva Hopi, Voegelin considers two possible analyses of the stop inventory: either voiceless nasals /m̥/ and pre-aspirated plosives /^{h}p/, or the phoneme sequences /mh/ and /hp/.
- Klamath (Barker 1964) has a series of nasals that are "preaspirated and voiceless throughout", and postaspirated plosives.
- In Tenango Otomi (Blight & Pike 1976), initial sequences of /h+m/ are realized as [m̥m] and /m+h/ often as [mm̥]. Medial plosives are "frequently preaspirated".
- In Temoayan Otomi (Andrews 1949), both nasals and plosives may "unite with h or ʔ to form a sequence", meaning /hm/ and /p^{h}/, respectively.
- In Welsh, post-aspirated nasals alternate with post-aspirated plosives: /ən + p^{h}ɔrθmadɔg/ → /əm^{h}ɔrθmadɔg/ 'in Porthmadog'.
- In Iaai (Maddieson & Anderson 1994), voiceless nasals may be analysed as /hm/ sequences phonetically (because voicing starts halfway the closure) as well as phonologically (because they alternate with /m/ in the same way as vowels alternate with /hV/ sequences). Still, all voiceless plosives, except the dental, have long releases and long voice-onset times (i.e., they are aspirated).
- Jalapa Mazatec (Silverman, Blankenship, Kirk & Ladefoged 1994) has, besides voiceless nasals, full series of plain voiceless as well as aspirated plosives.

[25] If we had added the *GESTURE (+voi / obs) constraint, which can be ranked below *GESTURE (–voi / obs), we would have generated the inventories { m, p, p^{h} } and { m, m̥, p, p^{h} }; if we had restricted the categorization of the voicing dimension, we would have gotten { m, b, p^{h} } and { m, m̥, b, p^{h} } as well.

[26] We must make an exception for final voiceless sonorants as may occur after voiceless obstruents in French, which has no aspirated plosives. Like final devoicing of obstruents (section 10.4.7), this is caused by the universal spreading of the breathing position of the vocal folds after the utterance.

- In Burmese (Cornyn 1944, Sprigg 1965, Okell 1969), there are "voiceless" or "preaspirated" nasals, with a voiced second half, as confirmed by the measurements by Dantsuji (1984, 1986) and Bhaskararao & Ladefoged (1991), contrasting and morphologically alternating with voiced nasals in much the same way as aspirated plosives do with plain ones.
- In Tee (Ladefoged 1995), the only voiceless nasal is /n̥/. Ladefoged is not explicit about the VOT of the voiceless plosives (there are voiced ones, too), though he transcribes them as /p/ etc.
- Angami (Bhaskararao & Ladefoged 1991) has completely voiceless nasals whose second part also has oral airflow (no information about plosives).
- Xi-de Yi (Dantsuji 1982) has voiceless nasals, and an aspiration contrast in plosives.
- Mizo (= Lushai) (Weidert 1975), has a series of nasals whose first part is voiceless. Bhaskararao & Ladefoged (1991) call them "voiceless (unaspirated) nasals", in order to contrast them with the voiceless and postaspirated nasals of Angami (no information about plosives).

Thus, most of these languages with voiceless nasals also have aspirated plosives, whereas less than 30% of the 317 languages of Maddieson's (1984) database have aspirated plosives[27]. To what extent this supports our prediction, is hard to find out precisely, because many of the above languages belong to one family (Tibeto-Burman), which may have a skewed distribution of aspirated plosives. Furthermore, in many of these languages the timing of the glottal gestures with respect to the oral gestures often differs between nasals and plosives. Thus, most of these languages use different glottal-oral coordinations for voiceless nasals and aspirated plosives, which is a somewhat surprising phenomenon. According to Ohala (1975), "voiceless nasals should be partly voiced, because otherwise we would hear no place distinctions".

– The activity argument: "the feature [+voice] can spread, but only from obstruents; sonorants, therefore, do not contain the feature [+voice]".

This argument is due to a failure to appreciate the difference between articulatory and perceptual features. Voiced obstruents are implemented with active gestures to facilitate voicing under the adverse conditions of a supralaryngeal obstruction, such as an extra adduction of the vocal folds to compensate for the raised intraglottal pressure, a slackening of the pharyngeal and oral walls, and a lowering gesture of the larynx. Whatever combination of these tricks is used by the speaker (or the language), this "obstruent-voiced" gesture may spread to a preceding obstruent, making that one voiced as well: /s + b/ → [zb]. For sonorants, by contrast, such a gesture is less needed, and if the gesture is not there, it does not spread: /s + m/ → [sm]. The *perceptual* feature [voice], however, is present in both [b] and [m], because the vocal folds vibrate in both

[27] The 30% is probably an underestimation caused by the common linguistic practice of transcribing aspirates as plain voiceless stops in languages without aspiration contrasts.

sounds, which leads to the perceptual impression of periodicity. If we make a distinction between articulatory gestures and perceptual features, there is no need to assume an underlying [+voice] only in voiced obstruents and a redundancy rule that should assign [+voice] to sonorants at the end of the derivation.

In a framework with underspecification and rule ordering, we would expect the default rule to be able to occur before the spreading rule. Thus, spreading of [+voice] from sonorants is expected to occur, and because of this, Steriade (1995) proposes a feature [expanded larynx] and a feature [voice], both of which should be able to spread. In a framework with a distinction between articulatory and perceptual features, this would not be expected. We must review, therefore, the evidence that has been brought up for the spreading of [voice] from sonorants.

First, Steriade (1995) mentions the English morpheme plural morpheme, which shows up as [+voiced] after voiced obstruents and sonorants ([bʌg-z] ‘bugs’, [k^{h}ɔːl-z] ‘calls’), but as [–voiced] after voiceless obstruents ([tʃ^{h}ɪk-s] ‘chicks’). This can be analysed, however, with a voiced morpheme /z/, with spreading of [–voice] from voiceless obstruents. Confirmation of this analysis is found with the morpheme /θ/, which, being specified as [–voice] shows no voicing after sonorants ([hɛl-θ] ‘health’), nor, for that matter, after voiced obstruents ([brɛd-θ] ‘breadth’).

Another example is final voice neutralization. In Sanskrit, word-final obstruents assimilate their voicing features to those of any following sound, be it an obstruent, a sonorant consonant (but /k+m/ → [ŋm]), or a vowel. In Limburgian, word-final obstruents “assimilate” to following plosives and vowels; before fricatives and sonorant consonants, they are voiceless. Neither of these cases has to be described as spreading from a sonorant, because in both Sanskrit and Limburgian, utterance-final obstruents devoice, which, together with the “assimilations” mentioned earlier, leads to a complete voice neutralization of word-final obstruents. Therefore, PARSE (±voi / $]_W$) must be ranked very low, probably as a generalization of utterance-final voice neutralization: words are often utterance-final, so their final obstruents are less likely to show up as voiced than their initial obstruents, even if a voicing contrast is maintained at the end of a word but not at the end of an utterance, so PARSE (±voi / $]_W$) must be ranked lower than PARSE (±voi / $_W[$), and the local-ranking principle does the rest. The data of Limburgian can now be explained by the following ranking (the interaction with fricative devoicing is too complex to discuss here):

PARSE (±voice / $_W[$)
|
*[–voi / obs / V_V]
|
*[+voi / obs / V_V]
|
PARSE (±voice / $]_W$)

(13.3)

The Sanskrit data are found by generalizing the right-hand environment to all sonorants. The typology suggested by the two languages derives from the near-universal ranking *[–voi / obs / V_C son] >> *[–voi / obs / V_V]. If sonorants could spread their voicing gesture, we would have to realize that sonorant consonants need a stronger voicing gesture than vowels, so that we should expect the ranking *[+voi / obs] >> *[+voi / C son] >> *[+voi / V] to be active. The typology that can be derived from this ranking would predict that there are languages where sonorant consonants spread, but vowels do not: the reverse situation from the Limburgian case. Only if such languages exist, I am ready to believe in the spreading of [+voice] from sonorants.

13.3 Weak features

In our account, *specifications are constraints*. Some features, like [coronal], are less likely to surface than some other features, like labial. For instance, /n/ is specified as being coronal from the beginning, but a higher-ranked gesture-minimizing constraint can cause the underlying value not to surface (§11.4). So, Dutch /n/ only surfaces as a coronal if it cannot get its place specification from a following consonant. Underspecification theories “explain” this by stating that /n/ is not specified at all for place underlyingly, so that its place specification does not have to be erased by the following consonant, which would be one of those unwanted structure-changing processes. Afterwards, a *default rule* would fill in the coronal place specification. Kiparsky (1985), who analysed the similar data of Catalan, would describe this situation with the following ordered set of rules:

1. (Underlying specifications:) /ŋ/ is specified as using the dorsal articulator and the velar place of articulation, /m/ is specified as using the labial articulator (lower lip) but has no specification for place of articulation, and /n/ is not specified for any articulator or place at all.
2. (Feature-filling assimilation rule:) every nasal consonant, if not yet specified, takes on the articulator and place of the following consonant.

3. (Feature-filling default rules:) a labial stop (plosive or nasal) that is not yet specified for place of articulation, is bilabial, and a consonant not yet specified for place at all, is coronal and alveolar.

A major drawback of such an approach is that rule 2 produces a result that can be expressed as a well-formedness condition on clusters of a nasal plus a consonant, i.e., it ensures that clusters originating when two words are concatenated, adhere to the same phonotactic constraints that hold inside morphemes. Thus, rule 2 seems to be goal-oriented (the goal being the fulfilment of the preference for homorganic clusters), but does not refer explicitly to that goal. Optimality Theory and other constraint-based theories promote these goals to the status of the actual building blocks of phonological description. In the approach of §11.4, underspecification is taken care of in a natural way: /n/ is not really unspecified for place, but the place specification for /n/ just ranks much lower than many other constraints, likewise, bilabiality of /m/ emerges although its specification is weak.

Thus, underspecification is not a separate device, but an automatic result from the general theory.

13.4 The lexicon

There is one area where underspecification is still useful: the efficient storage of forms in the lexicon. For instance, a morpheme with /m/ will only contain the minimal information needed to reconstruct this segment: perhaps the specification /nasal + labial/ or just the specification /m/. In both cases, these specifications must be pointers to a fully specified list of the perceptual features that are desired in the output, like [voice].

In Chomsky & Halle (1968), the specification of the most common (or *unmarked*) values of all features could be left out of the underlying representation (*m* for “marked”), for the sake of even more efficiency of lexical representation, :

	/t/	/ɛ/	/n/	/s/
coronal	+			
voiced				
continuant				*m*
strident				
nasal			*m*	
vocalic				*m*
sonorant				
high		*m*		
back				

(13.4)

The empty cells would be filled in by redundancy rules, such as [+son] → [+voi], [+nas] → [+son], ∅ → [–voi], etcetera (note the subtle difference between “plus” and “marked”; also note that the values for [vocalic] for the first two segments could be filled in on the basis of the default CV syllable). It was not suggested that the marked values were phonologically more active than the unmarked values. The phonetic form [t^{h}ɛn^{t}s] is derived by late rules that govern the aspiration of plosives in onsets of stressed syllables, and the insertion of an intrusive stop in sequences of sonorant plus /s/.

Our example /tɛns/ would in such a theory be specified as a sequence of “oral plosive” plus “mid vowel” plus “nasal” plus “fricative”, in this order and without overlap. We could run this specification through the English constraint system. All the consonants would become coronal, not labial, because *GESTURE (coronal) is ranked below *GESTURE (labial), or because FILL (coronal) is ranked below FILL (labial). The resulting output would be [[t^{h}ɛɛ̃n_ts]], like in the real world. So we could ask whether the underspecified input is real or not. The question cannot be answered in general, because it depends on what criteria of simplicity you apply. As always in phonology, there is a trade here: the simplicity of the underlying form shifts the burden of stress to the recognition phase: many FILL constraints are violated in deriving an actual output from an underspecified input. If the simplicity of recognition is the criterion, the underlying form should be maximally similar to the surface form. If the underlying form is /t^{h}ɛɛ̃n_ts/, no constraints are violated in the resulting tableau. With a “tableau of tableaux” criterion of lexicon formation (Prince & Smolensky 1993), this underlying form would be optimal.

Opting for /t^{h}ɛɛ̃n_ts/ as the underlying form, however, does not take account of the speaker’s intuitions as to the phonological make-up of this morpheme. Spoken backwards, for instance, the word is not [[st_nɛ̃ɛ^{h}t]], but [[snɛ^{ʔ}t]], which suggests an underlying /snɛt/, with an appreciable degree of segmental organization (i.e., high path constraints).

13.5 Optionality and stylistic variation

In rule-based theories, rules either do or do not apply. If a rule does not apply, it is not in the grammar. If a speaker sometimes does apply the rule, and sometimes does not, it has to be marked in her grammar as *optional*. This is a separate device again.

In a theory based on constraint ranking, there is no such built-in phenomenon as optionality. A constraint does not have to leave the grammar if it becomes weaker. It may even still be active, but less visibly so. The rule-based counterpart of this change in visibility would be a *change* in the environment of the rule, a change which can not be related in any principled way to the function of the rules.

In §11.5, I showed that even within a language, constraint ranking can show variation, and that (it will come as no surprise) the division between articulatory and perceptual constraints plays an interesting role there.

13.6 Privative features

Steriade (1995) states that the need for underspecification theories is much diminished if most features are seen as privative. For instance, if [nasal] is considered a privative feature, this would “explain” the fact that nasality can spread, but non-nasality cannot. But as seen in §8, this effect is related to a listener strategy based on commonness, and is expressed in the grammar as the fixed ranking *DELETE (+nasal) >> *INSERTPATH (+nasal & place). The same goes for the /ɛ/ in /tɛns/: it can be underlyingly specified as [–nasal], but *REPLACE (ɛ, ɛ̃) is ranked below *SYNC (blade: open | closed, velum: closed | open).

13.7 “Trivial” underspecification

According to Steriade (1995), “plain coronals are trivially, inherently, and permanently lacking in specifications for the features [labial] or [tongue root]”. But coronals are specified for [labial] in the sense that the lips cannot be closed during the burst of [t]: as we saw in §2.4, the articulatory underspecification is restricted by the needs of perceptual invariance, i.e. the variable α in a dominated *REPLACE (t, α) cannot be perceptually too far away from [t]. Because the spectrum of the burst of [t] is determined most prominently by the location of the release, and less so by secondary constrictions, the illusion of underspecification comes to the surface.

13.8 Invisible specifications

In §7, I argued that the /s/ in /tɛns/, though not rounded at the surface, may be underlyingly unspecified for the feature [round]. In /usu/, the lips may stay rounded throughout the coronal constriction, and in /isi/, they may stay spread, so there is no empirical difference between specifying /s/ as [+round] or [–round]. Even the fact that an isolated utterance /s/ is pronounced without rounding, can be attributed to the ranking *GESTURE (lips: rounded) >> PARSE (±round / sibilant)[28]. In a sense, the grammar that uses an underlyingly unrounded /s/ is simpler than the grammar that uses a rounded /s^w/, because the former grammar inflicts a smaller constraint violation for a maximally faithful rendering of the underlying form. However, no empirical effects are associated with this “minimization of grammatical stress”.

13.9 Conclusion

In functional phonology, listener-based constraint rankings replace the “unmarkedness” that other theories ascribe to certain features or feature values, and that they try to build into their representation of phonology. The explanation for each of these rankings has to be sought in the coincidental properties of the human speech channels and the human ear, not in the human language faculty. The default assumption must be that the input contains full specifications of all feature values, though some of these specifications are so weak that they can easily be overridden by articulatory constraints. These weaknesses cannot be stipulated, but can be derived instead from considerations of perceptual contrast.

[28] In English, this is not quite true, because an isolated utterance /ʃ/ is pronounced with lip rounding.

14 Empirical adequacy of Functional Phonology

In the previous sections, I developed a theory of how phonology would look like if it were governed by functional principles. This may be fine as an exercise of the bottom-up construction of an ideal world, but the resulting theory of Functional Phonology will be an acceptable alternative to generative approaches only if it is capable of describing phonological structures and processes with an equal or higher amount of empirical adequacy, efficiency, and explanatory force. In subsequent papers (Boersma, fc. a-e), I will show that Functional Phonology can stand up to this test and clarify many hitherto recalcitrant phonological issues; on the phonetic side, some of these papers profit from a computer simulation of a comprehensive physical model of the vocal apparatus (Boersma 1993, 1995, in progress), which was devised with the intent of studying the "automatic" relation between articulation and acoustics.

In order not to keep the reader wondering what directions these investigations have taken, the following sections concisely describe their results.

14.1 Spreading

According to functional principles, only articulatory gestures can spread. Spreading of a perceptual feature would diminish the contrast between utterances; this would always be worse than no spreading[29].

There are at least three basic types of spreading. The first is a change in the timing of an articulatory gesture, needed to satisfy an articulatory constraint, most often *SYNC. Thus, /εn/ is pronounced [[εε̃n]] because [[εn]] would violate *SYNC (nasal, coronal) and because a shift in the other direction would give [[ε^{t}n]], with an offensive nasal plosive; likewise, /ns/ may be pronounced [[n_ts]] because its alternative, [[nə̃s]] would violate a stronger FILL constraint.

The second type of spreading occurs when a concatenation process causes the adjacency of two incompatible articulatory gestures. For instance, if [spread glottis] meets [constricted glottis], one of them will have to leave.

The third type of spreading occurs when a concatenation process causes the overlap of two gestures with conflicting perceptual correlates. One of the two gestures is then bound to be deleted. For instance, in [anpa], the slower labial gesture overlaps the coronal gesture, which diminishes the contrast between it and [ampa]. Thus, PARSE (place) will fall down the constraint hierarchy, perhaps below *GESTURE (tip), which would result in the deletion of the tip gesture.

The role of perception in spreading branches into two *blocking* effects. First, faithfulness constraints will have to allow spreading to take place at all. In Dutch, for instance, the labial gesture is allowed to spread in /n+p/, but not in /t+p/, though the articulatory gain would be equal in both cases (namely, the loss of a coronal gesture); the difference must be caused by a different ranking of PARSE (place) or FILL (place × nasal) constraints for nasals and plosives.

The second blocking effect of faithfulness constraints works at a distance: the demarcation of the domain of spreading. In harmony systems, the spreading of a feature can be blocked by a perceptual specification that is incompatible with the spreading gesture. In nasal-harmony systems, for instance, the [lowered velum] gesture is incompatible with the perceptual specifications of most consonants: in decreasing order of perceptual incompatibility, we find plosives, fricatives, liquids, oral glides, and laryngeal glides; this order reflects implicational universals of transparency of consonants to nasal harmony.[30]

The predicted correlations between articulatory gestures and spreading, and between faithfulness constraints and blocking, are verified in Boersma (fc. c).

14.2 OCP

In Functional Phonology, the OCP branches into two fundamentally different principles.

The first is a general principle of human perception, not confined to phonology. In designing a country map of Europe, the cartographer can choose to fill in the countries with the minimal number of four colours that are needed to give every pair of adjacent countries different colours. If she decided to paint both the Netherlands and Belgium red, the reader of the map would not be able to identify them as separate countries; thus, in cartography, adjacent identically coloured countries are avoided.

Likewise, if a morph ending in /-m/ is concatenated with a morph starting with /m-/, the usual timing of syllable-crossing clusters will result in the long consonant [-mː-]. The perceptual identity of one of its constituents is therefore lost, violating PARSE (root). Some of the information about the existence of two morphemes is kept in the timing, but if the language is adverse to geminates, it may just end up with [-m-], violating PARSE (X) in addition.

The problem of the long perceptually homogeneous sound can be levied by inserting a pause between the two consonants (i.e., drawing a black border between the

[29] Perceptually motivated 'spreading' could improve the probability of recognition of the feature. It would be associated with stem-affix vowel harmony, whole-word domains, etc. (the F-domain of Cole & Kisseberth 1994). But it is not spreading (as Cole & Kisseberth note). 'Transparent' segments with incompatible articulations are expected, not 'opaque' ones. The problem of Wolof, which shows both transparency and opacity, is treated in Boersma (fc. a).

[30] Guarani-type nasal-harmony systems, where plosives are transparent to the spreading of [+nasal] but are still pronounced as plosives, must be analysed in a different way. Analogously to the situation in most other languages, where nasality can be seen as superposed on an oral string and implemented with a [lowered velum] gesture, these harmony systems may consider orality (in half of their morphemes) as being superposed on a nasal string and implemented with a [raised velum] gesture.

Netherlands and Belgium): giving [[-m_m-]]. This violates a FILL (pause) constraint: a pause can be perceived as a phrase boundary. Another strategy would be to insert a segment (declaring the independence of the southern provinces of the Netherlands, and painting them blue), which will give [-məm-] or so: another FILL violation. Language-specific rankings of all the faithfulness constraints involved will determine the result.

The perceptual nature of this first functional correlate of the OCP is shown by the rules of vowel insertion in English, which are hard to capture with generalizations over single tiers in feature geometry. Thus, the insertion of /ɪ/ before the morpheme /-z/ occurs in *bridges* but not in *tents*, exactly because [dʒz] would contain a perceptually unclear boundary (The Netherlands in red, Belgium in purple), and [nts] would not; likewise, the insertion of /ɪ/ before the morpheme /-d/ occurs in *melted* but not in *canned*, because the boundary would be lost in [ltː] but not (or less so) in [nd].

The second functional correlate of the OCP is simply the tendency not to repeat the same articulatory gesture: an articulatory *REPEAT constraint. The features involved in this constraint are arguably of an articulatory nature: the Japanese constraint against two separate voiced obstruents within a morpheme obviously targets the articulatory gesture needed for the voicing of obstruents, not the perceptual voicing feature, which would also apply to sonorants. A clear difference with the first principle is exhibited by a morpheme-structure constraint in Arabic, which does not allow two labial consonants within a root; apart from disallowing two appearances of /b/, it does not even allow /m/ and /b/ to appear together. This generalization over plosives and nasals is typical of the articulatory labial gesture, which does not care whether the nasopharyngeal port is open or not, whereas the divergent behaviour of plosives and nasals in *melted* versus *canned* is exactly what is expected from a perceptually conditioned phenomenon.

The predicted correlations between near OCP effects and faithfulness constraints, and between distant OCP effects and articulatory constraints, are verified in Boersma (fc. d).

14.3 Feature geometry

In theories of *feature geometry*, features are placed in a hierarchical tree to express the fact that groups of features co-occur in phonological rules and structures. In the folowing, I will show that this tree is a hallucination caused by a confusion of articulatory and perceptual features. In Functional Phonology, the only acceptable hierarchies are the implicational trees (2.4) and (2.5).

– No place node. The prototypical example of why we need the place node, is the language in which a nasal consonant that is followed by another consonant, is constrained to have the same place of articulation as that other consonant, or, to put it shorter, all nasal-nonnasal consonant clusters are homorganic. Such a language may also show the *active* process of place assimilation of nasals, in which every nasal takes on the place of the following consonant. The conclusion is that the labial, coronal, and dorsal articulators undergo the same rule, and have thus to be subsumed under one node, the *place node*.

But there is no articulatory reason why the three articulators should act as a group: articulatorily, the labial, coronal, and dorsal articulators can be moved independently from each other. Also, there is some uneasiness (or controversy) as to where the pharyngeal articulator belongs; in some respects, it is an articulator like the labial, coronal, and dorsal articulators, and in some respects it is not: now, does it belong under the place node, or not?

The answer is: there is no place node. For place assimilation of nasals, the pharyngeal articulator is not included in the rule; for rules involving fricatives, it *is* included. Therefore, the focus has been wrong: instead of identifying the common denominator of the targets of the place assimilation rule for nasals as “oral place”, we should just stay with the feature [nasal]. The pertaining sounds are specified for [nasal], and as long as there is a constriction anywhere in the mouth, these sounds will be heard as nasal consonants, i.e., sounds characterized by an airstream that travels exclusively throught the nose. So, there is nothing common to a labial closure, a coronal closure, and a dorsal closure, except that they all can be trusted to bring a perceptual [nasal] specification to the surface of the utterance. It was not the idea of the theory of feature geometry to have its nodes supervised by a single perceptual specification; rather, it considered the nodes as universal, perhaps innate, groupings of features. But reality seems to be simpler: to implement the *perceptual* feature [nasal consonant], we can choose from the *articulatory* gestures [lips: closed], [blade: closed], and [body: closed].

There remains the problem of why the coronal gesture should delete when the labial gesture spreads. Such phenomena may have two causes:

- Double spreading: the spreading of an articulatory gesture obscures the perceptual result of the overlapped gesture, which can subsequently be deleted with less of a problem.
- The deletion of [coronal] is the cause, and the spreading of [labial] is a consequence which preserves the faithfulness of a perceptual feature like non-orality or timing.

Often, these two causes cannot be separated: a process like /an+p/ → /amp/ may only be possible if spreading and deletion occur simultaneously, because /am͡np/ involves a perceptual loss without any articulatory gain, and /aãp/ involves a perceptual loss that may not outweigh the gain of satisfying *GESTURE (coronal). The Optimality-Theoretic approach, of course, serves well in the evaluation of this kind of tunnelling processes.

– The feature [continuant]. Continuancy is an unsolved problem in theories of feature geometry. Again, this problem rests on a failure to distinguish between articulatory and perceptual features. The articulatory feature [continuant] would refer to a certain of stricture, and can be independently implemented for every articulator, so it should be located under every articulator node separately. In the geometry of (2.5), it could be considered to refer to every degree of constriction more open than “closed”. However, such an articulator-dependent position of the feature [cont] is not usually considered appropriate, because [cont] does not have to spread when its articulator spreads: in /n+f/ → /ɱf/, the place of /f/ spreads, but its degree of constriction does not; likewise, in an *i*-umlaut system /o/ becomes /ø/, not /y/, before /i/. Therefore, [cont] must be located somewhere else: Sagey (1986) hangs a single [cont] feature directly under the root node, and goes in great lengths to manufacture a projection of it on the “primary articulator”: an arrow which points to the relevant articulator, or functionally speaking, to the articulatory gesture that implements the perceptual feature [cont]. Problems arise, then, because clicks in Nama can have friction contrasts for the anterior closure as well as for the velar closure, and so on. Clearly, the situation is in want of a better idea.

First, the feature [cont] often does spread: even people with missing or irregular teeth can produce a reasonable /ɱ/, because some oral leak is no problem perceptually (in contrast with the case of the labiodental plosive, which, because of the population’s average dental health, does not occur as a speech sound). Therefore, /ɱf/ may well involve assimilation of degree of closure.

The Sanskrit processes /s+t/ → [st] and /s+ʈ/ → [ʂʈ] is a better example: apparently, place can spread without dragging continuancy along. This process can be seen as spreading of place with conservation of frication: a minimal blade gesture is needed between [ʂ] and [ʈ] in order to preserve the perceptual features [fricative] and [plosive]. But it is clear that the sequence [ʂʈ] is much easier to implement than [sʈ]: in the former, the sides of the tongue remain fixed throughout the cluster. Everything is explained by the quite acceptable ranking PARSE (sibilant) >> *GESTURE (blade: grooved | retroflex) >> *GESTURE (blade: critical | closed) >> PARSE (place). Note that in this formulation, “sibilant” and “critical” are the two keywords that replace the hybrid feature [cont].

– ***The laryngeal node***. Evidence for the laryngeal node is found in processes where voicing and aspiration seem to act as a single natural class (McCarthy 1988): in Greek, [+voice] and [aspirated] spread throughout obstruents clusters; and in Thai, voiced and aspirated plosives become voiceless word-finally. However, proving the existence of a laryngeal node would involve showing the interdependence of these processes. For instance, if 70% of voiced consonants show final devoicing cross-linguistically, evidence for a laryngeal node would involve proving that the proportion of voiced consonants that show devoicing in languages that also show final neutralization of aspiration contrasts, is higher than 70%. In absence of such evidence, we should not stipulate a laryngeal node. If, by contrast, these percentages will prove to be equal, final devoicing and final deaspiration must be uncorrelated and, therefore, probably independent processes.

– ***The root node.*** The root node is thought to contain all the features that are not phonologically active, like [sonorant] and [consonantal]. We now know that these are *perceptual* features, so we are not surprised that they do not spread. However, the root node is a node because it can spread as a whole. This is *total assimilation*, e.g., Arabic /al/ ‘the’ + /daːr/ ‘house’ → /adːaːr/. The constraints that are satisfied, however, in this process, are *GESTURE (lateral) and the perceptually motivated PARSE (timing): most of the segment has to disappear, but its coronality (in Arabic) and, above all, its *timing* remain.

Perceptually, the root node is the location where the identity of adjacent identical elements is lost completely: if all the perceptual features stay the same, no boundary is perceived at all (§12). This combination of all the perceptual features available to the listener at a given time, thus has a special status, but not as a result of a built-in hierarchy.

Within a feature-based framework, the main effects of the root node (timing and complete identity) can be handled with prosodic units (morae) and an OCP that is ranked according to §12.3 (and inviolable if its arguments involve two equal values on the maximal combination of tiers).

The conclusion must be that feature geometry is superfluous. Some features and gestures form classes because they happen to work in the same perceptual domains. Several aspects of feature geometry are addressed in Boersma (fc. a, b).

14.4 Inventories

The Frisian short-vowel system is

i			y			u	
	e		ø		o		
		ɛ		ɔ			
			a				(14.1)

This system has been drawn in a somewhat triangular shape to stress the fact that the perceptual front-back contrast is smaller for lower vowels than for higher vowels.

Phonological approaches to sound systems like this (radical or contrastive underspecification) try to express sound systems with the minimal number of distinctive features and their combinations. Starting with a finite number of distinctive features, they derive the “gaps” (here, the gap at /œ/ and the restricted distribution of the low vowels) with the help of redundancy rules or default rules. No explanatory adequacy is arrived at.

Phonetic attempts to explain sound inventories have used only a few functional principles. Kawasaki (1982) restricted her explanations to the two perceptual principles of maximization of distinction and salience. Stevens (1989) tried to explain the commonness of some sounds as the minimization of precision and the simultaneous maximization of acoustical reproducibility. Liljencrants & Lindblom (1972) investigated how vowel systems would look like if they were built according of the principle of maximum perceptual contrast in a multi-dimensional formant space. Lindblom (1990a) sought the solution in auditory and proprioceptive distinctivity, adding them to each other with a kind of global bookkeeper’s strategy. Ten Bosch (1991) explained vowel systems on the basis of maximal distinctions within an articulatory space bounded by an effort limit based on the distance from the neutral vocal-tract shape; similar approaches are found in Boë, Perrier, Guérin & Schwartz (1989), Schwartz, Boë, Perrier, Guérin & Escudier (1989), Boë, Schwartz & Vallée (1994), and Schwartz, Boë & Vallée (1995). None of these approaches derives the symmetry that is visible in (14.1).

Surpringly, none of the ‘phonetic’ approaches took into account the symmetrizing principles of perceptual categorization, which would explain, among other things, the perspicuous phenomenon found in Frisian and in many other languages with a lot of vowels, namely, that back vowels tend to be on equal heights with front vowels. In (14.1), we see four height categories and three colour categories. This reflects a general phenomenon of human perception, and is not necessarily due to any special property of phonology. On the other hand, the ‘phonological’ approaches ignore the explanatory power of phonetics, which predicts that faithfulness constraints are ranked by perceptual contrast, and in the Frisian case this means that PARSE (round) is ranked lower for /œ/ than for /ø/, which explains the gap in the Frisian lower-mid-short-vowel system.

So, relying on a single principle to explain everything is not enough. This may be a defensible approach in physics, but not in linguistics. We should use all our functional principles. This means that we do not use a single effort or contrast measure, but take energy, organization, and synchronization into account, as well as perceptual distinctivity and salience. Another aspect of the comprehensive approach is that our standpoint on the question of the continuity versus the discreteness of the articulatory and perceptual spaces is that they are both discrete *and* continuous: discrete in the sense that only a few values or regions are used within a language, continuous in the sense that the categories are taken from a continuous scale on a language-particular basis.

While all the approaches mentioned above aimed at explaining vowels systems only Boersma (fc. c) attacks inventories of consonants as well.

14.5 Sound change

In Boersma (1989), I developed a “decision regime that ‘only’ requires knowledge of rank orderings of the articulatory ease and the perceptual salience of sound sequences and knowledge of the orderings of dissimilarities of pairs of words. Under this regime the sound patterns of languages will keep changing forever, even if there are no external influences on them”. In the strategy used, “it is possible not to refer to any data measured in numbers. Instead, we can do with a number of rank orderings”. Indeed, the present paper can be seen as an exploded and OT-ized version of this earlier work.

The OT counterpart of the decision regime used in Boersma (1989), which decided “the interaction between the optimization principles” by means of majority vote, is the pressure that arises in a constraint system if some constraints are reranked randomly. If we have three constraints, we can rank them in six ways, and if four of these rankings prefer sound system B and two of them prefer sound system A, the ultimate result will be that the system becomes B, even if it used to be A. In other words, a language with system A experiences a *pressure* towards system C, and a sound change will eventually take place if there is some *temperature* (random reranking) in the system.

With an additive evaluation of functional principles, the sound system will eventually settle down in an optimal state. But with an evaluation procedure based on counting votes (as in Boersma 1989) or on strict ranking (as in Optimality Theory), it is possible that system B is better than A, and system C is better than B, and system A is better than C. This causes an eternal circular optimization loop, reminiscent of some circular sound changes as the Germanic consonant shifts and the Middle-English vowel shift, which I use as examples to show that all this actually works (Boersma, fc. d).

14.6 One or two levels? Containment or stratification?

Besides the input-output relations identified in §8, there is another possible interpretation of faithfulness constraints: they can be seen as direct output constraints, e.g., PARSE (*f*) could be replaced directly by “∃ [*f*]” (which could again be abbreviated as the specification “/*f*/”). This declarative formulation, which says “the output contains the feature [*f*]”, is explicitly output-oriented, just like the articulatory constraints work directly on the articulatory representation. Now, because the relation between the articulation and the acoustic output is automatic, we are left with only one level for constraints to work on: the output.

However, there is a cost involved. If the specifications are constraints themselves, there are no underlying forms to change, so that morphemes must be bundles of constraints, and the constraints must be part of the lexicon; this is the standpoint of Declarative Phonology (Bird & Klein 1991, Scobbie 1993, Bird & Ellison 1994), and has been proposed within Optimality Theory by Hammond (1995). This means that many of the constraints are language-specific and have to be ranked somewhere between the universal articulatory constraint families. Such a theory is both more and less restrictive than a two-level approach.

In the one-level approach, every instance of /nasal/ can in principle be ranked individually, thus, ranking is morphologically conditioned by default, in sharp contrast with the two-level approach, where the morphology determines the shape of the specification, and a phonological constraint ranking determines the output (by default). In order to restrict all instances of /nasal/ in the same phonological environment to an equally high ranking, the one-level version of the lexicon would have to link this specification to a location where information is stored about that ranking. Thus, the lexicon contains a part of the grammar.

In the two-level approach, faithfulness concerns relations between input and output; every thinkable input, therefore, results in a well-formed output, and many different inputs will result in the same output. This indeterminacy of the input will have to be restricted by output-centred criteria like choosing the input that incurs the least dramatic constraint violations given the output (Prince & Smolensky 1993), and/or by output-ignoring criteria as sparsity of specification in the input. The latter possibility will lead to underspecifications in the input reminiscent of Chomsky & Halle's (1968); the constraint system will fill in the blanks with the articulatorily and perceptually least offensive material, thus replicating the markedness conventions, redundancy rules, and default rules, of traditional underspecification theories.

In another sense, our version of OT is a one-level version: all the constraints that simplify the structure of the utterance, work on the level of articulatory implementation, i.e., they are output-oriented. Also, the environment clause in faithfulness constraints refers to the output, becuse perceptual contrast must be evaluated on pairs of possible outputs. This overall output-orientedness is in contradistinction with the original idea of containment in OT (Prince & Smolensky 1993) and with the multi-level approaches of Cognitive Phonology (Lakoff 1993) and Harmonic Phonology (Goldsmith 1993). Consider, for instance, the prototypical example of counterfeeding crucial rule ordering: the American English words [raiɾɚ] 'writer' and [raːiɾɚ] 'rider' from underlying /rait+ər/ and /raid+ər/ show that vowel lengthening before a voiced consonant precedes flapping. The three non-derivational approaches mentioned solve the problem by having their rules or constraints refer to the underlying form ("/ai/ is lengthened before an underlying voiced stop"): with this strategy, vowel lengthening and flapping can be evaluated in parallel. The original containment approach, however, could refer *only* to the underlying form, because the output contained the input. This, then, has problems with representing feeding or bleeding rule ordering, a difficulty not found with the two-level approaches, in which cross-level rules may refer to the environments at both levels. In McCarthy & Prince (1995), containment was abandoned, and the resulting system became comparable to the one advocated in the present paper, but not with respect to the material allowed in the environment clause of constraints.

It is hard to see how our output-oriented one-level approach could handle [raiɾɚ] and [raːiɾɚ] other than with ordered levels of representation. Transparent rule orders, on the other hand, are handled in a natural way. For instance, in a hypothetical English dialect where [raːiɾɚ] corresponds to both /raitər/ and /raidər/, we could accomodate this phenomenon by imposing two constraints on the output only, rather than assuming the feeding order of flapping before vowel lengthening: vowels are long before voiced consonants, and post-tonic intervocalic coronal plosives are flaps; [raːiɾɚ] will automatically emerge as the optimal result. A perspicuous cross-linguistic property of rule ordering supports this approach: since transparent (feeding and bleeding) rule ordering is much more common than opaque (counterfeeding and counterbleeding) ordering, our approach will involve much more parallel (within-level) evaluation than serial (cross-level) derivation. This is a favourable result that combines the need for models of speech production as being computationally fast (i.e., parallel), with a functional (i.e., output-oriented) view of phonological simplification.

If we accept the presence of a comparison loop (figure 2.1), we have a two-level phonology with input and output, and violation of faithfulness. In Functional Phonology, all faithfulness constraints and all articulatory constraints are output-oriented, because perceptual contrast is evaluated between pairs of perceptual results, and articulatory effort is evaluated on articulatory implementations.

Attempts to parallellize crucial rule ordering were made by Orgun (1995), Cole & Kisseberth (1995), and McCarthy (1995).

In Orgun's approach, the ranking of faithfulness constraints may depend on the input environment. This is able to capture a single instance of counterfeeding or counterbleeding rule ordering.

In Cole & Kisseberth's approach, a single input level may occur in the output in a special way, and this output is then evaluated: from the features that occur in the output (respecting MAX-F), some do surface (respecting EXPRESS-F), and some do not (violating EXPRESS-F). Their example captures the counterbleeding order of Harmony and Lowering in Yawelmani. Cole & Kisseberth leave out the equally counterbleeding order of Lowering and Shortening, perhaps because including it would force them to allow *three* levels in the output, for which the combined actions of MAX-F and EXPRESS-F would not suffice.

In McCarthy's (1995) approach, constraints may refer to material that occurs (a) in the input, or (b) in the output, or (c) in either. In the light of new data, his framework would probably have to be extended. To capture the facts of Mohawk stress assignment, for instance, his three possible environments would have to be supplement with a fourth, namely, "in both": Mohawk penultimate stress assignement disregards vowels deleted from the input as well as vowels epenthesized into the output, so the only vowels that determine where the stress falls, are those that occur both in the input as well as in the output. Even this extension, however, would still not be able to capture all instances of crucial rule ordering.

Level ordering is needed in our theory of grammar because phonologies can work with the results of long series of sound changes without adapting themselves to any theory of how abstract underlying forms should correspond to surface forms in a system of parallel constraint evaluation. With an output-only approach as presented in this paper, even very complicated rule systems can be captured in only a few sequential levels of parallel evaluation (Boersma, fc. e). After abandoning the axiom of serial rule ordering, the phonological world experiences an axiom of parallel constraint evaluation. Eventually, a synthesis will emerge.

15 Conclusion

Starting from the functional principles of maximization of ease and contrast, we identified four articulatory constraint families and four perceptual (faithfulness) constraint families, and there are probably some more. We also developed a strategy for finding universal rankings and predicting at what points languages are allowed freedom. Optimality Theory seems to be very suitable for expressing function.

Functional Phonology can solve several hitherto controversial issues:

1. Phonetic implementation and perceptual categorization can be described with the interaction of continuous constraint families (§6 and §10).
2. Spreading is not a separate phonological device (§14.1; Boersma fc. a). Assimilation effects result from the interaction between articulatory and perceptual constraints. Only articulatory features can spread. Only perceptual features can block spreading.
3. The OCP is not a separate phonological device (§14.2; Boersma fc. b). Its effects result from the interaction of a constraint against loss of perceptual identity with articulatory and perceptual constraints.
4. Feature Geometry is not a separate phonological device (§14.3). Nodes only combine articulatory gestures that have cancelling perceptual results.
5. Underspecification is not a separate phonological device (§13). Specifications are constraints, and as with all other constraints, some are strong and some are weak.
6. In segment inventories, symmetries and gaps are predicted by the same constraint-ranking system (§14.4, Boersma fc. c).
7. Randomly varying constraint ranking produces a pressure in the direction of preferred sound change (§14.5, Boersma fc. d). An eternally optimizing sequence of sound change can be circular.
8. The stratification of the grammar is limited to processes that used to be described with counterfeeding or counterbleeding rule ordering (§14.6, Boersma fc. e).

Remaining problems include the role of additivity versus strict ranking: acoustic cues for perceptual features and aspects of energy have been presented as additive, and segments are obviously categorical. In between, there is the realm of the separate features; these have been presented as categorical, but could also be considered as cues for the categorization of segments.

We explained some language-independent constraint rankings with phonetic principles, but others will have to be derived from the data of the languages of the world. This situation may be less than ideal, but the possibility of bridging the gap between phonology and phonetics at all is such a good prospect that we should not be afraid of a few initial holes in our knowledge. More positively, if more than one phonetic

explanation for a given language fact has been advanced (as is often the case), the phonology may well tell us which of them is correct.

References

(ROA = Rutgers Optimality Archive, http://ruccs.rutgers.edu/roa.html)

Anderson, J.M., & Colin Ewen (1987): *Principles of Dependency Phonology*. Cambridge University Press.

Andrews, Henrietta (1949): "Phonemes and morphophonemes of Temoayan Otomi", *International Journal of American Linguistics* **15**: 213-222.

Archangeli, Diana, & Douglas Pulleyblank (1994): *Grounded Phonology*. MIT Press, Cambridge.

Avery, P., & Keren Rice (1989): "Segment structure and coronal underspecification", *Phonology* **6**: 179-200.

Barker, M.A.R. (1964): *Klamath Grammar*. University of California Press, Berkeley and Los Angeles.

Beach, D.M. (1938): *The Phonetics of the Hottentot Language*. Heffer, Cambridge.

Bergem, Dick R. van (1993): "Acoustic vowel reduction as a func tion of sentence accent, word stress and word class", *Speech Communication* **12**: 1-23.

Bhaskararao, Peri, & Peter Ladefoged (1991): "Two types of voiceless nasals", *Journal of the International Phonetic Association* 80-88.

Bird, Steven, & T. Mark Ellison (1994): "One-Level Phonology: autosegmental representations and rules as finite automata", *Computational Linguistics* **20-1**: 55-90.

Bird, Steven, & Ewan Klein (1990): "Phonological events", *Journal of Linguistics* **26**: 33-56.

Blevins, Juliette (1995): "The syllable in phonological theory", in: John A. Goldsmith (ed.), pp. 206-244.

Blight, Richard C., & Eunice V. Pike (1976): "The phonology of Tenango Otomi", *International Journal of Amercian Linguistics* **42**: 51-57.

Boë, L.J., Pascal Perrier, B. Guérin, & J.L. Schwartz (1989): "Maximal vowel space", *EuroSpeech '89*, **2**: 281-284.

Boë, Louis-Jean, Jean-Luc Schwartz, & Nathalie Vallée (1994): "The prediction of vowel systems: perceptual contrast and stability", in: Eric Keller (ed.): *Fundamentals of Speech Synthesis and Speech Recognition*, pp. 185-213. Wiley, London & New York.

Boersma, Paul (1989): "Modelling the distribution of consonant inventories by taking a functionalist approach to sound change", *Proceedings of the Institute of Phonetic Sciences of the University of Amsterdam* **13**: 107-123.

Boersma, Paul (1993): "An articulatory synthesizer for the simulation of consonants", *Proceedings Eurospeech '93*: 1907-1910.

Boersma, Paul (1995): "Interaction between glottal and vocal-tract aerodynamics in a comprehensive model of the speech apparatus", *International Congress of the Phonetic Sciences* **2**: 430-433.

Boersma, Paul (fc. a): "Spreading in Functional Phonology", ms. [to appear on the ROA]

Boersma, Paul (fc. b): "The OCP in Functional Phonology", ms. [to appear on the ROA]

Boersma, Paul (forthcoming c): "Inventories in Functional Phonology", ms. [to appear on the ROA]

Boersma, Paul (fc. d): "Sound change in Functional Phonology", ms. [to appear on the ROA]

Boersma, Paul (fc. e): "Stratification in Functional Phonology", ms. [to appear on the ROA]

Boersma, Paul (in progress): *Functional Phonology*. Doctoral thesis, Universiteit van Amsterdam. To be published in Studies in Language and Language Use.

Bosch, Louis ten (1991): *On the Structure of Vowel Systems. Aspects of an extended vowel model using effort and contrast*. Doctoral thesis, Universiteit van Amsterdam.

Brentari, Diane (1995): "Sign language phonology: ASL", in: John A. Goldsmith (ed.), pp. 615-639.

Browman, Catherine P., & Louis Goldstein (1984): "Dynamic modeling of phonetic structure", *Haskins Status Report on Speech Research* **79/80**: 1-17. Also in Victoria A. Fromkin (ed., 1985): *Phonetic Linguistics: Essays in Honour of Peter Ladefoged*, Academic Press, pp. 35-53.

Browman, Catherine P., & Louis Goldstein (1986): "Towards an articulatory phonology", in: C. Ewan & J. Anderson (eds.): *Phonology Yearbook* **3**, Cambridge University Press, pp. 219-252.

Browman, Catherine P., & Louis Goldstein (1989): "Articulatory gestures as phonological units", *Phonology* **6**: 201-251.

Browman, Catherine P., & Louis Goldstein (1990a): "Tiers in articulatory phonology, with some applications for casual speech", in: Kingston & Beckman (eds.), pp. 341-376.

Browman, Catherine P., & Louis Goldstein (1990b): "Gestural specification using dynamically-defined articulatory structures", *Journal of Phonetics* **18**: 299-320.

Browman, Catherine P., & Louis Goldstein (1992): "Articulatory Phonology: an overview", *Phonetica* **49**: 155-180.

Browman, Catherine P., & Louis Goldstein (1993): "Dynamics and articulatory phonology", *Haskins Status Report on Speech Research* **113**: 51-62.

Chomsky, Noam, & Morris Halle (1968): *The Sound Pattern of English*. Harper and Row, New York.

Clements, G. N. (1985): "The geometry of phonological features", *Phonology Yearbook* **2**: 225-252.

Clements, G. N. (1987): “Phonological feature representation and the description of intrusive stops”, *Chicago Linguistic Society Parasession* **23**: 29-50.

Clements, G. N., & Elizabeth V. Hume (1995): “The internal organization of speech sounds”, in: John A. Goldsmith (ed.): *The Handbook of Phonological Theory*, Blackwell, Cambridge & Oxford, pp. 245-306.

Clements, G.N., & S.J. Keyser (1983): *CV Phonology: A Generative Theory of the Syllable*. MIT Press, Cambridge (USA).

Cole, Jennifer S., & Charles W. Kisseberth (1994): “An optimal domains theory of harmony”, *Studies in the Linguistic Sciences* **24-2**. [ROA-22]

Cole, Jennifer S., & Charles W. Kisseberth (1995): “Restricting multi-level constraint evaluation: opaque rule interaction in Yawelmani vowel harmony”, ms. University of Illinois. [ROA-98]

Cornyn, William S. (1944): *Outline of Burmese Grammar*. *Language*, Supplement 20-4.

Dantsuji, Masatake (1982): “An acoustic study on glottalized vowels in the Yi (Lolo) language – a preliminary report –”, *Studia Phonologica* **XVI**: 1-11.

Dantsuji, Masatake (1984): “A study on voiceless nasals in Burmese”, *Studia Phonologica* **XVIII**: 1-14.

Dantsuji, Masatake (1986): “Some acoustic observations on the distinction of place of articulation for voiceless nasals in Burmese”, *Studia Phonologica* **XX**: 1-11.

Dik, Simon C. (1989): *The Theory of Functional Grammar. Part I: The Structure of the Clause*. Foris, Dordrecht.

Ewen, Colin, & Harry G. van der Hulst (1987): “Single-valued features and the distinction between [–F] and [ØF]”, in: F. Beukema & Peter Coopmans (eds.): *Linguistics in the Netherlands 1987*, Foris, Dordrecht, pp. 51-60.

Fourakis, Marios, & Robert Port (1986): “Stop epenthesis in English”, *Journal of Phonetics* **14**: 197-221.

Gandour, Jack (1974): “The features of the larynx: n-ary or binary?”, *UCLA Working Papers in Phonetics* **27**: 147-159.

Gentil, Michèle (1990): “Organization of the articulatory system: peripheral mechanisms and central coordination”, in: William J. Hardcastle & Alain Marchal (eds.): *Speech Production and Speech Modelling*, Kluwer, Dordrecht, pp. 1-22.

Goldsmith, John (1976a): *Autosegmental Phonology*. Doctoral thesis, MIT, Cambridge. [Indiana University Linguistics Club, Garland Press, New York 1979]

Goldsmith, John (1993): “Harmonic Phonology”, in: Goldsmith (ed.), pp. 21-60.

Goldsmith, John (ed., 1993): *The Last Phonological Rule: Reflections on Constraints and Derivations*. University of Chicago Press.

Goldsmith, John A. (ed., 1995): *The Handbook of Phonological Theory*. Blackwell, Cambridge (USA) & Oxford.

Gupta, J.P., S.S. Agrawal, & Rais Ahmed (1968): “Perception of (Hindi) consonants in clipped speech”, *Journal of the Acoustical Society of America* **45**: 770-773.

Hammond, Mike (1995): “There is no lexicon!”, ms. University of Arizona, Tucson. [ROA-43]

Hardcastle, William J. (1976): Physiology of Speech Production. An Introduction for Speech Scientists. Academic Press, London.

Hayes, Bruce (1989): “Compensatory lengthening in moraic phonology”, *LI* **20**: 253-306.

Hayes, Bruce (1995): “A phonetically-driven, optimality-theoretic account of post-nasal voicing”, handout Tilburg Derivational Residue Conference. [ROA-93g]

Hayes, Bruce (1996a): “Phonetically driven optimality-theoretic phonology”, handout of a LOT course, Utrecht.

Hayes, Bruce (1996b): “Phonetically driven phonology: the role of Optimality Theory and inductive grounding”, to appear in: *Proceedings of the 1996 Milwaukee Conference on Formalism and Functionalism in Linguistics*. [ROA-158]

Hulst, Harry G. van der (1988): “The dual interpretation of |I|, |U|, |A|”, *Proceedings of NELS* **18**: 208-222. GLSA, University of Massachusetts, Amherst.

Hulst, Harry G. van der (1989): “Atoms of segmental structure: components, gestures, and dependency”, *Phonology* **6**: 253-284.

Hyman, Larry (1985): *A Theory of Phonological Weight*. Foris, Dordrecht.

Itô, Junko, Armin Mester, & Jaye Padgett (1995): “NC: Licensing and underspecification in Optimality Theory”, *Linguistic Inquiry* **26**: 571-613.

Jun, Jongho (1995): “Place assimilation as the result of conflicting perceptual and articulatory constraints”, *Proceedings of the West Coast Conference on Formal Linguistics* **14**.

Kawasaki, Haruko (1982): *An Acoustical Basis for Universal Constraints on Sound Sequences*. Doctoral thesis, University of California, Berkeley.

Keating, Patricia A. (1990): “The window model of coarticulation: articulatory evidence”, in: Kingston & Beckman (eds.), pp. 451-470.

Kelso, J.A.S., E.L. Saltzman, & B. Tuller (1986): “The dynamical perspective on speech production: data and theory”, *Journal of Phonetics* **14**: 29-59.

Kenstowicz, Michael (1994): *Phonology in Generative Grammar*. Blackwell, Cambridge (USA) & Oxford.

Kingston, John, & Mary E. Beckman (eds., 1990): *Papers in Laboratory Phonology I: Between the Grammar and Physics of Speech*. Cambridge University Press.

Kiparsky, Paul (1985): “Some consequences of lexical phonology”, *Phonology Yearbook* **2**: 85-138.

Koopmans-van Beinum, Florien J. (1980): *Vowel Contrast Reduction. An Acoustic and Perceptual Study of Dutch Vowels in Various Speech Conditions*. Doctoral thesis, University of Amsterdam.

Labov, William (1994): *Principles of Linguistic Change. Volume I: Internal Factors*. Blackwell, Oxford.

Ladefoged, Peter (1971): *Preliminaries to Linguistic Phonetics*. University of Chicago Press.

Ladefoged, Peter (1973): “The features of the larynx”, *Journal of Phonetics* **1**: 73-83.

Ladefoged, Peter (1990a): “On dividing phonetics and phonology: comments on the papers by Clements and by Browman and Goldstein”, in: John Kingston & Mary E. Beckman (eds.), pp. 398-405.

Ladefoged, Peter (1990b): “Some reflections on the IPA”, *UCLA Working Papers in Phonetics* **74**: 61-76.

Ladefoged, Peter (1995): “Voiceless approximants in Tee”, *UCLA Working Papers in Phonetics* **91**: 85-88.

Lakoff, George (1993): “Cognitive Phonology”, in: Goldsmith (ed.), pp. 117-145.

Leben, William (1973): *Suprasegmental Phonology*. Doctoral thesis, MIT, Cambridge. Garland Press, New York 1980.

Levin, Juliette (1985): *A Metrical Theory of Syllabicity*. Doctoral thesis, MIT, Cambridge (USA).

Liljencrants, Johan, & Björn Lindblom (1972): “Numerical simulation of vowel quality systems: the role of perceptual contrast”, *Language* **48**: 839-862.

Lindau, Mona (1975): *[Features] for vowels*. University of California Los Angeles Working Papers in Phonetics 30.

Lindblom, Björn (1963): “Spectrographic study of vowel reduction”, *Journal of the Acoustical Society of America* **35**: 1773-1781.

Lindblom, Björn (1990a): “Models of phonetic variation and selection”, *Phonetic Experimental Research, Institute of Linguistics, University of Stockholm* **XI**: 65-100.

Lindblom, Björn (1990b): “Explaining phonetic variation: a sketch of the H&H theory”, in” William J. Hardcastle & Alain Marchal (eds.): *Speech Production and Speech Modelling*, pp. 403-439. Kluwer, Dordrecht.

Lindblom, Björn, James Lubker & Thomas Gay (1979): “Formant frequencies of some fixed-mandible vowels and a model of speech motor programming by predictive simulation”, *Journal of Phonetics* **7**: 147-161.

Lindblom, Björn, & Michael Studdert-Kennedy (1967): “On the rôle of formant transitions in vowel recognition”, *Journal of the Acoustical Society of America* **42**: 830-843.

Lombardi, Linda (1995): “Why Place and Voice are different: constraint interactions and featural faithfulness in Optimality Theory”, ms. University of Maryland. [ROA-105]

Maddieson, Ian (1984): *Patterns of Sounds*. Cambridge University Press.

Maddieson, Ian, & Victoria Balboa Anderson (1994): “Phonetic structures of Iaai”, *UCLA Working Papers in Phonetics* **87**: 163-182.

McCarthy, John (1988): “Feature geometry and dependency: a review”, *Phonetica* **45**: 84-108.

McCarthy, John J. (1995): “Remarks on phonological opacity in Optimality Theory”, to appear in: Jacqueline Lecarme, Jean Lowenstamm, & Ur Shlonsky (eds.): *Proceedings of the Second Colloquium on Afro-Asiatic Linguistics*. [ROA-79]

McCarthy, John, J. & Prince, Alan (1986): *Prosodic Morphology*. ms, University of Massachusetts and Brandeis University.

McCarthy, John, & Alan Prince (1993a): *Prosodic Morphology I: Constraint Interaction and Satisfaction*. Ms., University of Massachusetts, Amherst, and Rutgers University, New Brunswick.

McCarthy, John, & Alan Prince (1993b): “Generalized alignment”, in: Geert Booij & Jaap van Marle (eds.): *Yearbook of Morphology 1993*, pp. 79-153. Kluwer, Dordrecht. [ROA-7]

McCarthy, John, & Alan Prince (1994): “The emergence of the unmarked: optimality in prosodic morphology”, in: Mercè Gonzàlez (ed.): *Proceedings of the North East Linguistic Society* **24**, Graduate Linguistic Student Association, Amherst, pp. 333-379. [ROA-13]

McCarthy, John, & Alan Prince (1995): “Faithfulness and reduplicative identity”, in: J. Beckman, S. Urbanczyk, & L. Walsh (eds.): *Papers in Optimality Theory*, Occasional Papers **18**, University of Massachusetts, Amherst, pp. 249-384. [ROA-60]

Mohanan, K.P. (1986): *The Theory of Lexical Phonology*. Reidel, Dordrecht.

Mohanan, K.P. (1993): “Fields of attraction in phonology”, in: John Goldsmith (ed.), pp. 61-116.

Newman, Stanley (1944): *Yokuts Language of California*. Viking Fund Publications in Antrhopology **2**, New York.

Ohala, John J. (1975): “Phonetic explanations for nasal sound patterns”, in: C.A. Ferguson, L.M. Hyman & J.J. Hyman (eds.): *Nasálfest*, Stanford University, pp. 289-316.

Ohala, John J. (1976): “A model of speech aerodynamics”, *Report of the Phonology Laboratory at Berkeley* **1**: 93-107.

Okell, J. (1969): *A Reference Grammar of Colloquial Burmese*. Oxford University Press.

Orgun, Cemil Organ (1995): “Correspondence and identity constraints in two-level Optimality Theory”, *West-Coast Conference on Formal Linguistics* **14**.

Padgett, Jaye (1995): “Partial class behaviour and nasal place assimilation”, *Proceedings of the Arizona Phonology Conference: Workshop on Features in Optimality Theory*. Coyote Working Papers, University of Arizona, Tucson. [ROA-113]

Passy, Paul (1890): *Etude sur les changements phonétiques et leurs caractères généraux*. Librairie Firmin - Didot, Paris.

Pater, Joe (1996): “Austronesian nasal substitution and other NC̥ effects”, to appear in the *Proceedings of the Utrecht Prosodic Morphology Workshop*. [ROA-160]

Pols, Louis C.W. (1983): “Three-mode principal component analysis of confusion matrices, based on the identification of Dutch consonants, under various conditions of noise and reverberation”, *Speech Communication* **2**: 275-293.

Prince, Alan, & Paul Smolensky (1993): *Optimality Theory: Constraint Interaction in Generative Grammar*. Rutgers University Center for Cognitive Science Technical Report 2.

Recasens, Daniel (1991): *Fonètica Descriptiva del Català*. Biblioteca Filològica XXI, Institut d’Estudis Catalans, Barcelona.

Reenen, Pieter Thomas van (1981): *Phonetic Feature Definitions. Their Integration into Phonology and their Relation to Speech. A Case Study of the Feature NASAL*. Doctoral thesis, Vrije Universiteit, Amsterdam. Foris, Dordrecht.

Sagey, Elizabeth (1986): *The Representation of Features and Relations in Nonlinear Phonology*. Doctoral thesis, MIT, Cambridge.

Schachter, Paul, & Fe T. Otanes (1972): *Tagalog Reference Grammar*. University of California Press, Berkeley.

Schwartz, Jean-Luc, Louis-Jean Boë, Pascal Perrier, B. Guérin, & P. Escudier (1989): “Perceptual contrast and stability in vowel systems: a 3-D simulation study”, *Eurospeech ’89* **2**: 63-66.

Schwartz, Jean-Luc, Louis-Jean Boë, & Nathalie Vallée (1995): “Testing the dispersion-focalization theory: phase spaces for vowel systems”, *International Congress of the Phonetic Sciences* **13-1**: 412-415.

Scobbie, James (1993): “Issues in constraint violation and conflict”, in: T. Mark Ellison & James M. Scobbie (eds.): *Computational Phonology*. Edinburgh Working Papers in Cognitive Science **8**.

Siewierska, Anna (1991): *Functional Grammar*. Routledge, London.

Silverman, Daniel, Barbara Blankenship, Paul Kirk, & Peter Ladefoged (1994): “Phonetic structures in Jalapa Mazatec”, *UCLA Working Papers in Phonetics* **87**: 113-130.

Son, Rob J.J.H. van, & Louis C.W. Pols (1992): “Formant movements of Dutch vowels in a text, read at normal and fast rate”, *Journal of the Acoustical Society of America* **92**: 121-127.

Sprigg, R.K. (1965): “Prosodic analysis and Burmese syllable initial features”, *Anthroplogical Linguistics* **7-6**: 59-81.

Steriade, Donca (1995): “Underspecification and markedness”, in: Goldsmith (ed.), pp. 114-174.

Stevens, Kenneth N. (1989): “On the quantal nature of speech”, *Journal of Phonetics* **17**: 3-45.

Stevens, Kenneth N. (1990): “Some factors influencing the precision required for articulatory targets: comments on Keating’s paper”, in: Kingston & Beckman (eds.), pp. 471-475.

Stevens, Kenneth N., Samuel Jay Keyser, & Haruko Kawasaki (1986): “Toward a phonetic and phonological theory of redundant features”, in: Joseph S. Perkell & Dennis H. Klatt (eds.): *Invariance and Variability in Speech Processes*, Lawrence Erlbaum, Hillsdale, pp. 426-449.

Tesar, Bruce, & Paul Smolensky (1995): “The learnability of Optimality Theory: an algorithm and some basic complexity results”, ms. Department of Computer Science & Institute of Cognitive Science, University of Colorado at Boulder. [ROA-2]

Trask, R.L. (1996): *A Dictionary of Phonetics and Phonology*. Routledge, London.

Vallée, Nathalie (1994): *Systèmes vocaliques: de la typologie aux prédictions*. Doctoral thesis, Institut de la Communication Parlée, Université Stendhal, Grenoble.

Voegelin, C.F. (1956): “Phonemicizing for dialect study with reference to Hopi”, *Language* **32**: 116-135.

Weidert, Alfons (1975): *Componential Analysis of Lushai Phonology*. John Benjamins, Amsterdam.

Westbury, John R., & Patricia A. Keating (1986): “On the naturalness of stop voicing”, *Journal of Linguistics* **22**: 145-166. Also in *UCLA Working Papers in Phonetics* **60**: 1-19 (1985).

Yip, Moira (1988): “The obligatory contour principle and phonological rules: a loss of identity”, *LI* **19**: 65-100.

Zoll, Cheryl S. (1994): “Subsegmental parsing: floating features in Chaha and Yawelmani”, *Phonology at Santa Cruz* **3**. [ROA-29]

Zoll, Cheryl S. (1996): *Parsing below the Segment in a Constraint-Based Framework*. Doctoral thesis, University of California, Berkeley. [ROA-143]