CHAPTER 1

Introduction

1. CENTRAL THESIS
This work elucidates the nature of the notion of locality in phonology, describing the conditions under which a sound can assimilate to another sound. The central thesis is that articulatory contiguity between gestures is the correct notion of locality in phonology. The main body of this work is devoted to defending this point.

The argument supporting this central thesis builds on knowledge of the articulatory events taking place in the vocal tract during the production of speech. Consider, for example, the production of a simple CVC utterance, like [mat]. The basic fact of interest is that the articulation of the vowel extends throughout the whole syllable, overlapping with the surrounding consonant articulations. Thus in a bisyllabic CV₁.C.CV₂.C utterance, the vowel articulations are contiguous, much as they would be in a VV sequence. This is depicted in (1).

1. Vowel-to-vowel contiguity

Vowel-to-vowel contiguity is a universal property of language, known from various acoustic and articulatory studies, going as far back as Öhman’s seminal investigation of coarticulatory effects in Swedish, English, and Russian VCV sequences (Öhman 1966).
Consonant articulations, on the other hand, crucially lack the corresponding property of contiguity when separated by a vowel. In a CVC utterance, the consonantal articulations are not contiguous in the same sense as the two vowels in the VCV utterance are. This is because of a fundamental asymmetry between vowels and consonants. The invariant characteristic of consonantal gestures is the formation of a constriction at some small area in the vocal tract. On the other hand, the primary characteristic of vowel gestures is lack of such a constriction. Vowels are produced by movements of the tongue body and jaw, with relatively slow changes in the global shape of the vocal tract because of the larger mass involved. During the production of a VCV utterance, because a consonantal constriction is localized at some area in the vocal tract, the free part of the tongue is allowed to assume different shapes and thus move smoothly from the shape of the first vowel to that of the second vowel. In a CVC utterance, on the other hand, the intervening vowel will undo the constriction of the preceding consonant. A smooth transition between the two consonants would require maintaining a constriction throughout the articulation of the vowel, which would suppress the vowel completely, and is thus impossible. Hence, we have the following asymmetry: the vowel gestures are contiguous in a VCV utterance, but the consonant gestures are not contiguous in a CVC utterance. I refer to this asymmetry as V-to-V contiguity versus no C-to-C contiguity.

Using the gestural framework of Browman & Goldstein (1986 et seq.), the above asymmetry can be represented in the following way. In (2), we see the gestures of a simple VCV utterance, [ipa], and in (3) we see an example of a simple CVC utterance, [mat] (for /m/, I only show the oral gesture here). TB and TT stand for the major articulators involved in these sounds, the tongue body and tongue tip respectively. Gestures have duration indicated by the length of the boxes.

2. VCV: [ipa]

| TB | i | a |
| LIPS | p |
3. CVC: [mat]

These diagrams show that the gestures of the two vowels in the VCV are contiguous, while the gestures of the two consonants in a CVC are not.

I propose that articulatory contiguity is the correct notion of locality in phonology, as stated in (4) below. According to Articulatory Locality, the two vowels in [ita] are phonologically in a local relation, but the two consonants in [mat] are not.

4. Articulatory Locality

Two gestures of segments S1 and S2 are in a phonologically local relation if and only if their articulations are contiguous.

Note that Articulatory Locality is crucially different from the notion of locality as ‘strict-adjacency’ of phonological segments. In [ipa], because the C overlaps with the the two Vs, there are three simultaneous locality relations: i-p, p-a, i-a.

The most important prediction of Articulatory Locality is that all languages employ the same notion of locality in their phonologies. This is in contrast to various language-particular and in some cases rule-particular notions of locality that have been used in the past. One particularly popular notion of phonological locality has been formulated in terms of ‘tier-adjacency’ and is illustrated in two different variants shown in (5) below. In (5a), the Roots of vowels and consonants are segregated on different planes. This representation is known as the V/C Planar Segregation (McCarthy 1989). The second representation in (5b) shows the arrangement of the Place nodes of vowels and consonants on two different tiers (hence, Tier Segregation) in a particular model of feature geometry (Clements 1989). Planar segregation can be seen as a special case of Tier Segregation, where the two segregated autosegments are the Roots of vowels and consonants.
5. Two representations of a CVC

a. V/C Planar Segregation

```
<table>
<thead>
<tr>
<th>Root-V</th>
</tr>
</thead>
<tbody>
<tr>
<td>X</td>
</tr>
<tr>
<td>X</td>
</tr>
<tr>
<td>X</td>
</tr>
</tbody>
</table>
```

b. Tier (Place) Segregation

```
<table>
<thead>
<tr>
<th>X</th>
<th>X</th>
<th>X</th>
</tr>
</thead>
<tbody>
<tr>
<td>Root</td>
<td>Root</td>
<td>Root</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Root-C</td>
<td>Root-C</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>V-Place</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>C-Place</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>C-Place</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
```

There are important differences between Articulatory Locality and the tier-adjacency notion of locality. In terms of the latter, the two consonantal Roots in (5a) are adjacent on their plane and thus are in a local relation. Similarly, the two C-Place nodes of the consonants in (5b) are tier-adjacent and thus local relative to one another. This means, for example, that the second consonant in (5a) can fully assimilate to the first by spreading its Root node directly to the X slot of the first consonant, or, comparably in (5b), that any dependent of the C-Place of the first consonant (Labial, Coronal etc.) can spread to the C-Place node of the second consonant, yielding partial assimilations in terms of place of articulation. In contrast, Articulatory Locality treats the two consonants in the CVC as not in a local relation, because their articulations are not contiguous, and thus predicts that direct assimilations between the two consonants skipping over the vowel should not be found.

This dissertation argues that (a) the tier-adjacency notion of locality provides only an incomplete understanding of certain phonological phenomena, and that (b) each of these phenomena lead independently to the proposed notion of Articulatory Locality. As a brief illustration of the empirical range of phenomena that fall under the explanatory domain of Articulatory Locality, consider the following observations in (6a-b), indicating two incompletely understood areas of phonological patterning.

6. Some puzzles for previous notions of locality

a. To express certain regularities in the patterning of “nonconcatenative” languages (e.g. Arabic), phonological theory has claimed that an output such as C_iVC_iC where the two
Introduction

consonants are identical, may result from the familiar autosegmental operation that spreads the root of a single underlying consonant to the two C positions, as shown in (i) below. Spreading can proceed unobstructed by the intervening vowel because vowels and consonants lie on different planes.

\[
\begin{array}{c}
\text{i. V/C Segregation} \\
\text{V-Root} \\
\text{X X X} \\
\text{C}_1\text{-Root} \\
\text{C}_2\text{-Root} \\
\text{[cor]} \\
\text{[lab]} \\
\text{[voice]}
\end{array}
\quad
\begin{array}{c}
\text{ii. /nap/} \rightarrow [map] \\
\text{V-Root} \\
\text{X X X} \\
\text{C}_1\text{-Root} \\
\text{C}_2\text{-Root} \\
\text{[cor]} \\
\text{[lab]} \\
\text{[voice]}
\end{array}
\quad
\begin{array}{c}
\text{iii. /pad/} \rightarrow [bad] \\
\text{V-Root} \\
\text{X X X} \\
\text{C}_1\text{-Root} \\
\text{C}_2\text{-Root} \\
\text{[cor]} \\
\text{[lab]} \\
\text{[voice]}
\end{array}
\]

However, according to the representation in (i), we expect to find phonological interaction between a pair of consonants separated by a vowel of the same character as when the two consonants are string adjacent in a CC cluster. Specifically, we expect cases of long-distance assimilation as in the two examples in (ii) and (iii): (ii) shows a hypothetical process of place assimilation, /nap/ \rightarrow [map], and (iii) another process of voice assimilation, /pad/ \rightarrow [bad], both applying over the intervening vowel. Such processes are widely attested in CC clusters cross-linguistically, but are not attested over intervening vowels at all in either concatenative or nonconcatenative languages. Hence, the tier-adjacency notion of locality offers no answer to the question: \textit{why does “spreading” in a configuration like (i) always spread the whole consonant and never its place node as in (ii), or other individual features in the feature geometry as in (iii)?}

b. Vowel harmony is cross-linguistically well-attested, in sharp contrast to the more infrequent consonant harmony. Under the geometric notion of locality defined as tier-adjacency, this is unexpected: by placing consonant and vowel features on different tiers, and defining locality as tier-adjacency, vowel-to-vowel harmony over consonants and
consonant-to-consonant harmony over vowels are equally predicted.

Furthermore, it is also known that when consonant harmony is attested, it is apparently restricted to coronal consonants. This is the second set of issues which has defied a coherent explanation so far, raising the questions: *whence the asymmetry between vowel and consonant harmony, and why is consonant harmony restricted to coronal consonants?*

I argue that that each of the above dilemmas leads individually to the proposed revision of the notion of locality. Recall that the vowel articulations in a VCV sequence are contiguous while consonant articulations in a CVC sequence are not. Let us see how Articulatory Locality relates to the observations in (6).

The gap in phonological patterning noted in (6a) was that in a C1VC2 sequence the two consonants do not assimilate in the ways found in clusters, although the tier-adjacency notion of locality predicts that they should. According to Articulatory Locality, this follows from the lack of contiguity between the two consonants in the C1VC2 configuration. No contiguity implies no basis for a direct interaction, contra the prediction of the geometric notion of locality and apparently in accordance with the facts (including cases of ‘long-distance’ C-to-C assimilation discussed below). In other words, if spreading of features in assimilatory phenomena applies under Articulatory Locality, it follows that it could not be involved in a CVC configuration, spreading a feature of the consonant or the consonant itself over the vowel.

This leads us to the following question. If assimilation is not involved in creating the output C1VC1 (as in Arabic radadtu ‘I returned’) then how should we characterize the identity between the two consonants? As noted in (6a), when ever long-distance consonantal spreading is claimed to apply, it only ‘spreads’ the whole consonant and never less than that. This and other factors which I will discuss later suggest that what was seen as ‘spreading’ in C1VC1 is in fact a phenomenon of the same formal character as reduplication. This explains why the putative C-to-C ‘spreading’ always targets whole consonants: copying of segments, as in reduplication, only targets the whole segment, never individual features.
Turning to (6b), it was noted that vowel harmony is well-attested in contrast to the more infrequent consonant harmony. This is in direct accord with the articulatory facts motivating Articulatory Locality. On the one hand, V-to-V contiguity in VCV sequences predicts interaction between the two vowels, and hence the possibility of vowel harmony. On the other hand, absence of C-to-C contiguity in CVC sequences predicts absence of direct interaction between the two consonants, hence precluding the possibility of consonant harmony, a phenomenon which appears to involve such direct assimilations between two consonants skipping the intervening vowel (as in /saš/ → [saš] in Chumash-type harmonies). Consonant harmony, however, is attested only with a very specific class of consonants, whose properties as such as to make that phenomenon also consistent with (and predictable from) Articulatory Locality.

Consider that Articulatory Locality predicts that if assimilation is to apply between the two consonants in a CVC configuration it must be mediated by the vowel: the assimilating feature must propagate through the articulatorily intervening vowel. As noted in (6b), consonant harmony is attested only for coronal consonants. More specifically, a cross-linguistic investigation of consonant harmony reveals that the consonantal features which assimilate are limited to those which describe the mid-sagittal or cross-sectional shape of the tongue tip-blade, the major articulator of coronal consonants.

Two facts that are particular to the tip-blade allow us to understand this limitation. First, this articulator is independent from the tongue dorsum with which vowels are articulated. Second, the precise shape of the tip-blade, cross-sectional or mid-sagittal, does not have a significant effect on the acoustics of the intervening vowel. The restriction of consonant harmonies, then, to features which describe the shapes of the tip-blade (of coronal consonants) can be understood if the spreading of the tip-blade shape must pass through the intervening vowel, exactly as predicted by Articulatory Locality. If assimilation between the two consonants in the CVC configuration were not mediated by the vowel, contra Articulatory Locality and as predicted by the tier-adjacency notion of locality (in segregated representations), then there would be no explanation for the privileged status of coronals in consonant harmonies.

When talking about assimilation it is useful to keep in mind the
distinction between (a) the forces that require it, and (b) the conditions that allow it. In the discussion above, I am strictly concerned with what allows vowel harmonies or consonant harmonies to happen, and not with what requires them to happen. V-to-V contiguity allows vowel harmony to happen but does not require it. Other constraints of the grammar do so. Similarly, articulatory independence of the tip-blade from the tongue dorsum and acoustic irrelevance of the tip-blade posture for intervening vowels allows (V-mediated) coronal harmony to happen. There are other constraints which require assimilation between coronal consonants in a CVC. These and other related issues will be discussed at length in chapter 5.

2. THEORETICAL BACKGROUND

In this work, I employ the representations of Articulatory Phonology of Browman & Goldstein (1986) in combination with the conception of grammar as developed in Optimality Theory of Prince & Smolensky (1993). Articulatory Phonology and Optimality Theory are two originally independent research programs, whose combination I believe holds considerable promise for elucidating the phonetic and phonological aspects of speech. I assume some minimal familiarity on the part of the reader with the latter theory. Recent elaborations within Optimality Theory, whenever crucial to the proposed analyses, are introduced and laid out explicitly. Articulatory Phonology, less known within the practice of mainstream phonology, is introduced in the next section.

The types of representations employed in work within Optimality Theory so far are those proposed within autosegmental phonology (Goldsmith 1976). A segment, for example, is seen as a set of features linked to a Root node either directly or through further organizational nodes, such as Place, in what is called a feature geometry (Clements 1985, Sagey 1986, McCarthy 1988). In Articulatory Phonology, on the other hand, the primitive units of the representations are not features but gestures and their gestural parameters. Moreover, whereas most of the work done within Optimality Theory has concentrated on mainstream phonological areas of research such as prosody, harmony, morphophonological alternations etc., Articulatory Phonology has focused on other phenomena, such as connected speech processes, allophonic
variation, historical sound change etc.

In the approach I take in this dissertation, which relies on both Articulatory Phonology and Optimality Theory, it is necessary to bridge various gaps between the two by making specific assumptions. These assumptions reflect my intuition on how the two must fit together. In essence, Articulatory Phonology provides the representations which the constraints of an Optimality Theoretic grammar evaluate. After an introduction to the relevant basic aspects of Articulatory Phonology in 2.1, I introduce these more specific assumptions in 2.2.

2.1 Gestures in Articulatory Phonology

In this section, I succintly present the formalization of articulatory gestures in the model of Articulatory Phonology of Browman & Goldstein (1986, 1989, 1990a-b, 1991a-b, 1992a-b). The primary reason for adopting this model is that the notion of locality I propose relies crucially on articulatory considerations.

Articulatory Phonology presents a formal model of gestures, the linguistically significant articulatory movements, which generally speaking consist of the formation of a constriction by some articulator at some place in the vocal tract. A set of parameters, called vocal tract variables, specify the spatial goals of the constriction. This specification consists of the coordinative structure, the ‘articulator set,’ employed in producing the constriction, the constriction location (CL) and the constriction degree (CD). For example, a gesture involving the tongue body (TB) is parametrized by the values of two vocal tract variables: the constriction location, CL or TBCL, and the constriction degree, CD or TBCD. The temporal characteristics of a gesture are specified by a dynamic parameter called ‘stiffness,’ which roughly encodes the time it takes for the gesture to reach its spatial target.

In (7) below, I show the organization of the vocal tract variables according to the major articulator involved in the constriction. At the top level we see a division among three types of gestures: laryngeal, oral, and velic. For the gestures of the larynx and the velum there is no notion of constriction place, because their corresponding articulator also specifies the place of articulation; so only the value of the constriction degree variable needs to be specified. For every gesture, the parameter of stiffness
The Articulatory Basis of Locality in Phonology

is also specified but it is not shown in (7). Oral gestures divide over the two major oral articulators, the lips and the tongue. The latter further subdivides into the three lingual articulators, the tip, the body, and the root of the tongue. For example, an oral gesture that involves the tongue body (TB) must specify the values for two parameters: its constriction location, CL or TBCL, and its constriction degree, CD or TBCD.

7. Tract variable organization according to major articulator

---

There are additional variables which specify the constriction shape (CS) of oral articulators, and capture distinctions such as apical vs. laminal for the tongue tip. Constriction shapes and their role in consonant harmonies are discussed at length in chapters 4 and 5.

Gestures can contrast on the basis of their tract variables. The CD variable, for example, takes on a range of values with five categorical distinctions: [closed], [critical], [mid], [narrow], and [wide]. The first two values correspond to the binary feature [±continuant] of Chomsky & Halle (1968), with stops assigned the [closed] value and fricatives the [critical] value. The other three values are used for approximants and vowels. The value [mid] corresponds roughly to the [approximant] category of Catford (1977), but can also serve to provide distinctions of height between vowels. For example, front vowels differing in height may contrast in terms of the three values [mid], [narrow], and [wide] (see also Clements’ 1991b ‘aperture’ proposal; cf. Lindau 1978).

CL specifies the place of the constriction in the vocal tract. It takes the values of [labial], [dental], [alveolar], [postalveolar], [palatal], [velar], [uvular], and [pharyngeal] (based on proposals by Ladefoged 1989). CL serves to make the same distinctions in terms of place of articulation as do
the dependent features of the articulator nodes in feature geometric representations: [±anterior], [±distributed], [±back] etc.

Finally, the tract variable of stiffness is currently being used to encode the categorical distinction between consonantal and vocalic gestures. Browman & Goldstein (1992a) note that “consonantal gestures typically have . . . a shorter time constant (higher stiffness) than vocalic gestures” (p. 164). Stiffness, then, corresponds to the feature [±consonantal] of Chomsky & Halle (1968).

The actual continuous movement trajectory for a gesture unfolds from a dynamics model, the task dynamics of Saltzman & Kelso (1987). Part of the task dynamic characterization of a gesture is a mathematical model of the continuous articulatory trajectory involved in attaining the functional goal or ‘task’ of the gesture, that is, the formation of the constriction as specified by CD and CL. Each of these two tract variables of a gesture is modeled using a differential equation, the standard mathematical way for describing dynamical systems. The particular equation used models the tract variable as a damped mass-spring system (a “point-attractor”, Abraham & Shaw 1982). The solution to that equation describes the continuous trajectory movement of a gesture. An important property of the task dynamic model is that when gestures overlap over time their influences on common articulators are blended. This property of the model captures one of the most fundamental facts about speech, the context-dependency observed in the physical realization of gestural units (see Browman & Goldstein 1990a and Saltzman 1995 for illustrations).

Turning now to basic issues of representation, in Articulatory Phonology an utterance is represented as a sequence of gestures and their timing relations. The representation of a simple utterance, /pʰa/, is shown below. There are three boxes representing three gestures: the labial closing gesture for /p/, the laryngeal opening gesture for the aspiration, and the pharyngeal vocalic open gesture for the vowel /a/. The length of a box indicates the stretch of time during which the tract variables of the corresponding gesture are assumed to be under the active control of the dynamic system.
8. Gestural representation of /pʰa/

<table>
<thead>
<tr>
<th></th>
<th>CD wide</th>
</tr>
</thead>
<tbody>
<tr>
<td>Larynx</td>
<td></td>
</tr>
<tr>
<td>Lips</td>
<td>CL bilabial, CD closed</td>
</tr>
<tr>
<td>Tongue Body</td>
<td>CL pharyngeal, CD narrow</td>
</tr>
</tbody>
</table>

In this representation there are two lines connecting gestures which express the fact that gestures are temporally organized or coordinated. Coordination between two gestures is achieved by synchronizing one point in the temporal extent of one gesture with another point in the temporal extent of the other gesture. For example, for the gestures of /pʰa/ above, the onset of the movement for the labial and tongue body gestures are synchronized to begin at the same time; typically, this precise timing information is not shown in representations such as the above. The aspiration gesture is also coordinated with the labial closure. In this case, however, it is the point of the achievement of the target of the laryngeal gesture, namely, the attainment of the CD wide opening for the larynx, which is synchronized with the release of the labial closure. Identifying specific points in the temporal extent of a gesture that may be synchronized with other gestures is currently an area of active research.

Browman & Goldstein hypothesize that points such as achievement of target (in terms of CD and CL), onset of movement, and release of closure are part of a small set of points that are available for synchronization (Browman & Goldstein 1990b: p. 309).

Furthermore, not all gestures are coordinated with all others. Browman & Goldstein (1988) argue that consonants are coordinated with a vowel in a way that depends on their syllabic position. In a cluster like spl in [splats], the temporal midpoint of all underlined consonantal gestures gives the most stable measure of a relation with the vowel gesture. This point is called the C-center of the cluster. For postvocalic consonants, on the other hand, it is the achievement of the target of the leftmost consonant that is coordinated with the vowel (see Browman & Goldstein 1988 and Smith 1991 for further discussion).
In an item like [splats], then, the prevocalic consonant cluster behaves as a unit for the purposes of coordination with the following vowel. This type of coordination can be seen as an articulatory manifestation of the syllabic constituent onset. This allows for a characterization of syllables as patterns of coordination among gestures.

The same strategy is applied to the phonological unit of a segment. It is proposed that two modes of gestural overlap, complete and partial, are characteristic of subsegmental gestural coordination. For example, complete overlap between a velic and an oral gesture results in the percept of a single consonant [n]. When the velic gesture is slightly slid leftward (i.e. anticipated) the resulting overlap is partial and gives the percept of the so-called ‘pre-nasal’ stop [\textsuperscript{\textasciitilde}d]. Finally, a third degree of overlap, minimal overlap, would give rise to the percept of a velic gesture dissociated from the oral gesture, the cluster [nd] (Browman & Goldstein 1991a: pp. 319 ff.).

Browman & Goldstein’s hypothesis for why it is only those three degrees of overlap that are relevant to intra-segmental coordination makes connection with Stevens’ quantal theory of speech. Stevens (1972, 1989) has shown that the relation between acoustic output and corresponding articulatory gestures is quantal in the following sense. There are certain ranges of articulatory parameter variation within which the acoustic output remains relatively stable. In other ranges, however, small variations in the articulatory parameter cause large variations in the quality of the acoustic output. Browman & Goldstein then hypothesize that their three degrees of gestural overlap correspond to a partition of the continuum of overlap into three quantal regions so that “certain qualitative acoustic events may emerge at some critical degree of overlap” (1991a: p. 322). This is essentially an extension of Stevens’ quantal proposal to the dimension of time. Thus three categorical distinctions arise within a continuum of overlap. Within the range of synchronous gestures the resulting output is perceived as distinct from the output produced from gestures within the range of partial overlap. The latter is in turn perceived as distinct from the output produced by gestures in the range of minimal overlap.\textsuperscript{5}

Browman & Goldstein write that the “current research strategy is to see how much structure inheres directly in the relations among gestures, without recourse to higher level nodes” (1989: p. 233). Note that this
statement does not mean that the indispensable units of phonological structure, segments and syllables, are not there in the phonology. Rather, within the model of Articulatory Phonology, the aim is to elucidate the nature of these higher phonological units in terms of the primitives of the model, namely, gestures and their relations. In this connection, it should be noted that previous proposals to define the notion of syllable in terms of articularatory/aerodynamic parameters have proved ineffective. The most explicit of them, proposed by Stetson (1951), hypothesized that syllables correspond to the "pulses" created by contractions of the intercostal muscles, which control lung volume during speech. It was later shown, however, by studies of pulmonary air pressure during speech that this hypothesis could not be maintained (Draper, Ladefoged, and Whitteridge 1960, cf. Lieberman 1967). These studies revealed that lung pressure is kept relatively steady over the course of the production of a sentence and that the slight variations in pressure do not correspond perfectly to the "syllable pulses" as suggested by Stetson. Articulatory Phonology, in contrast, presents specific proposals, results, and predictions about how syllables emerge out of characteristic patterns of gestural coordination.

2.2 Specific Assumptions
Browman & Goldstein have discussed a number of speech production phenomena, such as casual speech processes, allophonic variations, and sound changes for which gestural representations provide significant insights (see below). The focus of the present work is on a somewhat different empirical domain. Specifically, I wish to explore the extent to which representations of overlap and gestural contiguity can shed light on certain phenomena that have been the concern of mainstream phonological research, such as vowel harmony, consonant harmony, and nonconcatenative morphophonology. This necessitates some extensions to the representations of Articulatory Phonology so that the constraints of Optimality Theory can apply to those representations.

To motivate these extensions, let me discuss briefly some of the phenomena that have been investigated within Articulatory Phonology. Consider the casual speech pronunciation of 'seven plus' /sevn plas/ as [sevm plas]. Browman & Goldstein (1989, 1991) have shown that this phenomenon does not actually involve the assimilation of the nasal to the
following labial with concomitant loss of the alveolar constriction of /n/. Instead, what actually happens is that the alveolar closure of the tongue tip for the nasal /n/ is completely overlapped by two other gestures, the dental fricative closure of /v/ and the labial closure of /p/. This is shown in (9), which is a schematic view of an actual articulatory tracing of [svmplas]. The alveolar gesture of the nasal is thus not deleted (although it may be slightly decreased in magnitude). All the gestures are present but their timing relations have been altered because of the fast speech rate.

9. Gestural representation of the underlined fragment of [svmplas]

<table>
<thead>
<tr>
<th>VEL</th>
<th>wide</th>
</tr>
</thead>
<tbody>
<tr>
<td>TT</td>
<td>n: closed alv</td>
</tr>
<tr>
<td>LIPS</td>
<td>v: crit dent</td>
</tr>
</tbody>
</table>

A similar account can be given for other casual speech phenomena, such as the apparent segmental deletion in the pronunciation of 'perfect memory' /p̩ʊ̱fəkt məməri/ as [p̩ʊ̱fəkm̩əməri]. Here the effect of deletion results from the fact that the alveolar closure of /t/ is completely overlapped by the constrictions of the preceding velar closure of /k/ and the following labial closure of /m/.

Browman & Goldstein argue convincingly that an account of these phenomena in terms of gestural overlap is superior to other accounts that have treated them in terms of rules of autosegmental spreading and delinking (e.g. Clements 1985). The reason is that an autosegmental account of such apparent deletions posits that a place of articulation in [svmplas] or a whole segment in [p̩ʊ̱fəkm̩əməri] gets deleted and thus predicts that the output should contain no trace of the deleted gestures. Browman & Goldstein, however, provide actual gestural tracings which clearly show that the putatively deleted gestures are still present. It is only their temporal relations with respect to other gestures that have been altered.

Many other seemingly disparate phenomena, ranging from historical changes (Latin intervocalic /b/ lenited to a fricative in modern Romance
languages, as in Latin habere ‘have’, Italian avere, French avoir), to synchronic weakenings (because pronounced as [pxøz]), and apparent ‘insertions’ (something pronounced as [sʌmpθɪŋ]), receive simple explanations in terms of the few gestural manipulations between the timing relations and the magnitude of the gestures involved (see Browman & Goldstein 1991a for demonstrations).

The repertoire of gestural manipulations that seem necessary to deal with the phenomena discussed above is rather limited. It includes a small set of operations such as temporal sliding, which causes increase in temporal overlap, magnitude reduction and temporal extension of a gesture. Note in particular that gestural insertions and deletions are not part of this inventory. Browman & Goldstein are of course aware that phonology cannot entirely be reduced to some set of gestural manipulations. As they acknowledge, “other principles and sources of constraints are no doubt required to completely explicate patterns in phonology” (Browman & Goldstein 1991a: p. 334). They cite examples of phenomena that have been dubbed as “crazy rules” (Bach & Hams 1972). These are rules which defy a coherent statement in terms of natural classes of sounds, and are usually the results of historical restructurings through rule inversion, telescoping etc. (Browman & Goldstein cite Wang 1969, Vennemann 1972, cf. Anderson 1981).

I believe that one does not have to resort to exotic cases of “crazy rules” to find domains of sound structure where the limited set of gestural manipulations such as overlap, reduction, and extension, are not sufficient. Much of current research in phonology deals with systematic morphophonological alternations which, I believe, are not reducible to gestural manipulations of overlap, reduction, and extension. Take for example the submodule of the grammar called Prosodic Morphology (McCarthy & Prince 1986 et seq.). Reduplication, a widespread and well-researched phenomenon in this submodule, has been handled by a set of universal constraints that have no apparent relation to the primitives of gestural extensions, reductions, and overlap (see McCarthy & Prince 1995a and references therein). Briefly, the reduplicant is a morpheme that stands in a systematic relationship to its base. The output of reduplicative morphology includes the segments of the base in addition to the segments of the reduplicant, which must be made up of copies of base segments.
Surely an extension of the gestural manipulations to include gestural insertions is needed here. Moreover, the exact form of the reduplicant is determined by a set of structural constraints. These constraints may require, for example, that the reduplicant have the prosodic shape of a light or a heavy syllable. Such constraints are also not reducible to notions of gestural manipulations.

In fact, extending the set of gestural manipulations by adding gestural insertions and deletions, motivated above solely on the basis of empirical considerations, is also a theoretical imperative under the approach I am pursuing here. In Optimality Theory a potential infinite set of candidates is generated (by the function GEN) from an input form, and each candidate is subject to evaluation by the constraints of the grammar. The set of candidates includes forms which are generated from the input via arbitrary deletions and insertions of segmental features (and other higher units). In the representations of Articulatory Phonology where gestures are the primitive units, we must allow for candidates which are generated from the input via deletions and insertions of gestures.

I now turn to extensions in the representations of Articulatory Phonology. The gestural manipulation that have been studied in Articulatory Phonology so far do not make reference to phonological units larger than the gesture. For example, the casual speech processes discussed above, refer to individual gestures, and not to the segments that these gestures belong to. Segments and syllables are not an explicit part of ‘gestural scores’, the representations that Browman & Goldstein have been using to demonstrate their model. There are, however, constraints in the grammar that make reference to such higher units (see chapter 3). To this extent they become indispensable parts of the phonological representation.

I assume, then, that segments and syllables, although not primitive notions of Articulatory Phonology, are part of the phonological representation. An example representation of the lexical item /bitmap/ is shown in (10) below. The gestural parameters of constriction degree, CD, and constriction location, CL, are shown for every gesture. Recall that velic and laryngeal gestures do not have a CL. Multiple gestures of a single segment, such as the velic and oral gestures of /m/, are indicated by prefixing the CD, CL specifications (in the gesture’s box) with the segment. Drawing association lines from the Root node to its constituent
gestures, exactly as with features in an autosegmental representation, could also be used but will not be adopted here for typographical reasons. The lengths of the boxes representing time are not meant to be an accurate depiction of the timing relations between gestures (and the height of the boxes is irrelevant). What is important for our purposes is that the asymmetry between vocalic and consonantal gestures, V-to-V contiguity versus no C-to-C contiguity, is represented explicitly.

10. Gestural representation of ‘bitmap’

A similar representation has in fact been proposed by Browman & Goldstein (1989), who point out that for the purposes of phonological constituency “the gestures can be grouped into higher-level units, using either association lines or a mapping on to prosodic structure, regardless of the simultaneity, sequentiality or partial overlap of the gestures” (Browman & Goldstein 1989: p. 234).

Finally, as noted earlier, most analyses in Optimality Theory assume some version of autosegmental representations. This is not an assumption of Optimality Theory per se. The constraints of an Optimality theoretic
grammar may well refer to some other representation. In the approach pursued here, constraints that refer to features in standard Optimality theoretic analyses correspond to constraints that refer to gestures and gestural parameters. To give a simple example, in the analysis of consonant harmony presented in chapter 5, the constraint enforcing the spreading of the assimilating gestural parameter is stated as a constraint of the Generalized Alignment theory of McCarthy & Prince (1993b), \text{ALIGN}(\text{Gestural Parameter, Domain of Spreading, Direction of Spreading}), where a gestural parameter is used instead of a feature.

3. ORGANIZATION OF THE DISSERTATION

The dissertation is organized as follows. Chapter 2 presents the main proposal of Articulatory Locality. Building on studies of coarticulation in simple VCV and CVC utterances, explored in the works of Öhman and Fowler, I establish the basic asymmetry between vocalic and consonantal gestures, V-to-V contiguity versus no C-to-C contiguity. Articulatory Locality is proposed and various sources of converging evidence from different areas of phonological theory are presented in its support. Previous notions of locality are discussed and contrasted with Articulatory Locality.

Chapter 3 focuses on the implications of Articulatory Locality for the area of nonconcatenative morpho-phonology. I reconsider cases of long-distance consonantal spreading in CVC sequences, which are putative instances of assimilation between two consonants, skipping the intervening vowel. Articulatory Locality predicts that such assimilations should not exist. I argue that, indeed, this type of alleged assimilation must and can be eliminated from the theory, by reducing it to segmental copying, as in reduplication. Crucial to this reduction is the notion of gradient violation of constraints in Optimality Theory (Prince & Smolensky 1993), and the notion of Correspondence with its particular application to reduplicative morphology (McCarthy & Prince 1995a). The reduction is demonstrated in detail for Temiar, one of the main indigenous languages of Malaysia, notorious for the complexity of its copying patterns. Extensions of the proposal to Semitic languages are also discussed. Two main theoretical implications of this reduction are then developed. First, the distinction between concatenative and nonconcatenative languages need not and
should not be encoded in terms of the special phonological mechanisms
of consonantal spreading over a vowel, applying under V/C Planar
Segregation (the geometric premise of long-distance consonantal
spreading). Second, the locus of the distinction is found, instead, in the
mode of affixation employed in nonconcatenative languages, namely, a-
templatic reduplicative affixation. This type of affixation is implicitly
predicted, though heretofore undocumented, in the typology of word
formation in Prosodic Morphology. Hence, exploring the implications of
Articulatory Locality has two surprising and welcome consequences for
the area of nonconcatenative morpho-phonology.

The following two chapters, 4 and 5, present two closely related
studies of the sounds participating in consonant harmonies, namely, the
coronal consonants, and, then, of the phenomenon of consonant harmony
itself. A recurring property of consonant harmony is its limitation to
coronal sounds. Other consonants such as labials and velars are neither
triggers nor targets of consonant harmonies. To understand this special
status of coronals with respect to consonant harmony, we must first gain
an understanding of the articulatory parameters under control in these
sounds. Chapter 4 is devoted to this purpose. This chapter explores the
two relevant dimensions of articulatory control in coronal sounds. The first
is the traditional mid-sagittal dimension. In this dimension, I discuss the
gestural parameter under control in the distinction between apical and
laminal gestures, a parameter which serves as the basis for phonological
contrasts attested in the coronal stop inventories of various languages. The
second dimension of articulatory control is the cross-sectional (or
transverse) dimension. By exploring the behavior of the tip-blade in the
cross-sectional view, we find that the area of the channel created by the
approximation of the tongue to the palate provides a consistent basis for
distinctions among different coronal fricatives of various languages. In
English, for example, [θ] corresponds to a [wide] value of the cross-
sectional channel, [ʃ] to a [mid], and [s] to a [narrow]. Other languages
like Chinese and Tohono O’odham do not exhibit the full triad of
contrasts, employing only two contrastive values of the cross-sectional
channel area. These two gestural parameters called TTCO for ‘Tongue-
Tip Constriction Orientation’ and TTCA for ‘Tongue-Tip Constriction
Area’ will turn out to be the assimilating parameters in consonant
harmonies.
Chapter 5 turns to a discussion of consonant harmony as attested in various American-Indian, African, Australian, and Indo-Aryan languages. A generous part of this discussion is devoted to the Athabaskan languages of North America and Alaska, where consonant harmony is a relatively common phenomenon. The emerging typology of consonant harmony in terms of the assimilating gestural parameters turns out to be rather limited. I propose that there are only two species of consonant harmony: TTCO harmony and TTCA harmony. The first involves assimilation in terms of the mid-sagittal shape of the tip-blade, and the second involves assimilation in terms of the cross-sectional shape of the tip-blade.

I identify two properties of TTCO and TTCA which are crucial in understanding the limitation of consonant harmony to these parameters and hence to coronal sounds. The shape of the tip-blade, as defined by TTCO and TTCA, can be sustained during the production of an intervening vowel. Moreover, this shape has no significant effect on the acoustic quality of that vowel. I thus argue that, in an assimilation between the two consonants in a CVC sequence, the assimilating parameter propagates through the articulatorily contiguous vowel. The transparency of that vowel, which gives the appearance of long-distance assimilation to consonant harmony, follows from the second property of TTCO and TTCA noted above. I also discuss previous analyses of consonant harmony, arguing that they miss the crucial generalizations about the phenomenon because they do not have the correct assimilating features (or gestural parameters).

Chapter 6 concludes the dissertation with a synopsis of main the points, stressing their relevance to the central thesis of Articulatory Locality.
NOTES

1. This section is not meant to be a comprehensive review of the model of Articulatory Phonology. In addition to the various references cited in the text, further useful discussions can be found in: Ladefoged (1990), Barry (1991), Fowler & Saltzman (1993), and Stevens (1990).

2. The term ‘(tongue) tip’ usually refers to the apex of the tongue. It should be made clear that what Browman & Goldstein call ‘tongue tip’ actually refers to the combination of the tip and the blade. See chapter 4 for the relevant details of the functional anatomy of the tongue.

3. There are striking similarities between Browman & Goldstein’s vocal tract ‘geometry’ and the feature geometric representations utilized in autosegmental phonology (Clements 1985, Sagey 1986, McCarthy 1988 and others). These similarities as well as some differences have been discussed extensively by several authors, e.g. see Browman & Goldstein (1989), Steriade (1990), and Clements (1992).

4. This does not mean that overlap of the consonants with the vocalic gesture begins at the C-center. The vocalic gesture starts the achievement of its target at the first consonant and thus completely overlaps the whole consonant cluster.

5. Browman & Goldstein (1991a: pp. 318 ff.) have also shown how the gestural model of the segment can also explain various gaps in segmental inventories. Relying on acoustic and aerodynamic effects of gestural combinations they explain, for example, why labio-velars are the most frequent double articulations in the languages of the world.

6. Epenthesis is of course another phonological phenomenon that makes clear the need for an operation of gestural insertion. The need for gestural deletion has been pointed out elsewhere by many authors (e.g. Clements 1992, Kohler 1992, Scully 1992).